**Esimerkki 5.428**

NSA:n yleisyyden osoittamiseksi esitämme lisäksi kokeita videotekstien teksteistä, konekääntämisestä ja visuaalisten kysymysten vastaamisesta VATEX-, WMT 2014 English-to-German- ja VQA-v2-tietokannoilla.

**Tulos**

Mitä tietokokonaisuuksia käytetään kolmea muuta tehtävää koskevissa kokeissa?

**Esimerkki 5.429**

Koulutamme LSTM:ää harjoitusjoukon kanssa ja ilman huomiota. Koulutuksen jälkeen otamme kehitysjoukon parhaan mallin BLEU-pistemäärän BIBREF16 perusteella ja laskemme BLEU-pistemäärän testijoukolle.

**Tulos**

Miten generatiivista mallia arvioidaan?

**Esimerkki 5.430**

Olemme keränneet yhteensä 3500 kysymystä Internetistä ja muista lähteistä, kuten yleistietokirjoista, historiasta jne. Korpus sisältää kysymykset ja luokat, joihin kukin kysymys kuuluu.

**Tulos**

mitä tietokokonaisuuksia he käyttivät?

**Esimerkki 5.431**

On huomattava, että BoW/SVM-malli on lineaarinen ennustaja, joka luottaa suoraan sanataajuustilastoihin, joten sen ilmaisuvoima on heikompi kuin CNN-mallilla, joka oppii lisäksi piilokerroksen välikerroksen representaatioita ja konvoluutiosuodattimia. Lisäksi CNN-malli voi hyödyntää semanttista samankaltaisuutta, joka on koodattu hajautettuihin word2vec-representaatioihin, kun taas BoW/SVM-mallissa kaikki sanat ovat "yhtä kaukana" sanasäkkien semanttisessa avaruudessa.

**Tulos**

Miksi CNN-malli on kirjoittajien mukaan selitettävyydeltään parempi?

**Esimerkki 5.432**

Analogisesti sanojen upotusten kanssa lauseiden upotukset, esimerkiksi Universal Sentence Encoder BIBREF8 ja Sentence-BERT BIBREF6, mahdollistavat erilaisten lauseiden kosinus-muotoisen samankaltaisuuden laskemisen, kuten esimerkiksi kysymyksen ja vastaavan vastauksen samankaltaisuuden. Mitä sopivampi tietty vastaus on tiettyyn kysymykseen, sitä vahvempi on sen kosinusarvoinen samankaltaisuus. Kun tarkastellaan kahta vastakkaista vastausta, on siis mahdollista määrittää vinoutusarvo: missä $\vec{q}$ on kysymyksen vektorimuotoinen esitys ja $\vec{a}$ ja $\vec{b}$ ovat kahden vastauksen/valinnan esitykset.

**Tulos**

Miten moraalista puolueellisuutta mitataan?

**Esimerkki 5.433**

Laajennamme Amazon Conversational Bot Toolkit (CoBot) BIBREF6 -ohjelmistoa, joka on joustava tapahtumapohjainen kehys. CoBot tarjoaa ASR-tuloksia ja luonnollisen kielen käsittelyputkia Alexa Skills Kit (ASK) BIBREF7 -paketin kautta. Gunrock korjaa ASR:n kontekstin mukaan (asr) ja luo luonnollisen kielen ymmärtämisen (NLU) (nlu) moduulin, jossa useat komponentit analysoivat käyttäjän lausumia. Dialoginhallinta (DM) (dm) käyttää NLU:n piirteitä aihekohtaisten dialogimoduulien valitsemiseen ja määrittelee yksilöllisen dialogivirran. Kukin dialogimoduuli hyödyntää useita tietopankkeja (knowledge). Tämän jälkeen luonnollisen kielen generointimoduuli (NLG) (nlg) tuottaa vastaavan vastauksen. Seuraavissa jaksoissa esitetään yleiskatsaus järjestelmästä, mutta yksityiskohtaiset tiedot järjestelmän toteutuksesta löytyvät teknisestä raportista BIBREF1.

**Tulos**

Määrittelevätkö he Gunrockin käyttämän mallin?

**Esimerkki 5.434**

Sitten se valitsee osajoukon 1700 tunnin ( INLINEFORM2 1,1M instanssit) merkitsemättömästä tietokokonaisuudesta.

**Tulos**

Kuinka monen datanäytteen perusteella he aloittavat ennen kuin he saavat alkuperäiset mallin merkinnät?

**Esimerkki 5.435**

Äskettäin julkaistussa tietokokonaisuudessa nimeltä SCAN BIBREF2 (Simplified version of the CommAI Navigation tasks) testataan koostumuksen yleistämistä sekvenssistä sekvenssiin (seq2seq) jättämällä systemaattisesti pois harjoitusjoukosta kaikki syötteet, jotka sisältävät primitiivisen perusverbin ("hyppää"), ja testaamalla sekvenssejä, jotka sisältävät kyseisen verbin.

**Tulos**

Miten SCAN-tietokannassa arvioidaan koostumuksen yleistämistä?

**Esimerkki 5.436**

Viimeaikaisten töiden mukaisesti vertailemme kahta laajalti käytettyä tietokokonaisuutta, Penn Treebank (PTB) BIBREF28 -tietokantaa, sellaisena kuin se on laadittu BIBREF:ssä29 , ja WikiText2 (WT-2) BIBREF20 -tietokantaa.

**Tulos**

mitä tietoja he käyttivät?

**Esimerkki 5.437**

Näin ollen arkkitehtuurimme perustuu Embeddings from Language Model eli ELMo BIBREF10 -malliin. Käytämme esivalmennettua ELMo-mallia, joka on saatu käyttämällä 1 Billion Word Benchmark -vertailumallia, joka sisältää noin 800 miljoonaa tokenia uutisdataa WMT 2011 BIBREF:stä24 . Tämän jälkeen kontekstualisoidut upotukset siirretään BiLSTM-malliin, jossa on 2 048 piilotettua yksikköä. Yhdistämme LSTM:n piilotetut tilat käyttämällä max-pooling-menetelmää, joka alustavissa kokeiluissamme tarjosi parempia tuloksia, ja syötämme tuloksena saadun vektorin 2-kerroksiseen feed-forward-verkkoon, jossa kussakin kerroksessa on 512 yksikköä. Tämän ulostulo syötetään sitten mallin viimeiselle kerrokselle, joka suorittaa binäärisen luokittelun.

**Tulos**

Minkä tyyppisessä mallissa ELMo-esityksiä käytetään?

**Esimerkki 5.438**

Kuten kohdassa SECREF15 selitetään, Doc2VecC:ssä käyttöön otettu korruptoituminen toimii datasta riippuvana regularisaationa, joka tukahduttaa usein esiintyvien mutta epäinformatiivisten sanojen upotukset. Sitä vastoin Doc2VecC onnistuu hillitsemään harjoitusjoukossa usein esiintyvien mutta epäinformatiivisten sanojen, kuten symbolien ja stop-sanojen, edustusta.

**Tulos**

Miten he määrittelevät, mitkä sanat ovat informatiivisia?

**Esimerkki 5.439**

Tässä jaksossa vertailemme empiirisesti automaattista differentiointia (AD, Cladiin perustuva toteutuksemme) ja numeerista differentiointia (ND, joka perustuu äärellinen differenssi -menetelmään) ROOTissa. Osoitamme, että AD voi parantaa huomattavasti derivaatan arvioinnin tarkkuutta ja suorituskykyä ND:hen verrattuna.

**Tulos**

Miten automaattisen derivoinnin oikeellisuus todistetaan?

**Esimerkki 5.440**

Käytimme joukon globaaleja verkkoindikaattoreita, joiden avulla voimme koodata jokaisen verkkokerroksen ominaisuustuplilla. Tämän jälkeen yksinkertaisesti yhdistimme tuplat toisiinsa, jotta jokainen monikerroksinen verkko voidaan esittää yhdellä ominaisuusvektorilla. Käytimme seuraavia verkon globaaleja ominaisuuksia: Strongly Connected Components (SCC): suunnatun graafin Strongly Connected Component on maksimaalinen (ali)graafi, jossa jokaiselle kärkiparille $u,v$ on polku kumpaankin suuntaan ($u\rightarrow v$, $v\rightarrow u$).Suurimman vahvasti kytkeytyneen komponentin koko (LSCC): tietyn graafin suurimman vahvasti kytkeytyneen komponentin solmujen lukumäärä.Heikosti kytkeytyneiden komponenttien lukumäärä (WCC): suunnatun graafin heikosti kytkeytynyt komponentti on maksimaalinen (osa)graafi, jossa jokaiselle kärkiparille $(u, v)$ on polku $u \ vasenoikeaoikeaoikeaoikeaoikeaoikea v$, ottamatta huomioon reunojen suuntaa.Suurimman heikosti kytketyn komponentin koko (LWCC): tietyn graafin suurimman heikosti kytketyn komponentin solmujen lukumäärä.Suurimman heikosti kytketyn komponentin halkaisija (DWCC): suurin etäisyys (lyhimmän polun pituus) kahden solmun välillä graafin suurimmassa heikosti kytketyssä komponentissa (suuntaamattomassa versiossa).Keskimääräinen klusterointikerroin (Average Clustering Coefficient, CC): graafin kaikkien solmujen paikallisten klusterointikertoimien keskiarvo; solmun paikallinen klusterointikerroin ilmaisee, kuinka lähellä sen naapurit ovat täydellistä graafia (tai klikkiä). K-ydinluku (KC): Graafin K-ydin BIBREF13 on maksimaalinen osa-graafi, joka sisältää solmuja, joiden sisäinen aste on vähintään $k$; K-ydinluku on $k$:n suurin arvo (suunnatuissa graafeissa otetaan huomioon kokonaisaste).Tiheys (d): suunnattujen graafien tiheys on $d=\frac{|E|}{|V||V-1|}$, jossa $|E|$ on reunojen lukumäärä ja $|N|$ on graafin kärkien lukumäärä; tiheys on 0, jos graafissa ei ole reunoja, ja 1, jos graafi on täydellinen.Suurimman heikosti kytketyn komponentin (SV) rakenteellinen viraliteetti: Tämä mitta määritellään BIBREF14:ssä kaskadipuun kaikkien solmuparien väliseksi keskimääräiseksi etäisyydeksi tai vastaavasti solmujen keskimääräiseksi syvyydeksi keskiarvona kaikista solmuista, jotka vuorostaan toimivat juurina; jos $|V| > 1$ kärkipistettä, $SV=\frac{1}{|V||V-1|}\sum \_i\sum \_j d\_{ij}$, jossa $d\_{ij}$ tarkoittaa lyhimmän polun pituutta solmupisteiden $i$ ja $j$ välillä. Tämä vastaa sitä, että lasketaan graafin Wiener-indeksi BIBREF29 ja kerrotaan se kertoimella $\frac{1}{|V||V-1|}$. Meidän tapauksessamme laskimme sen suurimman heikosti kytkeytyneen komponentin suuntaamattomalle vastaavalle graafille ja asetimme sen arvoksi 0 aina kun $V=1$.

**Tulos**

Mitkä ovat ne globaalin verkon piirteet, jotka määrittävät jakamisprosessin eri osatekijöitä?

**Esimerkki 5.441**

SVM: Määritellään 3 ominaisuusjoukkoa, jotka kuvaavat kutakin kysymystä. Ensimmäinen on pelkkä sanasäkkipiirteiden joukko kysymyksen osalta (SVM-BOW), toinen on kysymyksen sanasäkkipiirteet sekä kysymyksen pituus sanoina (SVM-BOW + LEN), ja viimeiseksi poimimme kysymyksen sanasäkkipiirteet, kysymyksen pituuden sanoina sekä part-of-speech-tunnisteet kysymykselle (SVM-BOW + LEN + POS). Näin saadaan 200, 201 ja 228 ulottuvuuden vektorit, jotka annetaan SVM:lle, jossa on lineaarinen ydin. Ei vastausta -perustilanne (NA) : Useimpiin saamiimme kysymyksiin on vaikea vastata oikeudellisesti järkevällä tavalla tietosuojaselosteessa olevien tietojen perusteella. Määritämme yksinkertaisen perustason, jonka avulla voimme arvioida, miten vaikuttaisi, jos jokainen kysymys määritettäisiin vastaamattomaksi. Sanojen lukumäärän perustaso: Jotta voisimme arvioida, miten yksinkertaisen leksikaalisen yhteensovittamisen käyttäminen kysymyksiin vastaamiseen vaikuttaa, etsimme kutakin kysymystä varten parhaat ehdokkaat politiikan lauseiksi käyttämällä sanojen lukumäärän perustasoa BIBREF53, joka laskee niiden kysymyksen sanojen lukumäärän, jotka esiintyvät myös lauseessa. Otamme mukaan 2, 3 ja 5 parasta ehdokasta perusviitekehyksenä. Ihmisen suorituskyky: Valitsemme jokaisen kommentoijan antaman vertailuvastauksen ja laskemme F1-arvon suhteessa jäljelle jääviin viitteisiin, kuten kohdassa 4.2.1 on kuvattu. Kutakin vertailuvastausta käsitellään ennusteena, ja jäljelle jääviä n-1 vastausta käsitellään kultaisena vertailuna. Kaikkien referenssivastausten suurimman F1-arvon keskiarvo lasketaan ihmisen perustasona.

**Tulos**

Testattiinko muita lähtötasoja, jotta niitä voitaisiin verrata neuraaliseen lähtötasoon?

**Esimerkki 5.442**

On syytä mainita, että kerätyissä teksteissä on paljon erityyppisiä virheitä: ortografisia ja syntaktisia virheitä, koodinvaihdettuja sanoja (eli sanoja, jotka eivät ole vaaditulla kielellä), vitsejä jne. Alkuperäisiä kirjoitettuja lauseita on siis käsitelty "puhtaampien" versioiden tuottamiseksi, jotta aineisto olisi käyttökelpoinen joihinkin tutkimustarkoituksiin (esim. kielimallien kouluttamiseen, piirteiden poimimiseen taitojen arviointia varten jne...).

**Tulos**

Ovatko kaikki lausumat epäkieliopillisia?

**Esimerkki 5.443**

Menetelmä ::: Se saa kysymyksen $q$ ja $K$ kohtaa $\mathcal {P} = \lbrace p\_1, p\_2 ... p\_K\rbrace $ ehdokkaiden joukosta ja tuottaa ketjun valituista kohdista. Menetelmä ::: Ehdollisen valinnan lisäksi ehdotamme lisäksi, että MatchLSTM-arkkitehtuurilla toteutettu yhteistyöhön perustuva Reasoner-malli (ks. liite SECREF6) ennustaa linkittävän kokonaisuuden valituista kohdista, jotta voidaan lieventää etävalvontasignaalin $\mathcal {C}$ kohinaa.

**Tulos**

Mitkä ovat ehdotetun ratkaisun kaksi malliarkkitehtuuria?

**Esimerkki 5.444**

Ehdotamme laajennettua keskikontekstia, uutta kontekstiedustusta CNN:lle relaatioiden luokittelua varten. Kontekstit jaetaan kahteen relaatioargumenttiin perustuvaan kolmeen erilliseen alueeseen: vasempaan kontekstiin, keskimmäiseen kontekstiin ja oikeaan kontekstiin. Näin ollen ehdotamme kahden kontekstin käyttöä: (1) vasemman kontekstin, vasemman entiteetin ja keskimmäisen kontekstin yhdistelmä ja (2) keskimmäisen kontekstin, oikean entiteetin ja oikean kontekstin yhdistelmä. Nämä kaksi kontekstia käsitellään kahdella itsenäisellä konvoluutio- ja maksimipoolauskerroksella. Yhdistämisen jälkeen tulokset ketjutetaan lauseen esityksen muodostamiseksi.

**Tulos**

Miten he saavat uuden kontekstin edustuksen?

**Esimerkki 5.445**

Samoin kuin BIBREF5 , ensimmäisessä aliverkossa pinyin-sekvenssi ennustetaan videosta. Toisin kuin BIBREF5 , joka ennustaa pinyin-hahmoja videosta, CSSMCM:ssä pinyin otetaan kokonaisuutena, joka tunnetaan myös tavuina. Kuten tiedämme, mandariinikiina on tavuihin perustuva kieli, ja tavut ovat niiden looginen ääntämisyksikkö.

**Tulos**

Mitä syntaktista rakennetta käytetään sävyjen mallintamiseen?

**Esimerkki 5.446**

Oikeudenmukaisten järjestelmien luomiseksi on siis otettava huomioon yhteiskunnan edustusongelmat, jotka on tarkoitus sisällyttää tietoihin.

**Tulos**

Mikä on NLP:n sukupuoleen perustuvan ennakkoluulon tutkimisen tavoite erityisesti uutislähetysten alalla ja uutisankkurin roolissa?

**Esimerkki 5.447**

Järjestelmä koostuu seuraavista osista: (i) paikallisista, pinnallisista piirteistä, jotka perustuvat pääasiassa ortografisiin, sananmuoto- ja n-grammiominaisuuksiin sekä niiden yhteyteen; ja ii) kolmenlaisista yksinkertaisista klusterointipiirteistä, jotka perustuvat unigrammaattiseen yhteensovittamiseen: i) Brown BIBREF32 -klustereista, joissa käytetään polun 4., 8., 12. ja 20. solmua; ii) Clark BIBREF33 -klustereista ja iii) Word2vec BIBREF34 -klustereista, jotka perustuvat K-menetelmään, jota sovelletaan poimittuihin sanavektoreihin skipgrammialgoritmia käyttäen.

**Tulos**

Mitä matalia paikallisia piirteitä poimitaan?

**Esimerkki 5.448**

Arvioimme tietokokonaisuuttamme käyttämällä perinteisiä ja syväoppimismenetelmiä. Yksinkertaisin mallimme on lineaarinen SVM, joka on koulutettu sanojen unigrammeilla. SVM:t ovat tuottaneet huipputuloksia monissa tekstiluokittelutehtävissä BIBREF13 . Koulutamme myös kaksisuuntaisen pitkän lyhytaikaisen muistin (BiLSTM) mallin, jonka sovelsimme sentimentSystem,rasooli2018crossin sentimenttianalyysijärjestelmästä ja muutimme sen sijaan ennustamaan loukkaavia leimoja. Se koostuu (1) syötteen upotuskerroksesta, (2) kaksisuuntaisesta LSTM-kerroksesta, (3) syötteen piirteiden keskimääräisen pooling-kerroksesta. LSTM:n ja keskiarvopoolauskerroksen ketjutus johdetaan tiheän kerroksen läpi ja ulostulo johdetaan softmax-funktion läpi. Asetimme kaksi tulokanavaa tulon upotuskerroksille: valmiiksi koulutetut FastText- upotukset BIBREF14 sekä päivitettävät upotukset, jotka malli oppii koulutuksen aikana. Lopuksi sovellamme myös BIBREF15 -arkkitehtuuriin perustuvaa konvoluutiohermoverkkomallia (Convolutional Neural Network, CNN), jossa käytetään samoja monikanavaisia syötteitä kuin edellä mainitussa BiLSTM:ssä.

**Tulos**

Mitä malleja kokeessa käytetään?

**Esimerkki 5.449**

Mukautimme BERTNLU:n ConvLab-2:sta. Toteutimme sääntöpohjaisen mallin (RuleDST) ja mukautimme TRADE:n (Transferable Dialogue State Generator) BIBREF19:stä tähän kokeeseen. Mukautimme ConvLab-2:sta valvotusti koulutetun vaniljakäytännön (SL-politiikka).

**Tulos**

Mitkä ovat vertailumallit?

**Esimerkki 5.450**

Erityisesti kootaan CommonCrawl-tietokannasta ne asiakirjat, joissa on eniten päällekkäisiä n-grammeja kysymysten kanssa. Nimeämme tämän tietokokonaisuuden STORIES, koska suurin osa sen sisältämistä asiakirjoista on tarinan muotoisia ja sisältää pitkän ketjun yhtenäisiä tapahtumia. Kuva 5 - vasemmalla ja keskellä osoittaa, että STORIES tuottaa aina korkeimman tarkkuuden molemmilla syötteen käsittelytavoilla.

**Tulos**

Mikä niiden koulutusaloista parantaa suorituskykyä eniten?

**Esimerkki 5.451**

Ironian luokittelija: Toteutamme CNN-luokittimen, joka on koulutettu ironia-aineistollamme. Ironian aiheluokittelija: Toteutamme ensin yksikerroksisen LSTM-verkon, joka luokittelee aineistossamme olevat ironiset lauseet positiiviseen ja negatiiviseen ironiaan. Tunteiden luokittelija ei-ironiaa varten: Samoin kuin ironian tunteiden luokittelijan koulutusprosessi, toteutamme ensin yksikerroksisen LSTM-verkon, joka on koulutettu tavallisten twiittaajien tunteiden analysointia varten laaditulla tietokannalla, jotta ei-ironiset lauseet voidaan luokitella positiivisiin ja negatiivisiin ei-ironisiin lauseisiin. Tässä jaksossa kuvaamme joitakin lisäkokeita ironisten lauseiden muuntamisesta ei-ironisiksi lauseiksi.

**Tulos**

Mitä kokeita tehdään?

**Esimerkki 5.452**

Aineisto sisältää yhteensä 353 keskustelua 40 puhujalta (11 sairaanhoitajalta, 16 potilaalta ja 13 hoitajalta), jotka ovat antaneet suostumuksensa anonymisoitujen tietojen käyttöön tutkimuksessa. Jaksossa SECREF16 muodostamme mallit ja ilmaisupoolit kielellisen analyysin avulla, jota seuraa manuaalinen verifiointi.

**Tulos**

Kuinka suuri on niiden luoma tietokokonaisuus?

**Esimerkki 5.453**

Kymmenkertainen ristiinvalidointi tällä asetuksella antoi noin 71 %:n tarkkuuden token-tasolla.

**Tulos**

Ilmoitetaanko artikkelissa automaattisen käännösmallin käännöstarkkuus tunisian ja arabian sanojen välillä?

**Esimerkki 5.454**

Koestusmenetelmät ja mallinnus ::: : Tehtävän määrittely ja mallintaminen ::: Synteettisiä tietokokonaisuuksia luotaessa on tärkeää varmistaa, että systemaattiset vääristymät tai annotaatio-artefaktit BIBREF41 eivät päädy tuloksena syntyviin koettimiin ja että kohdetietokokonaisuudet ovat riittävän haastavia (tai hyviä, BIBREF42:n merkityksessä). Tämän testaamiseksi käytämme useita MCQA:n perusmalleja, jotka esiteltiin ensimmäisen kerran BIBREF0:ssa ja jotka perustuvat BIBREF43:ssa NLI:lle käytettyihin LSTM-pohjaisiin malleihin, sekä erilaisia näihin malleihin perustuvia osittaissyöttöisiä perusmalleja.

**Tulos**

Miten he valvovat merkintöjen keinotekoisuutta?

**Esimerkki 5.455**

Kuten sen nimestä käy ilmi, Recurrent Deep Stacking Network pinoaa ja yhdistää edellisten kehysten tuotokset nykyisen kehyksen tulo-ominaisuuksiin.

**Tulos**

Mitä rekursiivinen syvä pinoamisverkko tekee?

**Esimerkki 5.456**

Kysymysten luontimalli antaa kullekin vastausehdokkaalle pistemäärän mittaamalla kysymyksen ja luodun kysymyksen semanttista vastaavuutta vastausehdokkaan semantiikan perusteella.

**Tulos**

Missä käytetään kysymyksenmuodostusmallia?

**Esimerkki 5.457**

ASR-järjestelmä ylläpitää sisäisesti runsaasti hypoteesiavaruutta puheverkkojen tai sekaannusverkkojen (cnets) muodossa.

**Tulos**

Mikä on sanasekaannusverkko?

**Esimerkki 5.458**

Tässä työssä arvioimme kolmea julkisesti saatavilla olevaa valmista ydinviittausten erottelujärjestelmää, jotka edustavat kolmea erilaista koneoppimisparadigmaa: sääntöpohjaisia järjestelmiä, ominaisuuksiin perustuvia tilastollisia järjestelmiä ja neuraalisia järjestelmiä. Arvioimme esimerkkejä kustakin kolmesta ydinviittausjärjestelmäarkkitehtuurista, jotka on kuvattu kohdassa "Coreference Systems": BIBREF5-sieve-järjestelmä sääntöpohjaisesta paradigmasta (jäljempänä RULE), BIBREF6 tilastollisesta paradigmasta (STAT) ja BIBREF11-syvä vahvistusjärjestelmä neuraalisesta paradigmasta (NEURAL).

**Tulos**

Mitä coreference resolution -järjestelmiä testataan?

**Esimerkki 5.459**

Tarkastelemme myös seuraavia peruslinjoja. BOW-Tags edustaa sijainnit sanasäkkiedustuksen avulla käyttäen samaa tunnisteiden painotusta kuin upotusmallissa. BOW-KL(Tags) käyttää samaa esitystä, mutta termien valinnan jälkeen, käyttäen samaa KL-pohjaista menetelmää kuin upotusmalli. BOW-All yhdistää sanasäkkiedustuksen ja strukturoitua tietoa, joka on koodattu BIBREF:ssä7 ehdotetulla tavalla. GloVe käyttää alkuperäisen GloVe-mallin tavoitetta sijaintivektoreiden oppimiseen, eli tämä muunnos eroaa EGEL-Tagsista siten, että INLINEFORM1:n sijasta käytämme INLINEFORM4:llä mitattua INLINEFORM2-tunnisteen INLINEFORM2 samanaikaisten esiintymien lukumäärää sijainnin INLINEFORM3 lähellä.

**Tulos**

Mitkä ovat nykyiset lähestymistavat?

**Esimerkki 5.460**

Lisäksi raportoimme sentimenttitarkkuuden (Senti ACC), joka mittaa, onko tulostettu lause sama sentimenttipolariteetti kuin syöttölause, joka perustuu standardoituihin sentimenttiluokittelijoihimme. Sisällön säilyttämisen suorituskyvyn arvioimiseksi lasketaan BLEU-pistemäärä BIBREF25 syöttölauseiden ja tulostuslauseiden välillä. Eri mallien kokonaissuorituskyvyn arvioimiseksi ilmoitamme myös sentimenttitarkkuuden ja BLEU-pisteiden geometrisen keskiarvon (G2) ja harmonisen keskiarvon (H2). Ironiatarkkuus ilmoitetaan vain ihmisen tekemän arvioinnin tuloksissa, koska ihmisen on tarkempi arvioida ironian laatua, koska se on hyvin monimutkainen.

**Tulos**

Kuka arvioi ironian tarkkuutta, tunteiden säilymistä ja sisällön säilymistä?

**Esimerkki 5.461**

Syylliset asiakirjat valittiin siten, että ne sisälsivät vain yhden täsmällisen unigrammin esiintymän: `caused', `causing' tai `causes'.

**Tulos**

Mitkä ovat "syy-seuraussuhdelausuntojen" valintakriteerit?

**Tulos**

Miten he poimivat tekstistä kausaliteetin?

**Esimerkki 5.462**

Ratkaisemme tämän jännitteen kouluttamalla datan esittämiseksi vektoriavaruusmalleja, joissa jokainen suuren korpuksen sana esitetään vektorina (upotettuna) korkea-ulotteisessa avaruudessa. Tuloksena syntyvän vektoriavaruuden geometria kuvaa monia sanojen välisiä semanttisia suhteita. Ratkaisemme tämän jännitteen kouluttamalla vektoriavaruusmalleja aineiston esittämiseksi, jossa jokainen yksittäinen sana suuressa korpuksessa esitetään vektorina (upotuksena) korkea-ulotteisessa avaruudessa. Tuloksena syntyvän vektoriavaruuden geometria kuvaa monia sanojen välisiä semanttisia suhteita.

**Tulos**

Mallintavatko ne semantiikkaa

**Esimerkki 5.463**

Verrattuna ulkoisten mallien käyttämiseen luottamuksen mallintamiseen ehdotetun menetelmän etuna on se, että perusmalli ei muutu: binäärinen luokitteluhäviö tarjoaa vain lisävalvontaa. Valvotuissa avoimissa IE-järjestelmissä väitteen luottamuspistemäärä lasketaan tyypillisesti mallin BIBREF3 , BIBREF5 antaman uuttamiskelpoisuuden perusteella. Seuraamme Stanovsky:2016:OIE2016:ssa kuvattuja arviointimittareita: tarkkuus-uudelleenvalintakäyrän alapuolella oleva pinta-ala (Area under the precision-recall curve, AUC) ja F1-pistemäärä.

**Tulos**

Miten tämä vertautuu perinteisiin kalibrointimenetelmiin, kuten Platt Scalingiin?

**Esimerkki 5.464**

Lopulta saamme tasapainoisen testitietokannan, jossa kullakin termi-merkitysparilla on noin 15 näytettä testausta varten (keskimäärin kullakin parilla on 14,56 näytettä ja näytteiden mediaani on 15), ja vertailu harjoitustietokantaan on esitetty kuvassa KUVA 11. Koska testitietokannan kerääminen on vaikeaa, päätimme kerätä vain satunnaisesti valitut 30 termiä.

**Tulos**

Kuinka suuri on testauksessa käytettävä tietokokonaisuus?

**Esimerkki 5.465**

Depechemood on leksikoniin perustuva tunteiden tunnistusmenetelmä, joka on kerätty joukkorekisteröidyistä uutisista BIBREF24. Tutkijat käyttivät noin 23,5 000 asiakirjaa, joissa oli keskimäärin 500 sanaa asiakirjaa kohti, ja pyysivät koehenkilöitä ilmoittamaan tunteistaan kunkin artikkelin lukemisen jälkeen. Sen jälkeen he kertoivat asiakirja-tunne-matriisin ja sana-asiakirja-matriisin saadakseen näille sanoille tunne-sana-matriisin.

**Tulos**

Miten id Depechemood koulutettu?

**Esimerkki 5.466**

Rakennamme syötteenä rinnakkaisen yksikielisen korpuksen kääntämällä sekakielisen sekvenssin Google NMT:n avulla englannin ( INLINEFORM0 ) ja mandariinin ( INLINEFORM1 ) sekvensseiksi.

**Tulos**

Mitä rinnakkaista korpusta he käyttivät?

**Esimerkki 5.467**

Kolmen kommentoijan, joilla ei ole tietoa siitä, mistä järjestelmästä vastaus on peräisin, on sitten arvioitava itsenäisesti voitto (vastaus$\_1$ on parempi), tappio (vastaus$\_2$ on parempi) ja tasapeli (ne ovat yhtä hyviä tai huonoja) ottaen huomioon neljä näkökohtaa: johdonmukaisuus, looginen johdonmukaisuus, sujuvuus ja monipuolisuus.

**Tulos**

Mitä inhimillisen arvioinnin mittareita käytetään?

**Esimerkki 5.468**

Tunteiden havaitsemisjärjestelmien kehittämisessä ja arvioinnissa käytetään yleisesti kolmea tunteilla annotoitua tietokokonaisuutta: Affective Text -tietokokonaisuutta, Fairy Tales -tietokokonaisuutta ja ISEAR-tietokokonaisuutta. Yhteenveto on esitetty taulukossa TABREF8 , jonka alimmalla rivillä näkyy myös, mikä rooli kullakin tietokokonaisuudella on kokeiluissamme: lukuun ottamatta Affective Text -tietokannan kehitysosaa, jota käytimme mallien kehittämiseen (SECREF4 ), kaikkia kolmea tietokantaa on käytetty arvioinnissamme vertailukohtina.

**Tulos**

Mihin olemassa oleviin vertailuarvoihin niitä verrattiin?

**Esimerkki 5.469**

Vertailemme kolmea perusmenetelmää. $({1})$ SC-LSTM BIBREF3 on kanoninen malli ja vahva perusmalli, joka käyttää ylimääräistä dialogin tekovektoria ja lukuporttia ohjaamaan lausuman tuottamista. $({2})$ GPT-2 BIBREF6:ta käytetään suoraan hienosäätöön aluespesifisten merkintöjen perusteella ilman esiharjoittelua laajamittaisella (dialogi teko, vastaus) -parien korpuksella. $({3})$ HDSA BIBREF7 on huippuluokan malli MultiWOZ:ssa. Se hyödyntää dialogin tekorakenteita, jotta siirto olisi mahdollista monialueen ympäristössä, ja osoittaa SC-LSTM:ää parempaa suorituskykyä.

**Tulos**

Mihin nykyisiin menetelmiin SC-GPT:tä verrataan?

**Esimerkki 5.470**

Taulukossa TABREF16 esitetään BLSTMP- ja VGG-mallia käyttävän naiivin monikielisen lähestymistavan tunnistussuorituskyky BLSTMP:llä koulutettua yksikielistä mallia vastaan. Tulokset osoittavat selvästi, että parempi arkkitehtuuri, kuten VGG-BLSTM, auttaa parantamaan monikielistä suorituskykyä. Käytimme merkkitason RNNLM:ää, joka koulutettiin 2-kerroksisella LSTM:llä merkkisekvensseihin.

**Tulos**

Mitä arkkitehtuureja tutkitaan seq2seq-mallin parantamiseksi?

**Esimerkki 5.471**

Seuraavassa tarkastelemme eri tapoja perääntyä, kun ScRNN ennustaa UNK:n (yleinen tulos harvinaisille ja tuntemattomille sanoille): Pass-through: sanantunnistaja siirtää (mahdollisesti väärin kirjoitetun) sanan sellaisenaan.Backoff neutraaliin sanaan: Vaihtoehtoisesti, huomioiden, että kulkeminen $\colorbox {gray!20}{\texttt {UNK}}$ -ennustettujen sanojen läpivieminen muuttumattomana altistaa myöhemmän prosessin mallin mahdollisesti korruptoituneelle tekstille, harkitaan takaisinkytkentää neutraaliin sanaan, kuten `a', jolla on samanlainen jakauma eri luokissa. takaisinkytkentä taustamalliin: Harkitsemme myös turvautumista yleisempään sanantunnistusmalliin, joka on koulutettu suuremmalla, vähemmän erikoistuneella korpuksella, aina kun etualan sanantunnistusmalli ennustaa UNK:n. Kuvassa 1 esitetään tämä skenaario kuvallisesti.

**Tulos**

Miten backoff-strategiat toimivat?

**Esimerkki 5.472**

Tässä skenaariossa meillä on paritettu tietokokonaisuus, joka sisältää artikkelien ja kommenttien rinnakkaisen sisällön INLINEFORM0 , ja parittamaton tietokokonaisuus, joka sisältää asiakirjat (artikkelit tai kommentit) INLINEFORM1 . Valvottu malli koulutetaan INLINEFORM2:lla, jotta voimme oppia artikkeleiden ja kommenttien välisen vastaavuuden tai yhdistämisen. Jakamalla valvotun mallin ja valvomattoman mallin koodaajan voimme kouluttaa molemmat mallit yhteisellä tavoitefunktiolla: DISPLAYFORM0

**Tulos**

Mitä paritettuja korpuksia he käyttivät toisessa kokeessa?

**Esimerkki 5.473**

Englanninkielisen version osalta suoritimme sekä perusteellisen manuaalisen analyysin että automaattisen arvioinnin kolmella yleisesti käytetyllä TS-tietokannalla kahdelta eri alalta arvioidaksemme kehyksemme suorituskykyä lauseiden jakamisen osatehtävän osalta. Tulokset osoittavat, että ehdotettu lauseenjakomenetelmämme on rakenteellisen TS:n nykytilaa parempi, sillä se palauttaa hienojakoisia yksinkertaistettuja lauseita, jotka ovat kieliopillisesti hyvin täsmällisiä ja säilyttävät syötteen merkityksen. Koko arviointimenetelmä ja yksityiskohtaiset tulokset on raportoitu asiakirjassa niklaus-etal-2019-transforming. Lisäksi vertaileva analyysi RST Discourse Treebankin BIBREF6 sisältämien annotaatioiden kanssa osoittaa, että pystymme vangitsemaan jaettujen lauseiden välisen kontekstuaalisen hierarkian lähes 90 prosentin tarkkuudella ja saavuttamaan noin 70 prosentin keskimääräisen tarkkuuden lauseiden välisten retoristen suhteiden luokittelussa. Saksankielisen version arviointi on käynnissä.

**Tulos**

Onko malli arvioitu?

**Esimerkki 5.474**

Tietojen kerääminen: Tässä vaiheessa $p$ koostuu $n$ lauseesta ja nykyinen havainto vastaa lausetta $s\_k,~1 \le k \le n$:edellinen: hyppää kohtaan $ \small {\left\lbrace \begin{array}{ll} s\_n & \text{jos $k = 1$,}\\\ s\_{k-1} & \text{muussa tapauksessa;} \end{array}\right.} $sext: hyppää kohtaan $ \small {\left\lbrace \begin{array}{ll} s\_1 & \text{if $k = n$,}\\\ s\_{k+1} & \text{muussa tapauksessa;} \end{array}\right.} $Ctrl+F $<$kysely$>$: hyppää lauseeseen, joka sisältää "kyselyn" seuraavan esiintymän;stop: lopeta tiedonkeruuvaihe.

**Tulos**

Millaisia komentoja niiden asetukset tarjoavat tietoa etsiville malleille?

**Esimerkki 5.475**

Kokeissamme käytämme samoja tietokokonaisuuksia kuin muissa teoksissa BIBREF32, BIBREF23 ja BIBREF33 sekä muita tietokokonaisuuksia, jotka olemme itse keränneet käyttämällä samankaltaista kriteeriä (kuvattu kohdassa SECREF4). Käytämme kolmekymmentä erilaista keskustelua, jotka käytiin maaliskuun 2015 ja kesäkuun 2019 välisenä aikana, puolet niistä kiistanalaisia ja puolet kiistattomia. Otimme huomioon neljällä eri kielellä käytyjä keskusteluja: Ne tapahtuivat viidellä alueella eri puolilla maailmaa: Etelä- ja Pohjois-Amerikassa, Länsi-Euroopassa, Keski- ja Etelä-Aasiassa. Tutkimme näitä keskusteluja myös ottamalla ensin 140 merkkiä ja sitten 280 merkkiä kustakin twiitistä analysoidaksemme eroa suorituskyvyssä ja laskenta-ajassa viestien pituuden suhteen.

**Tulos**

Mitä tietokokonaisuuksia he käyttivät?

**Esimerkki 5.476**

Oletetaan, että meille annetaan tietty kielimalli, kuten GPT-2 BIBREF6, GROVER BIBREF8 tai CTRL BIBREF7, ja sitä luonnehditaan joko risti-entropian $H(P,Q)$ tai perpleksisyyden $\mathrm {PPL}(P,Q)$ estimaateilla.Näemme suoraan, että Neyman-Pearsonin havaintovirhe i.i.d. merkkien tapauksessa on: ja vastaavat tulokset pätevät ergodisille havainnoille.Koska pidämme $H(P)$ vakiona, havaitsemme, että päätösongelman virheeksponentti on nimenomaan risti-entropian affiininen siirtymä. Risti-entropian tai perpleksisyyden kannalta parempien mallien tuotoksia on vaikeampi erottaa aidosta tekstistä.Näin ollen näemme, että intuitiiviset generatiivisen tekstin laadun mittarit vastaavat muodollista erottamattomuuden operatiivista mittaria, joka tulee hypoteesin testauksen raja-arvosta.

**Tulos**

Mitkä kielimallit tuottavat tekstiä, joka on helpompi luokitella aidoksi tai generoiduksi?

**Esimerkki 5.477**

Tässä työssä käytämme BIBREF1:n julkaisemia tietokokonaisuuksia ja BIBREF0:n toimittamaa HEOT-tietokokonaisuutta. Sekä HEOT- että BIBREF1-tietokannat sisältävät twiittejä, jotka on annotoitu kolmeen luokkaan: loukkaavat, loukkaavat ja ei mitään (tai hyvänlaatuiset) twiitit.

**Tulos**

Suorittavatko ne jotain merkintöjä?

**Esimerkki 5.478**

DNN-mallin joustavuus mahdollisti monien muiden pintatason piirteiden, kuten affiksien, sanojen ja kantojen etu- ja loppumerkkien sekä sanojen esiintymisen suurissa nimettyjen yksiköiden hakemistoissa, sisällyttämisen. Kuten myöhemmin osoitamme, nämä lisäominaisuudet alensivat CEER-arvoa merkittävästi.

**Tulos**

mitä pintatason ominaisuuksia käytetään?

**Esimerkki 5.479**

Analyyttisen approksimaation sijasta AEM käyttää diskriminaattoriverkkoa erottamaan latenttien tapahtumien perusteella rekonstruoidut asiakirjat ja alkuperäiset syöttöasiakirjat toisistaan. Tämä lähinnä auttaa generaattoria rakentamaan realistisemman asiakirjan Dirichlet-jakaumasta poimitusta satunnaiskohinasta. Neuroverkkojen joustavuuden ansiosta generaattori pystyy oppimaan monimutkaisia epälineaarisia jakaumia. Ja diskriminaattorin antama valvontasignaali auttaa generaattoria vangitsemaan tapahtumiin liittyvät mallit.

**Tulos**

Miten tämä malli voittaa oletuksen, jonka mukaan kaikki asiakirjan sanat syntyvät yhdestä tapahtumasta?

**Esimerkki 5.480**

Taulukosta TABREF46 käy ilmi, että Open-mallimme f1-tulos on yli 3 pistettä korkeampi kuin uusimman tekniikan tason tulos, ja RelAwe ja DepPath&RelPath ovat parhaita sekä suljetuissa että avoimissa asetuksissa.

**Tulos**

Kuinka suuri on parannus vanhaan huipputason suorituskykyyn verrattuna CoNLL-2009-tietokannassa?

**Esimerkki 5.481**

Aluksi kokeilimme graafipohjaista yhteisön havaitsemisalgoritmia, joka maksimoi klusterien modulaarisuuden BIBREF20 , mutta emme löytäneet yksinkertaista tapaa rajoittaa tätä menetelmää tuottamaan tietty määrä samankokoisia klustereita. Kaikkien mahdollisten klusterimääritysten luetteleminen raa'alla voimalla on hankalaa, koska hakuavaruus on suuri ( INLINEFORM0 mahdollista määritystä). Kehitämme yksinkertaisen klusterointialgoritmin, jolla tätä prosessia voidaan lähentää. Aluksi aloitamme satunnaisilla klusterijaoilla ja määrittelemme klusterin vahvuuden ryhmän sisäisen euklidisen etäisyyden ja ryhmän välisen euklidisen etäisyyden suhteelliseksi erotukseksi. Sitten ehdotamme iteratiivisesti satunnaisia jäsenyyksien vaihtoja ja hyväksymme nämä ehdotukset vain, kun klusterin vahvuus kasvaa, kunnes konvergenssi on saavutettu. Arvioidaksemme laskennallisesti johdettujen klusterien laatua Calvinon klustereihin verrattuna mittaamme klusterien puhtautta BIBREF21 : kun on annettu joukko ennustettuja klustereita INLINEFORM1 ja todellisia klustereita INLINEFORM2, jotka molemmat jakavat joukon INLINEFORM3 datapisteitä, INLINEFORM4

**Tulos**

Mitä klusterointimenetelmää he käyttävät kaupunkikuvausten klusterointiin?

**Esimerkki 5.482**

Tässä työssä tarkastelemme monia erilaisia ominaisuuksia ja raportoimme ennustustuloksia 62 demografisesta ominaisuudesta.

**Tulos**

Kuinka monta demografista ominaisuutta he yrittävät ennustaa?

**Esimerkki 5.483**

Tämän jakson päätteeksi todettakoon, että mallimme korjaa luotettavasti kielioppi-, oikeinkirjoitus- ja sanajärjestysvirheet, kun taas leksikaalisten valintavirheiden ja syötteen tarpeettoman parafrasoinnin tulokset ovat vaihtelevampia.

**Tulos**

Mitä virhetyyppejä varten heidän mallinsa on luotettavampi?

**Esimerkki 5.484**

Aiempien töiden BIBREF13, BIBREF12, BIBREF10, BIBREF17, BIBREF9 ja BIBREF7 mukaisesti valitsemme SemCor3.0-korpuksen harjoituskorpukseksi, joka on laajin manuaalisesti annotoitu korpus, johon on lisätty WSD:n WordNet-merkitys.

**Tulos**

Kuvastaako SemCor3.0 englanninkielisiä tietoja yleisesti?

**Esimerkki 5.485**

BIBREF3-perusmalli on toteutettu rekursiiviseen neuroverkkoon perustuvalla koodaaja-dekooderi-kehyksellä.

**Tulos**

Vertailevatko ne Noraset et al. 2017:ää?

**Esimerkki 5.486**

Käytetyt luokittelijat ovat tukivektorikone (SVM), logistinen regressio (Log.Reg), satunnaismetsä (RF) ja gradienttihäviö (XGB).

**Tulos**

Mitä klassisia koneoppimisalgoritmeja käytetään?

**Esimerkki 5.487**

INLINEFORM4 -vastaustyyppejä kuvaamaan on määritelty viisi ominaisuutta, jotka määrittelevät tiettyjä kliinisesti merkittäviä yksityiskohtia: (1) aika, jonka potilas on kokenut oireen, (2) toiminnot, jotka laukaisevat oireen (esiintyminen tai paheneminen), (3) oireen vakavuusaste, (4) oireen esiintymistiheys ja (5) oireen sijainti. Kullekin oireelle/ominaisuudelle voidaan antaa erilaisia kielellisiä ilmaisuja, jotka määritellään entiteeteiksi. Huomaa, että jos kysyttyä oiretta tai ominaisuutta ei mainita vuoropuhelussa, perustotuuden tuloste on "No Answer", kuten BIBREF:ssä6 .

**Tulos**

Mitä merkintöjä he luovat tietokokonaisuuteensa?

**Esimerkki 5.488**

Toiseksi tekstit käyvät läpi annotaatiotyökalujen kaskadin, joka rikastuttaa niitä seuraavilla tiedoilla: morfosyntaktiset tulkinnat (tagien joukot), käyttäen Morfeusz 0.82 BIBREF25 , taggaus (todennäköisimmän tulkinnan valinta), käyttäen transformaatioon perustuvaa oppivaa taggeria, PANTERA 0.9.1 BIBREF26 ,Syntaktiset ryhmät (mahdollisesti sisäkkäiset), joilla on syntaktiset ja semanttiset päät, käyttäen sääntöpohjaista matalaa jäsentäjää Spejd 1.3.7 BIBREF27 ja puolankielistä kielioppia, mukaan lukien parannettu versio BIBREF28 :n tekemistä muutoksista, joka mahdollistaa nimellisten syntaktisten ryhmien lemmatisoinnin,Nimetyt entiteetit, käyttäen kahta saatavilla olevaa työkalua: NERF 0.1 BIBREF29 ja Liner2 2.3 BIBREF30 .

**Tulos**

Miten RAFAELin tiedot on merkitty?

**Esimerkki 5.489**

Tämän puutteen korjaamiseksi ROUGE:ssa ehdotamme uutta arviointimittaria: Critical Information Completeness (CIC). Muodollisesti CIC on ehdokasyhteenvedon ja viiteyhteenvedon välisten semanttisten tietojen muistaminen. CIC määritellään seuraavasti: missä $V$ tarkoittaa vertailun tiivistelmän delexikalisoitujen arvojen joukkoa, $Count\_{match}(v)$ on niiden arvojen lukumäärä, jotka esiintyvät yhdessä ehdokastiivistelmässä ja vertailun tiivistelmässä, ja $m$ on arvojen lukumäärä joukossa $V$. Kokeiluissamme CIC lasketaan aritmeettisena keskiarvona kaikista dialogialueista kokonaissuorituskyvyn säilyttämiseksi.CIC on sopiva täydentävä mittari ROUGE:lle, koska se ottaa huomioon tärkeimmän tiedon kullakin dialogialueella. CIC:tä voidaan soveltaa mihin tahansa tiivistämistehtävään, jossa on ennalta määriteltyjä olennaisia kokonaisuuksia.

**Tulos**

Miten uudessa arviointimittarissa otetaan huomioon kriittiset informatiiviset yksiköt?

**Esimerkki 5.490**

Arvioimme huomiomuunnoksia kolmella kieliparilla. Käytämme IWSLT 2014 -korpusta De-En:n osalta, KFTT-korpusta Ja-En BIBREF19:n osalta ja WMT 2016 -datakokonaisuutta Ro-En:n osalta.

**Tulos**

Mitä kielipareja tässä tutkielmassa tarkastellaan?

**Esimerkki 5.491**

EIN:n lisäksi loimme mallin (Emotion-based Model), joka käyttää vain tunnepiirteitä, ja vertasimme sitä kahteen perusmalliin. Tavoitteenamme on tutkia, voidaanko tunnepiirteiden avulla itsenäisesti havaita vääriä uutisia. Tämän mallin kaksi peruslinjaa ovat enemmistöluokan peruslinja (MC) ja satunnaisvalinnan peruslinja (RAN).

**Tulos**

Mikä on lähtötaso?

**Esimerkki 5.492**

Sisäinen tietokokonaisuutemme sisältää manuaalisesti annotoituja RE-tietoja 6 kielestä: kielet: englanti, saksa, espanja, italia, japani ja portugali. ACE05-tietokanta sisältää manuaalisesti annotoitua RE-dataa kolmelle kielelle: Englannin, arabian ja kiinan kielillä. Se määrittelee 7 oliotyyppiä (henkilö, organisaatio, geopoliittinen yksikkö, sijainti, laitos, ase, ajoneuvo) ja 6 olioiden välistä relaatiotyyppiä (agentti-artefakti, yleinen yhteys, ORG-yhteys, osa-kokonaisuus, henkilö-sosiaalinen, fyysinen).

**Tulos**

Mitä kieliä he kokeilevat?

**Esimerkki 5.493**

Nämä 16 luokkaa perustuvat OntoNotes5 -korpukseen BIBREF7 sekä ACE (Automatic Content Extraction) English Annotation Guidelines for Entities Version 6.6 2008.06.13 BIBREF8.

**Tulos**

Miten he määrittelivät eri luokat?

**Esimerkki 5.494**

Jokaisen diagnoosin määrittelevän twiitin osalta haemme vastaavan Twitter-käyttäjän aikajanan Twitter user\_timeline API-päätepisteen avulla. Tämän jälkeen poistamme kaikki muut kuin englanninkieliset twiitit (Twitter API:n koneellisesti havaitsema "lang"-kenttä), kaikki uudelleentwiittaukset ja twiitit, jotka sisältävät "diagnos\*" tai "depress\*", mutta eivät kelvollista diagnoosilausumaa. Tuloksena saatu Depressed-kohortti sisältää 1 207 henkilöä ja 1 759 644 twiittiä toukokuusta 2008 syyskuuhun 2018.

**Tulos**

Raportoidaanko tulokset vain englanninkielisistä tietokokonaisuuksista?

**Esimerkki 5.495**

Käsittelemme kestävyysongelmaa GE-FL BIBREF0 -menetelmällä, joka hyödyntää merkittyjä piirteitä ennakkotietona.

**Tulos**

Mitä taustatietoja he hyödyntävät?

**Esimerkki 5.496**

Artikkelitason merkintöjen keräämiseksi hyödynsimme yrityksessä käytössä olevaa alustaa, jota markkinatutkimusryhmä on käyttänyt kerätäkseen kyselyitä eri uutiskustantajien tilaajilta. Kysely toimii seuraavasti: Käyttäjälle esitetään ensin joukko valittuja sivuja (yleensä 4 sivua ja noin 20 artikkelia) edellisen päivän painetusta lehdestä. Käyttäjä voi valita joka kerta artikkelin, jonka hän on lukenut, ja vastata siihen liittyviin kysymyksiin. Lisäsimme olemassa olevaan kyselyyn kolme kysymystä, joissa kysyttiin puolueellisuuden tasoa, puolueellisuuden napaisuutta ja sitä, mitä pro- tai anti-kokonaisuuksia artikkeli esittää. Kysyimme myös käyttäjän poliittista kantaa. Täydellinen kysely löytyy liitteistä.

**Tulos**

Hankkivatko he kommentit joukkorahoituksella?

**Esimerkki 5.497**

Mittasin ensin upotusten tehokkuutta skip-grammin koulutustehtävässä, jossa ennustettiin kontekstisanoja INLINEFORM0 annettujen tulosanojen INLINEFORM1 perusteella. Tämä tehtävä mittaa menetelmien suorituskykyä ennakoivassa kielimallinnuksessa. Käytin neljää yhteiskuntapoliittisesti, tieteellisesti ja kirjallisesti kiinnostavaa tietokokonaisuutta: NIPS-artikkeleiden korpus vuosilta 1987-1999 ( INLINEFORM2 miljoonaa), Yhdysvaltain presidentin puheet vuosina 1790-2015 ( INLINEFORM3 ), Shakespearen koko teos ( INLINEFORM4 ; tämä versio ei sisältänyt sonetteja) ja mustien oppineiden ja aktivistien W.E.B. Du Bois'n kirjoitukset, sellaisina kuin ne on digitoitu Project Gutenbergin toimesta ( INLINEFORM5 ). Kunkin aineiston osalta pidin 10 000 INLINEFORM6 -paria tasaisesti satunnaisesti, kun INLINEFORM7 , ja pyrin ennustamaan INLINEFORM8 -parin INLINEFORM9 (ja valinnaisesti INLINEFORM10 ) perusteella. Koska luokkia on suuri määrä, käsittelen tätä luokitusongelmana ja ilmoitan keskimääräisen vastavuoroisen sijoituksen. Kokeet toistettiin ja niiden keskiarvo laskettiin 5 train/test-jakojen aikana.

**Tulos**

Mikä on MRR?

**Esimerkki 5.498**

Olemme arvioineet mallejamme F1-pistemäärän perusteella, joka on tarkkuuden ja palautuksen harmoninen keskiarvo. Olemme suorittaneet kymmenen kertaa kokeen kullekin mallille ja tarkastelleet F1-pisteiden keskiarvoa. Tulokset on esitetty taulukossa TABREF11. Kun otetaan huomioon F1-makro, mallit, jotka sisältävät monihuomiointimekanismin, ovat muita parempia, ja erityisesti malli, jossa on projisoitu kerros, on suorituskyvyltään paras. Kolmessa malliparissa neljästä malliparista Projected Layer -mallin sisältävä malli saavutti paremman suorituskyvyn, joten useimmissa tapauksissa Projected Layer -mallin lisääminen paransi suorituskykyä merkittävästi.

**Tulos**

Millä muunnelmalla saadaan parhaat tulokset tässä tietokokonaisuudessa?

**Esimerkki 5.499**

Laajennamme ensin aiempaa työtä ja yleisiä dialogitoimintataksonomioita, kehitämme hienojakoisen joukon asiakaspalvelun dialogitoimintoja ja suoritamme järjestelmällisen käyttäjätutkimuksen näiden toimien tunnistamiseksi 800 keskustelun aineistosta, joka on peräisin neljältä Twitterin asiakaspalvelutililtä (eli neljältä eri yritykseltä televiestintä-, elektroniikka- ja vakuutusalalta).

**Tulos**

Mitä Twitterin asiakaspalvelualoja tutkitaan?

**Esimerkki 5.500**

Perusjärjestelmämme (Baseline\_1850K) on otettu BIBREF-tietokannasta13 . Se koostuu DNN:stä, joka on koulutettu ennustamaan avainsanojen sisällä olevia alasanoja. DNN:n syötteenä on sekvenssi, jossa on INLINEFORM0-kehyksiä vasemmanpuoleisesta ja INLINEFORM1-kehyksiä oikeanpuoleisesta kontekstista; kummassakin on INLINEFORM2 . Topologia koostuu 1-D-konvoluutiokerroksesta, jossa on 92 suodatinta (muodoltaan 8x8 ja askelpituudeltaan 8x8), ja sen jälkeen kolmesta täysin yhdistetystä kerroksesta, joissa kussakin on 512 solmua ja oikaistu lineaarinen yksikköaktivointi. Lopullinen softmax-ulostulo ennustaa 7 osasanatavoitetta, jotka on saatu samasta pakotetusta kohdistusprosessista, joka on kuvattu SECREF5 -ohjelmassa. Tämä johtaa siihen, että perus-DNN sisältää 1,7 miljoonaa parametria ja suorittaa 1,8 miljoonaa kerrannais-kertymäoperaatiota päätelmää kohden (joka 30 ms:n äänivirran aikana). Avainsanojen havaitsemisen pistemäärä, joka on 0 ja 1 välillä, lasketaan tasoittamalla ensin posterioriset arvot keskiarvona 100 edellisen kehyksen liukuvassa ikkunassa suhteessa nykyiseen INLINEFORM3 ; pistemäärä määritellään sitten liukuvassa ikkunassa olevien tasoitettujen posterioristen arvojen suurimpana tulona, kuten alun perin ehdotettiin BIBREF:ssä6 .

**Tulos**

Mitä aiempia lähestymistapoja on otettu huomioon?

**Esimerkki 5.501**

BLEU: BLEU (Bilingual Evaluation Understudy) BIBREF34 -mittaria, jota käytetään yleisesti konekäännöstehtävissä. BLEU-mittaristoa voidaan käyttää dialogin luomismallien arviointiin, kuten BIBREF5, BIBREF35. BLEU-mittari on sanojen päällekkäisyysmittari, joka laskee N-grammien samanaikaisen esiintymisen vertailussa ja generoidussa vastauksessa ja soveltaa myös lyhyyssakkoa, jolla pyritään rankaisemaan liian lyhyistä vastauksista, joita ei yleensä haluta tehtäväkeskeisissä chat-roboteissa. Laskemme BLEU-pisteet käyttämällä kaikkia järjestelmiemme tuottamia vastauksia: Kierroskohtainen tarkkuus mittaa järjestelmän tuottaman vastauksen ja kohdevastauksen samankaltaisuutta. Eric ja Manning eric2017copy käyttivät tätä mittaria arvioidessaan järjestelmiään, joissa he pitivät vastausta oikeana, jos kaikki järjestelmän tuottaman vastauksen merkit vastasivat kohdevastauksen vastaavaa merkkiä. Tämä metriikka on hieman ankara, ja tulokset voivat jäädä alhaisiksi, koska kaikkien generoidun vastauksen sisältämien merkkien on oltava täsmälleen samassa kohdassa kuin kohdevastauksessa.Per-Dialogue Accuracy: Laskemme dialogikohtaisen tarkkuuden, jota käytetään BIBREF8:ssa ja BIBREF5:ssä. Tätä mittaria varten tarkastelemme kaikkia järjestelmän tuottamia vastauksia ja vertaamme niitä kohdevastauksiin. Dialogin katsotaan olevan oikea, jos kaikki järjestelmän tuottamien vastausten käänteet vastaavat kohdevastausten vastauksia. Huomaa, että tämä on hyvin tiukka metriikka, jossa kaikkien dialogin lausumien on oltava samoja kuin kohdevastausten ja oikeassa järjestyksessä. F1-Entity Score: Tehtäväsuuntautuneissa askareissa käytetyissä tietokannoissa on joukko entiteettejä, jotka edustavat käyttäjän mieltymyksiä. Esimerkiksi ravintola-alalla chatbottien yleisiä entiteettejä ovat ateria, ravintolan nimi, päivämäärä, kellonaika ja henkilömäärä (nämä ovat yleensä pakollisia entiteettejä, jotka ovat ratkaisevia varausten tekemisen kannalta, mutta voi olla myös valinnaisia entiteettejä, kuten sijainti tai luokitus). Jokaisessa kohdevastauksessa on joukko entiteettejä, joita järjestelmä kysyy tai ilmoittaa käyttäjälle. Mallimme on kyettävä erottamaan nämä erityiset kokonaisuudet ja sisällyttämään ne tuotettuun vastaukseen. Mallien arviointiin voidaan käyttää nimettyjen entiteettien tunnistamisen arviointimittareita BIBREF36. F1-pistemäärä on yleisimmin käytetty metriikka, jota käytetään nimitystunnistusmallien arvioinnissa ja joka on mallin tarkkuuden ja palautuksen harmoninen keskiarvo. Laskemme tämän mittarin mikrokeskiarvona kaikista järjestelmän tuottamista vastauksista.

**Tulos**

Suoritetaanko inhimillinen arviointi?

**Esimerkki 5.502**

Yhä useammat todisteet osoittavat, että nykyaikaiset mallit oppivat hyödyntämään tietokokonaisuuksien virheellisiä tilastollisia malleja BIBREF12, BIBREF13, BIBREF14, BIBREF15, BIBREF16, BIBREF17 sen sijaan, että ne oppisivat merkityksiä joustavasti ja yleistettävällä tavalla, kuten ihmiset tekevät. Tämän vuoksi ihmisen annotoijat - kokeneet NLP-tutkijat tai ei-asiantuntijat - voivat helposti rakentaa esimerkkejä, jotka paljastavat mallin haurauden.

**Tulos**

Mitä heikkouksia ei-asiantuntija-annotoijat ovat havainneet nykyisissä uusimmissa NLI-malleissa?

**Esimerkki 5.503**

Taulukoiden TABREF17 , TABREF18 ja TABREF19 tulosten analyysi osoittaa, että 12 parasta tulosta 15:stä saatiin käyttämällä uusia sanojen upotuksia. Taulukossa TABREF20 esitetyt arviointitulokset (taulukossa TABREF19 valitut parhaat upotusmallit ) osoittavat, että paras sanojen upotusten ryhmä on EC. Korkein F1-pistemäärä saatiin EC1-mallille, joka rakennettiin käyttämällä binääristä FastText Skip-gram -menetelmää, jossa hyödynnetään alasanatietoa, vektorin ulottuvuus on 300 ja negatiivinen näytteenotto on 10. Mallin kyky tarjota vektoriedustus tuntemattomille sanoille näyttää olevan tärkein.

**Tulos**

Mitä johtopäätöksiä näistä kokeista tehdään?

**Esimerkki 5.504**

Teimme kokeita neljällä kiinalaisella NLP-tehtävällä, joihin kuuluvat tunteiden luokittelu (EC), nimettyjen entiteettien tunnistus (NER), lauseparien yhdistäminen (SPM) ja luonnollisen kielen päättely (NLI).

**Tulos**

Millä vertailuarvoilla he tekivät kokeita?

**Esimerkki 5.505**

Esimerkkien tarkoituksena on tarjota asianmukainen konteksti. Ymmärtääksemme kontekstia paremmin analysoimme kokeellisesti esimerkin avulla tuotettuja kysymyksiä. Havaitsimme, että tukeva esimerkki voi todellakin tunnistaa merkitykselliset tunnisteet (lehmät kuvassa FIGREF3 ) kysymysten luomiseksi.Parannamme esimerkkien käyttöä käyttämällä triplettiverkkoa. Tämä verkko varmistaa, että kuvan ja kuvatekstin yhteinen upotus tukevan esimerkin osalta on lähempänä kohdekuvan kuvatekstin upotusta ja päinvastoin.

**Tulos**

Miten kirjoittajat määrittelevät esimerkit?

**Esimerkki 5.506**

Kokeidemme avulla teemme hienovaraisia huomioita, jotka liittyvät seuraaviin seikkoihin: (a) ominaisuuksiemme suorituskykyyn, (b) siihen, miten lähestymistapamme vertautuu ihmisen kykyyn havaita juopuneita tekstejä, (c) kaikkein erottelevimpiin tyylillisiin ominaisuuksiin ja (d) virheanalyysiin, joka antaa viitteitä tulevaan työhön.

**Tulos**

Mainitsevatko kirjoittajat tutkimuksessaan esiintyviä häiriötekijöitä?

**Esimerkki 5.507**

FastTextin, GloVen ja word2vecin arviointi osoittaa, että muihin sanojen esittämisen oppimisalgoritmeihin verrattuna FastText toimii parhaiten.

**Tulos**

Mitä käytät sanojen/sub-sanojen upotusten laskemiseen?

**Esimerkki 5.508**

GENIA Corpus BIBREF3 sisältää biolääketieteellisiä tiivistelmiä PubMed-tietokannasta. Käytämme GENIAn teknisten termien annotaatioita 3.02, jotka kattavat molekyylibiologian kannalta kiinnostavien yksiköiden, kuten proteiinien, geenien ja solujen, kielelliset ilmaisut. CoNLL2003 BIBREF14 on vakiomuotoinen NER-tietokanta, joka perustuu Reutersin RCV-1-uutiskorpukseen. Se kattaa nimettyjä entiteettejä, jotka ovat tyypiltään henkilö, paikka, organisaatio ja muu.Yleisen annotaation suorituskyvyn testaamiseen käytämme CoNLL2003-testA:ta ja 50 asiakirjan jakoa GENIAsta.

**Tulos**

mitä vakiotietoaineistoa käytettiin?

**Esimerkki 5.509**

Artikkelin viimeisessä osassa pohditaan, miten mielipiteen dynamiikan teknisestä havainnoinnista saadut havainnot voivat olla hyödyksi käsitteellisessä mallintamisessa ja lähestymistavoissa, jotka koskevat mielipiteiden välittämistä verkossa. Näin ollen asiakirjassa tuodaan esiin ja arvioidaan kriittisesti perustavaa laatua oleva käsitteellinen harppaus koneohjatusta havainnoinnista keskustelun fasilitointiin ja siihen puuttumiseen.

**Tulos**

Raportoidaanko asiakirjassa samoihin tehtäviin sovellettujen aiempien mallien tulokset?

**Esimerkki 5.510**

noudatamme Lapata2005Automaattista mitata koherenssia lauseiden samankaltaisuutena.

**Tulos**

Arvioivatko kirjoittajat järjestelmänsä tuotoksen johdonmukaisuutta?

**Esimerkki 5.511**

Koulutustavoitteena on minimoida vastaamattoman kysymyksen $\tilde{q}$ negatiivinen todennäköisyys, kun vastauskysymys $q$ ja sitä vastaava kohta $p$, joka sisältää vastauksen $a$ : L=-(q,q,p,a)DP(q|q,p,a;) missä $\mathcal {D}$ on koulutuskorpus ja $\theta $ tarkoittaa kaikkia parametreja. Sekvenssistä sekvenssiin ja parista sekvenssiin -mallit koulutetaan samalla tavoitteella.

**Tulos**

Mikä on heidän parisekvenssimallinsa koulutustavoite?

**Esimerkki 5.512**

Yllä oleva pseudokoodi on riippumaton pirstoutumis- ja ympäristötoimintojen valinnasta; tehtäväkohtaiset valinnat kuvataan yksityiskohtaisemmin kunkin kokeen osalta jäljempänä. keskustelu.

**Tulos**

Mitä kieliä he testaavat?

**Esimerkki 5.513**

Lopuksi arvioimme lähestymistapojamme 9:llä yleisesti käytetyllä tekstiluokitustietokannalla. Arvioimme menetelmiämme useilla yleisesti käytetyillä tietokokonaisuuksilla, joiden aiheet vaihtelevat tunnetiedoista, verkkosivuista, tieteestä lääketieteeseen ja terveydenhuoltoon.

**Tulos**

Mitä NLP-tehtäviä he pitävät mielessä?

**Esimerkki 5.514**

Perinteiset tekstistä puheeksi (TTS) -järjestelmät koostuvat monimutkaisista putkistoista BIBREF0 , joihin kuuluu usein akustisia frontendeja, kestomalli, akustinen ennustemalli ja vokooderimallit TTS-mallimme kouluttamiseen käytettiin avoimen lähdekoodin LJSpeech-tietoaineistoa. Mallimme arkkitehtuurissa käytetään RNN-pohjaista Seq2Seq-mallia melaspektrogrammin tuottamiseen tekstistä.

**Tulos**

Millä tietokokonaisuudella (tietokokonaisuuksilla) ne arvioidaan?

**Esimerkki 5.515**

BIBREF11-funktiota käytetään laskemaan tekstien ja kysymysten välinen samankaltaisuuspistemäärä seuraavasti: INLINEFORM2

**Tulos**

Käyttävätkö he huomiota?

**Esimerkki 5.516**

Aineiston kehittämiseksi poimimme 200 lausetta Irakin Kurdistanin alueella sijaitsevan peruskoulun ensimmäisestä kolmeen luokkaan kuuluvista sorani-kurdinkielisistä kirjoista. Loimme satunnaisesti 2000 lausetta poimituista lauseista.

**Tulos**

Mikä on tietokokonaisuuden koko?

**Tulos**

Miten aineisto kerättiin?

**Esimerkki 5.517**

Stanford - Twitter Sentiment Corpus (STS Corpus): STS-korpus sisältää 1 600 000 harjoitustwiittiä, jotka on kerätty BIBREF0-tietokannasta. BIBREF0 rakensi manuaalisesti testijoukon, jossa oli 177 negatiivista ja 182 positiivista twiittiä. Stanfordin testijoukko on pieni. Sitä on kuitenkin käytetty laajalti eri arviointitehtävissä BIBREF0 BIBREF5 BIBREF13 .Sanders - Twitter Sentiment Corpus: Tämä tietokokonaisuus koostuu käsin luokitelluista twiiteistä, jotka on kerätty hakusanojen avulla: INLINEFORM0 , #google, #microsoft ja #twitter. Rakennamme tietokokonaisuuden BIBREF14 binääriluokittelua varten.Terveydenhuollon uudistus (HCR): Tämä tietokokonaisuus rakennettiin indeksoimalla twiitit, jotka sisälsivät hashtagin #hcr BIBREF15 . Tehtävänä on ennustaa positiiviset/negatiiviset twiitit BIBREF14 . Taulukossa IV esitetään tunteisiin perustuvan luokittelumallimme tulokset verrattuna muihin malleihin. Vertaamme mallimme suorituskykyä BIBREF0 BIBREF5:n lähestymistapoihin STS-korpuksella. Sanders- ja HCR-tietokokonaisuuksien osalta vertaamme tuloksia BIBREF14-malliin, jossa käytettiin useiden perusluokittelijoiden (ENS) yhdistelmää, kuten NB, Random Forest (RF), SVM ja Logistic Regression (LR).

**Tulos**

Raportoidaanko tulokset vain englanninkielisistä tietokokonaisuuksista?

**Esimerkki 5.518**

Jotta voidaan valita sopivimmat lauseet suuresta määrästä merkitsemättömiä korpuksia, ehdotamme aktiivisen oppimisen näytteenottostrategiana informaatioentropiaan ja neuroverkkoon perustuvaa pisteytysmallia, joka on Cain ja Zhaon BIBREF32 innoittama. Segmentoidun lauseen pistemäärä lasketaan seuraavasti. Ensin segmentoitu lause kartoitetaan ehdollisten sanojen upotusten sarjaksi. Tämän jälkeen pisteytysmalli ottaa sanojen upotussarjan syötteenä ja pisteyttää jokaisen yksittäisen ehdokkaan sanan kahdesta näkökulmasta: (1) mahdollisuus, että ehdokkaan sanaa itsessään voidaan pitää laillisena sanana; (2) sen linkin järkevyys, että ehdokkaan sana seuraa suoraan aiempaa segmentointihistoriaa. Kuvassa FIGREF10 havainnollistetaan koko pisteytysmalli. Merkkien upotusten yli käytetään porttiverkkoa luomaan hajautettuja esityksiä ehdokassanoista, jotka lähetetään LSTM-malliin.

**Tulos**

Miten pisteytysmalli toimii?

**Esimerkki 5.519**

D2V-malli on arvioitu 80 kertaa "huonosti relevantiksi", kun taas pmra palautti vain 24 kertaa huonosti relevantteja asiakirjoja.

**Tulos**

Kuinka paremmat tulokset pmra-algoritmi kuin Doc2Vec antaa ihmisen arvioinnissa?

**Esimerkki 5.520**

Ottaaksemme huomioon sanojen ja merkkien sarjat, jotka saattavat sisältää hyödyllistä tietoa, poimimme ominaisuuksiksi sanojen unigrammit, bigrammit ja trigrammit.

**Tulos**

Mitä ominaisuuksia he käyttävät Twitter-viestien mallintamiseen?

**Esimerkki 5.521**

Sen vuoksi otamme käyttöön differentiaalisen yksityisyyden suojan (DP) menetelmän, jotta ennustemallin kestävyyttä voidaan parantaa. DP on järjestelmä, jossa tietoa tietokokonaisuudesta jaetaan julkisesti kuvaamalla ryhmien malleja tietokokonaisuuden sisällä ja salaamalla samalla tietoa tietokokonaisuuteen kuuluvista yksilöistä. DP voidaan saavuttaa, jos olemme valmiita lisäämään tulokseen satunnaista kohinaa. Esimerkiksi sen sijaan, että ilmoitettaisiin vain summa, voidaan lisätä kohinaa Laplace- tai gaussin jakaumasta, jolloin saadaan tulos, joka ei ole aivan tarkka ja joka peittää minkä tahansa rivin sisällön. Se edellyttää intuitiivisesti, että mekanismi, joka tuottaa tietoa taustalla olevasta tietokokonaisuudesta, on kestävä yhden näytteen mille tahansa muutokselle, mikä suojaa yksityisyyttä. Mekanismi ${f}$ on satunnaisfunktio, joka ottaa syötteenä tietokokonaisuuden $\mathcal {N}$ ja tuottaa satunnaismuuttujan ${f}(\mathcal {N})$. Jos esimerkiksi oletetaan, että $\mathcal {N}$ on uutisartikkelitietokanta, niin funktio, joka antaa $\mathcal {N}$:n artikkelien yhdistelmäpisteet ja normaalijakauman kohinan, on mekanismi [7].

**Tulos**

Miten eriytetty yksityisyysmekanismi toimii?

**Esimerkki 5.522**

Reddit-tiedot kerättiin Redditin julkisen API:n avulla, ja niihin kerättiin viimeisimmät vitsit. Aina kun scraper ajettiin, se päivitti myös aiemmin kerättyjen vitsien upvote-pisteet. Tämä tiedonkeruu tapahtui tunnin välein maalis- ja huhtikuun 2019 aikana. Koska tiedot oli jo jaettu Redditistä runko- ja iskulauseosioihin, loimme erilliset tietokokonaisuudet, jotka sisälsivät yksinomaan vitsin rungon ja yksinomaan vitsin iskulauseen. Lisäksi loimme tietokokonaisuuden, joka yhdisti vitsin rungon ja iskulauseen yhdessä." Joitakin esimerkkivitsejä on esitetty taulukossa 1, yllä. Vitsien pistemäärien jakauma vaihtelee rajusti, vaihdellen 0:sta 136 354 upvoteen. Huomasimme, että 0-200 upvote-arvon välillä on suuri hyppäys 200:n ja siitä ylöspäin, ja vain 6 prosenttia vitseistä sai 200-20 000 pistettä. Käytimme tätä luonnollista jakoa raja-arvona päättäessämme, mikä katsotaan hauskaksi vitsiksi, ja saimme tulokseksi 13884 ei-hauskaa vitsiä ja 2025 hauskaa vitsiä.

**Tulos**

Miten he arvioivat, onko vitsi humoristinen vai ei?

**Esimerkki 5.523**

(2) LEM:stä ja DPEMM:stä poiketen AEM käyttää generaattoriverkkoa tapahtumiin liittyvien mallien tallentamiseen ja pystyy louhimaan tapahtumia eri tekstilähteistä (lyhyistä ja pitkistä). Lisäksi toisin kuin LEM:ssä ja DPEMM:ssä käytetyt perinteiset päättelymenetelmät, kuten Gibbsin näytteenotto, AEM pystyy louhimaan tapahtumia tehokkaammin CUDA-kiihdytyksen ansiosta;

**Tulos**

Mitä vaihtoehtoa Gibbsin otannalle käytetään?

**Esimerkki 5.524**

Keräsimme Streaming API:n avulla kymmeneen yhdysvaltalaiseen valtavirtauutissivustoon eli BIBREF18:ssa kuvattuihin luotettavimpiin lähteisiin liittyviä twiittejä, ja viittasimme Hoaxy API:n BIBREF16:een niiden twiittien osalta, jotka sisälsivät linkkejä yli 100 yhdysvaltalaiseen disinformaatiopisteeseen. Suodatimme pois alle 50 twiittiin liittyvät artikkelit. Tuloksena saatu tietokokonaisuus sisältää yhteensä $\sim 1,7 miljoonaa twiittiä valtavirtauutisia varten, jotka on kerätty kolmen viikon aikana (25. helmikuuta 2019 - 18. maaliskuuta 2019) ja jotka liittyvät 6 978 uutisartikkeliin, ja $\sim 1,6 miljoonaa twiittiä disinformaatiota varten, jotka on kerätty kolmen kuukauden aikana (1. tammikuuta 2019 - 18. maaliskuuta 2019) näiden kahden luokan tasapainon vuoksi, ja niihin sisältyy 5 775 erillistä artikkelia. Diffuusion sensurointivaikutukset BIBREF14 otettiin oikein huomioon molemmissa keruumenetelmissä. Kuvassa FIGREF4 esitetään artikkelien jakautuminen lähteen ja poliittisen puolueellisuuden mukaan molemmilla uutisalueilla. Italian skenaarion osalta keräsimme ensin twiittejä Streaming API:n avulla kolmen viikon aikana (19. huhtikuuta 2019 - 5. toukokuuta 2019) ja suodatimme ne twiitit, jotka sisälsivät URL-osoitteita, jotka viittasivat italialaisten virallisten sanomalehtien verkkosivuille, kuten BIBREF22 kuvailee; nämä vastaavat Italian sanomalehtien levikkiä tarkistavan yhdistyksen (Accertamenti Diffusione Stampa) toimittamaa luetteloa. Sen sijaan viittasimme BIBREF23:n tarjoamaan tietokokonaisuuteen saadaksemme joukon twiittejä, jotka on kerätty jatkuvasti tammikuusta 2019 lähtien käyttämällä samaa Twitter-päätepistettä ja jotka sisältävät URL-osoitteita yli 60 italialaiselle disinformaatiosivustolle. Tasapainoisten luokkien saamiseksi (5. huhtikuuta 2019 - 5. toukokuuta 2019) säilytimme tiedot, jotka oli kerätty pidemmällä ajanjaksolla valtavirtauutisten osalta. Molemmissa tapauksissa suodatimme pois artikkelit, joissa oli alle 50 twiittiä; kaiken kaikkiaan tämä tietokokonaisuus sisältää $\sim $160k valtavirran twiittejä, jotka vastaavat 227 uutisartikkelia, ja $\sim $100k disinformaatioviestien twiittejä, jotka vastaavat 237 uutisartikkelia. Kuvassa FIGREF5 esitetään artikkelien jakautuminen eri lähteiden mukaan molemmilla uutisalueilla.

**Tulos**

Mitkä kaksi uutisaluetta ovat maakohtaisesti riippumattomia?

**Esimerkki 5.525**

Asiakirja $d$ voidaan määritellä termien (itsenäiset tekstikokonaisuudet asiakirjan sisällä, esimerkiksi sanat) rajallisena sarjana, nimittäin $d=(t\_1,t\_2,\dots ,t\_n)$.

**Tulos**

Millaisia esitystapoja ne käyttävät tekstidokumenteille?

**Esimerkki 5.526**

Toinen on ensimmäinen satunnainen siemen, joka on valittu luokittimelle, joka tuottaa tuloksen 0,8083. Vaikka se on parempi kuin NBSVM-ratkaisu, valitsemme parhaan validoinnin F1:n kokeilemistamme 20 siemenestä. Tämä tuotti lopullisen tuloksemme 0,8099. Paras mallimme saavutti kuvassa FIGREF27 esitetyllä validointijoukolla viisinkertaisen keskimääräisen F1-arvon 0,8254, mutta testijoukon F1-arvo oli 0,8099 - F1-arvon lasku oli 0,0155 todellista otoksen ulkopuolista dataa varten.

**Tulos**

Mitkä olivat heidän tuloksensa luokittelu- ja regressiotehtävissä?

**Esimerkki 5.527**

Yahoo! Answersista poimituilla termeillä on yleensä enemmän korreloivien termien lukumäärän perusteella yhteyttä uskontoon tai etnisyyteen liittyviin ominaisuuksiin kuin Twitteristä poimituilla termeillä. Kahden tietyn ominaisuuden (hinta ja buddhalainen) osalta Twitteristä saatujen korreloivien termien määrä on kuitenkin suurempi kuin Yahoo! Answersista saatujen termien määrä.

**Tulos**

Onko Twitterissä demografisten ominaisuuksien ja vastausten välillä enemmän korrelaatioita kuin Yahoo! Answersissa?

**Esimerkki 5.528**

Arvioinnissa on käytetty seuraavia eri aistiluetteloita: Watlink, automaattisesti rakennettu sanojen merkitysverkko. Se käyttää Watset[CWnolog, MCL]-menetelmällä BIBREF2 valvomattomasti indusoituja synsettejä ja semanttisia suhteita sellaisista sanakirjoista kuin Wiktionary, joihin viitataan Ustalov:17:dialogissa Joint INLINEFORM0 Exp INLINEFORM1 SWN.

**Tulos**

Mitä synsettien korpusta käytetään?

**Esimerkki 5.529**

Käytimme alustavissa kokeissa fastText- ja SVM BIBREF16 -menetelmää.

**Tulos**

Mitä luokittelumalleja käytettiin?

**Esimerkki 5.530**

Annotoitujen twiittien perusteella halusimme selvittää, miten loukkaava kielenkäyttö jakautuu: loukkaavan kielenkäytön tyypit, genret, joissa sitä käytetään, käytetyt murteet ja tällaista kieltä käyttävien käyttäjien sukupuoli. Kuten kuviosta käy ilmi, urheilussa ja politiikassa käytetään eniten loukkaavaa kieltä, mukaan lukien mautonta ja vihapuhetta. Murteiden osalta tarkasteltiin MSA:ta ja neljää päämurretta, nimittäin egyptiläistä (EGY), leventiiniläistä (LEV), maghrebilaista (MGR) ja gulfilaista (GLF). Kuvasta FIGREF14 käy ilmi, että 71 prosenttia mauttomista twiiteistä oli kirjoitettu EGY:llä ja seuraavaksi eniten GLF:llä, jonka osuus oli 13 prosenttia mauttomista twiiteistä. MSA:ta ei käytetty yhdessäkään mauttomassa twiitissä. Mitä tulee loukkaaviin twiitteihin yleensä, EGY:tä käytettiin 36 prosentissa ja GLF:ää 35 prosentissa loukkaavista twiiteistä. Toisin kuin mauttoman kielen tapauksessa, jossa MSA:ta ei ollut lainkaan, 15 prosenttia loukkaavista twiiteistä oli itse asiassa kirjoitettu MSA:lla. Vihapuheen osalta GLF ja EGY olivat jälleen hallitsevia, ja MSA:n osuus oli 21 prosenttia twiiteistä. Tämä on yhdenmukaista muiden kielten, kuten englannin ja italian, havaintojen kanssa, joissa mauton kieli liittyi useammin puhekieleen BIBREF24, BIBREF25. Sukupuolen osalta kuvio FIGREF15 osoittaa, että valtaosa loukkaavista twiiteistä, mukaan lukien mauttomuus ja vihapuhe, oli miesten kirjoittamia. Naispuolisten Twitter-käyttäjien osuus oli 14 prosenttia loukkaavista twiiteistä yleensä ja vastaavasti 6 prosenttia mauttomista twiiteistä ja 9 prosenttia vihapuheista. Kuviossa FIGREF16 esitetään vihapuheiden tyyppien yksityiskohtainen luokittelu, jossa kolmeen tärkeimpään kuuluvat poliittisen ideologian, alkuperän ja urheiluun kuulumisen perusteella loukkaavat ryhmät. Uskonnollista vihapuhetta esiintyi vain 15 prosentissa kaikista vihapuheviestien twiiteistä.

**Tulos**

Miten he analysoivat, mitkä aiheet, murteet ja sukupuoli liittyvät eniten twiitteihin?

**Esimerkki 5.531**

Käytimme seuraavia eri tunnemalleihin perustuvia affektiivisia resursseja: Emolex: se sisältää 14 182 sanaa, jotka liittyvät kahdeksaan ensisijaiseen tunteeseen Plutchikin mallin BIBREF10 , BIBREF11 perusteella.EmoSenticNet(EmoSN): se on rikastettu versio SenticNetistä BIBREF12 , joka sisältää 13 189 sanaa, jotka on merkitty kuuden Ekmanin perustunteen mukaan BIBREF13 , BIBREF14 .Dictionary of Affect in Language (DAL): se sisältää 8742 englanninkielistä sanaa, jotka on merkitty kolmella pisteytyksellä, jotka edustavat kolmea ulottuvuutta: Affective Norms for English Words (ANEW): sisältää 1 034 englanninkielistä sanaa BIBREF16 , jotka on luokiteltu Valence-Arousal-Dominance (VAD) -malliin perustuvilla luokituksilla BIBREF17 Linguistic Inquiry and Word Count (LIWC): tämä psykolingvistinen resurssi BIBREF18 sisältää 4 500 sanaa, jotka on jaettu 64 tunnekategoriaan, mukaan lukien positiiviset (PosEMO) ja negatiiviset (NegEMO).

**Tulos**

Millaisia affektiivisia piirteitä käytetään?

**Esimerkki 5.532**

Taulukossa TABREF19 ilmoitettu kokonaistarkkuuden kokonaispistemäärä koko ominaisuusjoukon avulla on 549.

**Tulos**

Kokeilevatko he tietokokonaisuutta?

**Esimerkki 5.533**

Tässä vaiheessa laaditaan vaatimuksesta kysely ja tehdään hakukoneelle (tässä kokeilemme Googlen ja Bingin käyttöä) hakuasiakirjojen hakemista varten. Sen sijaan, että hakukoneelta kysyttäisiin koko väitettä (koska väite on keskimäärin kaksi lausetta pitkä), luodaan lyhyempi kysely BIBREF0:ssa esitettyjen opetusten mukaisesti. Asetamme sanat paremmuusjärjestykseen tf-idf:n avulla. Laskemme idf-arvot vuoden 2015 Wikipedia-dumpista ja englanninkielisestä Gigawordista. BIBREF0 ehdotti, että hyvä tapa suorittaa laadukas haku on ottaa huomioon vain väitteen verbit, substantiivit ja adjektiivit; jätämme siis väitteestä pois kaikki sanat, jotka kuuluvat muihin sananosiin. Lisäksi väittämät sisältävät usein nimettyjä entiteettejä (esim. henkilöiden, paikkojen ja organisaatioiden nimiä), joten täydennämme alkuperäistä kyselyä kaikilla väittämän tekstissä olevilla nimetyillä entiteeteillä. Käytämme IBM:n AlchemyAPI:tä nimettyjen entiteettien tunnistamiseen. Lopulta luomme 5-10 merkin kyselyitä, jotka suoritamme hakukoneella. Sen jälkeen keräämme hakutulosten pätkät ja URL-osoitteet ja jätämme väliin kaikki tulokset, jotka viittaavat epäluotettavaksi katsottuun verkkotunnukseen.

**Tulos**

Miten mahdollisesti merkitykselliset tekstikatkelmat tunnistetaan?

**Esimerkki 5.534**

Toinen näistä kaavioista saatu oivallus on, että satunnainen yhteenvetomalli johti yli 50 prosentin tuloksiin kaikissa mittareissa, ja ilman dokumenttitietoisia ominaisuuksia malli saavuttaa vain pienen parannuksen satunnaiseen yhteenvetomalliin verrattuna.

**Tulos**

Onko uusi lähestymistapa testattu verrattuna uusimpaan tekniikkaan?

**Esimerkki 5.535**

Tämä malli koostuu kahdesta identtisestä RNN-kooderista, joilla ei ole yhteisiä parametreja, sekä tavallisesta RNN-dekooderista. Kutakin kohdelausetta varten käytetään kahta versiota lähdelauseesta: peräkkäistä (vakioversiota) ja linearisoitua jäsennystä (leksikalisoitua tai leksikalisoimatonta).

**Tulos**

Minkälaisia koodaajia käytetään jäsennellyssä lähdelauseessa?

**Esimerkki 5.536**

Aineistomme koostuu kahdesta aineistosta, joita käytetään automaattisen puheentunnistusjärjestelmän harjoitteluun ja arviointiin. Neljä suurta arviointikampanjaa on mahdollistanut laajan korpuksen luomisen ranskankielisestä lähetyspuheesta: ESTER1 BIBREF13, ESTER2 BIBREF14, ETAPE BIBREF15 ja REPERE BIBREF16.

**Tulos**

Mitä korpuksia tässä asiakirjassa analysoidaan?

**Esimerkki 5.537**

Lisäksi käytämme ominaisuuksina väitteen, parhaimman pistemäärän saaneen pätkän ja parhaimman pistemäärän saaneen lauseen kolmion upotuksia verkkosivulta. Laskemme nämä upotukset (i) tekstin sanojen upotusten keskiarvona ja (ii) käyttämällä LSTM-koodauksia, jotka koulutamme tehtävään osana syvää neuroverkkoa (NN). Käytämme myös tehtäväkohtaista väitteen upotusta yhdessä kaikkien edellä mainittujen sitä koskevien todisteiden kanssa, jotka tulevat NN:n viimeiseltä piilokerrokselta.

**Tulos**

Mitä algoritmia ja sulautusulottuvuuksia käytetään tehtäväkohtaisten sulautusten muodostamiseen?

**Esimerkki 5.538**

BIBREF21:ssä ja BIBREF22:ssa ehdotettu RKS-lähestymistapa kartoittaa tietovektorit avaruuteen, jossa lineaarinen erottelu on mahdollista. RKS-menetelmällä saadaan likimääräinen ydinfunktio eksplisiittisen kartoituksen avulla.

**Tulos**

Mikä on Random Kitchen Sink -lähestymistapa?

**Esimerkki 5.539**

BIBREF17: Lemming-malli on log-lineaarinen malli, joka suorittaa yhteisen morfologisen merkinnän ja lemmatisoinnin.

**Tulos**

Mitä muita kuin neurologisia perusarvoja käytettiin tehtävässä?

**Esimerkki 5.540**

Tarkastelimme UD1.2-korporaatioita seuraaville 16 kielelle: Englanti, bulgaria, englanti, espanja, indonesia, italia, kroatia, norja, persia, portugali, ranska, ruotsi, portugali, ranska, saksa, sloveeni, tanska, tšekki ja ruotsi.

**Tulos**

mitä kieliä tutkitaan?

**Esimerkki 5.541**

Yhdistämme vakiomuotoisia kielellisiä piirteitä, kuten POS (Part-Of-Speech) ja chunk tag, useisiin erityisesti klassista musiikkia varten kehitettyihin hakemistoihin sekä useisiin piirteisiin, jotka kuvaavat merkkien vasenta ja oikeaa kontekstia.

**Tulos**

Millaisia korpuspohjaisia ominaisuuksia otetaan huomioon?

**Esimerkki 5.542**

Lyhyesti sanottuna B4MSA:ssa käytetään ensin tekstimuunnoksia viesteihin, sitten muunnettu teksti esitetään vektoriavaruusmallissa (ks. SECREF13 ), ja lopuksi luokittelijana käytetään tukivektorikonetta (lineaarisella ytimellä).

**Tulos**

Mitkä ovat monikielisen kehyksen osatekijät?

**Esimerkki 5.543**

Tutkimuksessamme käytimme uudelleentwiittausten lukumäärää poimimaan ne, jotka muuttuivat virukseksi otoksessamme. Tähän osajoukkoon kuuluvat twiitit (jäljempänä "virustwiitit") ovat erilaisia ja liittyvät eri aiheisiin. Katsomme, että twiitti sisältää valeuutisia, jos sen teksti kuuluu johonkin seuraavista Rubin et al. BIBREF7 kuvaamista kategorioista (ks. seuraavassa jaksossa yksityiskohtaiset tiedot näistä kategorioista): vakava väärennös, laajamittaiset huijaukset, nimensä mukaisesti otetut vitsit, todellisten tosiseikkojen vinoutunut raportointi ja tarinat, joissa totuus on kiistanalainen. Asiantuntijan manuaalisesti merkitsemä tietokokonaisuus BIBREF8 on julkaistu julkisesti, ja se on tutkijoiden ja asianosaisten käytettävissä.

**Tulos**

Miten he määrittelivät valeuutistwiitit?

**Esimerkki 5.544**

Saavutetun parannuksen luotettava arviointi edellyttää määrällistä mittausta. Yksi tulkittavuuden mittaamiseen ehdotetuista menetelmistä on sanojen tunkeutumistesti BIBREF41 . Tämän menetelmän soveltaminen on kuitenkin kallista, koska se edellyttää useiden ihmisarvioijien arviointeja kunkin upotusulottuvuuden osalta. Tässä tutkimuksessa käytämme tulkittavuuden kvantifioimiseksi semanttiseen kategoriaan perustuvaa lähestymistapaa, joka perustuu BIBREF:ssä27 esiteltyyn menetelmä- ja kategoria-aineistoon (SEMCAT). Sovellamme erityisesti muutettua versiota BIBREF40:ssä esitetystä lähestymistavasta, jotta voimme ottaa huomioon mahdolliset alaryhmittelyt SEMCAT:n luokkien sisällä.

**Tulos**

Millaisia kokeita he käyttävät tulkinnanvaraisuuden laajuuden kvantifioimiseksi?

**Esimerkki 5.545**

Tästä huolimatta kaikki kirjallisuudessa esitetyt mallit perustuvat sanatason esityksiin, minkä vuoksi mallit eivät pysty helposti tallentamaan joitakin leksikaalisia ja morfosyntaktisia vihjeitä, joiden tiedetään ilmaisevan ironiaa, kuten isoja isoja kirjaimia, lainausmerkkejä ja hymiöitä sekä Twitterissä myös hymiöitä ja hashtageja.

**Tulos**

Minkä morfosyntaktisten piirteiden ajatellaan osoittavan ironiaa tai sarkasmia?

**Esimerkki 5.546**

Myrkyllisten kommenttien havaitsemisen vertailemiseksi Wikipedian myrkyllisiä kommentteja koskeva tietokokonaisuus (jota tutkimme tässä työssä) kerättiin ja poimittiin Wikipedian keskustelusivuilta ja esiteltiin Kaggle-kilpailussa BIBREF12, BIBREF15.

**Tulos**

Mitä tietokokonaisuuksia käytetään?

**Esimerkki 5.547**

Wikipedia-aineisto koostuu englanninkielisen Wikipedian artikkeleista, joille Wikipedia-yhteisö on antanut laatuluokkamerkinnät.

**Tulos**

Mitä kieliä he käyttävät?

**Esimerkki 5.548**

Ymmärtääksemme, mitä mallissa saattaa tapahtua, käytimme vain runko- ja iskulauseaineistoja selvittääksemme, mikä osa vitsistä oli huumorin kannalta tärkein. Havaitsimme, että kaikki mallit, myös ihmiset, tukeutuivat ennusteissaan enemmän vitsin punchlineen (taulukko 2). Vaikuttaa siis siltä, että vaikka vitsin molemmat osat ovat välttämättömiä, jotta se olisi humoristinen, iskulauseella on suurempi painoarvo kuin vartalolla. Oletamme, että tämä johtuu eri vitsirunkojen vaihtelusta: joidenkin vitsien kohdalla vitsin rakentamiseen tarvitaan kappaleita, kun taas toisissa vitsi on alle lauseen mittainen.

**Tulos**

Kumpi osa vitsistä on tärkeämpi huumorissa?

**Esimerkki 5.549**

Tässä menetelmässä yksinkertaisesti yhdistämme useiden eri toimialueiden korporaatiot kahdella pienellä muutoksella: a. Lisäämme toimialuetunnisteen "<2domain>" vastaavien korporaatioiden lähdelauseisiin.

**Tulos**

Miten he käyttivät verkkotunnuksia?

**Esimerkki 5.550**

Perusarvot. Vertaamme lähestymistapaamme (FacTweet) seuraaviin perusjoukkoihin:[leftmargin=4mm]LR + Bag-of-words: Tweet2vec: Yhdistämme syötteen twiitit ja käytämme sanasäkki-edustusta logistisen regression (LR) luokittelijan kanssa: Käytämme BIBREF20:ssä ehdotettua Bidirectional Gated recurrent neural network -mallia. Pidämme oletusarvoiset parametrit, jotka toimitettiin toteutuksen mukana. Twiittien esittämiseen käytämme mallin tuottamaa dekoodattua upotusta. Tämän perustason avulla pyrimme arvioimaan, voivatko twiittien hashtagit auttaa havaitsemaan ei-todennäköiset tilit.LR + All Features (twiittitaso): Poimimme jokaisesta twiitistä kaikki piirteemme ja syötämme ne LR-luokittimeen. Tässä tapauksessa emme aggregoi twiittejä ja tarkastelemme näin ollen jokaista twiittiä itsenäisesti.LR + All Features (chunk-level): Yhdistämme chunkissa olevien twiittien piirrevektorit ja syötämme ne LR-luokittimeen.FacTweet (twiittitaso): Samanlainen kuin FacTweet-lähestymistapa, mutta twiittitasolla; twiittien peräkkäistä kulkua ei hyödynnetä. Tavoitteena on tutkia twiittien peräkkäisen virtauksen merkitystä. top-$k$ vastaukset, tykkäykset tai uudelleentwiittaukset: Jotkut huhujen havaitsemiseen liittyvät lähestymistavat käyttävät vastausten, tykkäysten ja uudelleentwiittausten määrää huhujen havaitsemiseen BIBREF21. Näin ollen poimimme kustakin tilistä parhaat $k$ vastatut, tykätyt tai uudelleentwiittaamat twiitit arvioidaksemme tilien asiallisuutta. Testasimme eri $k$-arvoja, jotka vaihtelivat 10 twiitin ja kunkin tilin twiittien enimmäismäärän välillä. Kuvassa FIGREF24 esitetään makro-F1-arvot eri $k$-arvoilla. Vaikuttaa siltä, että $k=500$ vastaa parhaiten vastattujen twiittien osalta saavuttaa parhaan tuloksen. Siksi pidämme tätä perustasona.

**Tulos**

Mihin perusarvoihin niitä verrataan?

**Esimerkki 5.551**

Raportoimme kokeelliset tulokset kahdesta tehtävästä oikean luokittelun osuuden (CCR) osalta. Tunneanalyysissä meillä on kolmen luokan ongelma (positiivinen, negatiivinen ja neutraali), jossa luokat ovat toisensa poissulkevia. CCR, joka on keskiarvo twiittien joukosta, määritellään oikein ennustettujen tunnetilojen lukumääräksi suhteessa näissä twiiteissä esiintyvien perustodellisuuden mukaisten tunnetilojen lukumäärään. NER:ää varten katsomme, että kukin twiitti voi viitata enintään neljään ehdokkaaseen eli kohdejoukkoon. CCR, joka on keskiarvo twiittien joukosta, on oikein ennustettujen entiteettien (ehdokkaiden) määrä suhteessa kyseisessä joukossa olevien groundtruth-entiteettien (ehdokkaiden) määrään.

**Tulos**

Mitä toimenpiteitä käytetään arvioinnissa?

**Esimerkki 5.552**

Kasvu on havaittavissa Yhdysvaltain presidentinvaalien esivaalien päätyttyä sekä demokraattien ja republikaanien puoluekokousten aikana, eikä se vähene edes 8. marraskuuta pidettyjen Yhdysvaltain presidentinvaalien päätyttyä.

**Tulos**

Mitä muita poliittisia tapahtumia tietokanta sisältää?

**Esimerkki 5.553**

Aiemmin Twitterin väärinkäytösten havaitsemista koskevassa tutkimuksessa avoimesti saatavilla olleet tietokokonaisuudet olivat kooltaan 10 000 - 35 000 BIBREF9 , BIBREF11 . Tämä määrä ei riitä syväoppimismallien merkittävän määrän parametrien kouluttamiseen. Tästä syystä näitä tietokokonaisuuksia on tutkittu pääasiassa perinteisillä koneoppimismenetelmillä. Viimeisimpänä Founta et al. founta2018large esitteli Hate and Abusive Speech on Twitter -tietokannan, joka sisältää 100 000 twiittiä ristiinvalidoitujen merkintöjen kanssa. Vaikka tällä korpuksella on merkittävän kokonsa ansiosta suuri potentiaali syvämallien kouluttamiseen, tähän mennessä ei ole olemassa perusraportteja.Tässä artikkelissa tutkitaan erilaisten oppimismallien tehokkuutta loukkaavan kielen havaitsemisessa. Vertailemme tarkkuutta käyttämällä yleisimmin tutkittuja koneoppimisen luokittelijoita sekä uusimpia neuroverkkomalleja. Luotettavat perustulokset esitetään ensimmäisenä vertailututkimuksena tällä tietokokonaisuudella. Lisäksi osoitamme erilaisten ominaisuuksien ja varianttien vaikutuksen ja kuvaamme mahdollisuutta lisäparannuksiin ensemble-mallien avulla.

**Tulos**

Sisältääkö tietokokonaisuus vain englanninkielisiä tietoja?

**Esimerkki 5.554**

Kehittääksemme yleistettävän mallin, jossa vältetään ylisovittaminen, suoritamme ominaisuuksien valinnan käyttämällä tilastollisia testejä ja kaikkia asiaankuuluvia ensemble-oppimismalleja. Se lisää satunnaisuutta dataan luomalla sekoitettuja kopioita kaikista piirteistä (varjo-ominaisuus) ja kouluttaa sitten Random Forest -luokittimen laajennettuun dataan.

**Tulos**

Minkä mallin avulla saavutetaan 5 prosentin parannus F1:hen masentuneiden henkilöiden tunnistamisessa Twitterissä?

**Esimerkki 5.555**

Tietoaineistot: Kokeilimme neljällä vakiotietoaineistolla: WN18 ja FB15k on poimittu BIBREF5:llä Wordnet BIBREF32:sta Freebase BIBREF33:sta.

**Tulos**

Mitä tietokokonaisuuksia käytetään mallin arvioinnissa?

**Esimerkki 5.556**

Käytimme BIBREF-tietokannan Reuters-8-tietokokonaisuutta ilman stop-sanoja27 , jonka tavoitteena on yhden merkin luokittelu ja joka on esikäsitelty Reuters-21578-tietokokonaisuuden muoto.

**Tulos**

Mitä tietokokonaisuutta tässä työssä on käytetty?

**Esimerkki 5.557**

Tieto-ongelman ratkaisemiseksi esittelemme Taskmaster-1:n, joka koostuu 13 215 dialogista, joista 5 507 on puhuttua ja 7 708 kirjoitettua dialogia, jotka on luotu kahdella eri menettelyllä. Kukin keskustelu kuuluu johonkin kuudesta alueesta: pizzan tilaaminen, autokorjaamon ajanvarauksen luominen, kyytipalvelun järjestäminen, elokuvalipun tilaaminen, kahvijuomien tilaaminen ja ravintolavarausten tekeminen.

**Tulos**

Mitkä kuusi osa-aluetta sisältyvät tietokokonaisuuteen?

**Esimerkki 5.558**

Keskitymme ensisijaisesti sanavektoriedustuksiin (word embeddings), jotka on luotu nimenomaan Twitter-tietokannan avulla. GloVe BIBREF13 on valvomaton oppimisalgoritmi sanojen vektoriedustusten saamiseksi. Integroidaan 200-dimensionaaliset GloVe-sanojen upotukset, jotka on koulutettu 2 miljardin twiitin perusteella. Edinburghin upotukset BIBREF14 saadaan kouluttamalla skip-gram-malli Edinburghin korpukseen BIBREF15 . Koska twiitit sisältävät runsaasti emojeja, on käytetty emoji-kuvauksista opittuja Emoji embeddings BIBREF16 . Kunkin twiitin upotukset saadaan laskemalla yksittäiset sanavektorit yhteen ja jakamalla ne twiitin sisältämien merkkien lukumäärällä.

**Tulos**

mitä esivalmisteltuja sanasulkeumia käytettiin?

**Esimerkki 5.559**

Tekstin esittäminen.Syvänä neuroverkkona toteutettu malli oppii vastaamaan harjoittelemalla satoja miljoonia konteksti-vastaus $(c,r)$ -pareja. Ensin, samoin kuin Henderson:2017arxiv, sekä $c$:n että $r$:n raakateksti muunnetaan unigrammiksi ja bigrammiksi. Valokuvien esittäminen: Valokuvat esitetään käyttämällä CNN-malleja (convolutional neural net), jotka on esivalmennettu ImageNet BIBREF17 -tietokannassa. Käytämme MobileNet-mallia, jonka syvyyskerroin on 1,4 ja syöttöulottuvuus on $224 \ kertaa 224$ pikseliä, kuten BIBREF18:ssa.

**Tulos**

Miltä PolyResponsen arkkitehtuuri näyttää?

**Esimerkki 5.560**

Kukin annotoija merkitsi 90 referenssilauseeseen (eli harjoitusjoukkoon) sen, minkä tyylin hän uskoi lauseen olevan peräisin. Annotaattoreiden A1, A2 ja A3 tarkkuus tässä perustehtävässä oli 80 %, 88 % ja 80 %, mikä antaa odotetun ylärajan ihmisen suorittamalle arvioinnille.

**Tulos**

Miten he suorittavat manuaalisen arvioinnin, mitkä ovat kriteerit?

**Esimerkki 5.561**

Kysymys on koodattu 1024-ulotteisella LSTM-mallilla, joka ottaa vastaan yhden hots-kuvauksen jokaisesta kysymyksen sanasta. Kuva kuvataan 4096-ulotteisella ulostulolla, joka on peräisin konvolutiohermoverkon (CNN) viimeisestä täysin kytketystä kerroksesta, VGG16 BIBREF25 . Järjestelmä suorittaa kuvan ja kysymyksen ominaisuuksien elementtiviisaasti kerrottuna sen jälkeen, kun kuvan kuvaaja on muunnettu lineaarisesti 1024 ulottuvuuteen. Arkkitehtuurin viimeinen kerros on softmax-kerros.

**Tulos**

Mitä malliarkkitehtuuria käytetään?

**Esimerkki 5.562**

On kuitenkin mahdollista, että tässä mitatut tarkat tulkittavuuspisteet ovat vääristyneitä käytetyn aineiston vuoksi. On kuitenkin huomattava, että tulkinnanvaraisuuspisteiden muutokseen eri sanapeittävyyksillä saattaa vaikuttaa luokan sanojen epäideaalinen osavalinta. Vaikka sanojen etäisyyksiin kategorioiden keskuksista perustuvan sanojen otantamenetelmämme odotetaan tuottavan paremmin edustettuja kategorioita verrattuna kategoriasanojen satunnaisotantaan, kategorioiden representaatiot saattavat olla epäoptimaalisia verrattuna ihmisen suunnittelemiin kategorioihin.

**Tulos**

Mitkä ovat heidän ehdottamansa tulkittavuuden kvantifiointimenetelmän heikkoudet?

**Esimerkki 5.563**

Kun otetaan huomioon joukko arviointidokumentteja, joilla kullakin on tunnetusti oikea merkintä suljetusta merkintäjoukosta (jota usein kutsutaan "kultaiseksi standardiksi"), ja jokaiselle dokumentille ennustettu merkintä samasta joukosta, dokumenttitason tarkkuus on niiden dokumenttien osuus, jotka on merkitty oikein koko arviointikokoelmasta. Tämä on useimmin raportoitu mittari, ja se välittää saman tiedon kuin virheprosentti, joka on yksinkertaisesti niiden asiakirjojen osuus, jotka on merkitty väärin (eli INLINEFORM0 ). Tulokset esitetään yleensä kielikohtaisesti kahdella eri tavalla: (1) tarkkuus, jossa asiakirjat ryhmitellään niiden ennustetun kielen mukaan, ja (2) palautus, jossa asiakirjat ryhmitellään sen mukaan, millä kielellä ne on todellisuudessa kirjoitettu. Yleinen käytäntö on myös ilmoittaa F-pistemäärä INLINEFORM0 , joka on tarkkuuden ja palautuksen harmoninen keskiarvo.

**Tulos**

mitä arviointimenetelmiä käsitellään?

**Esimerkki 5.564**

Saamme aineistomme uutisjulkaisuista, viinikritiikeistä ja Redditistä, joiden suuren volyymin lisäksi voimme myös luonnehtia binomeja uusilla tavoilla ja analysoida binomijärjestelyjen eroja eri yhteisöjen välillä ja ajassa.

**Tulos**

Mitä verkkotekstilähteitä käytetään binomilistojen testaamiseen?

**Esimerkki 5.565**

Keräämme lausumia $\mathbf {C}$hiinan $\mathbf {A}$tekoäly $\mathbf {I}$älykielisiltä $\mathbf {S}$puhujilta (CAIS) ja annotoimme niihin slot-tagit ja intent-merkinnät. Harjoitus-, validointi- ja testijoukot on jaettu aikomusten jakautumisen mukaan, ja yksityiskohtaiset tilastotiedot on esitetty lisäaineistossa. Koska lausumat on kerätty reaalimaailman kaiutinjärjestelmistä, aikomusmerkinnät ovat osittaisia PlayMusic-vaihtoehdolle. Otamme käyttöön BIOES-merkintäjärjestelmän lähtö- ja saapumisaikoihin ATIS:ssä käytetyn BIO2:n sijasta, koska aiemmissa tutkimuksissa on tuotu esiin, että tällä järjestelmällä BIBREF30 on saavutettu merkittäviä parannuksia sekvenssien merkitsemisessä.

**Tulos**

Mikä on heidän keräämänsä korpuksen ala?

**Tulos**

Mikä on CAIS-tietokannan lähde?

**Esimerkki 5.566**

Saadaksemme tietoa siitä, miksi hienojakoisempi bi-sense-emoji-kuviointi auttaa ymmärtämään twiittien taustalla olevia monimutkaisia tunteita, visualisoimme ATT-E-LSTM:n ja MATT-BiE-LSTM:n huomiopainot vertailun vuoksi. Rakennamme siis uuden syötteen INLINEFORM0 kullekin LSTM-yksikölle yhdistämällä alkuperäisen sanan upotuksen ja huomiovektorin yhtälössä EQREF21, jotta senti-emoji-tieto voidaan jakaa kullekin vaiheelle. Tätä mallia kutsutaan monitasoiseksi Attention-based LSTM with Bi-sense Emoji Embedding (MATT-BiE-LSTM) Attention-based LSTM with emojis: Käytämme myös sana-emoji-kasautusta emoji-sanan huomion laskemiseen yhtälön EQREF20 ja EQREF21 mukaisesti, ja ainoa ero on se, että korvaamme huomiosta johdetun sentti-emoji-kasautuksen nopean tekstin avulla valmiiksi koulutetulla sana-emoji-kasautuksella, jota kutsutaan nimellä ATT-E-LSTM. Kuvassa FIGREF27 (a) ATT-E-LSTM-malli (peruslinja) antaa suhteellisesti enemmän painoja sanoille "ei" ja "paine", kun taas MATT-BiE-LSTM kiinnittää huomiota lähinnä sanoihin "happy" ja "lovely". Erilaiset huomiojakaumat viittaavat siihen, että ehdotettu senti-emoji-kasauttaminen pystyy tunnistamaan sanoja, joilla on vahva tunne, joka liittyy läheisesti todelliseen tunteeseen, vaikka läsnä olisi sanoja, joilla on ristiriitaisia tunteita, kuten "paine" ja "onnellinen". kun taas ATT-E-LSTM pyrkii poimimaan kaikki tunnepitoiset sanat, jotka voivat aiheuttaa sekaannuksia. Senti-emoji-kasauttaminen kykenee poimimaan monimutkaisen semantiikan ja tunteiden representaatioita, jotka auttavat ohjaamaan huomioita myös silloin, kun sanan sentimentti ja emoji-tentimentti ovat hieman ristiriidassa keskenään. Kuvasta FIGREF27 (b) ja (c) voidaan havaita, että ATT-E-LSTM antaa enemmän painoja tunteen kannalta epäolennaisille sanoille kuin MATT-BiE-LSTM, kuten "hupparit", "odota" ja "jälkeen", mikä osoittaa, että ehdotettu malli on kestävämpi epäolennaisille sanoille ja keskittyy paremmin tärkeisiin sanoihin. Koska bi-sense-emoji-koodauksen avulla saatu senti-emoji-koodaus ja lauseen tason LSTM-koodaus tekstinsyötteessä (kuvattu kohdassa SECREF13 ), pystymme rakentamaan vankemman koodauksen, joka perustuu koko kontekstin semanttiseen ja tunnetietoon verrattuna ATT-E-LSTM:ssä käytettyyn sana-emoji-koodaukseen, joka ottaa huomioon vain sanatason tiedon.

**Tulos**

Mitä todisteita huomion visualisointi antaa siitä, että se auttaa saamaan vankemman käsityksen semantiikasta ja tunteista?

**Esimerkki 5.567**

Vaikka muissa NLG-tehtävissä, kuten tekstin tiivistämisessä (BIBREF24 ), kuvatekstien tuottamisessa (BIBREF25 ) ja taulukoiden tuottamisessa tekstiksi (BIBREF26 ), on otettu huomioon monia eri syötemodaliteetteja, perinteisessä QG:ssä on keskitytty lähinnä tekstimuotoisiin syötteisiin, erityisesti deklaratiivisiin lauseisiin, mikä selittyy alkuperäisillä sovellusalueilla, kuten kysymyksiin vastaamisessa ja koulutuksessa, joissa myös syötteet ovat tyypillisesti tekstimuotoisia.Viime aikoina erilaisten kysymyksenasettelusovellusten, kuten tietopohjakysymyksiin vastaamisen (KBQA, Knowledge Base Question Answering) BIBREF27 ja visuaalisten kysymysten vastaamisen (VQA, Visual Question Answering) BIBREF28 , lisääntyessä myös NQG-tutkimus on laajentanut lähteiden kirjoa sisällyttämällä tietopohjat BIBREF29 ja kuvat BIBREF10 .

**Tulos**

Mitkä ovat kaikki kysymysten tuottamiseen liittyvässä aiemmassa työssä huomioon otetut syöttömuodot?

**Esimerkki 5.568**

Mallin suorituskyvyn arvioinnissa käytetään BLEU-mittaria.

**Tulos**

Mitä arviointimittaria käytetään?

**Esimerkki 5.569**

Arvioimme malliamme kahdella vertailutietoaineistolla BIBREF9 . Homografinen tietokokonaisuus sisältää 2 250 kontekstia, joista 1 607 sisältää sanaleikin. Heterografinen tietokokonaisuus koostuu 1 780 kontekstista, joista 1 271 sisältää sanaleikin. Huomaamme, että kummastakaan tietokokonaisuudesta ei ole annettu vakiomuotoista jakotietoa. Näin ollen käytämme 10-kertaista ristiinvalidointia. Jotta voisimme tehdä suoria vertailuja aiempiin tutkimuksiin, BIBREFin4 mukaisesti keräsimme kaikkien kymmenen kertauksen ennusteet ja laskimme lopuksi pisteet.

**Tulos**

Mitä tietokokonaisuuksia käytetään arvioinnissa?

**Esimerkki 5.570**

Visual Dialog -tehtävässä käytetään arvioinnissa neljää mittaria. NDCG on visuaalisen dialogin tehtävän ensisijainen mittari, jossa useita samanlaisia vastauksia pidetään oikeina. Muut kolme ovat MRR, recall@k ja mean rank, joissa otetaan huomioon vain yhden vastauksen arvo.

**Tulos**

Mitä mittareita haasteessa käytetään?

**Esimerkki 5.571**

Mallimme syötteenä ovat syötetekstin $x[1], ... sanat. , x[n]$ ja kysely $q[1], ... , q[n]$ . Yhdistämme GloVe BIBREF40 -ohjelmasta valmiiksi koulutetut sanasulkeumat ja CharCNN BIBREF41 -ohjelmalla koulutetut merkkisulkeumat edustamaan syötesanoja. Syötetyn tekstin $x\_1, ... , x\_n$ ja kyselyn $q\_1, ... , q\_n$ syötetään sitten Highway-kerrokseen BIBREF42 sanojen upotusten ja merkkien upotusten kyvykkyyden parantamiseksi seuraavasti: $$$\begin{split} g\_t &= {\rm sigmoid}(W\_gx\_t+b\_g) \\\\ s\_t &= {\rm relu } (W\_xx\_t+b\_x) \\ u\_t &= g\_t \odot s\_t + (1 - g\_t) \odot x\_t~. \end{split}$$ (Yht. 18) Samaa Highway-kerrosta sovelletaan $q\_t$:hen ja saadaan $v\_t$ . Seuraavaksi $u\_t$ ja $v\_t$ syötetään vastaavasti kaksisuuntaiseen BiLSTM-verkkoon (BiLSTM, Bi-Directional Long Short-Term Memory Network) BIBREF44 , jotta voidaan mallintaa sekvenssisanojen väliset ajalliset vuorovaikutukset: Sitten syötetään $\mathbf {U}$ ja $\mathbf {V}$ huomiovirtakerrokseen BIBREF27, jolla mallinnetaan syöttötekstin ja kyselyn välisiä vuorovaikutuksia. Siksi otamme malliimme käyttöön itsesovituskerroksen BIBREF29 seuraavasti: $$$\begin{split} o\_t &= {\rm BiLSTM}(o\_{t-1}, [h\_t, c\_t]) \\\ s\_j^t &= w^T {\rm tanh}(W\_hh\_j+\tilde{W\_h}h\_t)\\\ \alpha \_i^t &= {\rm exp}(s\_i^t)/\Sigma \_{j=1}^n{\rm exp}(s\_j^t)\\\\ c\_t &= \Sigma \_{i=1}^n\alpha \_i^th\_i ~. \end{split}$$ (yhtälö 20) Lopuksi syötetään sulautumat $\mathbf {O} = [o\_1, ... , o\_n]$ osoitinverkkoon BIBREF39 vastaussekvenssin dekoodaamiseksi seuraavasti$$\\begin{split} p\_t &= {\rm LSTM}(p\_{t-1}, c\_t) \\\ s\_j^t &= w^T {\rm tanh}(W\_oo\_j+W\_pp\_{t-1})\\ \beta \_i^t &= {\rm exp}(s\_i^t)/\Sigma \_{j=1}^n{\rm exp}(s\_j^t)\\\ c\_t &= \Sigma \_{i=1}^n\beta \_i^to\_i~. \end{split}$$ (Yht. 21) Näin ollen todennäköisyys tuottaa vastausjakso $\textbf {a}$ on seuraava$$${\rm P}(\textbf {a}|\mathbf {O}) = \prod \_t {\rm P}(a^t | a^1, ... , a^{t-1}, \mathbf {O})~.$$ (Yht. 23)

**Tulos**

Mitä laadunvarmistusmalleja käytettiin?

**Esimerkki 5.572**

Koska kaikkia kommentteja ei ole mahdollista kommentoida manuaalisesti, käsittelemme tätä tietokokonaisuutta poimimalla niistä viestiketjut, joihin liittyy epäiltyjä trollausyrityksiä, ja niihin annetut suorat vastaukset. Käytössämme oli kaksi inhimillistä annotoijaa, jotka koulutettiin 200 keskustelusta otetuilla katkelmilla (eli (epäilty trollausyritys, vastaukset) -pareilla), ja he saivat keskustella havainnoistaan. Tämän koulutusvaiheen jälkeen pyysimme heitä merkitsemään itsenäisesti neljä näkökohtaa kustakin katkelmasta.

**Tulos**

Käyttävätkö he kommentointiin joukkoistamisalustaa?

**Esimerkki 5.573**

Tässä työssä kehitämme tekniikan, jolla olemassa oleva esivalmennettu malli voidaan nopeasti siirtää englannista muille kielille energiatehokkaasti BIBREF8. Ensimmäisessä vaiheessa keskitymme rakentamaan kaksikielisen kielimallin englannin ja kohdekielen välillä. Englanninkielisestä esivalmennetusta LM-mallista lähtien opitaan kohdekielelle ominaiset parametrit (esim. sanojen upotukset) samalla kun englanninkielisen esivalmennetun LM-mallin koodauskerrokset pidetään kiinteinä. Sen jälkeen hienosäädämme sekä englannin- että kohdemallia, jotta saamme kaksikielisen LM:n.

**Tulos**

Miten malli siirretään muihin kieliin?

**Esimerkki 5.574**

Tuoreessa turkkilaisen BIBREF10-julkaisun tunneanalyysia käsittelevässä työssä he opettelevat upotuksia käyttäen turkkilaista sosiaalista mediaa. He käyttävät word2vec-algoritmia, luovat useita käsin muodostettuja, valvomattomia piirteitä, luovat asiakirjavektoreita ja syöttävät ne syötteenä tukivektorikoneisiin (SVM). Me päihitämme tämän peruslähestymistavan käyttämällä tehokkaampia sanojen upotuksia ja valvottuja käsin luotuja piirteitä.

**Tulos**

Mitä perustason menetelmää käytetään?

**Esimerkki 5.575**

Käytämme 300-ulotteista Glove Common Crawl Embeddings (840B Token) BIBREF11 -ohjelmaa ja hienosäädämme ne tehtävään sopiviksi.

**Tulos**

Millä datalla sulautumat on koulutettu?

**Esimerkki 5.576**

Seuraavaksi vertasimme VTN-malliamme RNN-pohjaiseen seq2seq VC-malliin nimeltä ATTS2S BIBREF8. Tämä malli perustuu Tacotron-malliin BIBREF32, jossa käytetään kontekstin säilyttämishäviötä ja ohjattua huomiohäviötä harjoittelun vakauttamiseksi ja kielellisen johdonmukaisuuden säilyttämiseksi muuntamisen jälkeen. Noudatimme BIBREF8:n konfiguraatioita, mutta käytimme WORLD-piirteiden sijasta melaspektrogrammeja.

**Tulos**

Mikä on perusmalli?

**Esimerkki 5.577**

Keskitymme kolmeen pääkysymykseen, kuten siihen, miten sisällyttää affektiivista tietoa chat-robotteihin, mitä resursseja on saatavilla ja mitä voidaan käyttää EAC:n rakentamiseen ja miten EAC:n suorituskykyä voidaan arvioida.

**Tulos**

Mitkä ovat asiakirjassa esitetyt tutkimuskysymykset, jotka koskevat EAC-tutkimuksia?

**Esimerkki 5.578**

Taulukossa TABREF14 esitetään pääkokeidemme tulokset vuosien 2016 ja 2018 testijoukoissa ranskan ja saksan osalta. Käytämme päämittarina Meteor BIBREF31 -mittaria, kuten WMT-tehtävissä BIBREF25 . Vertaamme muuntajiemme perustasoa kuvainformaatiolla rikastettuihin muuntajamalleihin sekä deliberaatiomalleihin, kuvainformaatiolla tai ilman sitä. toteamme ensinnäkin, että multimodaaliset mallimme saavuttavat muuntajaverkkojen (rajoitetut mallit) uusimman tason suorituskyvyn englannin ja saksan tietokokonaisuudessa verrattuna BIBREF30 -malliin. Toiseksi harkintamallimme johtavat merkittäviin parannuksiin tähän perustasoon verrattuna kaikissa testijoukoissa (keskimäärin INLINEFORM0 , INLINEFORM1 ).

**Tulos**

Millä tietokokonaisuudella tämä lähestymistapa saavuttaa uusimmat tulokset?

**Esimerkki 5.579**

Käytämme perusmallina siamilaisneuroverkkoa, jonka on osoitettu suoriutuvan huippuluokan muutaman otoksen oppimisesta BIBREF11. Muokkaamme alkuperäistä mallia siten, että se ottaa huomioon peräkkäiset tiedot, ja jokainen kaksonen koostuu sulauttamiskerroksesta, LSTM-kerroksesta (Long-Short Term Memory) BIBREF12 ja syöttö eteenpäin -kerroksesta, jossa on ReLU-aktivoinnit (ReLU = Rectified Linear Unit).

**Tulos**

Mitkä olivat lähtötasot?

**Esimerkki 5.580**

Ihmisten suorittamassa arvioinnissa noudatamme konekäännösjärjestelmien BIBREF23 -arvioinnissa vakiintunutta lähestymistapaa, jota BIBREF9 käyttää kysymysten luomisessa. Pyysimme kolmea työntekijää arvioimaan 300 generoitua kysymystä välillä 1 (huono) ja 5 (hyvä) kahdella erillisellä kriteerillä: käytetyn kielen sujuvuus ja kysymyksen relevanssi asiayhteysasiakirjan ja vastauksen kannalta.

**Tulos**

Mitä inhimillisen arvioinnin mittareita paperissa käytettiin?

**Esimerkki 5.581**

DBpedian SPARQL-päätepistettä käytetään kyselyihin vastaamiseen ja päättelyyn.

**Tulos**

Mitä päättelymenetelmää käytetään?

**Esimerkki 5.582**

Parhaiten menestyvät lähestymistavat Caravel, COAV ja NNCD ansaitsevat tarkempaa huomiota.

**Tulos**

Mikä on parhaiten toimiva menetelmä?

**Esimerkki 5.583**

Tässä jaksossa analysoimme enwik9-tietokannan lauseiden projektioiden ja samojen lauseiden vastaavien projektioiden välistä Hammingin etäisyyttä sen jälkeen, kun merkkitason häiriöitä on sovellettu. Kokeilemme kolmenlaisia merkkitason häiriötoimintoja BIBREF11 ja kahdenlaisia sanatason häiriötoimintoja.Häiriötutkimus ::: Merkkitason häiriötoiminnotinsert(word, n) : Valitaan satunnaisesti n merkkiä merkkisanastosta ja lisätään ne satunnaisiin paikkoihin syötettyyn sanaan. Säilytämme kuitenkin sanan ensimmäisen ja viimeisen merkin sellaisenaan. Esim. muunnos: $sample \rightarrow samnple$.swap(word, n): Vaihdamme satunnaisesti kahden merkin paikkaa sanassa n kertaa. Kuten insert-operaatiossa, säilytämme sanan ensimmäisen ja viimeisen merkin sellaisenaan ja sovellamme swap-operaatiota vain jäljellä oleviin merkkeihin. Esim. muunnos: $sample \rightarrow sapmle$.duplicate(word, n): Kaksinkertaistamme satunnaisesti yhden merkin sanassa n kertaa. Esim. muunnos: $sample \rightarrow saample$.Perturbation Study ::: Character Level Perturbation Operations ::: Pudotus(lause, n): Pudotamme satunnaisesti n sanaa lauseesta. Esim. muunnos: Tämä on iso kissa. $\rightarrow $ Tämä on kissa.duplicate(lause, n): Samanlainen kuin duplicate(word, n) edellä, toistamme satunnaisesti sanan lauseessa n kertaa. Esim. muunnos: Tämä on iso kissa. $\rightarrow $ Tämä on iso iso kissa.swap(lause, n): Vaihdetaan satunnaisesti kahden sanan paikkaa lauseessa n kertaa. Esim. muunnos: Tämä on iso kissa. $\rightarrow $ Tämä kissa on iso.

**Tulos**

Miten heidän häiriöalgorihminsa toimii?

**Esimerkki 5.584**

Käytämme BIBREF4:n julkaisemaa julkisia talousuutisia koskevaa tietokokonaisuutta, joka on kerätty Reutersista ja Bloombergista lokakuun 2006 ja marraskuun 2013 väliseltä ajalta. Teemme kokeita Standard & Poor's 500 -osakeindeksin (S&P 500) ja sen valittujen yksittäisten osakkeiden ennustamiseksi, ja saamme indeksit ja hinnat Yahoo Financesta. Koulutus-, kehitys- ja testijoukkojen yksityiskohtaiset tilastot esitetään taulukossa TABREF8. Raportoimme lopulliset tulokset testijoukosta sen jälkeen, kun olemme käyttäneet kehitysjoukkoa joidenkin hyperparametrien virittämiseen.

**Tulos**

Mitä tietokokonaisuutta käytetään uutisohjattujen osakkeiden liikkeiden ennustamiseen?

**Esimerkki 5.585**

Monitehtävämalli on parempi kuin yhden tehtävän malli kaikilla datakokoluokilla, ja suhteellinen suorituskyky kasvaa harjoitusdatan koon pienentyessä. Kun harjoitusdataa käytetään vain 200 lauseen verran, monitehtävämallin suorituskyky on noin 60 prosenttia parempi kuin yhden tehtävän mallin sekä Airbnb- että Greyhound-sovelluksissa. OpenTable-sovelluksen suhteellinen parannus on 26 %. Tulokset vaihtelevat tehtävien välillä, mutta yhdessäkään ei ole yleistä hyötyä avoimen sanaston järjestelmästä.

**Tulos**

Lisääntyykö suorituskyky heidän menetelmällään?

**Esimerkki 5.586**

(1) AutoJudge päihittää johdonmukaisesti ja merkittävästi kaikki vertailumallit, mukaan lukien RC-mallit ja muut neuraaliset tekstiluokitusmallit, mikä osoittaa mallimme tehokkuuden ja kestävyyden. (2) RC-mallit saavuttavat paremman suorituskyvyn kuin useimmat tekstinluokittelumallit (GRU+Attentionia lukuun ottamatta), mikä osoittaa, että lukumekanismi on parempi tapa integroida heterogeenisistä mutta toisiaan täydentävistä syötteistä saatua tietoa. (3) Verrattuna perinteisiin RC-malleihin AutoJudge saavuttaa merkittävän parannuksen, kun otetaan huomioon lisää lakiartikkeleita.

**Tulos**

mitkä ovat niiden tulokset rakennetulla tietokokonaisuudella?

**Esimerkki 5.587**

Sanamerkityksen induktio (WSI) voidaan nähdä WSD:n valvomattomana versiona. WSI:n tarkoituksena on klusteroida sanojen aistimuksia, eikä jokaisen klusterin tarvitse vastata ennalta määriteltyä aistimusta. Sen sijaan sanamerkitysluettelot muodostetaan automaattisesti klustereista siten, että kutakin klusteria käsitellään sanan yksittäisenä merkityksenä. Ehdotamme edistyneempää graafin rakentamismenettelyä, jossa käytetään vektorien yhteen- ja vähennysoperaatioiden tulkittavuutta sanojen sulautumisavaruudessa BIBREF6 , kun taas aiempi algoritmi perustuu vain lähimpien naapureiden luetteloon sanojen sulautumisavaruudessa. Algoritmimme keskeinen innovaatio on vektorin vähennyslaskun käyttö, jonka avulla löydetään parit, joissa on kaikkein erilaisimmat graafin solmut, ja rakennetaan graafi vain solmuista, jotka sisältyvät tällaisiin "antisärmiöihin".

**Tulos**

Onko tässä työssä kuvattu menetelmä klusterointiin perustuva menetelmä?

**Esimerkki 5.588**

Yhdessä nämä käyttäjät antoivat Gunrockille keskimäärin 3,65 arvosanan (mediaani: 4,0), joka saatiin keskustelun lopussa ("Asteikolla 1-5 tähteä, miltä tuntuisi keskustella tämän sosiaalibotin kanssa uudelleen?").

**Tulos**

Keräävätkö ne Gunrockista nimenomaisia käyttäjätyytyväisyystietoja?

**Esimerkki 5.589**

Varmistaaksemme olettamuksemme, että kohteen koodaus ja ortogonaalinen regularisointi auttavat lisäämään generoitujen sekvenssien monimuotoisuutta, käytämme kahta mittaria, yhtä kvantitatiivista ja toista kvalitatiivista, mittaamaan generoinnin monimuotoisuutta. Ensinnäkin laskemme yksinkertaisesti sekä INLINEFORM0:n että INLINEFORM1:n tuottamien ainutlaatuisten ennusteiden keskiarvon jaksossa SECREF36 esitetyissä kokeissa. Tässä esimerkissä on 29 pohjatodellisuuslausetta. Kumpikaan malleista ei pysty tuottamaan kaikkia avainsanoja, mutta on selvää, että INLINEFORM0:n ennusteet alkavat kaikki sanalla "test", kun taas INLINEFORM1:n ennusteet ovat monipuolisia.

**Tulos**

Miten avainsanojen monimuotoisuutta mitataan?

**Esimerkki 5.590**

Tämä arkkitehtuuri perustuu sogaard2016deepiin, mutta sitä on mukautettu kahdella tavalla: ensinnäkin lisäämme siihen valvomattoman sekvenssin merkintätehtävän (Language Modeling) ja toiseksi lisäämme hierarkiassa tehtävien väliin matalaulotteisen sulauttamiskerroksen, jonka avulla opimme tiheitä merkintätunnisteiden representaatioita.

**Tulos**

Mikä on viimeisen kerroksen valvomaton tehtävä?

**Esimerkki 5.591**

Ensin poimimme useita suhteita käyttämällä OpenIE-työkalupakettia (OpenIE, Open Information Extraction) BIBREF7 , ja sitten valitsemme vastauksen kannalta olennaisimman suhteen huolellisesti suunniteltujen heurististen sääntöjen avulla.

**Tulos**

Miten he poimivat "jäsennellyn vastauksen kannalta merkityksellisen suhteen"?

**Esimerkki 5.592**

Laskemme kunkin käyttäjän osalta niiden twiittien osuuden, jotka on arvioitu positiivisesti kussakin LIWC-kategoriassa.

**Tulos**

Miten LIWC on sisällytetty tähän järjestelmään?

**Esimerkki 5.593**

Arvioimme ehdotettuja mallejamme kolmella yleisesti käytetyllä tietograafitietoaineistolla: WN18RR BIBREF26, FB15k-237 BIBREF18 ja YAGO3-10 BIBREF27.

**Tulos**

Mitä vertailutietoaineistoja käytetään linkkien ennustamiseen?

**Esimerkki 5.594**

Nämä tunnepisteet saatiin automaattisesti aiemmassa työssämme BIBREF15 esitellyn aiheen poiminta- ja tunneanalyysiputken avulla, ja ne harjoitteltiin ennalta talon sisäisellä psykiatrisella EHR-tekstillä.

**Tulos**

Miten he poimivat aiheita?

**Esimerkki 5.595**

Jotta luotujen sekvenssien laatua voitaisiin arvioida sekä tarkkuuden että palautuksen osalta, arviointimittareina käytetään BLEU- ja ROUGE (1, 2, L) -pisteitä, joissa on useita viitteitä BIBREF22 .

**Tulos**

Mitä arviointimittareita käytetään?

**Esimerkki 5.596**

PlEWi toimitti 550 755 [virhe, korjaus]-paria, joista 298 715 oli yksilöllisiä.

**Tulos**

Miten PIEWi on kommentoitu?

**Esimerkki 5.597**

Ensimmäinen malli, jota kutsutaan seuraavissa jaksoissa nimellä `LSTM (PTB)', koulutettiin Penn Treebank BIBREF12 -tietokannan lauseilla. Toinen malli, jota kutsutaan nimellä `LSTM (FTB)', koulutettiin French Treebank BIBREF13 -tietokannan lauseilla. Käytämme kahta esivalmisteltua englanninkielistä mallia: toinen on koulutettu BIBREF14:n Billion Word -vertailumallilla (nimellä `LSTM (1B)') ja toinen BIBREF3:n englanninkielisellä Wikipedialla (nimellä `LSTM (enWiki)'). Ranskan kielen osalta koulutimme suuren LSTM-kielimallin (nimellä `LSTM (frWaC)') frWaC-tietokannan BIBREF15 satunnaisella osajoukolla (noin 4 miljoonaa lausetta, 138 miljoonaa sanamerkkiä).

**Tulos**

Mikä on käytettyjen tietokokonaisuuksien koko?

**Esimerkki 5.598**

Yhdistämällä moduuliosan ja vaiheosan HAKE kartoittaa kokonaisuudet polaarikoordinaatistoon, jossa säteittäinen koordinaatti vastaa moduuliosaa ja kulmakoordinaatit vaiheosaa.

**Tulos**

Miten entiteetit kartoitetaan napakoordinaattijärjestelmään?

**Esimerkki 5.599**

Viimeaikaiset edistysaskeleet syväoppimistekniikoissa ovat kuitenkin osoittaneet, että NER-tehtävässä voidaan hyödyntää neuroarkkitehtuureja, kuten biLSTM-verkkoja BIBREF3 , BIBREF4 . Käytämme BIBREF24:ssä ehdotettua toteutusta kolmen eri kokeen suorittamiseen.

**Tulos**

Mitä koneoppimisalgoritmeja tutkittiin?

**Esimerkki 5.600**

Käytimme Shortirin, indonesialaisen uutisaggregaattori- ja tiivistämisyrityksen, tarjoamaa aineistoa. Tietokanta sisältää noin 20 000 uutisartikkelia. Jokaisessa artikkelissa on otsikko, kategoria, lähde (esim. CNN Indonesia, Kumparan), alkuperäisen artikkelin URL-osoite ja tiivistelmä, jonka on laatinut manuaalisesti kaksi indonesian kielen äidinkielistä puhujaa.

**Tulos**

Käyttivätkö he yhteenvetojen tekemiseen joukkoistamisalustaa?

**Esimerkki 5.601**

Verrattaessa perusratkaisuihin, joissa yksi naapureista valitaan satunnaisesti tai joissa oletetaan, että alhaisimman pistemäärän omaava fakta on virheellinen, havaitaan, että virheiden havaitsemisessa saavutetaan 42 \%$ ja 55 \%$ tarkkuus, mikä on huomattavan suuri ero.

**Tulos**

Voidaanko tätä vastakkainasettelua käyttää suoraan parantamaan mallin tarkkuutta?

**Esimerkki 5.602**

Lisäksi seq2seq-perusmallissa oletetaan, että generoitujen merkkien välillä on tiukka järjestys, mutta todellisuudessa mallia ei pitäisi rangaista ankarasti siitä, että se ennustaa oikeat merkit väärässä järjestyksessä.

**Tulos**

Asettavatko ne mitään kieliopillisia rajoituksia tuotetulle tulosteelle?

**Esimerkki 5.603**

Käyttäjien havaittua käyttäytymistä paremmin vastaava esitys on käsitekartta BIBREF5 , joka on merkitty graafi, jossa käsitteet esitetään solmuina ja niiden väliset suhteet reunoina (kuva KUVA 2 ).

**Tulos**

Miten kirjoittajat määrittelevät käsitekartan?

**Esimerkki 5.604**

Toteutamme uudelleen BIBREF3:ssa ehdotetun mallin ja käytämme sitä ongelmamme lähtökohtana. Perusteena tämän mallin valinnalle perustasoksi on sen todistetusti hyvä ennustuskyky monikielisen tekstin luokittelussa.

**Tulos**

Mikä on heidän perusmallinsa?

**Esimerkki 5.605**

Käytimme uusimpia ominaisuuksia, jotka ovat osoittautuneet hyödyllisiksi tunnistamisessa: osa niistä on kielestä riippumattomia (esim. välimerkit, positiiviset ja negatiiviset hymiöt, sitaatit, persoonapronominit, twiitin pituus, nimetyt entiteetit), kun taas osa on kielestä riippuvaisia ja perustuu erityisiin sanakirjoihin (esim. negaatio, mielipidesanakirjat, oppositiosanat).

**Tulos**

Mitä tekstipohjaisia ominaisuuksia käytetään?

**Esimerkki 5.606**

Tässä luvussa esitämme neuroprojektorillemme käänteisehdon, jonka avulla voimme ratkaista optimointiongelman. Erityisesti rajoitamme neuraaliprojektoriamme kahdella vaatimuksella: (1) INLINEFORM0 ja (2) INLINEFORM1 on olemassa.

**Tulos**

Mikä on käänteisehto?

**Esimerkki 5.607**

Jotta voitiin testata ehdotetun menetelmän kykyä tuottaa sanoja ilman valvontaa, oli tarpeen kehittää menetelmä sanojen merkityksellisyyden mittaamiseksi. Käytettiin aihepiirien mallintamista, joka perustuu oletukseen, että samasta aihepiiristä löytyvät sanat ovat merkityksellisempiä toisilleen kuin eri aihepiirien sanat BIBREF14. Koko 200 000 otsikon aineisto BIBREF11 mallinnettiin Naïve Bayes -algoritmilla BIBREF15 sanojen ja luokkien yhteisesiintymismallin luomiseksi. Tämän jälkeen löydettiin 200 tärkeintä sanaa kustakin kategoriasta ja niitä käytettiin aihepiiritaulukon SECREF12 luomiseen. Oletettiin, että kukin luokka edustaa omaa ainutlaatuista aihettaan. Relevanttien tulossanojen määrä mitattiin otsikon luokkamerkinnän funktiona, ja se on esitetty kuvassa SECREF4. Tulokset osoittavat, että ehdotetulla menetelmällä pystyttiin tunnistamaan oikein uudet sanat, jotka liittyvät syötettyyn aiheeseen, kun signaali-kohinasuhde oli 4:1.

**Tulos**

Miten ne määrittävät samankaltaisuuden ennustetun sanan ja aiheiden välillä?

**Esimerkki 5.608**

Kokeilimme Glove BIBREF2- ja Twitter word2vec BIBREF3 -koodia käsiteltyjen twiittien upotusten kouluttamiseksi.

**Tulos**

Mitä sulautumia he käyttävät?

**Esimerkki 5.609**

Työhömme liittyy useita varoituksia: ensinnäkin twiitti-tunnelmat ovat harvoin binäärisiä (tätä työtä voitaisiin laajentaa moninomiaaliseen tai jatkuvaan malliin). Toiseksi, tuloksemme rajoittuvat Twitterin käyttäjiin, joiden tiedetään olevan negatiivisempia kuin Yhdysvaltain väestön BIBREF9 . Kolmanneksi emme ota huomioon jatkuvien luonnonkatastrofien kokonaisvaikutuksia ajan mittaan. Jatkossa on selvää kysyntää selvittää, voivatko sosiaaliset verkostot osoittaa ympäristöä koskevia mittareita "ennakoivalla" tavalla. Ilmastonmuutoksen muuttuessa yhä äärimmäisemmäksi jää nähtäväksi, millainen ennustevoima nykyisellä mallillamme on ilmastonmuutosta ja luonnonkatastrofeja koskevien mielipiteiden osalta.

**Tulos**

Mainitsevatko kirjoittajat tutkimuksessaan esiintyviä häiriötekijöitä?

**Esimerkki 5.610**

Syöttöasiakirjaa/yhteenvetoa, jossa voi olla järjestämättömiä lauseita, käsitellään siten, että lauseet ryhmitellään yhteen. Testataksemme lähestymistapaamme sekoitamme asiakirjan lauseiden järjestyksen, käsittelemme järjestämättömän asiakirjan ja vertaamme tulostusasiakirjan samankaltaisuutta alkuperäisen asiakirjan kanssa. Järjestämättömän asiakirjan jäsentäminen on olennainen tehtävä monissa sovelluksissa. Se on jälkivaatimus sellaisissa sovelluksissa kuin useiden asiakirjojen tekstin tiivistäminen, joissa on esitettävä yhteenveto useista asiakirjoista. Se on ennakkoedellytys sovelluksissa, kuten kysymysten vastaaminen useista asiakirjoista, joissa on esitettävä vastaus käsittelemällä useita asiakirjoja.

**Tulos**

Mikä on järjestämätön tekstidokumentti, esiintyykö niitä reaalimaailman korpuksissa?

**Esimerkki 5.611**

Vertailussa on mukana myös RNNLM BIBREF11 , joka on rekursiiviseen neuroverkkoon perustuva kielimalli.

**Tulos**

Mihin kielimalleihin niitä verrataan?

**Esimerkki 5.612**

Kun sanaa $y\_t$ luodaan aika-askeleella $t$, konteksti $\mathbf {X}$ koodataan kiinteän kokoiseksi dialogikontekstivektoriksi $\mathbf {c}\_t$ noudattamalla hierarkkista huomiorakennetta HRAN BIBREF13:ssa. Lisäksi poimimme tunteisiin liittyvän tiedon $\mathbf {X}$:n lausunnoista ulkoisen tekstianalyysiohjelman avulla ja koodaamme sen RNN:n avulla tunnekontekstivektoriksi $\mathbf {e}$, joka yhdistetään $\mathbf {c}\_t$:n kanssa jakauman tuottamiseksi.

**Tulos**

Miten monikierroksinen dialogijärjestelmä oppii?

**Esimerkki 5.613**

Määritellään STransE-pisteytysfunktio $\mathcal {R}$1 seuraavasti:$ f\_r(h, t) & = & \Vert \textbf {W}\_{r,1}\textbf {h} + \textbf {r} - \textbf {W}\_{r,2}\textbf {t}\Vert \_{\ell \_{1/2}} $ käyttäen joko $\ell \_1$- tai $\ell \_2$ -normia (valinta tehdään validointidatan avulla; kokeiluissamme havaitsimme, että $\ell \_1$ -normi antoi hieman parempia tuloksia).

**Tulos**

Mitä pisteytysfunktiota malli käyttää kolmikoiden pisteyttämiseen?

**Esimerkki 5.614**

Taulukossa TABREF14 esitetään twiittien jakautuminen maittain ennen ja jälkeen suodatusprosessin. Suuri osa näytteistä on Intiasta, koska MeToo-liike on saavuttanut huippunsa Intiassa vuoden 2018 loppupuolella. Venäjältä on hyvin vähän näytteitä, mikä johtuu todennäköisesti sisällön moderoinnista ja sosiaalisen median käyttöä koskevista säännöksistä maassa.

**Tulos**

Tulevatko twiitit tietyltä alueelta?

**Esimerkki 5.615**

Tässä osassa käsitellään kolmea BLEU:n ja ROUGE:n merkittävää rajoitusta. Näillä mittareilla voidaan määrittää: BLEU:n ja ROUGE:n haasteet ::: Oletetaan, että meillä on viiteyhteenveto s1. Lisäämällä s1:een muutamia negaatio-termejä voidaan luoda tiivistelmä s2, joka on semanttisesti vastakkainen kuin s1, mutta jolla on silti korkea BLEU/ROUGE-pistemäärä.Challenges with BLEU and ROUGE ::: Sen lisäksi, että BLEU- ja ROUGE-pisteet eivät ole herkkiä negaatiolle, BLEU- ja ROUGE-pisteet voivat antaa alhaiset pisteet merkitykseltään vastaaville lauseille. Jos s2 on parafraasi s1:stä, merkitys on sama ;s1:n ja s2:n sanojen päällekkäisyys ei kuitenkaan välttämättä ole merkittävä.BLEU:n ja ROUGE:n haasteet ::: Kolmas BLEU:n ja ROUGE:n heikkous on se, että yksinkertaisimmissa toteutuksissaan ne eivät ole herkkiä sanojen permutaatiolle, ja ne voivat antaa hyvin korkeat pisteet käsittämättömille lauseille. Olkoon s1 "Aamulla a näin miehen juoksevan kadulla." ja s2 "Aamulla a näin juoksevan miehen kadulla". s2 ei ole ymmärrettävä lause. ROUGE:n ja BLEU:n unigram-versio antaa näille kahdelle lauseelle pistemäärän 1.

**Tulos**

Mitkä ovat nämä kolme rajoitusta?

**Esimerkki 5.616**

Jaksossa SECREF5 käsitellään kysymysten, tekstien ja vastausten keräämistä Amazon Mechanical Turk -palvelun (jäljempänä MTurk) avulla.

**Tulos**

mitä joukkoistamisalustaa käytettiin?

**Esimerkki 5.617**

Kaikkien kielten osalta 145 ihmisannotoijaa pyydettiin pisteyttämään kaikki 1 888 paria (omalla kielellään). Lopulta keräsimme vähintään kymmenen pätevää annotaatiota jokaiselle sanaparille kullakin kielellä. Kaikkien annotoijien oli noudatettava seuraavia ohjeita: 1. Kunkin annotoijan oli annettava kokonaispistemäärä väliltä 0-6 (mukaan lukien), joka osoittaa, kuinka semanttisesti samankaltaisia tietyn sanaparin kaksi sanaa ovat. Pistemäärä 6 tarkoittaa erittäin suurta samankaltaisuutta (eli täydellistä synonymiaa), kun taas nolla tarkoittaa, että samankaltaisuutta ei ole.2. Kunkin kommentoijan on pisteytettävä koko aineiston 1 888 sanaparin joukko. Pareja ei saa jakaa eri annotoijien kesken.3. Annotoijat voivat jakaa työmäärän noin 2-3 viikon ajalle, ja he voivat tarvittaessa käyttää ulkoisia lähteitä (esim. sanakirjoja, tesauruksia, WordNet).4. Annotoijat pidetään nimettöminä, eivätkä he voi kommunikoida toistensa kanssa annotointiprosessin aikana.

**Tulos**

Miten tietokokonaisuuksia kommentoitiin?

**Esimerkki 5.618**

Olemme havainneet, että konfliktimalli on hyvin herkkä pienillekin eroille ja kompensoi niitä sellaisissa tapauksissa, joissa huomio kohdistuu voimakkaasti samankaltaisuuksiin, jotka ovat jo olemassa sekvensseissä.Sekvenssi 1: Mitkä ovat parhaat tavat oppia ranskaa?Sekvenssi 2: Miten opettelen ranskan sukupuolet?Vain huomio: 1Ahuomio+konflikti: 0Pohjatotuus: 0Sekvenssi 1: Miten estän rintasyövän?Sekvenssi 2: Onko rintasyöpä estettävissä?Vain huomio: 1Ahuomio+konflikti: 0Pohjatotuus: 0Toimitamme kaksi esimerkkiä, joissa on ennusteita malleista, joissa on vain huomio ja huomion ja konfliktin yhdistelmä. Kunkin esimerkin mukana on aineistomme pohjatotuus.

**Tulos**

Näyttävätkö he, missä esimerkeissä konflikti toimii paremmin kuin huomio?

**Esimerkki 5.619**

Olemme myös tehneet alustavia kokeita SemEval-2016 cQA -tehtävän arabian kielen osuudella.

**Tulos**

Kokeilivatko he muilla kielillä?

**Esimerkki 5.620**

Osoitimme kuitenkin myös, että joissakin tapauksissa saliaatiokartat eivät näytä vangitsevan tärkeitä syötteen piirteitä. Toinen havainto, jonka voimme tehdä, on se, että saliaatiokartta ei näytä korostavan syötteessä oikeita asioita sen tuottaman yhteenvedon kannalta.

**Tulos**

Onko selitys saliaatiokartasta oikea?

**Esimerkki 5.621**

Käytimme upotuksissa $AraVec$ BIBREF30 arabian kielelle, FastText BIBREF31 ranskan kielelle ja Word2vec Google News BIBREF32 englannin kielelle.

**Tulos**

Mitä yksikielisiä sanaesityksiä käytetään?

**Esimerkki 5.622**

Kontekstualisoidut sanojen upotukset, lauseiden upotukset, kuten syvät kontekstualisoidut sanaesitykset BIBREF20 , BERT BIBREF22 , koodaavat sanojen monimutkaiset ominaisuudet ja merkitykset eri yhteyksissä kouluttamalla yhdessä kaksisuuntaista kielimallia. Toinen menetelmä on käyttää lauseen upotuksia, BERT. Sen avulla luodaan yksi 786-ulotteinen lauseen sulautuma 10k-dimensionaalisesta yhden hotsin vektorista tai jakaumasta edellisiin sanoihin ja yhdistetään sitten yhdeksi kontekstivektoriksi kahdella eri yhdistämismenetelmällä. Tämän lähestymistavan avulla saadaan tiheämpi, informatiivisempi, kiinteäpituinen vektori koodaamaan keskusteluyhteyteen liittyvää informaatiota, $e^k\_{context}$, jota voidaan käyttää seuraavan $k$ -kymmenennen lausuman ennustamisessa. Käytämme dekooderiverkossamme kontekstuaalista gating-mekanismia yhdistääksemme keskustelukontekstin upotukset tehokkaasti puheen ja sanojen upotuksiin. Porttiarvomme on kontekstisidonnainen siinä mielessä, että useat upotukset laskevat porttiarvon, joka riippuu useiden keskustelussa esiintyvien lausumien kontekstista. Olkoon $e\_w = e\_w(y\_{u-1})$ edellinen sanan upotus sanalle $y\_{u-1}$ , ja olkoon $e\_s = e\_s(x^k\_{1:T})$ puheen upotus nykyisen $k$ -ennen lausuman $x^k\_{1:T}$ ja $e\_c = e\_c(s\_{k-1-n:k-1})$ on keskustelukontekstin upotus $n$ -lukua edeltäville lausumille ${s\_{k-1-n:k-1}}$ . Käytetään sitten porttimekanismia: $$g = \sigma (e\_c, e\_w, e\_s)$$ (Yht. 15)jossa $\sigma $ on 1 piilokerroksen DNN, jolla on $\texttt {sigmoid}$-aktivointi, porttikytkentä $e$ lasketaan seuraavasti: $$$e = g \odot (e\_c, e\_w, e\_s) \\\ h = \text{LSTM}(e)$$ (Yht. 16)$$ (Yht. 16)$$ ja syötetään LSTM-dekooderin piilokerrokseen. Dekooderin $h$ ulostulo yhdistetään sitten keskusteluyhteyden upottamiseen $e\_c$ jälleen porttausmekanismin avulla,$$g = \sigma (e\_C, h) \\\ \hat{h} = g \odot (e\_c, h)$$ (yhtälö 17)Sitten seuraava piilokerros ottaa nämä portatut aktivoinnit, $\hat{h}$ , ja niin edelleen.

**Tulos**

Miten lauseen sulautukset sisällytetään puheentunnistusjärjestelmään?

**Esimerkki 5.623**

Arvioimme malliamme julkisesti saatavilla olevalla aineistolla, KARA ONE BIBREF17 , joka koostuu multimodaalisesta aineistosta ärsykkeisiin perustuvasta, kuvitellusta ja artikuloidusta puhetilasta, joka vastaa 7:ää foneemista/syllabista ( /iy/, /piy/, /tiy/, /diy/, /uw/, /m/, /n/ ) sekä 4:ää sanaa (pat, pot, knew ja gnaw). Aineisto koostuu 14 osallistujasta, ja jokainen kehote esitettiin 11 kertaa kullekin yksilölle.

**Tulos**

Kuinka monesta koehenkilöstä EEG-tiedot ovat peräisin?

**Tulos**

Mitä tietoja koehenkilöille esitettiin tapahtumiin liittyvien reaktioiden aikaansaamiseksi?

**Esimerkki 5.624**

Lähestymistapamme parantaa LR:ää keskimäärin 5,17 % (tarkkuus) ja 18,38 % (AUC) ja MLP:tä 10,71 % (tarkkuus) ja 30,27 % (AUC). Tällaiset merkittävät parannukset osoittavat selvästi, että lähestymistapamme parantaa tehokkaasti mallien suorituskykyä.

**Tulos**

Miten lähestymistavan tarkkuus on osoitettu?

**Esimerkki 5.625**

DrQA on CRC-perustaso, joka on tulossa CoQA-tietokannan kanssa. Huomaa, että tämä DrQA-toteutus eroaa SQuAD BIBREF8:n DrQA-toteutuksesta siten, että se on muutettu tukemaan vastaamattomien kysymysten vastaamista siten, että asiakirjan lopussa on erityinen tuntematon merkki. DrQA+CoQA on edellä mainittu perustaso, joka on esiviritetty CoQA-tietokannalla ja sitten hienosäädetty $(\text{RC})\_2$ -tietokannalla. BERT on vanilla BERT-malli, joka on hienosäädetty suoraan $(\text{RC})\_2$:lle. Käytämme tätä perustasoa ablaatiotutkimukseen esivirityksen tehokkuudesta. BERT+review virittää ensin BERT:n toimialueen arvosteluihin käyttäen samoja tavoitteita kuin BERT:n esivalmennuksessa ja hienovirittää sitten $(\text{RC})\_2$ . Käytämme tätä perustasoa osoittaaksemme, että yhteistyöelimen pelkkä toimialueen mukauttaminen ei ole hyvä. BERT+CoQA hienosäätää ensin yhteistyöelimen valvotun CoQA-datan perusteella ja hienosäätää sitten $(\text{RC})\_2$:n perusteella. Käytämme tätä perustasoa osoittaaksemme, että esiviritys on erittäin kilpailukykyinen jopa verrattuna malleihin, jotka on koulutettu laajamittaisesta valvotusta datasta.

**Tulos**

Mikä on käytetty perusmalli?

**Esimerkki 5.626**

Suoritimme ehdotetun järjestelmän luotettavuustutkimuksen käyttäen kahta asiantuntijalausunnonantajaparia, P1 ja P2. Arvioinnissa käytettiin Cohenin kappaa, joka on painottamaton argumentoinnin ja tietämysalueen osalta, mutta kvadraattipainotettu spesifisyyden osalta, kun otetaan huomioon sen järjestetyt merkinnät.

**Tulos**

mitä kokeita tehdään?

**Esimerkki 5.627**

Sanan kontekstittomat ominaisuudet. Nämä ominaisuudet johdetaan suoraan sanasta, ja ne kuvaavat yleistä taipumusta siihen, että sana kaikuu selityksissä.

**Tulos**

Mihin kontekstittomiin ominaisuuksiin ne viittaavat?

**Esimerkki 5.628**

Lähes kaikissa genreissä DenseNMT-mallit ovat huomattavasti parempia kuin perusmallit.

**Tulos**

Ovatko ne olleet parempia kuin aiemmat menetelmät?

**Esimerkki 5.629**

Pyrimme löytämään tällaista sisältöä sosiaalisesta mediasta keskittyen twiitteihin.

**Tulos**

Sisältääkö tietokokonaisuus sisältöä eri sosiaalisen median alustoilta?

**Esimerkki 5.630**

LDA-mallia käytettäessä jokaisella henkilöllä aineistossa on aiheen todennäköisyysvektori $X\_i$ . Oletetaan, että $x\_{ik}\in X\_{i}$ tarkoittaa todennäköisyyttä sille, että $\emph {i}^{th}}$ twiittitili suosii $\emph {k}^{th}$ aihetta datasetissä. Aiheeseen perustuvat ominaisuutemme voidaan laskea seuraavasti. Global Outlier Standard Score mittaa sitä, missä määrin käyttäjän twiittisisältö liittyy tiettyyn aiheeseen verrattuna muihin käyttäjiin. Local Outlier Standard Score mittaa sitä, kuinka kiinnostunut joku on tietystä aiheesta, kun otetaan huomioon vain hänen oman kotisivunsa sisältö. Kolme perustason luokittelumenetelmää: SVM, Adaboost ja Random Forests -luokitusmenetelmiä käytetään arvioimaan poimittuja ominaisuuksia.

**Tulos**

Miten he havaitsevat roskapostittajat?

**Esimerkki 5.631**

Kehitämme nollamallin määrittääksemme, kuinka paljon vaihtelua binomijärjestelyissä voitaisiin odottaa eri yhteisöjen välillä ja eri aikoina, jos binomijärjestelyt järjestettäisiin satunnaisesti globaalien epäsymmetria-arvojen mukaan.

**Tulos**

Mitä uutta mallia ehdotetaan binomilistoille?

**Esimerkki 5.632**

Alustan käyttäjät antavat vaikuttavuusääniä arvioidakseen, kuinka vaikuttava tietty väite on. Käyttäjät voivat valita yhden viidestä mahdollisesta väitteen vaikutusmerkinnästä: ei vaikutusta, vähäinen vaikutus, keskisuuri vaikutus, suuri vaikutus ja erittäin suuri vaikutus. Arvioidessaan väitteen vaikutusta käyttäjät voivat tutustua väitteen koko asiayhteyteen, joten he voivat arvioida, kuinka vaikuttava väite on väitteen tietyssä asiayhteydessä.

**Tulos**

Mitä merkintöjä tietokokonaisuudessa on saatavilla?

**Esimerkki 5.633**

Analyysissämme havaittiin, että jengiläiset käyttävät mielellään vain pientä joukkoa emoji-symboleita, jotka välittävät vihaa ja väkivaltaista käyttäytymistä twiiteissään. Kuvassa FIGREF24 on havainnollistettu aineistomme jengiläisprofiilien 20 yleisimmin käyttämän emojin jakauma. Polttoainepumppuemoji oli jengiläisten useimmin käyttämä emoji, jota käytetään usein marihuanan myynnin tai kulutuksen yhteydessä. Pistooli-emoji on aineistossamme toiseksi yleisin emoji, jota käytetään usein yhdessä vartijaemojin tai poliisi-emojin kanssa "emojiketjussa". Kuvassa FIGREF28 esitetään joitakin prototyyppisiä jengiläisten käyttämiä emojien "ketjuja". Ketjut saattavat kuvastaa heidän vihaansa lainvalvontaviranomaisia kohtaan, sillä poliisi-emojia seuraa usein aseen, pommin tai räjähdyksen emoji. Havaitsimme, että 32,25 prosenttia aineistomme jengiläisistä on ketjuttanut poliisi- ja pistooli-emojin yhteen, kun taas jengiin kuulumattomista vain 1,14 prosenttia. Lisäksi vain 1,71 % ei-jengiläisistä on käyttänyt sadan pisteen emojia ja pistooli-emojia yhdessä twiiteissä, kun taas 53 % jengiläisistä on käyttänyt niitä. Erilaiset vihaiset kasvot-emojit, kuten pirun kasvot-emoji ja imp-emoji, olivat myös yleisiä jengiläisten twiiteissä.

**Tulos**

Mitä eroja emojien käytössä on jengiläisten ja muun Twitter-väestön välillä?

**Esimerkki 5.634**

Arvioidaksemme lähestymistapamme suorituskykyä käytimme osajoukkoa SNIPS BIBREF12 -tietokannasta, joka on helposti saatavilla RASA nlu -muodossa. Tehtävänä ja mittarina käytämme tarkoituksen ja entiteetin tunnistamisen tarkkuutta.

**Tulos**

Miten niiden muutoksia arvioidaan?

**Esimerkki 5.635**

Mallissa on noin 836 miljoonaa parametria, joista vain 66 000 on tavun upotuksia.

**Tulos**

Kuinka monta parametria mallissa on?

**Esimerkki 5.636**

Opi jokaiselle koodausjoukon muuttujalle uudet upotukset käyttäen upotusten harjoittelusarjaa . Koska upotusten koulutusmallissa on stokastisuutta, toistamme edellä esitetyn useita kertoja eri kokeita varten (ja ilmoitamme vastaavat keskiarvo- ja keskihajontatilastot).

**Tulos**

Miten he harjoittelevat upotuksiaan?

**Esimerkki 5.637**

Sen jälkeen koulutamme sensaatiohakuisuuden pisteytyksen luokittelemalla sensaatiohakuiset ja ei-sensaatiohakuiset otsikot käyttämällä yksikerroksista CNN:ää, jossa on binäärinen ristikkäisentropiahäviö $L\_{\text{sen}}$. Ensin käytetään 1-D-konvoluutiota sanaominaisuuksien poimimiseen otsikon syötteestä. Tämän jälkeen seuraa ReLU-aktivointikerros ja maksimipoolauskerros aikaulottuvuutta pitkin. Kaikki eri kanavista saadut piirteet yhdistetään ja projisoidaan sensationalismipisteytykseen lisäämällä toinen täysin kytketty kerros, jossa on sigmoidiaktivointi. Binääristä ristiinentropiaa käytetään tappion $L\_{\text{sen}}$ laskemiseen.

**Tulos**

Miten sensaatiohakuisuus on koulutettu?

**Esimerkki 5.638**

Kaavan mukaisesti voimme laskea jokaisen syötesanan osuuden jokaisesta tulossanasta, jolloin muodostuu osuustaulukko, jonka koko on $M \ kertaa N$, jossa $N$ on tulostetun lauseen pituus. Kun otetaan huomioon kontribuutiomatriisi, saadaan kunkin syötetyn sanan merkitys koko tuloslauseen kannalta. Tätä varten kunkin syötesanan osalta lasketaan ensin yhteen sen kontribuutioarvot kaikkiin tulossanoihin summaoperaatiolla ja normalisoidaan sitten kaikki summat Softmax-funktion avulla.

**Tulos**

Miten heidän mallinsa päättävät, kuinka paljon parannusta annetaan tulossanoille?

**Esimerkki 5.639**

Seq2seq-malli globaalilla huomiolla antaa parhaat tulokset, kun keskimääräinen BLEU-pistemäärä on 29,65 tyylinsiirtotietokannassa, kun keskimääräinen BLEU-pistemäärä on 26,97 käytettäessä seq2seq-mallia osoitinverkkojen kanssa.

**Tulos**

Mikä on paras BLEU-pistemäärä, jonka kielityylinsiirron tekijät saivat?

**Esimerkki 5.640**

Tässä asiakirjassa käytämme kolmea kirjallisuudesta löytyvää aineistoa oman luokittelijamme kouluttamiseen ja arviointiin. BIBREFin3 keräämät tiedot, joita kutsumme Sexist/Racist (SR) -datajoukoksi, kerättiin alustavalla Twitter-haulla, jonka jälkeen kirjoittajat ja heidän tiiminsä analysoivat ja suodattivat ne. He tunnistivat 17 yleistä lausetta, hashtagia ja käyttäjää, jotka viittasivat loukkaavaan puheeseen. BIBREF4 keräsi HATE-tietoaineiston etsimällä twiittejä Hatebase.orgin tarjoaman sanaston avulla. Lopullisen käyttämämme aineiston, jota kutsumme nimellä HAR, keräsi BIBREF9 ; poistimme kaikki uudelleentwiittaukset ja pienensimme aineiston 20 000 twiittiin. Twiitit merkittiin "häiritseviksi" tai "ei-häiritseviksi"; vihapuhetta ei nimenomaisesti merkitty, vaan sitä käsiteltiin merkitsemättömänä osajoukkona laajemmasta "häiritsevä"-luokasta BIBREF9 .

**Tulos**

Mitä julkisesti saatavilla olevia tietokokonaisuuksia käytetään?

**Esimerkki 5.641**

Huomasimme kuitenkin kaikkien multimodaalisten mallien kohdalla yhteisen virhekäyttäytymisen: jos huomio katoaa kuvassa olevista kohteista ja "eksyy", malli ottaa sen silti huomioon ja jotenkin ohittaa tekstipohjaisten merkintöjen tuoman informaation. Käännös on tällöin täysin harhaanjohtava.

**Tulos**

Mitä väärinkäytöksiä on havaittu?

**Esimerkki 5.642**

Yhdistämme 13 000 kysymystä, jotka on kerätty tästä putkesta, ja 3 000 kysymystä, joihin on annettu kyllä/ei-vastauksia NQ-harjoitusjoukosta, jotta saamme yhteensä 16 000 kysymystä.

**Tulos**

Mikä on BoolQ-tietokannan koko?

**Esimerkki 5.643**

Tässä artikkelissa ehdotamme, että jokainen sana esitetään ilmeikkäällä multimodaalisella jakaumalla, joka mahdollistaa useat eri merkitykset, seuraussuhteet, epävarmuuden ja paremman tulkittavuuden. Esimerkiksi sanan "pankki" yksi moodi voisi olla päällekkäinen sellaisten sanojen kuin "rahoitus" ja "raha" jakaumien kanssa, ja toinen moodi voisi olla päällekkäinen sanojen "joki" ja "puro" jakaumien kanssa. Väitämme, että tällainen joustavuus on ratkaisevan tärkeää sekä sanojen merkitysten laadullisen oppimisen että optimaalisen suorituskyvyn kannalta monissa ennustustehtävissä.

**Tulos**

Miten tämä vertautuu kontekstuaalisiin upotusmenetelmiin?

**Esimerkki 5.644**

Arvioimme lauseiden spesifisyyttä joukolla kolmen kohdealueen osalta: Twitter, Yelp-arvostelut ja elokuva-arvostelut.

**Tulos**

Millä aloilla he kokeilevat?

**Esimerkki 5.645**

Tässä yhteydessä päätimme tutkia kolmea tärkeintä arkkitehtuurityyppiä, jotka ovat osoittaneet lupaavia tuloksia perinteisiin järjestelmiin verrattuna: 1) Connectionist Temporal Classification (CTC) BIBREF5, BIBREF6 , joka käyttää Markov-oletuksia (ts. ehdollinen riippumattomuus ennusteiden välillä kullakin aika-askeleella) ratkaistakseen tehokkaasti peräkkäisiä ongelmia dynaamisen ohjelmoinnin avulla, 2) Huomioon perustuvat menetelmät BIBREF7, BIBREF8, jotka luottavat huomiomekanismiin suorittaakseen ei-monotonisen yhdenmukaistamisen akustisten kehysten ja tunnistettujen akustisten yksiköiden välillä, ja 3) RNN-tulkki BIBREF0, BIBREF9, BIBREF10, joka laajentaa CTC:tä siten, että se mallintaa lisäksi eri vaiheissa tapahtuvien ulostulojen väliset riippuvuudet käyttämällä kielimallin kaltaista ennustusverkkoa.

**Tulos**

Mitkä ovat nykyiset end-to-end ASR-lähestymistavat ranskan kieltä varten?

**Esimerkki 5.646**

Näin ollen kartoissa voidaan erottaa kaksi pääaluetta tai klusteria. Violetti tausta peittää suurimman osan kartasta ja edustaa maaseutualueita, joilla on pieni, hajallaan oleva väestö. Analyysimme osoittaa, että tällä soluryhmällä on enemmän erityisiä sanoja sanastossaan. Vihreät ja keltaiset solut sen sijaan muodostavat toisen klusterin, joka on suurelta osin keskittynyt keskustaan ja rannikolle, jotka vastaavat suurkaupunkeja ja teollisuusalueita. Näissä soluissa espanjan kielen vakiokielen käyttö on yleistä, mikä johtuu todennäköisesti kouluopetuksesta, tiedotusvälineistä, matkailijoista jne.

**Tulos**

Mainitsevatko kirjoittajat tutkimuksessaan mahdollisia sekaannuksia?

**Esimerkki 5.647**

Yleensä virheellisen puhuja- ja/tai yleisötiedon antaminen laskee BLEU-pistemäärää, kun taas oikean tiedon antaminen parantaa sitä huomattavasti - jopa 2,3 BLEU:n lisäys perustasoon verrattuna. Näemme, että perusjärjestelmä ennustaa verbien feminiinimuodon huomattavasti liian huonosti vertailuun verrattuna. "Hän sanoi" -ehdot vähentävät feminiinisten verbien määrää entisestään, kun taas "Minä sanoin" -ehdot palauttavat sen takaisin perustasolle. Lopuksi "Hän sanoi" -prefiksit lisäävät huomattavasti feminiinimerkittyjen verbien määrää, jolloin niiden osuus on paljon lähempänä vertailua (vaikkakin osa feminiinisistä tapauksista on edelleen aliennustettuja).

**Tulos**

Miten osoitetaan, että tässä järjestelmässä syötetään oikeat sukupuoli- ja numerotiedot?

**Esimerkki 5.648**

Testataksemme ehdotetun MWA-huomion soveltuvuutta valitsemme kolme julkisesti saatavilla olevaa kiinalaista esivalmennettua mallia peruskooderiksi: BERT, ERNIE ja BERT-wwm. Oikeudenmukaisen vertailun tekemiseksi pidämme samat hyperparametrit (kuten enimmäispituus, lämmittelyvaiheet, alkuperäinen oppimisnopeus jne.), joita ehdotetaan BERT-wwm BIBREF13 -mallissa, sekä perusmalleille että menetelmällemme kussakin tietokokonaisuudessa.

**Tulos**

Mihin esivalmennettuihin malleihin niitä verrattiin?

**Esimerkki 5.649**

Toisin kuin useimmat koodaaja-dekooderiarkkitehtuurit, jotka perustuvat puhtaasti rekursiivisiin neuroverkkoihin (RNN), me rakennamme koodaimen, jossa on useita konvoluutiokerroksia BIBREF14 ja sen jälkeen NIN-kerroksia BIBREF15 koodaimen alempana ja yhdistämme ne syvään kaksisuuntaiseen kaksisuuntaiseen pitkäkestoiseen lyhytkestoiseen muistiin (Bi-LSTM)16 ylemmässä osassa. Dekooderin puolella käytämme tavallista syvää yksisuuntaista LSTM:ää, jossa on globaali huomio BIBREF13 ja joka lasketaan monikerroksisen perceptronin (MLP) avulla yhtälössä EQREF2 kuvatulla tavalla.

**Tulos**

Mitä arkkitehtuuria ne käyttävät kooderissa ja dekooderissa?

**Esimerkki 5.650**

Tarkistimme manuaalisesti 1177 paria järjestelmän tuottamia entiteettejä ja viittausilmauksia. Kaiken kaikkiaan 73,3 prosenttia entiteettien muista kuin ensimmäisistä maininnoista korvattiin sopivilla lyhyemmillä ja sujuvammilla ilmauksilla.

**Tulos**

Miten tuotetun tekstin sujuvuutta arvioidaan?

**Esimerkki 5.651**

Kielen mallintamisen arviointia varten arvioimme myös perustasoa ilman tiedon tislausta (NoKD), jossa malli on parametrisoitu samalla tavalla kuin tislattujen oppilasmallien parametrit, mutta se on koulutettu suoraan opettajan mallin tavoitteeseen tyhjästä. Jälkikäsiteltävien tehtävien osalta vertaamme NoKD:hen sekä BIBREF34:n Patient Knowledge Distillation (PKD) -menetelmään, joka tislaa 12-kerroksisen BERTBASE-mallin 3- ja 6-kerroksisiksi BERT-malleiksi käyttämällä opettajamallin piilotettuja tiloja.

**Tulos**

Mitä uusimpia pakkaustekniikoita vertailussa käytettiin?

**Esimerkki 5.652**

Taulukossa TABREF11 on esitetty muutamia esimerkkejä englanninkielisistä neuvontasäännöistä (jotka on muunnettu ensimmäisen kertaluvun logiikan muotoon ja annettu algoritmimme syötteenä). Muokkasimme Odom et al. odomAIME15,odomAAAI15:n työtä oppiaksemme RDN-sääntöjä neuvojen läsnä ollessa. Keskeinen ajatus on edustaa neuvoja eksplisiittisesti gradientteja laskettaessa.

**Tulos**

Miten ne sisällyttävät inhimilliset neuvot?

**Esimerkki 5.653**

Yksi ratkaisu, jota on ehdotettu sukupuoleen perustuvan vinoutuman lieventämiseksi sanojen upottamisen tasolla, on Counterfactual Data Augmentation (CDA) BIBREF25. Sovellamme tätä menetelmää täydentämällä tietokokonaisuuttamme kopiolla jokaisesta vuoropuhelusta, jossa sukupuolittuneet sanat on vaihdettu käyttäen BIBREF21:n toimittamaa sukupuolittuneiden sanaparien luetteloa.

**Tulos**

Miten kontrafaktuaalisen tiedon lisäämisen avulla pyritään torjumaan harhaa?

**Esimerkki 5.654**

Tämän artikkelin koeasetukset ja Mboshi-korpuksen arviointiprotokolla (Boundary F-pisteet käyttäen ZRC-puheviitettä) ovat samat kuin BIBREF8:ssa. Taulukossa esitetään tulokset kaksikielisen UWS:n ja monikielisen vivutuksen osalta. Ensin mainitussa tapauksessa saavutamme parhaan tuloksemme käyttämällä yhdenmukaistettuna tietona ranskaa, joka on tämän tietokokonaisuuden alkuperäinen yhdenmukaistettu kieli. Monikielistä valintaa varten kokeilimme kielten yhdistämistä ylhäältä alaspäin, kuten ne näkyvät taulukossa (paremmuusjärjestyksessä suorituskyvyn mukaan; esim. 1-3 tarkoittaa FR(1), EN(2) ja PT(3) yhdistelmää). Huomasimme, että suorituskyvyn paraneminen on pienempi kuin aiemmassa BIBREF10-työssä, mikä johtuu siitä, että tietokokonaisuuttamme on lisätty keinotekoisesti. Lopuksi, BIBREF8:n menetelmää noudattaen, poimimme kaksikielisten mallien löytämät varmimmat linjaukset (ANE:n suhteen). Taulukossa esitetään 10 varminta (löydetty tyyppi, käännös) paria. Kun tarkastellaan kaksikielisten mallien varmimpia pareja, havaitaan, että kaikki kaksikieliset mallit ovat löytäneet joitakin tyyppejä (esim. Mboshi-sana itua ja yhdistelmä oboá+ngá).

**Tulos**

Miten mallin suorituskykyä arvioidaan?

**Esimerkki 5.655**

Valvontakriteerit lasketaan heuristisesti. Tämän jälkeen kaikki kontrollisanaluetteloista löytyneet sanat poistetaan vertailulauseesta. Jäljelle jääviä sanoja, jotka edustavat sisältöä, käytetään mallin syötteenä yhdessä niiden POS-tunnisteiden ja lemman kanssa.Tällä tavoin rohkaisemme malleja rakentamaan lauseen sisällön ja tyylin avulla itsenäisesti.

**Tulos**

Mistä he tietävät, mitkä ovat sisällön sanoja?

**Esimerkki 5.656**

E2ECM BIBREF11: Vuoropuhelupolitiikassa se käyttää klassista luokittelua luurankomallin lauseen mallin osalta. Toteutuksessamme rakennamme useita binääriluokituksia kullekin teolle lauseen mallin etsimistä varten BIBREF11:n ehdottaman työn mukaisesti. CDM BIBREF10: Tässä lähestymistavassa suunnitellaan joukko luokituksia (kaksi moniluokkaista luokitusta ja joitakin binääriluokituksia) dialogipolitiikan mallintamiseksi.

**Tulos**

Mitä ovat uusimmat peruslähtökohdat?

**Esimerkki 5.657**

ConMaskissa käytämme samankaltaista ideaa valitaksemme eniten toisiinsa liittyvät sanat, kun jokin suhde on annettu, ja peittääksemme epäolennaiset sanat antamalla suhteesta riippuvaisen samankaltaisuuspisteytyksen sanoille annetussa olion kuvauksessa. ConMask valitsee sanat, jotka liittyvät annettuun suhteeseen, jotta epäolennaisten ja häiriöalttiiden sanojen sisällyttämistä voidaan vähentää.

**Tulos**

Voiko malli lisätä tietämysgraafiin uusia suhteita vai vain uusia entiteettejä?

**Esimerkki 5.658**

Lisäksi arvioimme esivalmennuksen tehokkuutta vertaamalla sitä yhdessä koulutettuun malliin, joka koostuu etukääntämisestä ja takakääntämisestä. Kokeelliset tulokset osoittavat, että koodaaja-dekooderi-rekonstruktori tarjoaa merkittävän parannuksen BLEU-pisteisiin ja lieventää englannin ja japanin käännöstehtävän toistuvien ja puuttuvien sanojen ongelmaa käännöksessä, eikä koodaaja-dekooderi-rekonstruktoria voida kouluttaa hyvin ilman esivalmennusta, joten se osoittaa, että meidän on koulutettava eteenpäin kääntämisen malli perinteisen tarkkaavaisuuspohjaisen NMT:n kaltaisella tavalla esivalmennuksena.

**Tulos**

Onko esivalmennus tehokasta niiden arvioinnissa?

**Esimerkki 5.659**

Mallinnamme ympäristöstä, tavoitteesta ja asiakirjasta tehtyjen havaintojen välistä vuorovaikutusta kerrosten avulla. Ensin koodaamme tekstisisällöt käyttämällä kaksisuuntaisia LSTM:iä, sitten laskemme yhteenvedot käyttämällä itsehuomiota ja ehdolliset yhteenvedot käyttämällä huomiota. Yhdistämme tekstin yhteenvedot tekstin ominaisuuksiksi, joita käsitellään yhdessä visuaalisten ominaisuuksien kanssa peräkkäisten kerrosten kautta. Tässä tekstiympäristön tapauksessa pidämme sanojen upotusten ruudukkoa visuaalisina piirteinä . Lopullinen tuloste käsitellään edelleen MLP:llä toimintatapajakauman laskemiseksi toimille ja perustason laskemiseksi etujen arviointia varten.

**Tulos**

Miten ehdotetaan mallia, joka kaappaa kolmisuuntaiset vuorovaikutussuhteet?

**Esimerkki 5.660**

Tässä asiakirjassa käytämme kolmea koodaajaa (NBOW, LSTM ja Attentive LSTM) tekstikuvausten mallintamiseen.

**Tulos**

Mitä neuraalisia malleja käytetään tekstin koodaamiseen?

**Esimerkki 5.661**

Twitter-tiedot: Käytimme Twitterin API:ta hakusanoja sisältävien twiittien keräämiseen. API suodatti kaikki muut kuin englanninkieliset twiitit pois.

**Tulos**

Arvioidaanko ne vain englanninkielisillä tietokokonaisuuksilla?

**Esimerkki 5.662**

Käytämme ceccarelli2013learning-ohjelmalla luotua vertailutietoaineistoa CoNLL 2003 -aineistosta.

**Tulos**

Mikä on vertailutietokanta?

**Esimerkki 5.663**

(Tiedonhaku). Tätä perustasoa on käytetty menestyksekkäästi vastaavissa tehtävissä, kuten kysymyksiin vastaamisessa BIBREF39 . Luomme kaksi versiota tästä perustasosta: toinen, jossa on INLINEFORM0-näkökulmia ja toinen, jossa on INLINEFORM1-todisteet.

**Tulos**

Mitä koneen peruslinjoja käytetään?

**Esimerkki 5.664**

Taulukossa IV esitetään tunteisiin perustuvan luokittelumallimme tulokset verrattuna muihin malleihin. Vertaamme mallimme suorituskykyä BIBREF0- ja BIBREF5-menetelmiin STS-korpuksella. BIBREF0 raportoi Maximum Entropy (MaxEnt), NB ja SVM -mallien tulokset STS-korpuksen osalta, ja niiden suorituskyky oli hyvä aiemmalla kerralla. BIBREF5:n malli on toistaiseksi huippuluokkaa käyttämällä CharSCNN:ää. Kuten voidaan nähdä, 86,63 on toistaiseksi paras ennustustarkkuus, jonka mallimme on saavuttanut STS-korpuksen osalta. Sanders- ja HCR-tietokokonaisuuksien osalta vertaamme tuloksia BIBREF14-malliin, jossa käytettiin useiden perusluokittelijoiden (ENS), kuten NB, Random Forest (RF), SVM ja Logistic Regression (LR), muodostamaa kokonaisuutta. ENS-malliin on yhdistetty sanasäkki (BoW), ominaisuuksien hassaus (FH) ja leksikonit. BIBREF14-malli on huippuluokkaa Sanders- ja HCR-tietoaineistoissa. Mallimme ovat parempia kuin BIBREF14:n malli Sanders- ja HCR-tietokannoissa.

**Tulos**

Mikä oli lähtötaso?

**Esimerkki 5.665**

Molemmissa tietokokonaisuuksissa noudatamme alkuperäisissä BIBREF13 -arviointitehtävissä käytettyjä arviointimittareita. DBQA:n osalta käytetään P@1 (Precision@1), MAP (Mean Average Precision) ja MRR (Mean Reciprocal Rank). Koska KBRE:ssä kullekin kysymykselle merkitään vain yksi kultainen ehdokas, käytetään vain P@1- ja MRR-arvoja.

**Tulos**

Mitä mittareita he käyttävät arvioitaessa yhteensopivuutta?

**Esimerkki 5.666**

Teimme myös laadullisen tutkimuksen Starbucksin (SBUX) osakkeen liikkeistä tämän tapahtuman aikana. Kuvio FIGREF12 on SBUXin ja NASDAQ-indeksin päivittäinen prosentuaalinen muutos 11. huhtikuuta ja 20. huhtikuuta välisenä aikana. SBUX ei seurannut koko markkinoiden noususuuntausta ennen 17. huhtikuuta, ja sitten sen muutos 20. huhtikuuta, INLINEFORM0 , on melko merkittävä historiallisista normeista. Keräsimme historialliset 52 viikon osakekurssit ennen tätä tapahtumaa ja laskimme päivittäisen osakekurssimuutoksen. Edellisten 52 viikon päivittäisen kurssimuutoksen jakauma on kuvio KUVIO FIGREF13, jonka keskiarvo on INLINEFORM1 ja keskihajonta INLINEFORM2 .

**Tulos**

Miten menetelmällä mitataan tapahtuman vaikutusta markkinahintoihin?

**Esimerkki 5.667**

Enintään kolme osallistujaa kuuntelee minkä tahansa äänileikkeen. Jos $<$audio,transkriptio$>$ -parille annetaan ensin kaksi ääntä ylöspäin, leike merkitään kelvolliseksi. Jos sen sijaan klippi saa ensin kaksi kielteistä ääntä, se merkitään mitättömäksi. Osallistuja voi halutessaan vaihtaa tallennuksen ja validoinnin välillä. Vain kelvollisiksi merkityt klipit sisällytetään kunkin kielen virallisiin koulutus-, kehitys- ja testisarjoihin. Klipit, jotka eivät ole saaneet riittävästi ääniä validointia tai mitätöintiä varten julkaisuhetkellä, julkaistaan nimellä "muut". Koulutus-, testi- ja kehitysjoukot on jaoteltu siten, että yksittäinen puhuja voi esiintyä vain yhdessä niistä. Näin varmistetaan, että harjoitusaikana nähtyjä avustajia ei nähdä testiaikana, mikä vääristäisi tuloksia. Lisäksi tekstilauseiden toistot poistetaan korpuksen harjoitus-, testi- ja kehitysjoukoista.

**Tulos**

Miten tiedot validoidaan?

**Esimerkki 5.668**

Tämän mielenkiintoisen ominaisuuden tallentamiseksi ehdotamme uutta merkintäjärjestelmää, joka koostuu kolmesta tunnisteesta, nimittäin { INLINEFORM0 }.INLINEFORM0-tunniste osoittaa, että nykyinen sana esiintyy ennen sanaleikkiä kyseisessä kontekstissa.INLINEFORM0-tunniste korostaa, että nykyinen sana on sanaleikki.INLINEFORM0-tunniste osoittaa, että nykyinen sana esiintyy sanaleikin jälkeen.

**Tulos**

Mikä on käytetty merkintäjärjestelmä?

**Esimerkki 5.669**

Tarkemmin sanottuna käytämme kutakin Q\_\mathit {tr}$:n monivalintakysymystä $(q,A) \in Q\_\mathit {tr}$ ja kutakin A$:n valintaa $a \in A$ varten kaikkia $q$:n ja $a$:n sisältämiä muita kuin pysähtymättömiä sanamerkkejä ElasticSearch-kyselynä S:lle. Otamme 200 suurinta osumaa, ajamme Open IE v4:n ja yhdistämme tuloksena saadut tuplat kaikkien $a \in A$:n ja kaikkien Q\_\mathit {tr}$:n sisältämiin kysymyksiin luodaksemme tuplan KB (T).

**Tulos**

Mitä OpenIE-menetelmää käytettiin uutteiden tuottamiseen?

**Esimerkki 5.670**

Tärkeimmät tuloksemme esitetään taulukossa TABREF6 viranomaisilmoitusten testijoukon osalta ja taulukossa TABREF7 kiinteistöjen vuokrasopimusten testijoukon osalta; F$\_1$, tarkkuus ja palautus lasketaan edellä kuvatulla tavalla.

**Tulos**

Mitä arviointimittareita käytettiin tulosten esittämiseen?

**Esimerkki 5.671**

Se alkaa tavallisena turnausvalinnan evoluutioarkkitehtuurihakuna, jossa on varhainen pysäytys, jolloin jokainen lapsimalli harjoittelee suhteellisen pienen $s\_0$ määrän askelia ennen kuin sen soveltuvuus arvioidaan. Kun ennalta määrätty määrä lapsimalleja, $m$ , on arvioitu, luodaan kuitenkin este, $h\_0$ , laskemalla nykyisen populaation keskimääräinen kunto. Seuraaville $m$ tuotetuille lapsimalleille, joiden kunto on suurempi kuin $h\_0$ $s\_0$ koulutusvaiheen jälkeen, annetaan $s\_1$ lisäkoulutusvaihetta, minkä jälkeen ne arvioidaan uudelleen lopullisen kuntonsa määrittämiseksi. Kun toiset $m$ mallia on tarkasteltu tällä tavalla, muodostetaan toinen este, $h\_1$ , laskemalla kaikkien niiden nykyisen populaation jäsenten keskimääräinen kunto, joita on koulutettu maksimimäärän askelia. Seuraavien $m$$ lapsimallien kohdalla koulutus ja arviointi jatkuu samalla tavalla, paitsi että malleille, joiden kunto on suurempi kuin $m$0 $m$1 koulutusaskeleen jälkeen, annetaan $m$2 lisää koulutusaskelia, ennen kuin niiden lopullinen kunto arvioidaan. Tätä prosessia toistetaan, kunnes saavutetaan tyydyttävä määrä maksimikoulutusaskeleita.

**Tulos**

Miten Progressive Dynamic Hurdles toimii?

**Esimerkki 5.672**

Taulukon TABREF64 tulokset osoittavat näiden skenaarioiden oikeellisuusasteet. Käyttäjän oikeellisuus on 7,5 % parempi kuin perusanalysaattorin (37,1 %:sta 44,6 %:iin), kun taas hybridimenetelmä päihittää molemmat, sillä sen oikeellisuus on 48,7 % ja parantaa perusanalysaattoria 11,6 %.

**Tulos**

Mitä kyselyjen selitysmenetelmää käyttäjät suosivat oikeellisuuden kannalta?

**Esimerkki 5.673**

Ehdotetut fuusiotekniikat ::: Ensimmäinen ehdottamamme tekniikka on visuaalisten piirteiden vaiheittainen dekooderifuusio jokaisen ennustusvaiheen aikana, eli yhdistämme visuaalisen koodauksen kontekstiksi jokaisessa dekoodausprosessin vaiheessa. Ehdotetut fuusiotekniikat :::: Multimodaalinen huomiomodulaatioHyvin samanlainen kuin yleinen huomiomodulaatio BIBREF8, jossa vaihtelevan pituinen kohdistusvektori $a\_{th}(s)$, jonka koko on yhtä suuri kuin lähdepuolen aika-askeleiden määrä, saadaan vertaamalla nykyistä piilotettua tavoitetilaa $h\_{t}$ kuhunkin piilotettuun lähdetilaan $\overline{h\_{s}}}$; tarkastelemme muunnelmaa, jossa visuaalista koodausta $v\_{t}$ käytetään laskemaan huomiojakauma $a\_{ttv}(s)$ myös lähdekoodausten yli. Ehdotetut fuusiotekniikat ::: Visual-Semantic (VS) RegularizerVisuaalisen modaliteetin hyödyntämisessä valvonnassa BIBREF1 käyttää monitehtäväoppimista oppiakseen maadoitettuja representaatioita kuvarepresentaation ennustamisen avulla.

**Tulos**

Mitä kolmea uutta fuusiotekniikkaa on ehdotettu?

**Esimerkki 5.674**

Tietokokonaisuuksien luotaus ja rakentaminenLuotausmenetelmämme alkaa rakentamalla haastetietokokonaisuuksia (kuva FIGREF1, keltainen laatikko) tietoresurssien kohdejoukosta. Kukin koettelemamme tietokokonaisuus koostuu monivalintakysymyksistä, jotka sisältävät kysymyksen $\textbf {q}$ ja joukon vastausvaihtoehtoja tai ehdokkaita $\lbrace a\_{1},...a\_{N}\rbrace $. Tässä luvussa kuvataan yksityiskohtaisesti viisi erilaista tietokokonaisuutta, jotka muodostamme kahdesta asiantuntijatiedon lähteestä, nimittäin WordNet BIBREF35 ja GNU Collaborative International Dictionary of English (GCIDE). Kuvaamme kutakin tietolähdettä vuorollaan ja selitämme, miten tuloksena syntyvät tietokokonaisuudet, joita kutsumme WordNetQA:ksi ja DictionaryQA:ksi, rakennetaan.Yksinkertaisuuden vuoksi kuvaamme kutakin asiantuntijatiedon lähdettä suunnattuna, reunoilla merkitynä graafina $G$. Tämän graafin solmut ovat $\mathcal {V} = \mathcal {C} \cup \mathcal {W} \cup \mathcal {W} \cup \mathcal {S} \cup \mathcal {D}$, jossa $\mathcal {C}$ on joukko atomisia käsitteitä, $\mathcal {W}$ joukko sanoja, $\mathcal {S}$ joukko lauseita ja $\mathcal {D}$ joukko määritelmiä (ks. taulukko Taulukko TABREF4, jossa on yksityiskohtaiset tiedot WordNetistä ja GCIDE:stä). Jokainen $G$:n reuna on suunnattu $\mathcal {C}$:n atomisesta käsitteestä $\mathcal {C}$:n toiseen solmuun $V$:ssä, ja se on merkitty relaatiolla, kuten hypernymillä tai isa$^\uparrow $:lla, relaatiojoukosta $\mathcal {R}$ (ks. taulukko TABREF4).Määriteltäessä koekysymysmalleja on hyödyllistä tarkastella $G$:tä joukkona (relaatio-lähde-kohde-kohde) -kolmioita $\mathcal {T}} \subseteq \mathcal {R} \times \mathcal {C} \times \mathcal {V}$. Koska tällaiset kolmikot ovat peräisin asiantuntijatietolähteestä, niiden semanttinen johdonmukaisuus säilyy. Esimerkiksi kun kolmion relaatio on def, vastaava reuna kuvaa $\mathcal {C}$:n käsitettä $\mathcal {D}$:n määritelmään.Luotaustietokokonaisuuksien rakentamisessa turvaudumme kahteen heuristiseen funktioon, jotka määritellään jäljempänä kullekin yksittäiselle luotaukselle: $\textsc {gen}\_{\mathcal {Q}}(\tau )$, joka tuottaa kultaisia kysymys-vastauspareja $(\textbf {q},\textbf {a})$ joukosta kolmioita $\tau \subseteq \mathcal {T}$ ja kysymysmalleja $\mathcal {Q}$, ja $\textsc {distr}(\tau ^{\prime })$, joka tuottaa häiritseviä vastausvaihtoehtoja $\lbrace a^{\prime }\_{1},...a^{{\prime }\_{N-1} \rbrace $ perustuu toiseen kolmiojoukkoon $\tau ^{\prime }$ (jossa yleensä $\tau \subset \tau ^{\prime }$). Lyhyyden vuoksi käytämme $\textsc {gen}(\tau )$ kuvaamaan $\textsc {gen}\_{\mathcal {Q}}(\tau )$, jolloin kysymysmallit $\mathcal {Q}$ jäävät implisiittisiksi.Dataset Probes and Construction ::: WordNetQAWordNet on englanninkielinen sanastotietokanta, joka koostuu noin 117 000 käsitteestä, jotka on järjestetty synsettien ryhmiin, joista kukin sisältää glossin (eli kohdekäsitteen määritelmän), joukon edustavia englanninkielisiä sanoja (ns. lemmat) ja noin 33 000 synsetissä esimerkkilauseet. Lisäksi monissa synseteissä on ISA-linkkejä muihin synsetteihin, jotka ilmaisevat monimutkaisia taksonomisia suhteita. Kuvassa FIGREF6 on esimerkki, ja taulukossa TABREF4 on yhteenveto siitä, miten WordNet on muotoiltu erilaisten kolmioiden $\mathcal {T}$ joukoksi. Nämä kolmikot muodostavat yhdessä suunnatun, reunoilla merkityn graafin $G$. Tärkein motivaatiomme WordNetin käyttämiselle ConceptNetin BIBREF36 kaltaisen resurssin sijaan on se, että WordNetissä on saatavilla glosseja ($\mathcal {D}$) ja esimerkkilauseet ($\mathcal {S}$), joiden avulla voimme laatia luonnollisen kielen kysymyksiä, jotka kontekstualisoivat käsitteiden tyypit, joita haluamme kartoittaa.Tietokokonaisuuden kartoitukset ja rakentaminen ::: WordNetQA ::: Esimerkkien tuottaminen @!START@$\textsc {gen}(\tau )$@!END@.Rakennamme neljä yksittäistä tietokokonaisuutta, jotka perustuvat WordNetille ominaiseen semanttiseen suhteeseen (ks. BIBREF37): hypernymia (eli yleistäminen tai ISA-pohdinta taksonomiasta ylöspäin, ISA$^\uparrow $), hyponyymia (ISA$^{\downarrow }$), synonyymia ja määritelmät. Kussakin tapauksessa kysymysten tuottamiseksi käytämme useita sääntömalleja $\mathcal {Q}$, jotka toimivat tupleihin. Osajoukko tällaisista malleista on esitetty taulukossa TABREF8. Mallit on suunniteltu jäljittelemään luonnontieteellisissä vertailuanalyyseissä havaitsemiamme naturalistisia kysymyksiä.Oletetaan esimerkiksi, että haluamme luoda kysymyksen $\textbf {q}$ kohdekäsitteen $c \in \mathcal {C}$ määritelmästä. Valitaan ensin kysymysmalli $\mathcal {Q}$:sta, jossa ensin esitellään käsite $c$ ja sen lemma $l \in \mathcal {W}$ asiayhteydessä esimerkkilauseen $s \in \mathcal {S}$ avulla ja sitten pyydetään tunnistamaan vastaava WordNet-kiilto $d \in \mathcal {D}$, joka toimii kultaisena vastauksena $\textbf {a}$. Samoin toimitaan ISA-järkeilyn osalta; jokainen kysymys kahden käsitteen välisestä hypernym/hyponyymisuhteesta $c \rightarrow ^{\uparrow /\downarrow } c^{\prime } c^{\prime} \ in \mathcal {T}\_{i}$ (esim. $\texttt {koira} \rightarrow ^{\uparrow /\downarrow } \texttt {eläin/terrieri}$) esittelee ensin kontekstin $c$:lle ja kysyy sitten vastausta, joka tunnistaa $c^{\prime }$ (joka on myös varustettu glossilla, jotta se sisältää kaiken saatavilla olevan kontekstin).Jälkimmäisessä tapauksessa säännöt $(\texttt {isa}^{r},c,c^{\prime }) \ in \mathcal {T}\_i$ taulukossa TABREF8 kattavat vain suorat ISA-linkit $c$:stä $r:n suuntaan \ in \lbrace \uparrow ,\downarrow \rbrace $. Käytännössä jokaiselle $c$:lle ja suunnalle $r$ rakennamme testit, jotka kattavat kaikkien $c$:n suorien ja johdettujen ISA-suhteiden joukon HOPS$(c,r)$:Näin voimme arvioida, missä määrin mallit pystyvät käsittelemään monimutkaisia päättelymuotoja, jotka vaativat useita päättelyvaiheita tai hyppyjä: WordNetQA ::: Distractor Generation: Esimerkki distraktoreiden tuottamisesta on esitetty kuvassa FIGREF6, joka perustuu samanlaisiin periaatteisiin kuin edellä. Kullekin käsitteelle $c$ valitaan 4 häiritsevää vastausta, jotka ovat lähellä toisiaan WordNetin semanttisessa avaruudessa. Kun esimerkiksi rakennetaan hypernymiatestejä $c$:lle joukosta hops$(c,\uparrow )$, muodostetaan distraktorit poimimalla $\textsc {hopsc}(c,\downarrow )$:sta (ja päinvastoin) sekä $c$:n $\ell$-syvästä sisarperheestä, joka määritellään seuraavasti. 1-syvä sisarusperhe on yksinkertaisesti $c$:n sisarukset tai sisarukset eli muut lapset $\tilde{c} \ne c$ ja $c$:n vanhemman solmun $c^{\prime }$ muut lapset. Jos $\ell > 1$, $\ell $-syvään sisarperheeseen kuuluvat myös jokaisen $\tilde{c}$:n kaikki jälkeläiset aina $\ell -1$ -tasolle asti, joita merkitään $\textsc {hops}\_{\ell -1}(\tilde{c},\downarrow )$. Muodollisesti: Määritelmiä ja synonyymejä varten muodostamme distraktorit kaikista näistä joukoista (ja sisar distraktoreiden syvyyttä rajoitetaan samalla tavalla kuin edellä). Näin voimme systemaattisesti tutkia mallin suorituskykyä laajalla valikoimalla distraktorijoukkoja.Dataset Probes and Construction ::: WordNetQA ::: Häiriöt ja semanttiset klusteritKunkin käsitteen $c$ (eli atomisen WordNet-synsetin) ja koettimen tyypin (eli määritelmät, hypernymiat jne.) osalta meillä on aineiston tuottamistavan perusteella monenlaisia $c$:hen liittyviä kysymyksiä, jotka manipuloivat 1) johtopäätösten monimutkaisuutta (esim. johtopäätösten hyppyjen määrää) ja 2) käytettyjen häiriötekijöiden tyyppejä (tai häiriötekijöiden häiriötekijöitä). Kutsumme tällaisia joukkoja semanttisiksi klustereiksi. Kuten kuvaamme seuraavassa jaksossa, semanttisten klustereiden avulla voimme kehittää uudenlaisia arviointitapoja, jotka paljastavat, onko malleilla kattava ja johdonmukainen tietämys kohdekäsitteistä (esim. arvioimalla, pystyykö malli vastaamaan oikein useisiin käsitteeseen liittyviin kysymyksiin, toisin kuin muutamaan erilliseen tapaukseen).Yksittäisten tietokokonaisuuksien yksityiskohdat esitetään taulukossa TABREF12. Noudatamme BIBREF22:ta ja käytämme näistä joukoista enintään 3 000 esimerkkiä harjoitteluun ja varaamme loput kehitykseen ja testaukseen. Koska olemme kiinnostuneita koettelemisesta, suurten pidätettyjen joukkojen avulla voimme tehdä yksityiskohtaista analyysia ja klusteripohjaista arviointia. datasetin koetteleminen ja rakentaminen ::: Se on kattava avoimen lähdekoodin englanninkielinen sanakirja, joka on rakennettu pitkälti Webster's Revised Unabridged Dictionary BIBREF38 -sanakirjan pohjalta. Kukin merkintä koostuu sanasta, sen puhekielestä, määritelmästä ja valinnaisesta esimerkkilauseesta (ks. taulukko TABREF14). Kaiken kaikkiaan 33k merkintää (yhteensä 155k:sta) sisältää esimerkkilauseita/käyttötarkoituksia. Kuten WordNetin koeasetusten kohdalla, keskitymme tähän osajoukkoon, jotta voimme kontekstualisoida jokaisen koekäytettävän sanan. Toisin kuin WordNetissä, GCIDE:ssä ei ole ISA-suhteita tai eksplisiittisiä synsettejä, joten pidämme jokaista yksittäistä merkintää erillisenä merkityksenä. Tämän jälkeen käytämme sanakirjamerkintöjä luodaksemme koettimen, joka keskittyy sanan merkityksen erotteluun, kuten jäljempänä kuvataan. datasetin koettimet ja rakentaminen ::: DictionaryQA ::: Esimerkkien ja häiriötekijöiden tuottaminen: Kultakysymysten ja -vastausten tuottamiseen käytämme samoja määritelmien tuottamismalleja kuin kuvassa TABREF8 on esitetty WordNetQA:lle. Häiritsevien tekijöiden tuottamiseksi otamme yksinkertaisesti kohdesanoille vaihtoehtoisia määritelmiä, jotka edustavat eri sanamerkitystä (esim. taulukossa TABREF14 esitetyt lahjan vaihtoehtoiset määritelmät), sekä satunnaisesti valitut määritelmät, jos ne ovat tarpeen viisitahoisen monivalintakysymyksen luomiseksi. Kuten edellä, varaamme harjoitteluun enintään 3k esimerkkiä. Koska tässä tietokokonaisuudessa on yhteensä vain 9k esimerkkiä (ks. WordSense taulukossa TABREF12), varaamme myös 3k kutakin kehittämistä ja testausta varten. Huomaa, että ensimmäiset yritykset rakentaa tämä tietokokonaisuus tavanomaisen satunnaisen jakamisen avulla johtivat tiettyihin systemaattisiin vääristymiin, joita seuraavassa jaksossa kuvatut vain valintamahdollisuuksiin perustuvat perusmallit käyttivät hyväkseen ja jotka siten paisuttivat mallien kokonaispistemääriä. Useiden suodatusyritysten jälkeen havaitsimme, että muiden tekijöiden ohella määritelmien käyttäminen häiritsevinä tekijöinä merkinnöistä, joissa ei ole esimerkkilauseet (esim. taulukon TABREF14 kaksi ensimmäistä merkintää), korreloi yllättävästi tällaisten vääristymien kanssa. Tämä viittaa siihen, että mahdolliset vääristymät, jotka liittyvät esimerkkejä sisältävien ja ilman esimerkkejä olevien sanakirjamerkintöjen välisiin eroihin, voivat pilata tuloksena syntyvän automaattisesti luodun MCQA-aineiston (lisätietoja automaattiseen tietokokonaisuuksien luomiseen liittyvistä sudenkuopista on kohdassa SECREF5).

**Tulos**

Valvotaanko automaattisesti muodostettujen tietokokonaisuuksien laatua?

**Esimerkki 5.675**

Vastausten hakua varten luodaan INLINEFORM4 -tietokanta, joka antaa INLINEFORM5 -tarkkuuden ja INLINEFORM6 -kattavuuden.

**Tulos**

Käyttävätkö he indeksointiin perustuvaa menetelmäänsä luodakseen näytteen QA-Wikipedian tietokokonaisuudesta?

**Esimerkki 5.676**

Yksi esimerkki havainnollistaa, että mallin tekemät virheet voivat liittyä puoluepolitiikan muutoksiin.

**Tulos**

Selittävätkö poliittisten toimijoiden politiikkojen muutokset kaikki mallin tekemät virheet?

**Esimerkki 5.677**

GANN on uusi neuroverkkomalli APC-tehtävää varten, jonka tarkoituksena on ratkaista perinteisten RNN- ja CNN-verkkojen puutteet. GANN sovelsi Gate Truncation RNN (GTR) -menetelmää oppiakseen informatiivisia, aspektiriippuvaisia sentimenttivihjeiden representaatioita. GANN saavutti parhaan mahdollisen APC-suorituskyvyn kiinalaisissa arvostelutietoaineistoissa.

**Tulos**

Mikä oli neljän kiinalaisen arviointiaineiston aiempi taso?

**Esimerkki 5.678**

Raakakorpuksena käytimme japanilaista web-korpusta, joka koottiin BIBREF13:n ehdottamien menettelyjen avulla. Noin 100 miljoonan lauseen korpuksesta saatiin 1,4 miljoonaa tapahtumaparia AL:lle, 41 miljoonaa CA:lle ja 6 miljoonaa CO:lle.

**Tulos**

Kuinka suuri on koulutukseen käytettävä raakakorpus?

**Esimerkki 5.679**

Tietokokonaisuutemme on kommentoitu Ruijinin sairaalan ruoansulatuskanavan kirurgian osaston toimittamien kiinalaisten patologisten raporttien perusteella. Se sisältää 17 833 lausetta, 826 987 merkkiä ja 2 714 kysymys-vastausparia.

**Tulos**

Kuinka monta kysymystä tietokokonaisuudessa on?

**Tulos**

Kuinka suuri on Ruijingin sairaalasta kerättyjen patologisten raporttien tietokanta?

**Esimerkki 5.680**

Kaikkien sarjojen tulokset osoittavat merkittävän eron tarkkuuden ja palautuksen välillä, sillä tarkkuus on huomattavasti suurempi kuin palautus. Myös kehitys- ja testijoukosta saatujen tulosten (F1 = 89,60, F1 = 87,82) ja täydentävästä testijoukosta saatujen tulosten (F1 = 71,49) välillä on merkittävä ero. Nämä erot selittyvät todennäköisesti lisätestijoukon ja kehitys- ja testijoukon välisellä aikaerolla (lisätestijoukon otsikot ovat eri ajanjaksolta kuin harjoitusjoukko).

**Tulos**

Mikä on CRF-mallin suorituskyky kuvatussa tehtävässä?

**Esimerkki 5.681**

Korpusten monipuolisuuden lisäämiseksi valitsimme 4 englanninkielistä korpusta ja 4 mandariininkielistä korpusta vähäisten resurssien kielikorpusten lisäksi.

**Tulos**

Testaavatko ne lähestymistapaansa suurten resurssien tehtävissä?

**Esimerkki 5.682**

Jotta koulutettua mallia voidaan käyttää, testisarjat käännetään ensin englanniksi konekääntämisen avulla, minkä jälkeen tehdään päättely. Konekääntämisessä käytetään Googlen käännösrajapintaa.

**Tulos**

miten kirjoittajat käänsivät arvostelut muille kielille?

**Esimerkki 5.683**

Kokeilemme viittä vertailuanalyysimenetelmää tekstien hyökkäämiseksi: FGM, FGVM, DeepFool BIBREF5, HotFlip BIBREF3) ja TYC BIBREF4.

**Tulos**

Mitkä ovat vertailukohtana käytettävät hyökkäysmenetelmät?

**Esimerkki 5.684**

Työskentelemme tietokokonaisuuden kanssa, joka koostuu 3 206 uutisartikkelista, joista jokainen on merkitty joko oikeaksi tai väärennetyksi, ja 1 603 oikeaa ja väärennettyä artikkelia on jaettu täydellisesti 50/50.

**Tulos**

Mikä on tietokokonaisuuden koko?

**Esimerkki 5.685**

Raportoimme segmentointituloksen tarkkuuden, palautuksen ja F-mitan avulla rajojen (BP, BR, BF) ja merkkien (WP, WR, WF) osalta. Raportoimme myös tarkan vastaavuuden (X) mittarin, joka laskee oikein segmentoitujen lausumien osuuden.

**Tulos**

Miten sanojen segmentointitehtävä arvioidaan?

**Esimerkki 5.686**

Luotettavaa usean matkan päättelyä varten OpenIE-tupeleiden avulla voimme lisätä tukigraafihakuun kuplettien välisiä yhteyksiä, joita ohjataan pienellä määrällä OpenIE-predikaatteja koskevia sääntöjä. Tällaisten sääntöjen oppiminen tieteen alaa varten on avoin ongelma ja mahdollinen tulevan työn kohde.

**Tulos**

Voidaanko menetelmän avulla vastata monihyppykysymyksiin?

**Tulos**

Pystyykö heidän menetelmänsä päättelemään usean matkan kautta?

**Esimerkki 5.687**

Käytimme kokeissa kahta eri korpusta: small\_parallel\_enja ja Asian Scientific Paper Excerpt Corpus (ASPEC) BIBREF5.

**Tulos**

Mitä tietokokonaisuutta he käyttävät?

**Esimerkki 5.688**

Molemmissa tapauksissa käytämme sanaa malli, joka on yksi klassisen/tilastollisen LSA:n ja luokittelijan mahdollisista yhdistelmistä. Käytetyt luokittelijat ovat tukivektorikone (SVM), logistinen regressio (Log.Reg), satunnainen metsä (RF) ja gradienttihäviö (XGB). Tarvittaessa tuloksia verrataan kirjallisuudessa jo esitettyihin tuloksiin.

**Tulos**

Mitä eri menetelmiä käytetään eri korporaatioihin?

**Esimerkki 5.689**

Käytämme Gaussin prosesseja, koska tämä todennäköisyysperusteinen kerneloitu kehys välttää kalliin ristiinvalidoinnin tarpeen hyperparametrien valinnassa.

**Tulos**

Miksi Gaussin prosessi on erityisen sopiva menetelmä tähän luokitusongelmaan?

**Esimerkki 5.690**

Osoittaaksemme, että yhteistyöelimen ranskankielisen version luominen on hyödyllistä, vertaamme ensin CamemBERTiä monikieliseen yhteistyöelimen koteloituun versioon (mBERT).

**Tulos**

Verrattiinko CamemBERTiä monikieliseen yhteistyöelimeen näissä tehtävissä?

**Esimerkki 5.691**

Vertailemme ja analysoimme huolellisesti eri kontekstin mallintamismenetelmien suorituskykyä kahdella suurella, monimutkaisella ja monialaisella tietokokonaisuudella, SParC BIBREF2:lla ja CoSQL BIBREF6:lla, tehtyjen kokeiden avulla.

**Tulos**

Mitä kahta suurta tietokokonaisuutta käytetään arvioinnissa?

**Esimerkki 5.692**

Julkisessa tulostaulukossa ilmoitettu lopputulos on 0,73019 ja yksityisessä tulostaulukossa 0,58455. Se on melko erilainen huonolla tavalla. Tämä saattaa johtua siitä, että malli on sovitettu liikaa harjoitusjoukkoon ja viritetty julkiseen testijoukkoon.

**Tulos**

Mikä on julkinen kojelauta?

**Tulos**

Mikä on yksityinen kojelauta?

**Esimerkki 5.693**

hierarkkinen Tässä asiakirjassa käytämme vuoden 2019 GermEvalin hierarkkisen tekstiluokittelun jaetun tehtävän BIBREF0 tietokokonaisuutta ja käytämme ennalta määriteltyä merkintäjoukkoa arvioidaksemme lähestymistapaamme tähän luokittelutehtävään.

**Tulos**

Mitä tietokokonaisuutta he käyttävät?

**Esimerkki 5.694**

n-grammi: Tarkastelemme 3-grammaista ja 4-grammaista ehdollista kielimallin perustasoa interpoloinnin kanssa. Interpoloidun mallin parhaat kertoimet etsitään satunnaisruutujen avulla: Käytämme fconv-arkkitehtuuria BIBREF24 ja oletushyperparametreja fairseq BIBREF25 -kehyksestä. Koulutamme verkon ADAM-optimointilaitteella BIBREF26 oppimisnopeuden ollessa 0,25 ja keskeytystodennäköisyyden ollessa 0,2. LSTM: Tarkastelemme LSTM-malleja BIBREF27 huomion kanssa ja ilman huomiota BIBREF28 ja käytämme tensor2tensor BIBREF29 -puitteistoa LSTM-peruslinjoihin. Käytämme kaksikerroksista LSTM-verkkoa sekä koodaimessa että dekoodaimessa 128-ulotteisilla piilovektoreilla: Kuten LSTM-verkoissa, käytämme tensor2tensor-kehystä Transformer-malliin. Transformer BIBREF21 -mallimme käyttää 256 ulottuvuutta sekä syötteen upottamiseen että piilotetilaan, 2 kerrosta ja 4 huomiopäätä. Sekä LSTM:n että Transformerin osalta koulutamme mallin ADAM-optimoijalla ($\beta \_{1} = 0.85$, $\beta \_{2} = 0.997$) ja pudotustodennäköisyys on asetettu 0.2.GPT-2: Valvottujen seq2seq-mallien lisäksi otamme mukaan myös tuloksia valmiiksi koulutetusta GPT-2 BIBREF30 -mallista, joka sisältää 117 miljoonaa parametria.

**Tulos**

Mitä perusmalleja on tarjolla?

**Esimerkki 5.695**

Tieteellisen julkaisun koko tekstin perusteella haluamme asettaa sen viittaukset paremmuusjärjestykseen kirjoittajan arvioiden mukaan. Keräsimme viimeaikaisia julkaisuja avoimista PLoS-lehdistä ja pyysimme kirjoittajia asettamaan läheisyyden mukaan paremmuusjärjestykseen viisi viittausta, jotka valitsimme heidän julkaisustaan.

**Tulos**

mitä joukkoistamisalustaa käytetään?

**Esimerkki 5.696**

Tässä tutkimuksessa keskitytään Switchboard-300:een, joka on 300 tuntia kestävä englanninkielinen puheentunnistustehtävä. Switchboard-300:lla saamiemme parhaiden tulosten vastakohtana koulutamme seq2seq-mallin myös 2000 tunnin Switchboard+Fisher -aineistolla.

**Tulos**

Kuinka paljon suurempi Switchboard-2000 on kuin Switchboard-300-tietokanta?

**Esimerkki 5.697**

Teksti- ja kuvakoodaukset yhdistettiin ketjuttamalla, jolloin saatiin 4 864 ulottuvuuden ominaisvektori. Tämä multimodaalinen esitys syötettiin sen jälkeen syötteenä monikerroksiseen perceptroniin (MLP), jossa oli kaksi piilokerrosta, joissa oli 100 neuronia ja ReLU-aktivointifunktio.

**Tulos**

Onko tietokokonaisuus multimodaalinen?

**Esimerkki 5.698**

Pyramidipisteytyksessä kultaisen ihmisen kirjallisten tiivistelmien sisältöyksiköt järjestetään pyramidin muotoon. Tässä pyramidissa sisältöyksiköt on järjestetty tasoihin, ja pyramidin korkeammat tasot merkitsevät suurempaa tärkeyttä.

**Tulos**

Mitä manuaalisia Pyramid-pisteitä käytetään?

**Esimerkki 5.699**

SemEval-2010-vertailutietokanta BIBREF0 koostuu 244 tieteellisestä artikkelista, jotka on kerätty ACM Digital Library -kirjastosta (konferenssi- ja seminaariasiakirjat).

**Tulos**

kuinka monta artikkelia aineistossa on?

**Esimerkki 5.700**

Esittelemme uuden 23 700 kyselystä koostuvan joukkoistetun tietokokonaisuuden, johon sisältyy 22 500 kattavaa kyselyä, jotka kattavat 150 tarkoitusta ja jotka voidaan ryhmitellä 10 yleiseen alaan.

**Tulos**

Mistä tiedot ovat peräisin?

**Esimerkki 5.701**

Katunäkymäympäristömme integroitiin ParlAI BIBREF6 -ohjelmaan, ja sitä käytettiin Mechanical Turk -palvelussa laajamittaisen tietokokonaisuuden keräämiseen, joka koski ihmisten havaitsemista, toimintaa ja viestintää.

**Tulos**

Mitä tietoja he käyttivät?

**Esimerkki 5.702**

BIBREF5 käyttää koneoppimista NER:ää varten. Heidän NER-järjestelmänsä poimii lääketieteellisiä ongelmia, testejä ja hoitoja kotiutusyhteenvedoista ja potilaskertomuksista. Käytetty tietokokonaisuus on i2b2 2010 challenge -tietokokonaisuus.

**Tulos**

Tutkitaanko asiakirjassa sähköisistä terveyskertomuksista poimittua tietoa?

**Esimerkki 5.703**

Tässä asiakirjassa kokeillaan kahta eri mallia ja verrataan niitä toisiinsa.

**Tulos**

Mikä oli lähtötaso?

**Esimerkki 5.704**

Parafraasin tunnistaminen (PI) on tehtävä, jolla määritetään, ovatko kaksi lausetta parafraasia vai eivät. Olemme toteuttaneet kernelfunktioiden kieltenvälisen muunnoksen PI- ja RE-tehtäviä varten SECREF3-jaksossa kuvatulla tavalla ja mitanneet mallien tarkkuutta testaamalla niitä rinnakkaisella aineistolla.

**Tulos**

Mitä luokittelutehtävää käytettiin tässä työssä kuvatun monikielisen mukauttamismenetelmän arvioinnissa?

**Esimerkki 5.705**

Toisen esityksemme tulos on 82,4 $F\_1$ virallisessa testijoukossa, ja se sijoittuu neljänneksi $4$ tässä jaetussa tehtävässä.

**Tulos**

Mihin tämä malli sijoittui jaetun tehtävän lopullisessa arvioinnissa?

**Esimerkki 5.706**

Tästä syystä otamme käyttöön ratkaisutapamuuttujat $\Pi = \lbrace \pi \_1, \ldots , \pi \_n\rbrace $ , jossa kunkin maininnan $j$ osalta muuttuja $\pi \_j \in \lbrace str, prec, attr\rbrace $ osoittaa, missä tilassa maininta olisi ratkaistava. Mallissamme määritellään kolme ratkaisutapaa - merkkijonojen täsmäytys (str), täsmällinen rakenne (prec) ja attribuuttien täsmäytys (attr) - ja $\Pi $ on deterministinen, kun $D$ on annettu (eli $P(\Pi |D)$ on pistejakauma). Määritämme $\pi \_j$ kullekin maininnalle $m\_j$ seuraavasti: $\pi \_j = str$ , jos on olemassa sellainen maininta $m\_i, i < j$ , että nämä kaksi mainintaa täyttävät Stanfordin BIBREF1 -monisieppijärjestelmän merkkijono-oikaisun seulan (String Match sieve), rennon merkkijono-oikaisun (Relaxed String Match seulan) tai tiukan pää-oikaisun (Strict Head Match A seulan).$\pi \_j = prec$ , jos on olemassa maininta $m\_i, i < j$ , jonka molemmat maininnat täyttävät Speaker Identification -seulan tai Precise Constructs -seulan.$\pi \_j = attr$ , jos mikään maininta $m\_i, i < j$ ei täytä edellä mainittuja kahta ehtoa.

**Tulos**

Mitä ovat resoluutiomallin muuttujat?

**Tulos**

Ovatko resoluutiotilan muuttujat käsityönä tehtyjä?

**Esimerkki 5.707**

Pystyimme jo yhdellä tekstillä ennustamaan sähkönkulutuksen alle 5 prosentin suhteellisella virheellä molemmissa aineistoissa.

**Tulos**

Kuinka tarkka malli on koulutettu yksinomaan tekstin perusteella?

**Esimerkki 5.708**

Tämä dekoodausmenetelmä vastaa hyvin pitkälti algoritmia SECREF7 , mutta pehmeiden takaisinkytkentöjen lisäksi lasketaan myös kovat takaisinkytkennät jokaisella aika-askeleella. Kun kaikki merkitykselliset suureet, kuten mallin pisteet, häviöt jne. on laskettu, seuraamme kovia takaisinsoittimia saadaksemme parhaan sekvenssin INLINEFORM0 .

**Tulos**

Vertailevatko ne osittain valmiita sekvenssejä (jotka on luotu säteenhaun vaiheiden aikana) kulta-/kohdesekvensseihin?

**Esimerkki 5.709**

Lauseiden luokittelussa yhdistetään luonnollisen kielen käsittely (NLP) ja koneoppiminen lauseiden rakenteen suuntausten tunnistamiseksi, BIBREF14 , BIBREF15 . Kukin twiitti muunnetaan numeeriseksi sanavektoriksi, jotta voidaan tunnistaa erottelevat piirteet kouluttamalla NLP-luokittimen validoidulle joukolle relevantteja twiittejä. Luokittelija toimii välineenä, jolla seulotaan mainokset, uutiset ja kommentit, jotka eivät liity potilaisiin. Järjestelmässämme yhdistetään logistinen regressioluokittelija, BIBREF16 , ja konvolutiivinen neuroverkko (Convolutional Neural Network, CNN), BIBREF17 , BIBREF18 , tunnistamaan itse ilmoitetut diagnoositwiititit.On tärkeää varoa automaattisia tilejä (esim. botteja, roskapostia), joiden suuri määrä twiittejä saastuttaa asiaankuuluvan orgaanisen sisällön, BIBREF19 , ja voi vääristää sentimenttianalyysejä, BIBREF20 . Ennen lauseiden luokittelua poistimme hyperlinkkejä sisältävät twiitit automaattisen sisällön poistamiseksi (osa orgaanisesta sisällöstä häviää väistämättä tällä tiukalla rajoituksella). Twitter antaa käyttäjille mahdollisuuden levittää toisten käyttäjien sisältöä "uudelleentwiittaamalla". Poistimme myös nämä viestit ennen luokittelua, jotta potilaiden kirjoittamat twiitit saatiin eristettyä. Otimme huomioon myös epäolennaisen astrologisen sisällön poistamalla kaikki twiitit, jotka sisälsivät jonkin seuraavista horoskooppi-indikaattoreista: `astrologia', `zodiac', `astronomia', `horoskooppi', `aquarius', `pisces', `aries', `taurus', `leo', `virgo', `libra' ja `scorpio'. Esikäsittelimme twiitit pienentämällä kirjaimia ja poistamalla välimerkit. Analysoimme myös vain sellaiset twiitit, joiden kohdalla Twitter oli tunnistanut `en' englannin kieleksi.

**Tulos**

Mitä koneoppimis- ja NLP-menetelmiä käytettiin rintasyöpäkokemuksiin liittyvien twiittien seulomiseen?

**Esimerkki 5.710**

Olemme arvioineet Nefnirin tulosta 21 093 merkkiä ja niiden oikeita lekseemejä sisältävää vertailukorpusta vastaan.Vertailukorpuksen näytteet poimittiin kahdesta suuremmasta korpuksesta, jotta saatiin monipuolinen sanavarasto:

**Tulos**

Mitä tietokokonaisuutta he käyttävät?

**Esimerkki 5.711**

Käytämme BIBREF12:n huipputason lähestymistavan ohjeita, sillä se saavutti parhaan taivutustarkkuuden viimeisimmässä SIGMORPHON 2019 -morfologisen taivutuksen jaetussa tehtävässä. Mallimme on toteutettu DyNet BIBREF13 -ohjelmalla. Koulutimme koodaaja-dekooderimallin, joka vastaanottaa lomakkeen merkkitason syötteenä, koodaa sen BiLSTM-koodaajalla ja sitten huomiolla parannettu dekooderi BIBREF14 tuottaa vastaavan morfologisten tunnisteiden sarjan, joka on toteutettu DyNetissä. Koulutimme merkkitason koodaaja-dekooderimallin, joka on samanlainen kuin edellä mainittu taivutusjärjestelmä ja joka on toteutettu DyNetissä.

**Tulos**

Mitä arkkitehtuureja käytetään näihin kolmeen tehtävään?

**Tulos**

Mitä järjestelmää käytetään perustasona?

**Esimerkki 5.712**

Ariston nykyisessä kokoonpanossa on kahdeksan ratkaisijaa, jotka kuvataan lyhyesti ja joista jokainen yrittää vastata monivalintakysymykseen. Tiettyjen ilmiöiden tutkimiseksi ja ratkaisijoiden kehittämiseksi hankkeessa on luotu suurempia tietokokonaisuuksia erilaisten ongelmien monistamiseksi ja tutkimiseksi, minkä tuloksena on syntynyt 10 uutta tietokokonaisuutta ja 5 suurta tietovarantoa yhteisölle.ratkaisijat voidaan ryhmitellä löyhästi seuraaviin ryhmiin:Tilastolliset ja tiedonhakumenetelmätJärkeilevät menetelmätSuurten kielimallien menetelmät Hankkeen keston aikana menetelmien suhteellinen painoarvo on siirtynyt suurten kielimallien menetelmiin. NLP:n ala on kehittynyt huomattavasti, kun on tullut käyttöön suuren mittakaavan kielimalleja, kuten ELMo (BID6), ULMFit (BID37), GPT (BID38), BERT (BID7) ja RoBERTa (BID8). Nämä mallit on koulutettu suorittamaan erilaisia kielen ennustustehtäviä, kuten ennustamaan puuttuva sana tai seuraava lause, käyttäen suuria tekstimääriä (esim. BERT koulutettiin Wikipedian ja Googlen 10 000 kirjasta koostuvan kirjakorpuksen avulla). Niitä voidaan myös hienosäätää uusiin kielen ennustustehtäviin, kuten kysymyksiin vastaamiseen, ja ne ovat menestyneet erinomaisesti niiden muutaman kuukauden aikana, jotka ne ovat olleet käytettävissä. Sovellamme BERTiä monivalintakysymyksiin käsittelemällä tehtävää luokitteluna: Kun annetaan kysymys $q$, jossa on vastausvaihtoehdot $a\_{i}$ ja valinnainen taustatieto $K\_{i}$, annamme sen BERT:lle muodossa: [CLS] $K\_i$ [SEP] $q$ [SEP] $a\_{i}$ [SEP] AristoBERT-ratkaisija käyttää kolmea menetelmää BERT:n soveltamiseksi tehokkaammin. Ensinnäkin haemme ja annamme taustatietoa kysymyksen mukana, kun käytämme BERTiä. Tämä antaa yhteistyöelimelle mahdollisuuden "lukea" tätä taustatietoa ja soveltaa sitä kysymykseen, vaikka taustatiedon käytön tarkka luonne onkin monimutkaisempi ja vaikeammin tulkittavissa. Toiseksi hienosäädämme yhteistyöelintä käyttämällä useiden tietokokonaisuuksien opetussuunnitelmaa, mukaan lukien joitakin, jotka eivät liity luonnontieteisiin. Lopuksi kootaan yhteen yhteistyöjärjestelmän eri muunnelmia.

**Tulos**

Onko Aristo vain jokin moderni NLP-malli (esim. BERT), joka on hienosäädetty tätä tehtävää varten?

**Esimerkki 5.713**

Lisäksi käytämme BLEU-pistemäärää BIBREF21-sukupolvien tarkkuuden arvioimiseksi ja erillisten n-grammien lukumäärää BIBREF6-sukupolvien monimuotoisuuden arvioimiseksi.

**Tulos**

Miten ne mittaavat johtopäätösten monimuotoisuutta?

**Esimerkki 5.714**

Apulauseen rakentaminen Yksinkertaisuuden vuoksi kuvaamme menetelmäämme pääasiassa TABSA:n avulla.Tarkastelemme seuraavia neljää menetelmää TABSA-tehtävän muuntamiseksi lauseparien luokittelutehtäväksi:Lause, jonka haluamme tuottaa kohde- ja näkökulmaparista, on kysymys, ja sen muodon on oltava sama. Esimerkiksi kohde-näkökulmaparin (SIJAINTI1, turvallisuus) joukosta generoimamme lause on "mitä mieltä olet sijainnin - 1 turvallisuudesta?" NLI-tehtävässä ehdot, jotka asetamme lauseita generoitaessa, eivät ole yhtä tiukkoja, ja muoto on paljon yksinkertaisempi. Tällä kertaa luotu lause ei ole tavallinen lause, vaan yksinkertainen pseudolause, jonka esimerkkinä on (SIJAINTI1, turvallisuus) -pari: apulause on: "sijainti - 1 - turvallisuus". QA-B:tä varten lisätään etikettitieto ja muunnetaan TABSA väliaikaisesti binääriseksi luokitusongelmaksi ( INLINEFORM0 ) todennäköisyysjakauman saamiseksi. Tällä hetkellä kukin kohde-aspekti - pari tuottaa kolme sekvenssiä, kuten "sijainti - 1:n turvallisuusaspektin polariteetti on positiivinen", "sijainti - 1:n turvallisuusaspektin polariteetti on negatiivinen", "sijainti - 1:n turvallisuusaspektin polariteetti ei ole mikään". Käytämme INLINEFORM1:n todennäköisyysarvoa vastaavuuspistemääränä. Sellaisen kohde-aspekti-parin kohdalla, joka tuottaa kolme sekvenssiä ( INLINEFORM2 ), otamme sen sekvenssin luokan, jolla on korkein vastaavuuspistemäärä ennustettua luokkaa varten. NLI-B:n ja QA-B:n ero on siinä, että apulause muuttuu kysymyksestä pseudolauseeksi. Apulauseet ovat: "sijainti - 1 - turvallisuus - positiivinen", "sijainti - 1 - turvallisuus - negatiivinen" ja "sijainti - 1 - turvallisuus - ei mitään".Kun olemme rakentaneet apulauseen, voimme muuttaa TABSA-tehtävän yhden lauseen luokittelutehtävästä lauseparien luokittelutehtäväksi. Kuten taulukosta TABREF19 käy ilmi, tämä on välttämätön toimenpide, joka voi parantaa merkittävästi TABSA-tehtävän kokeellisia tuloksia.

**Tulos**

Miten ne tuottavat apulauseen?

**Esimerkki 5.715**

Olemme ladanneet julkisesti saatavilla olevasta resurssista 1 873 Twitter-keskusteluketjua, noin 14 000 twiittiä, jotka on aiemmin esikäsitelty ja joista on poimittu keskusteluketjut.

**Tulos**

Kuinka suuri Twitter-tietokanta on?

**Esimerkki 5.716**

Lisäksi testaamme heidän suorituskykyään vain Zürichin ja Vispin murteella, koska suurin osa näytteistä on näistä kahdesta murteesta. Taulukossa TABREF15 esitetään näiden kahden mallin vertailun tulokset.

**Tulos**

Arvioidaanko mallia grafeemien ja foneemien välisessä tehtävässä?

**Esimerkki 5.717**

Perusarvot. Vertaamme lähestymistapaamme (FacTweet) seuraaviin perusasetuksiin: LR + Bag-of-words: Tweet2vec: Käytämme BIBREF20:ssä ehdotettua Bidirectional Gated recurrent neural network -mallia. Pidämme oletusarvoiset parametrit, jotka toimitettiin toteutuksen mukana. Twiittien esittämiseen käytämme mallin tuottamaa dekoodattua upotusta. Tämän perustason avulla pyrimme arvioimaan, voivatko twiittien hashtagit auttaa havaitsemaan ei-todennäköiset tilit.LR + All Features (twiittitaso): Poimimme jokaisesta twiitistä kaikki piirteemme ja syötämme ne LR-luokittimeen. Tässä tapauksessa emme aggregoi twiittejä ja tarkastelemme näin ollen jokaista twiittiä itsenäisesti.LR + All Features (chunk-level): Yhdistämme chunkissa olevien twiittien piirrevektorit ja syötämme ne LR-luokittimeen.FacTweet (twiittitaso): Samanlainen kuin FacTweet-lähestymistapa, mutta twiittitasolla; twiittien peräkkäistä kulkua ei hyödynnetä. Tavoitteena on tutkia twiittien peräkkäisen virtauksen merkitystä. top-$k$ vastaukset, tykkäykset tai uudelleentwiittaukset: Jotkut huhujen havaitsemiseen liittyvät lähestymistavat käyttävät vastausten, tykkäysten ja uudelleentwiittausten määrää huhujen havaitsemiseen BIBREF21. Näin ollen poimimme kustakin tilistä parhaat $k$ vastatut, tykätyt tai uudelleentwiittaamat twiitit arvioidaksemme tilien asiallisuutta. Testasimme eri $k$-arvoja, jotka vaihtelivat 10 twiitin ja kunkin tilin twiittien enimmäismäärän välillä. Kuvassa FIGREF24 esitetään makro-F1-arvot eri $k$-arvoilla. Vaikuttaa siltä, että $k=500$ vastaa parhaiten vastattujen twiittien osalta saavuttaa parhaan tuloksen. Siksi pidämme tätä perustasona.

**Tulos**

Mitä perustasoja tässä työssä käytettiin?

**Esimerkki 5.718**

Perinteisen VAE-rakenteen lisäksi otamme CWVAE:ssa käyttöön ylimääräisen kontekstitietoisen latentin muuttujan, jolla opitaan tapahtumien taustatietoa. Esiharjoitteluvaiheessa CWVAE koulutetaan aputietoaineistolla (joka koostuu kolmesta tarinakorpuksesta ja sisältää runsaasti tapahtumataustatietoa) tapahtumataustatiedon oppimiseksi kontekstin huomioon ottavan latentin muuttujan avulla. Tämän jälkeen, hienosäätövaiheessa, CWVAE:tä koulutetaan tehtäväkohtaisella tietokokonaisuudella, jotta tapahtumataustatieto voidaan mukauttaa kuhunkin erityiseen Jos-Tällöin-tapauksen päättelykohteen osa-alueeseen (esim. aikomukset, reaktiot jne.).

**Tulos**

Miten kontekstitietoinen varioiva autokooderi oppii tapahtuman taustatiedot?

**Esimerkki 5.719**

Tässä työssä käytämme BIBREF1:n julkaisemia tietokokonaisuuksia ja BIBREF0:n toimittamaa HEOT-tietokokonaisuutta. Tämän jälkeen tehtiin translitterointi Xlit-Crowd-muunnossanakirjan avulla ja kunkin sanan kääntäminen englanniksi hindi-englanti-sanakirjan avulla.

**Tulos**

Sisältävätkö kaikki tapaukset koodinvaihtoa?

**Esimerkki 5.720**

Twitterissä olevien viestien haku poimittujen leksikoiden avulla johti 19 300 dollarin twiittikokoelmaan. Leksikaalisen monimuotoisuuden saavuttamiseksi lisäsimme tietokantaamme 2500 satunnaisesti poimittua twiittiä.

**Tulos**

Kuinka monta twiittiä kerättiin?

**Esimerkki 5.721**

Kuten alaluvussa: tietokokonaisuudet mainittiin, kaikki sanojen samankaltaisuutta koskevat tietokokonaisuudet sisältävät sanapareja, joihin on merkitty samankaltaisuus- tai sukulaisuuspisteet, vaikka tämä ero ei aina olekaan selvä. MEN sisältää 3000 annotoitua sanaparia, joiden pistemäärät vaihtelevat 0:sta 50:een. Sanat vastaavat ESP-Game- ja MIRFLICKR-1M-kuva-aineistoissa esiintyviä kuvamerkintöjä.MTurk287 sisältää 287 annotoitua sanaparia, joiden pistemäärät vaihtelevat välillä 1,0-5,0. Se luotiin sanoista, jotka esiintyvät sekä DBpedian että The New York Timesin uutisartikkeleissa.MTurk771 sisältää 771 annotoitua sanaparia, joiden pisteet vaihtelevat 1,0-5,0. Sanat, joilla on synonyymi-, holonyymi- tai meronyymisuhteita, on poimittu WordNet BIBREF56:stä.RG sisältää 65 annotoitua sanaparia, joiden pisteet vaihtelevat 0,0-4,0 ja jotka edustavat "merkityksen samankaltaisuutta".RW sisältää 2034 sanaparia, joiden samankaltaisuuspisteet on annotoitu asteikolla 0-10. Tähän tietokokonaisuuteen sisältyvät sanat saatiin Wikipediasta niiden esiintymistiheyden perusteella, ja myöhemmin ne suodatettiin WordNet-synsettien mukaan, mukaan lukien synonymia, hyperonymia, hyponymia, holonymia ja meronymia. Tämä tietokokonaisuus luotiin testaamaan, kuinka hyvin mallit pystyvät edustamaan harvinaisia ja monimutkaisia sanoja.SimLex999 sisältää 999 sanaparia, jotka on merkitty samankaltaisuuspisteillä, jotka vaihtelevat välillä 0-10. Tässä tapauksessa kirjoittajat ottivat nimenomaisesti huomioon samankaltaisuuden eikä sukulaisuutta, mikä on vastoin niiden tietokokonaisuuksien puutteita, joissa ei oteta huomioon samankaltaisuutta, kuten MEN ja WS353. Sanat sisältävät substantiivit, adjektiivit ja verbit. SimVerb3500 sisältää 3500 verbiparia, joiden samankaltaisuuspisteet vaihtelevat 0-10. Verbit saatiin USF:n vapaiden assosiaatioiden tietokannasta BIBREF66 ja VerbNet BIBREF63 . Tämä tietokokonaisuus luotiin, jotta voitaisiin puuttua verbien puutteelliseen edustavuuteen SimLex999:ssä ja siihen, että luomisajankohtana parhaiten suoriutuvat mallit olivat jo ylittäneet kommentoijien välisen yhteisymmärryksen verbien samankaltaisuuden arviointiresursseissa. SimLex999:n tavoin myös tässä tietokokonaisuudessa tarkastellaan nimenomaisesti samankaltaisuutta eikä sukulaisuutta.WS353 sisältää 353 sanaparia, jotka on annotoitu samankaltaisuuspisteillä 0-10.WS353R on WS353:n osajoukko, joka sisältää 252 sanaparia, jotka on annotoitu sukulaisuuspisteillä. Tämä tietokokonaisuus luotiin pyytämällä ihmisiä luokittelemaan jokainen WS353-sanapari johonkin seuraavista luokista: synonyymit, antonyymit, identtiset, hyperonyymi-hyponimi, hyponimi-hyperonyymi, holonyymi-meronyymi, meronyymi-holonyymi ja mikään edellä mainituista. Näitä merkintöjä käytettiin myöhemmin parien ryhmittelyyn seuraaviin ryhmiin: samankaltaiset parit (synonyymit, antonyymit, identtiset, hyperonyymi-hyponimi ja hyponimi-hyperonyymi), toisiinsa liittyvät parit (holonyymi-meronyymi, meronyymi-holonyymi ja none-of-the-above, joiden inhimillinen samankaltaisuuspistemäärä oli suurempi kuin 5) ja toisiinsa liittymättömät parit (jotka luokiteltiin none-of-the-above-luokkaan ja joiden samankaltaisuuspistemäärä oli enintään 5). WS353S on WS353:n toinen osajoukko, joka sisältää 203 sanaparia, jotka on merkitty samankaltaisuuspisteillä. Tämä tietokokonaisuus muodostuu samankaltaisten ja epäsuhtaisten sanaparien liitosta, kuten edellä on kuvattu.

**Tulos**

Mitä samankaltaisuustietokantoja ne käyttävät?

**Esimerkki 5.722**

Vaikka neuraalinen konekääntäminen (NMT) on saavuttanut vaikuttavan suorituskyvyn suuren tietoresurssin olosuhteissa, ja siitä on tullut alan hallitseva tekijä BIBREF0 , BIBREF1 , BIBREF2 , viimeaikaisessa tutkimuksessa on väitetty, että nämä mallit ovat erittäin datatehottomia ja että ne alittavat suorituskyvyn fraasipohjaisessa tilastollisessa konekääntämisessä (PBSMT) tai valvomattomissa menetelmissä pienen datan olosuhteissa BIBREF3 , BIBREF4 .

**Tulos**

mitä sudenkuoppia asiakirjassa mainitaan?

**Esimerkki 5.723**

Kokeilussamme käytämme keskusteluaineistona mandariini-englanti-koodinvaihtopuhekorpusta nimeltä SEAME Phase II (South East Asia Mandarin-English).

**Tulos**

Mitä kieliä tässä asiakirjassa tarkastellaan?

**Esimerkki 5.724**

UutisartikkelitTietokantamme uutisartikkelien lähde on kuvattu BIBREF2:ssa. Tämä tietokokonaisuus muodostettiin kahdesta eri lähteestä: luotettavia uutisia (oikeita uutisia) varten otettiin uutisartikkeleita englanninkielisestä Gigaword-korpuksesta. Vääriä uutisia varten kerättiin artikkeleita seitsemältä eri epäluotettavalta uutissivustolta. TwitterTässä tietokokonaisuudessa käytettiin BIBREF6:sta saatua luetteloa useista Twitter-tileistä kutakin väärän tiedon tyyppiä varten. Tämä luettelo luotiin julkisten resurssien perusteella, joissa kommentoitiin epäilyttäviä Twitter-tilejä. BIBREF6:n kirjoittajat ovat rakentaneet tietokokonaisuuden keräämällä twiittejä näistä tileistä ja antaneet sen saataville.

**Tulos**

Mitä tietokokonaisuuksia he käyttivät?

**Esimerkki 5.725**

Käsittelemme kutakin yksittäistä kieltä tehtävänä ja koulutamme yhteisen mallin kaikille tehtäville. Pyrimme oppimaan joukon yhteisiä merkkien upotuksia molempien kielten taggereille yhdessä optimoimalla yhteistä häviöfunktiota, joka yhdistää korkean resurssin taggerin ja matalan resurssin taggerin.

**Tulos**

Miten eri kielten merkkiedustukset ovat yhteisiä?

**Esimerkki 5.726**

Mallimme saavuttaa CNN/Daily Mail -lehdessä ROUGE-1:n, ROUGE-2:n ja ROUGE-L:n keskiarvon 33,33, mikä on huippuluokkaa.

**Tulos**

Mikä on suorituskykyisimmän mallin ROUGE-pistemäärä?

**Esimerkki 5.727**

LRP antaa jokaiselle syötetietopisteelle ja mahdolliselle kohdeluokalle yhden skalaarisen relevanssiarvon jokaista syötemuuttujaa kohti, mikä osoittaa, onko kyseinen syötteen osa luokittelijan tietyn päätöksen puolesta vai sitä vastaan vai onko kyseinen syötemuuttuja luokittelutehtävän kannalta melko vähämerkityksinen ja merkityksetön.

**Tulos**

Toimiiko LRP-menetelmä tilanteissa, joissa sanat suhteutetaan toisiinsa?

**Esimerkki 5.728**

Puolivalvottu lähestymistapamme on melko yksinkertainen: ensin koulutetaan malli harjoitusjoukkoon, ja sitten tätä mallia käytetään ennustamaan hopeadatan merkinnät. Tämä hopeadata yksinkertaisesti lisätään koulutusjoukkoon, minkä jälkeen malli koulutetaan uudelleen. On kuitenkin otettava ylimääräinen vaihe sen varmistamiseksi, että hopeadata on laadultaan kohtuullista.

**Tulos**

Mitä puolivalvottua oppimista sovelletaan?

**Esimerkki 5.729**

esivalmennetut sanasulkeumat (WE): PDT-kokeita varten luomme word2vec-ohjelmalla sanojen upotukset LINDAT/CLARIN-tietokannasta saatavilla olevien suurten tšekkiläisten raakakorpusten yhdistelmällä. UD Tšekin kielessä käytämme FastText-sanan upotuksia BIBREF27, joiden ulottuvuus on 300 ja jotka esivalmennamme Tšekin Wikipediassa käyttämällä UD-tietueesta koulutettua segmentointia ja tokenisointia.BERT BIBREF1: Esikoulutetut kontekstuaaliset sanojen upotukset, joiden ulottuvuus on 768 Base-mallista. BERT-mallin neljä viimeistä kerrosta keskiarvoistetaan upotusten tuottamiseksi. Koska BERT käyttää sanapaloja, hajotamme UD-sanat sopiviksi alasanoiksi ja sitten keskiarvoistamme tuotetut upotukset samaan sanaan kuuluvien alasanojen osalta.Flair BIBREF2: Esiharjoittelemamme kontekstuaaliset sanojen upotukset, joiden ulottuvuus on 4096.

**Tulos**

Mitä tietoja käytetään upotusten rakentamiseen?

**Esimerkki 5.730**

HotpotQA:ssa löydämme keskimäärin 6 ehdollista ketjua (2-hop), ja ihmisen leimaama todellinen päättelyketju on ainutlaatuinen. Ennustettu ketju on oikea, jos ketju sisältää vain kaikki sitä tukevat kohdat (kohdat täsmäävät täsmälleen).MedHop-ohjelmassa löydetään keskimäärin 30 ehdollista ketjua (3-hop). Jokaisen ehdokasketjun kohdalla inhimilliset kommentoijamme merkitsivät, onko se oikea vai ei, ja oikea päättelyketju ei ole ainutkertainen. Ennustettu ketju on oikea, jos se on yksi niistä ketjuista, jotka ihmiset ovat merkinneet oikeiksi.Tarkkuus määritellään suhdelukuna: Tarkkuus määritellään suhteena:

**Tulos**

Mitä vertailuarvoja luodaan?

**Esimerkki 5.731**

Määritämme näiden strategioiden käytön vertailemalla keskustelijoiden puheenaiheille käyttämää lähetysaikaa. Olkoon puolen INLINEFORM0 oma kattavuus INLINEFORM1 se osuus INLINEFORM2:n kierroksella INLINEFORM3 lausumista sisällön sanoista, jotka kuuluvat sen omiin puheenaiheisiin INLINEFORM4 , ja vastustajan kattavuus INLINEFORM5 se osuus sen sisällön sanoista, jotka kattavat vastustajan puheenaiheet INLINEFORM6 . Käytämme kaikkia edellä käsiteltyjä keskusteluominaisuuksia. Kunkin puolen INLINEFORM0:n osalta otamme mukaan INLINEFORM1 , INLINEFORM2 ja niiden summan. Käytämme myös itsekattavuuden laskua, joka saadaan vähentämällä vastaavat arvot INLINEFORM3 :sta ja kunkin osapuolen hyväksymien keskustelupisteiden määrästä.

**Tulos**

mitä keskustelun sujuvuuteen liittyviä näkökohtia he tarkastelevat?

**Esimerkki 5.732**

Syy-yhteyden määritystietokanta on kokoelma tekstipareja, jotka kuvastavat ihmisten ehdottamia syy-seuraussuhteita (esimerkiksi "virus aiheuttaa sairauden"). Nämä kirjalliset lausumat tunnistavat verkon solmut (katso myös graafin fuusioalgoritmi semanttisesti vastaavien lausumien käsittelyä varten), kun taas syy-seuraussuhteet muodostavat kausaalisen attribuutioverkon suunnatut reunat ("virus" $\rightarrow $ "sairaus").

**Tulos**

Mitä ovat syy-yhteysverkot?

**Esimerkki 5.733**

Arvioimme monitehtäväistä lähestymistapaa ja kolmea algoritmia, jotka mallintavat nimenomaisesti tehtävien riippuvuudet.

**Tulos**

Ovatko nämä havainnot kestäviä eri tietokokonaisuuksissa ja erilaisissa kysymysten vastausalgoritmeissa?

**Esimerkki 5.734**

Menetelmät ::: Keräsimme Streaming API:n avulla kymmeneen yhdysvaltalaiseen valtavirtaiseen uutissivustoon eli BIBREF18:ssa kuvattuihin luotettavimpiin lähteisiin liittyviä twiittejä, ja viittasimme Hoaxy API:n BIBREF16:een niiden twiittien osalta, jotka sisälsivät linkkejä yli 100 yhdysvaltalaiseen disinformaatiopalveluun. Menetelmä ::: Italian tietokokonaisuus Italian skenaarion osalta keräsimme ensin twiittejä Streaming API:n avulla kolmen viikon aikana (19. huhtikuuta 2019 - 5. toukokuuta 2019) ja suodatimme ne, jotka sisälsivät URL-osoitteita, jotka osoittavat italialaisten virallisten sanomalehtien verkkosivustoille, kuten BIBREF22:ssa on kuvattu; nämä vastaavat Italian sanomalehtien levikin todentamiseen tarkoitetun yhdistyksen (Accertamenti Diffusione Stampa) toimittamaa luetteloa. Sen sijaan viittasimme BIBREF23:n tarjoamaan tietokokonaisuuteen saadaksemme joukon twiittejä, jotka on kerätty jatkuvasti tammikuusta 2019 lähtien käyttämällä samaa Twitter-päätepistettä ja jotka sisältävät URL-osoitteita yli 60 italialaiselle disinformaatiosivustolle.

**Tulos**

Mitä kahta laajamittaista tietokokonaisuutta käytetään?

**Esimerkki 5.735**

Mittaamme myös ihmisten perusarvoihin liittyvien sanojen käyttöä Boydin ym. boyd2015 mukaan. Sanakokonaisuudet eli teemat kaivettiin esiin käyttämällä Meaning Extraction Method (MEM) BIBREF10 -menetelmää.

**Tulos**

Miten he saavat ihmisistä psykologisia ulottuvuuksia?

**Esimerkki 5.736**

Aloitamme yleensä määrittelemällä kysymykset, joita haluamme tutkia. Voiko tekstianalyysi tarjota uuden näkökulman "suureen kysymykseen", joka on herättänyt kiinnostusta jo vuosia? Vai voimmeko nostaa esiin uusia kysymyksiä, jotka ovat nousseet esiin vasta äskettäin, esimerkiksi sosiaalisesta mediasta? Laskennallisen analyysin parissa työskentelevien yhteiskuntatieteilijöiden kysymykset pohjautuvat usein teoriaan, jossa kysytään: Miten voimme selittää havaitsemamme? Näiden kysymysten motivoima tekstin laskennallinen analyysi on oivalluslähtöistä: pyrimme kuvaamaan ilmiötä tai selittämään, miten se on syntynyt. Mitä voimme esimerkiksi oppia siitä, miten ja miksi vihapuhetta käytetään tai miten se muuttuu ajan myötä? Onko vihapuhe yksi asia vai sisältääkö se useita ilmaisumuotoja? Onko vihapuheen ja muunlaisen puheen välillä selkeä raja, ja mitkä piirteet tekevät siitä enemmän tai vähemmän moniselitteistä? Joskus toivomme myös yhteyttä useisiin tieteenaloihin. Vaikka esimerkiksi keskitymme arkiston humanistisiin kysymyksiin, voisimme myös esittää yhteiskunnallisia kysymyksiä, kuten "onko tässä arkistossa kyse enemmän yhteistyöprosesseista, kulttuurin rakentamisesta vai normien luomisesta?" tai "kuinka hyvin tämä arkisto heijastaa yhteiskuntaa, johon se on upotettu?". BIBREF3:ssa käytettiin kvantitatiivisia menetelmiä kertomaan tarina Darwinin älyllisestä kehityksestä - olennainen elämäkerrallinen kysymys tieteen historian avainhenkilölle.

**Tulos**

Millaisiin kysymyksiin (jotka eivät ole tietokoneavusteisen tekstianalyysin eturintamassa) he puuttuvat?

**Esimerkki 5.737**

Koska aiemmat järjestelmät keräävät asiaankuuluvia tietoja tietopankeista sen jälkeen, kun ne ovat havainneet kysymyksiä arvioinnin aikana (BIBREF24 , BIBREF25 ), tutkimme myös tämän vaihtoehdon käyttöä. Rakennamme nimittäin räätälöidyn tekstikorpuksen, joka perustuu maalaisjärjellä päättelytehtävissä esitettyihin kysymyksiin.

**Tulos**

Hienosäätävätkö he malliaan lopputehtävän perusteella?

**Esimerkki 5.738**

Kokeet suoritettiin käyttäen AMI IHM:n kokouskorpusta BIBREF18 eri kielimallien puheentunnistussuorituskyvyn arvioimiseksi.

**Tulos**

Mitä tietokokonaisuutta he käyttävät?

**Esimerkki 5.739**

Hakuavaruutemme koostuu kahdesta pinottavasta solusta, joista toinen on mallin kooderia ja toinen dekooderia varten Kukin solu sisältää NASNet-tyyppisiä lohkoja, jotka saavat kaksi piilotettua tilaa syötteenä ja tuottavat uusia piilotettuja tiloja tuotoksina Hakuavaruutemme sisältää viisi haara-tason hakukenttää (syöttö, normalisointi, kerros, ulostulon ulottuvuus ja aktivaatio), yhden lohkotason hakukentän (yhdistinfunktio) ja yhden solutasoisen hakukentän (solujen lukumäärä).

**Tulos**

Mitä mallin hakuavaruudessa on?

**Esimerkki 5.740**

Teimme kuitenkin muutakin, ja saavutimme myös toisen ja kolmannen tason aistien osalta erinomaisia tuloksia, koska vaadimme itse korkeaa laatua. Saavutimme lopulta yhteisymmärryksen 0,85 ja Kappa-arvon 0,83 näillä kahdella syvemmällä aistitasolla.

**Tulos**

Kuinka suuri on kommentoijien välinen yhteisymmärrys?

**Esimerkki 5.741**

Käytämme kahta uusinta neuraalista ydinviittausten erottelumallia, joita kuvaavat BIBREF2 ja BIBREF1 .

**Tulos**

Mikä on uusin neuraalinen ydinviittausten ratkaisumalli?

**Esimerkki 5.742**

Tarkastelemme QA PGNet- ja Multi-decoder QA PGNet -malleja, joissa on hakutaulukoiden upotus, perusmalleina ja parannamme perusmalleja muilla jäljempänä kuvatuilla muunnelmilla.

**Tulos**

Mikä on lähtötaso?

**Esimerkki 5.743**

Kokeet ::: Arviointimittarit ::: Automaattisen arvioinnin asetukset ::: Hienosäädimme GPT-2 keskipitkän mallin BIBREF51 kerättyihin otsikoihin ja käytimme sitä sitten mittaamaan tuotettujen tuotosten perpleksisyyttä (PPL).

**Tulos**

Miten sujuvuutta arvioidaan automaattisesti?

**Esimerkki 5.744**

Kaikki aineistot koostuvat videoista, joissa vain yksi puhuja on kameran edessä. Kullekin modaliteetille käyttämämme kuvaajat ovat seuraavat: Kieli Kaikki tietokokonaisuudet sisältävät manuaalisia transkriptioita. Visio Facet BIBREF26:n avulla poimitaan joukko piirteitä, joihin kuuluvat kehyskohtaiset perus- ja edistyneet tunteet ja kasvojen toimintayksiköt kasvolihasten liikkeiden indikaattoreina. Akustiikka Käytämme COVAREP BIBREF27 -ohjelmaa matalan tason akustisten piirteiden poimimiseen, mukaan lukien 12 melataajuuskeptraalikertoimen (MFCC), sävelkorkeuden seurannan ja äänteellisen/äänettömän segmentoinnin piirteet, glottaalilähteen parametrit, huipun kaltevuusparametrit ja maksimihajontakertoimet.

**Tulos**

Millaisia modaliteetteja käytetään eri tietokokonaisuuksissa?

**Esimerkki 5.745**

Kokeiden avulla tutkimme empiirisesti DIRL-analyysiämme ja ehdotetun ratkaisun tehokkuutta sen ongelman ratkaisemisessa. Lisäksi tutkittiin, miten kukin §SECREF10:ssä ja §SECREF14:ssä kuvattu vaihe vaikuttaa ehdotettuun ratkaisuun. Tutkimuksen toteuttamiseksi suoritimme suorituskykyvertailun seuraavien mallien välillä: SO: vain lähdealueelle koulutettu malli, jossa käytetään lähdealueen leimattua dataa ilman mitään alueellista mukauttamista CMD: alkuperäisen DIRL-kehyksen keskipisteeseen perustuva alueellinen mukauttamismalli BIBREF3, joka toteuttaa $\mathcal {L}\_{inv}$:n $\text{CMD}\_K$:lla.DANN: alkuperäisen DIRL-kehyksen BIBREF2:n vastakkaisoppimiseen perustuva toimialueen mukauttamismalli, joka toteuttaa $\mathcal {L}\_{inv}$ ja $\text{JSD}(\rm {P}\_S, \rm {P}\_T)$.$\text{CMD}^\dagger $: CMD-mallin painotettu versio, jossa sovelletaan vain ehdotetun menetelmän ensimmäistä vaihetta (kuvattu kohdassa SECREF10).$\text{DANN}^\dagger $: DANN-mallin painotettu versio, joka soveltaa vain ehdotetun menetelmän ensimmäistä vaihetta.$\text{CMD}^{\dagger \dagger }$: CMD-mallin painotettu versio, joka soveltaa sekä ehdotetun menetelmän ensimmäistä että toista vaihetta (kuvattu §SECREF14).$\text{DANN}^{\dagger \dagger }$: DANN-mallin painotettu versio, joka soveltaa sekä ehdotetun menetelmän ensimmäistä että toista vaihetta.$\text{CMD}^{\*}$: $\text{CMD}^{\dagger \dagger }$:n muunnos, jossa $\mathbf {w}^\*$ (kohdemerkityistä tiedoista saatu estimaatti) määritetään $\mathbf {w}$:lle ja vahvistetaan tämä arvo mallin harjoittelun aikana.$\text{DANN}^{\*}$: $\text{DANN}^{\dagger \dagger }$:n muunnos, joka määrittää $\mathbf {w}^\*$:n $\mathbf {w}$:lle ja vahvistaa tämän arvon mallin harjoittelun aikana. Suoritimme kokeita Amazonin arvostelutietokannalla BIBREF9, joka on vertailuarvotietokanta monialaisen tunneanalyysin alalla. Rakensimme tästä tietokokonaisuudesta 12 binääriluokiteltua toimialarajat ylittävää tunneanalyysitehtävää: B$\rightarrow $D, B$\rightarrow $E, B$\rightarrow $K, D$\rightarrow $B, D$\rightarrow $E, D$\rightarrow $K, D$\rightarrow $K, E$\rightarrow $B, E$\rightarrow $D, E$\rightarrow $K, K$\rightarrow $B, K$\rightarrow $D, K$\rightarrow $E. Lisäksi rakensimme 12 moniluokkaista, eri alojen välistä tunteiden luokittelutehtävää. Tehtävät suunniteltiin erottamaan 1 tai 2 tähteä (luokka 1) saaneet arvostelut 4 tähteä (luokka 2) ja 5 tähteä (luokka 3) saaneista arvosteluista. Taulukossa TABREF27 esitetään mallin suorituskyky 12 kaksiluokkaisessa ristikkäisaluetehtävässä. Taulukosta voidaan tehdä seuraavat havainnot.

**Tulos**

Miten DIRL arvioidaan?

**Esimerkki 5.746**

Arvioimme malliamme simuloidussa keräilytehtävässä, jossa robotin tehtävänä on sijoittaa kuutio kulhoon sanallisen käskyn mukaisesti. Kukin ympäristö sisältää kolmesta viiteen esinettä, jotka eroavat toisistaan koon (pieni, suuri), muodon (pyöreä, neliö) ja värin (punainen, vihreä, sininen, keltainen, vaaleanpunainen) mukaan, eli yhteensä 20 erilaista esinettä. Luodusta skenaariosta riippuen näiden kolmen ominaisuuden yhdistelmät ovat tarpeen kohteiden erottamiseksi toisistaan, mikä mahdollistaa eritasoisen monimutkaiset tehtävät.

**Tulos**

Millaisia simulaatioita kirjoittajat ovat tehneet lähestymistapansa validoimiseksi?

**Esimerkki 5.747**

Harjoittelussa käytetty epätäydellinen tietokokonaisuus koostuu pienaakkosista epätäydellisistä tiedoista, jotka on saatu alkuperäisiä korpuksia manipuloimalla.

**Tulos**

Testaavatko he lähestymistapaansa tietokokonaisuudella, jossa ei ole epätäydellisiä tietoja?

**Esimerkki 5.748**

Taulukossa TABREF19 verrataan BLEU-pistemäärän tuloksia mallimme ja Googlen neuraalisen konekäännöksen välillä.

**Tulos**

Mikä on heidän perustasonsa?

**Esimerkki 5.749**

Tutkimuksessamme keskityttiin ensisijaisesti englanninkielisiin twiitteihin, koska se oli diagnostisen koulutusnäytteen kieli.

**Tulos**

Raportoivatko kirjoittajat tulokset vain englanninkielisistä tietokokonaisuuksista?

**Esimerkki 5.750**

Taulukossa TABREF34 esitetään menetelmämme päättelynopeus, kun sekvenssimallinnuskerros toteutetaan LSTM-pohjaisella, CNN-pohjaisella ja Transformer-pohjaisella arkkitehtuurilla. Taulukosta nähdään, että menetelmämme päättelynopeus on paljon nopeampi kuin Lattice-LSTM:llä, kun käytetään LSTM-pohjaista sekvenssimallinnuskerrosta, ja se oli myös paljon nopeampi kuin LR-CNN, jossa sekvenssimallinnuskerros toteutettiin CNN-arkkitehtuurilla. Ja kuten odotettua, menetelmämme, jossa on CNN-pohjainen sekvenssimallinnuskerros, osoitti jonkin verran etua päättelynopeudessa verrattuna menetelmiin, joissa on LSTM-pohjainen ja Transformer-pohjainen sekvenssimallinnuskerros.

**Tulos**

Kuinka monta prosenttia ehdotettu menetelmä parantaa päätelmän nopeutta uusimpiin uusimpiin menetelmiin verrattuna?

**Esimerkki 5.751**

Monipuolisen harjoitusaineiston keräämiseksi olemme poimineet satunnaisotannalla 1000 viestiä politiikan, liike-elämän, tieteen ja AskRedditin alaryhmistä ja lisäksi 1000 viestiä Redditin etusivulta.

**Tulos**

mitä aiheita vedetään Redditistä?

**Esimerkki 5.752**

Peruslähestymistapa perustuu BIBREFiin20 . Se jakaa käyttäytymisnavigointia varten annettujen komentojen tulkinnan kahteen vaiheeseen: polun luomiseen ja polun todentamiseen. Polkujen luomisessa käytetään tavanomaista sekvenssistä sekvenssiin -mallia, jota on täydennetty huomiomekanismilla, joka on samanlainen kuin BIBREF23 ja BIBREF6 . Polun todentamiseen käytetään syvyyssuuntaista hakua, jolla graafista etsitään reitti, joka vastaa ennustettujen käyttäytymismallien sekvenssiä. Jos mikään reitti ei täsmää täydellisesti, peruslinja muuttaa enintään kolmea käyttäytymistä ennustetussa järjestyksessä yrittäen muuttaa sen kelvolliseksi poluksi.

**Tulos**

Mihin lähtökohtiin he vertasivat malliaan?

**Esimerkki 5.753**

Siirto-oppimiseen perustuvat lähestymistavatMathur et al. ehdottivat artikkelissaan loukkaavien twiittien havaitsemiseksi ternaarista Trans-CNN-mallia, jossa he kouluttavat malliarkkitehtuurin, joka koostuu kolmesta konvoluution 1D-kerroksesta, joiden suodatinkoot ovat 15, 12 ja 10 ja ytimen koko 3, ja sen jälkeen kahdesta tiheästä täysin kytketystä kerroksesta, joiden koko on 64 ja 3. Ensimmäisessä tiheässä FC-kerroksessa on ReLU-aktivointi ja viimeisessä tiheässä kerroksessa Softmax-aktivointi. He pystyivät kouluttamaan tämän verkon Davidsonin et al. toimittamalla rinnakkaisella englanninkielisellä tietokokonaisuudella. Kirjoittajat pystyivät saavuttamaan 83,9 %:n tarkkuuden, 80,2 %:n tarkkuuden ja 69,8 %:n palautuksen.Lähestymistapa näytti lupaavalta, kun otetaan huomioon, että tietokokonaisuus koostui vain 3189 lauseesta, jotka oli jaettu kolmeen luokkaan, ja näin ollen toistimme kokeen, mutta emme onnistuneet toistamaan tuloksia. Tulokset olivat heikompia kuin alkuperäisten tekijöiden saavuttamat tulokset. Suurin osa mallin hyperparametrien valinnoista oli kuitenkin peräisin tästä työstä: Nguyen et al. ehdottivat vuonna 2017 hybridiä monikanavaista CNN- ja LSTM-mallia, jossa he rakentavat vietnamin kielen ominaisuuskarttoja käyttämällä CNN:ää lyhytaikaisten riippuvuuksien ja LSTM:ää pitkäaikaisten riippuvuuksien tallentamiseen ja yhdistävät nämä molemmat ominaisuussarjat oppiakseen yhtenäisen ominaisuuksien joukon viesteistä. Nämä yhdistetyt ominaisuusvektorit lähetetään sitten muutamalle täysin yhdistetylle kerrokselle. Tällä arkkitehtuurilla saavutettiin 87,3 prosentin tarkkuus.

**Tulos**

Mitä malleja aiemmat työt käyttävät?

**Esimerkki 5.754**

Tämän työn panos on siis kaksitahoinen: (1) kehitämme kielellisen neuroverkkomallin, jolla voidaan luokitella sosiaalisen median viesteissä esiintyviä reaktioita, ja (2) sovellamme malliamme 10,8 miljoonan Twitter-viestin ja 6,2 miljoonan Reddit-kommentin merkitsemiseen arvioidaksemme eri uutislähteisiin kohdistuvien käyttäjien reaktioiden nopeutta ja tyyppiä. Kehitämme neuroverkkoarkkitehtuurin, joka perustuu reaktioista ja vanhemmista viesteistä poimittuihin sisältöön ja muihin kielellisiin signaaleihin, ja hyödynnämme "myöhäisfuusion" lähestymistapaa, jota on aiemmin käytetty tehokkaasti näkötehtävissä BIBREF13 , BIBREF14 . Tarkemmin sanottuna yhdistämme tekstisekvenssin aliverkon ja vektoriesityksen aliverkon, kuten kuvassa FIGREF5 on esitetty. Tekstisekvenssin aliverkko koostuu sulauttamiskerroksesta, joka on alustettu 200-ulotteisilla GloVe-sulautumilla BIBREF15 ja jota seuraa kaksi 1-ulotteista konvoluutiokerrosta, sitten max-pooling-kerros ja sen jälkeen tiheä kerros. Vektoriesityksen aliverkko koostuu kahdesta tiheästä kerroksesta. Sisällytämme molemmista aliverkoista saatua tietoa ketjutetuilla pehmustetuilla tekstisekvensseillä ja normalisoitujen LIWC-ominaisuuksien (Linguistic Inquiry and Word Count) vektoriedustuksilla BIBREF16 kunkin viestin ja sen vanhemman tekstin osalta.

**Tulos**

Mikä on heidän mallinsa arkkitehtuuri?

**Esimerkki 5.755**

300 testikyselyä valitaan satunnaisesti, ja kommentoijia pyydetään arvioimaan itsenäisesti näiden kyselyjen tulokset eri pisteillä niiden laadun perusteella: (1) Hyvä (3 pistettä): Vastaus on kieliopillinen, semanttisesti relevantti kyselyn kannalta ja ennen kaikkea informatiivinen ja mielenkiintoinen. (2) Hyväksyttävä (2 pistettä): Vastaus on kieliopillisesti ja semanttisesti relevantti, mutta liian triviaali tai yleinen (esim. "我不知道(En tiedä)", "我也是(Minäkin)", "我喜欢(Pidän siitä)" jne.); (3) Hylätty (1 piste): Vastauksessa on kielioppivirheitä tai se ei liity kyselyyn.

**Tulos**

Miten ihmisten arviointi suoritetaan, mitkä olivat kriteerit?

**Esimerkki 5.756**

Kaikissa tapauksissa tulokset ovat korkeat - valtaosassa tapauksista huomattavasti korkeammat kuin aiemmissa töissä on raportoitu. Erityisesti BIBREF1:ssä arvioidaan LSTM:ien kykyä oppia englannin kielen subjekti-verbisopimuskuvioita ja arvioidaan luonnossa esiintyvillä wikipedia-lauseilla. BIBREF2:ssa tarkastellaan myös subjekti-verbisopimusta, mutta "värittömissä vihreissä ideoissa", joissa luonnossa esiintyvien lauseiden sisältösanat korvataan satunnaisilla sanoilla, joilla on sama puhekielinen osa ja taivutus, jolloin varmistetaan, että keskitytään syntaksiin pikemminkin kuin valintaan perustuviin preferensseihin perustuviin vihjeisiin. BIBREF3:ssa tarkastellaan laajempaa valikoimaa syntaktisia ilmiöitä (subjektin ja verbin välinen sopimus, refleksiivinen anafora, negatiivisen polariteetin elementit) käyttämällä manuaalisesti rakennettuja ärsykkeitä, mikä mahdollistaa laajemman kattavuuden ja valvonnan kuin luonnollisessa ympäristössä.

**Tulos**

Onko näitä tehtäviä arvioitu aiemmissa töissä?

**Esimerkki 5.757**

Näin ollen SIM-mallissa on paljon vähemmän parametreja kuin nykyisissä dialogitilan seurantamalleissa. Kompensoidaksemme aukkokohtaisten parametrien poisjättämisen, käytämme parempaa ominaisuuksien esittämistä käyttäjän lausumasta ja dialogin tiloista käyttämällä syntaktista tietoa ja konvoluutio-neuraaliverkkoja (CNN, convolutional neural networks).

**Tulos**

Miten estetään se, että mallin monimutkaisuus kasvaa lähtö- ja saapumisaikojen määrän kasvaessa?

**Esimerkki 5.758**

Jokaisen testitapauksen parina oli 99 satunnaisesti poimittua negatiivista tapausta. Kukin suositusmalli asettaa 100 tapausta paremmuusjärjestykseen ennustettujen tulostensa mukaan. Järjestyksessä olevaa luetteloa arvioidaan osumaprosentilla (HR) BIBREF49 ja normalisoidulla kumulatiivisella voitolla (NDCG) BIBREF50 sijalla 10. HR@10 on recall-pohjainen metriikka, joka mittaa, kuinka monta prosenttia testattavasta kohteesta suositellaan oikein 10 parhaan joukossa. NDCG@10 on paremmuusjärjestykseen perustuva arviointimittari, jossa otetaan huomioon oikean osuman sijainti paremmuusjärjestykseen sijoitetussa tuloksessa. Kaiken kaikkiaan AMRAN-ohjelmamme on kaikkia perusratkaisuja parempi, sillä sen HR@10-arvo on 0,657 ja NDCG@10-arvo 0,410. Se parantaa HR@10:tä 5,3 % ja NDCG@10:tä 3 % verrattuna parhaaseen perustasoon (eli $NAIS\_{concat}$).

**Tulos**

Mikä on mallin tarkkuus?

**Esimerkki 5.759**

TietokannatKokeilemme kahdella kiinalaisella kysymysvastaustietokannalla NLPCC-2016-arviointitehtävästä BIBREF13 . DBQA on dokumenttipohjainen kysymysten vastaustietokanta. KBRE on tietoon perustuva relaatioiden poimintatietokanta.

**Tulos**

Millä tietokokonaisuudella (tietokokonaisuuksilla) ne arvioidaan?

**Esimerkki 5.760**

Toteuttaaksemme kehyksemme ensimmäisen tason ja havaitaksemme moraalisen relevanssin, täydennämme moraalisesti relevantteja siemensanoja vastaavalla siemensanojen joukolla, joka likimain vastaa moraalista epärelevanssia ja perustuu valenssin käsitteeseen eli ärsykkeen miellyttävyyden tai epämiellyttävyyden asteeseen. Viittaamme BIBREF-tietokannan28 keräämiin noin 14 000 englanninkielisen sanan emotionaaliseen valenssiluokitukseen ja valitsemme moraalisesti merkityksettömien ydinsanojen joukkoon sanat, joilla on neutraalimmat valenssiluokitukset ja jotka eivät esiinny MFD:ssä, jotta moraalisesti merkityksellisten ja moraalisesti merkityksettömien sanojen kokonaismäärä olisi yhtä suuri.

**Tulos**

Miten he määrittävät moraalisen merkityksen?

**Esimerkki 5.761**

Aiempien töiden mukaisesti raportoimme tarkkuuden ($P.$), palautuksen ($R.$) ja $F\_1$-pisteet kohteen tunnistusta ja kohdistettua tunnetta varten.

**Tulos**

Miten mallin tehokkuutta arvioidaan?

**Esimerkki 5.762**

IWSLT14:n saksan-englannin ja turkin-englannin perusmallina (BASE-4L) käytämme 4-kerroksista koodaajaa, 4-kerroksista dekooderia, jäännösyhteydellä varustettua mallia, jossa upotuksen ja piilotetun koon oletusarvona on 256.

**Tulos**

mitkä ovat peruslinjat?

**Esimerkki 5.763**

SPNetin tehokkuuden osoittamiseksi vertaamme sitä kahteen uusimpaan menetelmään, Pointer-Generator BIBREF5:een ja Transformer BIBREF6:een.

**Tulos**

Mitä aiempia nykyaikaisia asiakirjojen tiivistämismenetelmiä on käytetty?

**Esimerkki 5.764**

Hiljattain bhat-EtAl:2017:EACLshort toimitti CS-tietokannan jäsentelymalliensa arviointia varten, jonka he kouluttivat hindi- ja englanninkielisillä Universal Dependency (UD) -puupankeilla. Laajennamme tätä tietokokonaisuutta annotoimalla 1448 uutta lausetta.

**Tulos**

Kuinka suuri on toimitettu puupankki?

**Esimerkki 5.765**

Lisäksi käytämme vertailussa kahta erilaista jäsentäjää: 1) PCFGLA:han perustuvaa jäsentäjää eli Berkeley-jäsennystä BIBREF5 ja 2) minimaaliseen spaniin perustuvaa neurologista jäsentäjää BIBREF6 .

**Tulos**

Mikä on sopimukseen perustuvan tilan perusmalli?

**Esimerkki 5.766**

ACL-ARC on BIBREFin7 julkaisema tietokokonaisuus viittausintentioista. Tietokokonaisuus perustuu otokseen ACL Anthology Reference Corpus BIBREF15 -tietokannasta, ja se sisältää 1 941 viittaustapausta 186 artikkelista, ja NLP-alan asiantuntijat ovat kommentoineet sen.

**Tulos**

Mikä on ACL-ARC-tietoaineistojen koko?

**Esimerkki 5.767**

Mallimme koodaa ääni- ja tekstisekvensseistä saadun tiedon käyttämällä kaksois-RNN:iä ja yhdistää sitten näistä lähteistä saadun tiedon feed-forward-neuraalisen mallin avulla tunneluokan ennustamiseksi.

**Tulos**

Miten he yhdistävät ääni- ja tekstisekvenssit RNN:ssä?

**Esimerkki 5.768**

Abstrakti merkitysesitys BIBREF0 , tai lyhyesti AMR, mahdollistaa tämän sisällyttämällä siihen useimmat matalan semanttisen luonnollisen kielen prosessoinnin (NLP) tehtävät, joita käsitellään yleensä erikseen, kuten nimettyjen entiteettien tunnistaminen, semanttisten roolien merkitseminen ja rinnakkaisviittausten erottaminen.

**Tulos**

Mitä osatehtäviä he arvioivat?

**Esimerkki 5.769**

Tulokset osoittavat, että eniten suorituskykyä parantavat sanojen upotukset, tyyli ja moraalin piirteet. Muiden piirteiden (tunteet ja tunne) merkitys on vähäisempi, mutta ne parantavat silti järjestelmän kokonaissuorituskykyä (keskimäärin 0,35 % Macro-F$\_1$ parannus).

**Tulos**

Mikä on tämän asiakirjan perusteella ennakoivampi joukko ominaisuuksia valeuutisten havaitsemiseksi?

**Esimerkki 5.770**

Esimerkiksi lause "Tytöt lauloivat laulun ja tanssivat" on käännetty ranskaksi muotoon "Les filles ont chanté une chanson et ils ont dansé" Google Translate (GT), Bing Translate ja Yandex.

**Tulos**

Tekevätkö kirjoittajat kokeita mainituista tehtävistä?

**Esimerkki 5.771**

Suoritimme kokeilumme Universal Dependencies v1.2 -pankilla BIBREF21 , jäljempänä UD1.2, josta voidaan triviaalisti poimia morfosyntaktisesti annotoituja korpuksia. Tarkastelimme UD1.2:n korpuksia seuraavista 16 kielestä: Englanti, bulgaria, englanti, espanja, indonesia, italia, kroaatti, norja, persia, puola, portugali, ranska, ruotsi, saksa, sloveeni, tanska, tšekki ja ruotsi.

**Tulos**

millä tietokokonaisuuksilla he tekivät kokeita?

**Esimerkki 5.772**

Käytämme LSTM-verkkoa seuraavasti. Vektori $x$ syötetään LSTM-verkon läpi, joka tuottaa vektorin $\overrightarrow{h\_i}$ jokaiselle aika-askeleelle $i$ välillä 0 - $n-1$. Tämä on eteenpäin suuntautuva LSTM. Koska meillä on käytössämme koko vektori $x$, voimme käsitellä myös taaksepäin suuntautuvaa LSTM:ää. Tämä tapahtuu laskemalla vektori $\overleftarrow{h\_i}$ jokaiselle aika-askeleelle $i$ välillä $n-1$ ja 0. Lopuksi yhdistämme takaperin LSTM:n ja eteenpäin LSTM:n: Sekä $\overrightarrow{h\_i}$ että $\overleftarrow{h\_i}$:n ulottuvuus on $l$, joka on optimoitu hyperparametri. BiLSTM:n ulostulon $h$ ulottuvuus on siis $2l\ kertaa n$.

**Tulos**

Onko LSTM kaksisuuntainen?

**Esimerkki 5.773**

Kilpailun suurimmat haasteet johtuvat suoraan mahdollisuudesta käydä jatkuvia monivaiheisia vuoropuheluja, joissa ei noudateta tiettyä suunnitelmaa tai pyritä tiettyyn kiinteään tiedontarpeeseen. Tässä asiakirjassa kuvataan joitakin kokemuksia, joita saimme rakentaessamme SlugBotia vuoden 2017 Alexa-palkintoa varten, ja keskitytään erityisesti haasteisiin, jotka liittyvät haun kautta löydetyn sisällön integroimiseen strukturoidusta datasta peräisin olevaan sisältöön, jotta voidaan käydä jatkuvaa, johdonmukaista, avoimen alueen seka-aloitteista keskustelua Haasteellisempaa on se, että jokaisessa järjestelmäkierroksessa on suuri määrä mahdollisia keskustelun siirtoja. Lisäksi useimmilla muilla aloilla ei ole saatavilla yhtä korkealaatuista strukturoitua dataa, joten meidän on kehitettävä yleisempiä malleja diskurssin koherenssista tai yritettävä luottaa niihin. Hakua ei voida käyttää tehokkaasti tässä yhteydessä ilman, että laaditaan sopiva kysely tai tiedetään etukäteen, mistä juonitietoa on mahdollisesti saatavilla. Reaaliaikaisessa järjestelmässä reaaliaikaisella haulla ei välttämättä saavuteta vaadittua nopeutta ja tehokkuutta, joten olennaisen tiedon esikäsittely tai välimuistiin tallentaminen voi olla tarpeen.

**Tulos**

Miksi monivaiheiset dialogit ovat suurin haaste avoimen alueen keskusteluagenttien rakentamisessa?

**Esimerkki 5.774**

Tässä työssä sovellamme automatisoitua neuraalista annotaatioprosessia dialogin tekojen merkitsemiseen. Useita neuraalisia malleja koulutetaan Switchboard Dialogue Act (SwDA) -korpuksella BIBREF9, BIBREF10, ja niitä käytetään dialogitekojen päättelyyn emootiotietokannoissa. Kokoonpanemme viisi mallin tulostusmerkintää tarkistamalla enemmistöesiintymät (useimmat mallin merkinnät ovat samat) ja asettamalla mallien luottamusarvot paremmuusjärjestykseen. Otimme käyttöön neuraaliset arkkitehtuurit, jotka perustuvat Bothe et al. bothe2018discourseen, jossa on kaksi muunnelmaa: ei-kontekstimalli (luokittelu lausetasolla) ja kontekstimalli (nykyisen lausuman dialogitekojen tunnistaminen muutaman edeltävän lausuman perusteella).

**Tulos**

Kuinka monta mallia käytettiin?

**Esimerkki 5.775**

Koska sisällön ulkopuoliset sanat ovat riittävän erottuvia, jotta luokittelija voi määrittää tyylin, ehdotamme matalan tason kielellisten ominaisuuksien laskentaa (jäljempänä "kontrollit") muodolliseksi, sisällöltään sokeaksi tyylin määritelmäksi. Lauseen tyyli esitetään vektorina, joka koostuu suljettujen sanaluokkien (kuten persoonapronominien) lukumääristä sekä syntaktisten piirteiden lukumääristä, kuten SBAR-ei-terminaalien määrästä lauseenjäsennyksessä, sillä lauserakenteen on osoitettu olevan osoitus tyylistä BIBREF20. Ohjausmerkit poimitaan heuristisesti, ja lähes kaikki perustuvat ennalta määriteltyjen sanaluetteloiden lukumääriin. Käytämme vaalipiiriparsissa Stanford Parser BIBREF21. table:controlexamples sisältää kaikki kontrollit esimerkkeineen.

**Tulos**

Miten he mallintavat tyyliä matalan tason kielellisistä kontrolleista koostuvana kokonaisuutena, kuten pronominien, prepositioiden ja sivulauseen rakenteiden yleisyydestä?

**Esimerkki 5.776**

Käytämme Ultrax Typically Developing -tietokokonaisuutta (UXTD) julkisesti saatavilla olevasta UltraSuite-tietokannasta BIBREF19 . Tämä tietokokonaisuus sisältää synkronoituja akustisia ja ultraäänitietoja 58 tyypillisesti kehittyvältä 5-12-vuotiaalta lapselta (31 naista, 27 miestä).

**Tulos**

Kuinka monta puhujaa tietokannassa on?

**Esimerkki 5.777**

Käytämme BLEU BIBREF30 -mittaria validointijoukkoon VQG-mallin koulutuksessa. BLEU on luonnollisen kielen prosessoinnissa laajalti käytetty mittari, jolla mitataan tuotettujen ja kohdesanojen samankaltaisuutta. Siinä oletetaan, että kelvollisissa generoiduissa vastauksissa on merkittävää sanojen päällekkäisyyttä perustotuuden vastausten kanssa. Käytämme sitä, koska tässä tapauksessa meillä on viisi erilaista viitettä jokaiselle generoidulle kysymykselle. BLEU-pistemääräksi saadaan 2,07. Chatbot-mallissamme on sen sijaan vain yksi viite-maatotuus koulutuksessa, kun se luo sanasarjan. Katsoimme, että se ei ole hyvä metriikka, koska joissakin tapauksissa vastauksilla on sama merkitys, mutta niillä ei ole yhteisiä sanoja. Tallennamme siis useita malleja eri hyperparametreilla ja eri harjoitteluiteraatioiden määrällä ja vertailemme niitä ihmisen arvioinnin avulla valitaksemme mallin, joka toimii paremmin keskustelussa.

**Tulos**

Miten järjestelmän suorituskykyä mitataan?

**Esimerkki 5.778**

Opiskelijoiden pohdintatietokannan koko-ongelman ratkaisemiseksi tutkimme ensin, mitä vaikutuksia on toimialan siirron sisällyttämisellä äskettäiseen abstraktiiviseen yhteenvetomalliin: kattavuusmekanismilla varustettuihin osoitinverkkoihin (PG-net)BIBREF0.

**Tulos**

Mikä on tämän asiakirjan tuore abstrakti tiivistämismenetelmä?

**Esimerkki 5.779**

NLG-malli on seq2seq-malli, jossa on huomio, kuten kohdassa SECREF2 on kuvattu.

**Tulos**

Käyttävätkö he huomiota?

**Esimerkki 5.780**

Viimeaikaiset yritykset voidaan jakaa kahteen luokkaan: i) ne, joissa yritetään sisällyttää lisätietoa tietämysgraafin upottamisen suorituskyvyn parantamiseksi edelleen, esimerkiksi oliotyypit tai käsitteet BIBREF13, relaatiopolut BIBREF17, tekstikuvaukset BIBREF11, BIBREF12 ja loogiset säännöt BIBREF23; ii) ne, joissa yritetään suunnitella monimutkaisempia strategioita, esimerkiksi syvät neuroverkkomallit BIBREF24.

**Tulos**

Mitä viimeaikaisia teoksia knowedge-graafien sulautuksista kirjoittajat mainitsevat?

**Esimerkki 5.781**

Tässä asiakirjassa tutkimme todennettavissa olevaa kestävyyttä eli todistuksen antamista siitä, että tietyn verkon ja testitulon osalta mikään hyökkäys tai häiriö ei voi muuttaa ennusteita. Esimerkkinä käytetään tekstiluokittelutehtäviä, Stanford Sentiment Treebank (SST) BIBREF15 ja AG News BIBREF16.

**Tulos**

Mitä tietokokonaisuutta he käyttävät?

**Esimerkki 5.782**

Kasvojen läsnäolon tallentamisessa turvaudumme BIBREF56 -lähestymistapaan, jossa käytetään monitasoista konvolutiivista karkeasta hienoon -verkkokaskadia kasvojen maamerkkien paikantamiseen. Kasvojen ilmeet: BIBREF8:n lähestymistavan mukaisesti otamme käyttöön Ekmanin mallin kuudesta tunteesta: viha, inho, pelko, ilo, suru ja yllätys, ja käytämme Face++ API:ta niiden automaattiseen kuvaamiseen jaetuista kuvista. Yleiset kuvaominaisuudet: Tulkittavien laskennallisten esteettisten piirteiden merkitys käyttäjien verkkokäyttäytymisen tutkimisessa on korostunut useissa toimissa BIBREF55 , BIBREF8 , BIBREF57 . Laadullinen kielianalyysi: Tuoreessa LIWC-versiossa tiivistetään tekstisisältöä kielellisten muuttujien, kuten analyyttisen ajattelun, vaikuttavuuden, autenttisuuden ja emotionaalisen sävyn, perusteella. Se mittaa myös muita kielellisiä ulottuvuuksia, kuten kuvailuluokkia (esim. sanakirjasta poimittujen tai yli kuuden kirjaimen pituisten kohdesanojen prosenttiosuus - Sixltr) ja epävirallisia kielimerkkejä (esim. kirosanat, netspeak) sekä muita kielellisiä näkökohtia (esim. 1. persoonan yksikön pronominit).

**Tulos**

Minkälaisia ominaisuuksia kussakin tietotyypissä käytetään?

**Esimerkki 5.783**

Vaikka toimialueen siirto ei ole uutta, verrattuna aiempiin yhteenvetotutkimuksiin BIBREF6, BIBREF7, harjoitusalueemme (uutiset) ja viritysalueemme (opiskelijoiden pohdinta) ovat varsin erilaisia, ja toimialueen sisäiset tiedot ovat pieniä.

**Tulos**

Poikkeavatko opiskelijoiden pohdintatiedot suuresti sanomalehden tiedoista?

**Esimerkki 5.784**

Tehtävässä 1 käytämme arviointimittarina F1-tulosta. Tehtävässä 2 käytämme manuaalista arviointia. Tehtävää 2 varten on viisi arviointimittaria, jotka ovat seuraavat: Tehtävän suorittamisaste: Käyttäjätyytyväisyysaste: Suoritettujen tehtävien määrä jaettuna tehtävien kokonaismäärällä: On viisi pistemäärää -2, -1, 0, 1, 2, jotka tarkoittavat vastaavasti erittäin tyytymätöntä, tyytymätöntä, neutraalia, tyytyväistä ja erittäin tyytyväistä.Vastauksen sujuvuus: Kolme pistemäärää -1, 0, 1, jotka ilmaisevat sujumattomuutta, neutraaliutta ja sujuvuutta.Vuoropuhelukierrosten määrä: Tehtävässä suoritetun dialogin lausumien määrä.Guidance ability for out of scope input: On kaksi pistemäärää 0, 1, jotka kuvaavat ohjauskykyä tai ohjauskyvyttömyyttä.

**Tulos**

Mitä mittareita arvioinnissa käytetään?

**Esimerkki 5.785**

Käytimme Visual storytelling (VIST) -tietokokonaisuutta, joka koostuu Flickr-albumeista saaduista kuvasarjoista ja niihin liittyvistä kommentoiduista kuvauksista, jotka on kerätty Amazon Mechanical Turk BIBREF1 -palvelun kautta. Jokaisessa sarjassa on 5 kuvaa ja niitä vastaavat kuvaukset, jotka yhdessä muodostavat tarinan. Lisäksi jokaisessa Flickr-albumissa on 5 permutaatiota sen kuvista valitusta joukosta. Käytettävissä olevassa kokonaisaineistossa on 40 071 harjoitus-, 4 988 validointi- ja 5 050 käyttökelpoista testitarinaa.

**Tulos**

Mitä tilastoja VIST-tietokannasta raportoidaan?

**Esimerkki 5.786**

Nopeuden ja suorituskyvyn välisen kompromissin tutkimiseksi vertaamme tekniikkaamme vakiomalleihin, joissa on huomio ja joissa ei ole huomiointia, BIBREF:in14 kaltaisessa eri pituisessa sekvenssikopiointitehtävässä. Tähän tarkoitukseen käytämme WMT'17:n neljää suurta konekäännöstietokokonaisuutta, jotka koskevat seuraavia kielipareja: (en-cs, 52 miljoonaa esimerkkiä), englanti-saksa (en-de, 5,9 miljoonaa esimerkkiä), englanti-suomi (en-fi, 2,6 miljoonaa esimerkkiä) ja englanti-turkki (en-tr, 207 373 esimerkkiä).

**Tulos**

Mitä tietokokonaisuuksia kokeissa käytetään?

**Esimerkki 5.787**

Arvioimme kaikkia viittä mallia myös VecEval-paketin BIBREF13 myöhemmissä tehtävissä käyttäen vain sellaisia tehtäviä, joiden koulutus- ja arviointidata on vapaasti saatavilla: klusterien muodostaminen, tunteiden ja kysymysten luokittelu sekä luonnollisen kielen tunnistaminen (NLI). Käytetään paketin oletusasetuksia, mutta ajetaan vain kiinteitä asetuksia, joissa upotukset eivät ole mallien viritettävissä olevia parametreja, mikä pakottaa järjestelmän käyttämään vain upotuksissa jo olevaa tietoa. Lopuksi käytetään LV-N:ää, LV-M:ää ja FT:tä tuottamaan OOV-sanarepresentaatioita seuraaville sanoille: 1) "hellooo": pikaviesteissä yleisesti käytetty tervehdys, jossa korostetaan tavua. 2) "marvelicious": keksitty sana, joka saadaan yhdistämällä "marvelous" ja "delicious". 3) "louisana": oikean nimen "Louisiana" kirjoitusvirhe. 4) "rereread": etuliitteen "re" rekursiivinen käyttö. 5) "tuzread": keksitty etuliite "tuz".

**Tulos**

Miten he arvioivat tuloksena syntyviä sanojen upotuksia?

**Esimerkki 5.788**

Dekooderi on toinen LSTM-verkko, joka käyttää kooderista saatua tietoa sekvenssin tarinan luomiseen. Dekooderin ensimmäinen syötetieto $x\_0$ on kuva, jolle teksti luodaan. Kooderin viimeistä piilotettua tilaa $h\_e^{(t)}$ käytetään dekooderin ensimmäisen piilotetun tilan $h\_d^{(0)}$ alustamiseen. Tällä strategialla annamme dekooderille koko sekvenssin kontekstin ja nykyisen kuvan sisällön (eli globaalin ja paikallisen tiedon), jotta se voi tuottaa vastaavan tekstin, joka edistää kokonaistarinaa.

**Tulos**

Miten tarinan peräkkäinen luonne on kuvattu?

**Esimerkki 5.789**

Opimme, että useat tekijät voivat vaikuttaa vastahyökkäysten suorituskykyyn, kuten luokittelijan arkkitehtuuri, lauseen pituus ja syöttöalue.

**Tulos**

Mitkä muut tekijät vaikuttavat suorituskykyyn?

**Esimerkki 5.790**

Kehittyneiden merkkien syöttöön perustuvien neuroarkkitehtuurien (CNN, BPE jne.) oletetaan pystyvän oppimaan itse, miten oikeinkirjoituksen ja morfologian vaihtelut käsitellään, jopa kielissä, joissa on rikas morfologia: "lisää vain kerroksia!". Kontekstualisoidut upotusmallit noudattavat tätä perinnettä: ne koulutetaan pääsääntöisesti raakatekstikokoelmilla, joissa on vain vähän kielellistä esikäsittelyä. Jäljempänä osoitamme, että tämä ei ole täysin totta.

**Tulos**

Miksi lemmatisointi ei ole välttämätöntä englannissa?

**Esimerkki 5.791**

Muut Timex3-maininnat ovat epäselvempiä, joten käytämme etävalvontaa. Lauseet kuten "currently" ja "today's" esiintyvät yleensä lähellä Tapahtumia, jotka ovat päällekkäisiä nykyisen asiakirjan luomisen ajankohdan kanssa, kun taas "ago" tai " INLINEFORM0 -years" viittaavat menneisiin tapahtumiin. Nämä hallitsevat ajalliset assosiaatiot voidaan oppia harjoitusaineistosta ja käyttää sitten Timex3:n merkitsemiseen. Lopuksi määrittelemme logistisen regressiosäännön ennustamaan olioiden DocRelTime-arvoja sekä määrittelemme lineaarisen hyppyketjutekijän Tapahtumamainintojen ja niiden lähimmän Timex3-naapurin välille, mikä koodaa perusjärjestelmän heuristiikan suoraan päättelysääntönä.

**Tulos**

Miten he saavat kaukovalvontasäännöt suhteiden ennustamista varten?

**Esimerkki 5.792**

Tiedot kerättiin 10 prosentin yhtenäisestä otoksesta vuoden 2013 aikana tehdyistä Twitter-viesteistä, erityisesti Gardenhose API:sta. Lisäksi valittiin kontrolliasiakirjoja. Nämä asiakirjat eivät sisältäneet yhtään sanaa `caused', `causing' tai `causes', eivätkä mitään kaksisuuntaisia sanoja, ja ne sovitettiin edelleen ajallisesti yhteen, jotta saatiin sama määrä kontrolliasiakirjoja kuin kausaalisia asiakirjoja kullakin viidentoista minuutin jaksolla vuoden 2013 aikana. Kontrollidokumentit valittiin muuten satunnaisesti; kausaalisynonyymit voivat olla läsnä.

**Tulos**

Miten he keräävät valvontakorpuksen?

**Esimerkki 5.793**

Asiakirjan ominaisuuksien sisällyttäminen ::: Oppimisvaihe ::: Ominaisuuksien poiminta ::: Document-unware FeaturesOrdinal position: On osoitettu, että lauseen sisällyttäminen tiivistelmään on merkityksellistä sen sijainnin kannalta asiakirjassa tai jopa kappaleessa. Intuitiivisesti voidaan todeta, että tekstin alussa tai lopussa olevat lauseet sisällytetään todennäköisemmin tiivistelmään. Riippuen siitä, miten tämä ominaisuus määritellään, se voi olla asiakirjasta riippumaton tai ei. Esimerkiksi BIBREF29:ssä ja BIBREF37:ssä se on määritelty seuraavasti: $\frac{5}{5}$ ensimmäiselle lauseelle, $\frac{4}{5}$ toiselle lauseelle ja niin edelleen $\frac{1}{5}$ viidennelle lauseelle ja nolla muille lauseille. Toisessa Wong et al. BIBREF9:n tutkimuksessa se on määritelty muotoon $\frac{1}{lauseen numero}$. Tällaisella määritelmällä meillä voi olla useita lauseita, joiden asema on esimerkiksi $\frac{1}{5}$, mutta näillä lauseilla ei välttämättä ole samaa aseman merkitystä. Vaikka lause position=$\\frac{1}{5}}$ tarkoittaa "ensimmäisten joukossa" asiakirjassa, jossa on 40 lausetta, sillä on täysin erilainen merkitys "keskellä" toisessa asiakirjassa, jossa on 10 lausetta. Hyödyllisen ominaisuuskaavan tulisi siis sisältää asiakirjojen välisiä eroja, jotka voivat muuttaa siinä olevan tiedon merkitystä. Kokeissamme käytimme BIBREF9:n määritelmää. Asiakirjatietoinen versio sijainnista esitellään kohdassa (SECREF6). lauseen pituus: tämän ominaisuuden taustalla on intuitio, jonka mukaan liian pitkät tai liian lyhyet lauseet sisällytetään harvemmin tiivistelmään. Kuten lauseen sijainti, myös tämä ominaisuus on väärän määritelmän alainen, mikä tekee siitä asiakirjasta tietämättömän. Esimerkiksi BIBREF9:ssä se on määritelty lauseen sanojen lukumääräksi. Tällaisessa määritelmässä ei oteta huomioon sitä, että esimerkiksi 15 sanaa sisältävää lausetta voidaan pitää pitkänä, jos asiakirjan kaikissa muissa lauseissa on vähemmän sanoja. Toista lausetta, jossa on sama määrä sanoja, voidaan pitää lyhyenä, koska asiakirjan muissa lauseissa on enemmän kuin 15 sanaa. Tämä voi johtua erilaisista kirjoitustyyleistä. Otimme tämän kuitenkin mukaan kokeiluihimme, jotta voimme verrata sen vaikutusta sen dokumenttitietoisen vastineen vaikutukseen, joka luetellaan kohdassa (SECREF6). Substantiivien suhde: määritellään BIBREF30:ssä substantiivien lukumääräksi jaettuna lauseen sanojen kokonaislukumäärällä sen jälkeen, kun stop-sanat on poistettu. Kolme muuta ominaisuutta, verbejä, adjektiiveja ja adverbejä, määritellään samalla tavalla, ja niillä on osoittautunut olevan myönteinen vaikutus ranking-tulokseen. Meidän näkökulmastamme katsottuna lause, jonka substantiivien suhde on = 0,5, esimerkiksi asiakirjassa, jossa on paljon substantiiveja, on kuitenkin erotettava harjoitusjoukossa toisesta lauseesta, jossa on sama substantiivien suhde ja joka esiintyy toisessa asiakirjassa, jossa on vähemmän substantiiveja. Tämä ominaisuus ei kuvaa sitä, kuinka monta substantiivia asiakirjassa on, mikä on tärkeää lauseiden luokittelussa. Samassa keskustelussa perustellaan myös tarvetta ottaa huomioon verbien, adjektiivien ja adverbien määrä asiakirjassa. Näiden ominaisuuksien vaikutusta tutkitaan kokeiluissamme ja verrataan niiden dokumenttitietoisten vastineiden vaikutukseen. numeeristen yksiköiden suhde: Olettaen, että lauseet, jotka sisältävät enemmän numeerista tietoa, antavat meille todennäköisesti enemmän tietoa, tämä ominaisuus voi auttaa meitä sijoittamaan BIBREF31, BIBREF32. Laskennassa lasketaan numeroiden ja numeroiden esiintymät suhteessa lauseen pituuteen. Tätä ominaisuutta on painotettava vähemmän, jos lähes kaikissa asiakirjan lauseissa on numerotietoja. Se ei kuitenkaan laske numeroita ja numeroita asiakirjan muissa lauseissa. Cue Words: Jos lause sisältää erityisiä lauseita, kuten "in conclusion", "overall", "to summarize", "in a nutshell" ja niin edelleen, sen valinta tiivistelmän osaksi on todennäköisempää kuin muiden. Näiden lauseiden määrä lasketaan tätä ominaisuutta varten.Asiakirjan ominaisuuksien sisällyttäminen ::: Oppimisvaihe ::: Ominaisuuksien louhinta ::: Asiakirjatietoiset piirteetKosinuksen sijainti: Kuten kohdassa (SECREF5) mainittiin, hyvässä aseman määritelmässä olisi otettava huomioon asiakirjan pituus. Kirjallisuudessa käytetty tunnettu kaava BIBREF38, BIBREF7 on, jossa indeksi on kokonaisluku, joka kuvaa lauseiden järjestystä, ja T on asiakirjan lauseiden kokonaismäärä. Tämä ominaisuus vaihtelee välillä 0-1, ja mitä lähempänä alkua tai loppua se on, sitä suuremman arvon se saa. $\alpha $ on viritysparametri. Kun se kasvaa, tämän ominaisuuden arvo jakautuu tasaisemmin lauseiden kesken. Tällä tavoin tämän ominaisuuden yhtäläiset arvot harjoitusjoukossa edustavat yhtenäistä käsitystä sijainnista asiakirjassa, joten siitä tulee asiakirjatietoinen. suhteellinen pituus: Tämän ominaisuuden taustalla oleva intuitio on selitetty kohdassa (SECREF5). Siellä keskusteltiin siitä, että pelkkä sanojen laskeminen ei ota huomioon sitä, että tietyn sanamäärän sisältävää lausetta voidaan pitää pitkänä tai lyhyenä asiakirjan muiden lauseiden perusteella. Tämä huomioon ottaen jaoimme lauseen sanamäärän asiakirjan lauseiden keskimääräisellä pituudella. Muodollisemmin kaava on: jossa n on asiakirjan lauseiden lukumäärä ja $s\_i$ on asiakirjan i:s lause. Arvot, jotka ovat suurempia kuin 1, voidaan tulkita pitkiksi ja päinvastoin.TF-ISF: Tämä ominaisuus laskee termien frekvenssiä asiakirjassa ja antaa korkeampia arvoja lauseille, joissa on useammin esiintyviä termejä. Se myös vähentää termejä, jotka esiintyvät useammassa lauseessa. Koska se on selitetty hyvin kirjallisuudessa, emme ole ottaneet mukaan yksityiskohtia ja kaavaa, jotka löytyvät viitteistä BIBREF34 ja BIBREF39. Keskustelumme kannalta on kuitenkin olennaista, että sekä frekvenssi että käänteinen lauseen frekvenssi ovat termejä, joihin liittyy asiayhteyden ominaisuuksia ja jotka näin ollen ovat dokumenttitietoisia.POS-ominaisuuksia: Tässä esitellään toinen tapa sisällyttää POS-yksiköiden (part of speech) suhde piirteisiin ja pitää ne asiakirjanormalisoituina. Tätä varten kunkin POS-yksikön esiintymien määrä jaetaan niiden määrällä asiakirjassa sen sijaan, että ne esiintyisivät lauseessa. Uusien dokumenttitietoisten piirteiden muodollinen määritelmä on seuraava: Incorporating Document Features ::: Oppimisvaihe ::: Ominaisuuksien louhinta :::: Eksplisiittiset dokumenttiominaisuudetTutkittaessa tarkemmin, kuinka tehokkaita dokumenttikohtaiset ominaisuudet ovat lauseiden luokittelussa, määrittelimme useita dokumenttiominaisuuksia. Nämä ominaisuudet lasketaan kullekin asiakirjalle ja toistetaan kyseisen asiakirjan jokaisen lauseen ominaisuusvektorissa. Niiden muodollinen määritelmä kuvataan jäljempänä, ja niiden vaikutusta tarkastellaan tulos- ja keskusteluosassa (SECREF5): Asiakirjan lauseet: Tärkeä tiivistämiseen vaikuttava asiakirjan ominaisuus on lauseiden luokitteluun osallistuvien lauseiden kokonaismäärä. Kun tämä määrä kasvaa, tiivistelmän pitäisi olla valikoivampi ja tarkempi. Myös jotkin lauseiden ominaisuudet, kuten vihjesanat, saattavat painottua enemmän pidemmissä asiakirjoissa. Lisäksi tärkein asiayhteyteen liittyvä tieto on luultavasti jakautunut enemmän lauseisiin. Tällöin muiden ominaisuuksien pienempiä arvoja olisi pidettävä tärkeinä. Asiakirjan sanat: Asiakirjan sanojen määrä on toinen asiakirjan pituutta kuvaava käsite. Koska lauseiden määrä ei yksin riitä kuvaamaan asiakirjan pituutta, tämä ominaisuus olisi myös otettava huomioon. aihepiiriluokka: eri aihealueilla, kuten poliittisilla, taloudellisilla jne. aiheilla on erilaiset kirjoitustyylit, mikä saattaa vaikuttaa lauseiden sijoittumiseen. Esimerkiksi numeeriset yksiköt saattavat esiintyä enemmän talous- tai urheiluraporteissa kuin uskonnollisissa tai yhteiskunnallisissa uutisissa. Siksi tämän ominaisuuden painoarvon tulisi olla suurempi tai pienempi asiakirjan kategorian perusteella. Se on siis sisällytettävä.

**Tulos**

Mitä asiakirjan piirteitä on integroitu jokaisen lauseen vektoreihin?

**Esimerkki 5.794**

Kun suunnittelimme uusia arviointimittareita, joiden avulla vertaillaan viitteellisiä tiivistelmiä/käännöksiä hypoteeseihin, laadimme ensimmäiset periaatekriteerit sille, mitä hyvän arvioijan tulisi tehdä. Ensimmäinen niistä on se, että sen pitäisi korreloida vahvasti ihmisen tekemän samankaltaisuusarvion kanssa. Toinen kriteeri on, että sen pitäisi pystyä erottamaan lauseet, jotka ovat loogisesti ristiriidassa keskenään, jotka eivät liity loogisesti toisiinsa tai jotka ovat loogisesti samaa mieltä. Kolmanneksi vankan arvioijan pitäisi pystyä tunnistamaan myös käsittämättömät lauseet. Viimeisenä kriteerinä on, että hyvä arviointimittari ei saisi antaa korkeita pistemääriä semanttisesti kaukana toisistaan oleville lauseille ja alhaisia pistemääriä semanttisesti toisiinsa liittyville lauseille.

**Tulos**

Mitkä ovat hyvän mittarin kriteerit?

**Esimerkki 5.795**

Kokeilemme kahta kuvatekstimallia: Show&Tell-mallia BIBREF0 ja LRCN1u-mallia BIBREF1. Molemmissa malleissa noudatetaan koodaaja-dekooderi-arkkitehtuurin perusrakennetta, jossa CNN-kooderi tiivistää visuaalisen informaation kuvakoodaukseksi, joka puolestaan antaa LSTM-dekooderille edellytykset luonnollisen kielen kuvatekstin tuottamiseen. Suurin ero näiden kahden mallin välillä on tapa, jolla ne ehdollistavat dekooderia. Show&Tell-malli syöttää kuvan upotuksen "edeltävän sanan upotuksena" ensimmäiselle tuotetulle sanalle, kun taas LRCN1u-malli yhdistää kuvan piirteet ja upotetun edellisen sanan sekvenssimallin syötteeksi kullakin aika-askeleella.

**Tulos**

Mitä olemassa olevia malleja arvioidaan?

**Esimerkki 5.796**

Tässä työmuistiossa esitellään RNN- ja LSTM-pohjainen upotusjärjestelmä sosiaalisen median terveystekstien luokittelua varten. Vaikka tehtävän 1 ja tehtävän 2 tietokokonaisuudet ovat rajalliset, tässä työpaperissa ehdotetaan RNN- ja LSTM-pohjaista upotusmenetelmää.

**Tulos**

Minkä tyyppistä RNN:ää käytetään?

**Esimerkki 5.797**

Kaikissa kokeissamme käytimme BERT-Base BIBREF2:ta, joka tarjoaa uusinta mahdollista kontekstualisoitua mallinnusta syötetylle tekstille.Semanttinen haku: Käsittelimme neuraalista semanttista hakua sekä kappale- että lausetasolla binäärisinä luokitusongelmina, joissa mallien parametrit päivitetään minimoimalla binäärinen ristiinentropiahäviö.

**Tulos**

Miten ne mallintavat neuraalisia hakumoduuleja?

**Esimerkki 5.798**

MCTest-tietokannan osalta kuvassa FIGREF30 verrataan ehdottamiamme malleja käsin laadittujen syntaktisten ja kehyssemanttisten piirteiden nykyiseen huipputason kokonaisuuteen BIBREF16 sekä kirjallisuuden aiempiin neuromalleihin, jotka kaikki käyttävät huomiomekanismeja - BIBREF26 Attentive Reader, BIBREF27 Neural Reasoner ja BIBREF17 HABCNN-malliperhe.

**Tulos**

Mikä on mc-testien vastausten luokittelun nykytilanne?

**Esimerkki 5.799**

Kukin luokittelija on toteutettu seuraavilla määrittelyillä: Naïve Bayes (NB): Logistinen regressio (LR): Lineaarinen LR, L2-regularisointivakio 1 ja rajoitetun muistin BFGS-optimointiTukivektorikone (SVM): Lineaarinen SVM, L2-regularisointivakio 1 ja logistinen häviöfunktioSattumanvaraiset metsät (RF): 10 satunnaistetun päätöspuun todennäköisyysennusteiden keskiarvoistaminenGradient Boosted Trees (GBT): Tree boosting oppimisnopeudella 1 ja logistisella häviöfunktiolla Perinteisten koneoppimismenetelmien ohella tutkimme neuroverkkoihin perustuvia malleja arvioidaksemme niiden tehokkuutta suuremmassa tietokokonaisuudessa. Tutkimme erityisesti konvolutiivisia neuroverkkoja (Convolutional Neural Networks, CNN), toistuvia neuroverkkoja (Recurrent Neural Networks, RNN) ja niiden muunnosmalleja.

**Tulos**

Mitä oppimismalleja tietokokonaisuuteen käytetään?

**Esimerkki 5.800**

Taulukossa TABREF15 vertaillaan SimCluster-algoritmin ja K-means-algoritmin tuloksia. Tässä SimCluster-algoritmimme parantaa F1-pistemäärää 0,412:sta ja 0,417:stä 0,442:een ja 0,441:een kahdella alueella. Myös ARI-pisteet paranevat 0,176:sta ja 0,180:sta 0,203:een ja 0,204:een.

**Tulos**

Käytetäänkö SimCluster- ja K-means-algoritmissa samaa etäisyysmittaria?

**Esimerkki 5.801**

Tehtävä on laadittu jäljittelemään (vaikkakin liian yksinkertaistetusti) tulo- ja lähtösymbolien kohdistuksia ja paikallisia syntaktisia ominaisuuksia, jotka mallien on opittava monissa luonnollisen kielen tehtävissä, kuten käännöksissä, merkinnöissä ja yhteenvedoissa.

**Tulos**

Miksi ehdotettu tehtävä on hyvä korvike yleiskäyttöiselle sekvenssille sekvenssitehtäville?

**Esimerkki 5.802**

Teimme siis myös ihmiskokeita Amazon MTurk -palvelussa arvioidaksemme tuotettuja vastauksia pareittain vertailemalla vuoropuhelun laatua. Vertaamme mallejamme kehittyneeseen dekoodausalgoritmiin MMI BIBREF2 ja kahteen malliin, nimittäin LSTM BIBREF0 ja VHRED BIBREF7, jotka molemmat käyttävät additiivista huomiota. Parhaan tietämyksemme mukaan LSTM ja VHRED olivat ensisijaiset mallit, joilla F1-tuloksia raportoitiin Ubuntu-tietokannassa. BIBREF5:n (BIBREF5) mukaisesti käytämme kahta kriteeriä: Uskottavuus ja sisällön rikkaus. Ensimmäinen kriteeri mittaa sitä, onko vastaus uskottava asiayhteys huomioon ottaen, kun taas toinen mittaa sitä, onko vastaus monipuolinen ja informatiivinen.

**Tulos**

Miten ihmisten arviointi suoritetaan, mitkä olivat arviointiperusteet?

**Esimerkki 5.803**

Analysoimme laajennetuilla piirteillä koulutetun luokittimen virheet neljässä ennustustehtävässä, jotta voimme antaa suuntaa tulevalle työlle.Virheet aikomuksen (I) ennustamisessa: Taustan puute on suuri ongelma trollauskommenttien tunnistamisessa. Muut kuin kiroilevat hyökkäykset ja loukkaukset Tämä on haastava ongelma, koska suurin osa loukkaavista ja loukkaavista kommenteista perustuu kirosanoihin ja kiroiluun. Toinen virhelähde on trollien yleisesti käyttämät kiistanalaiset sanat, kuten "musta", "feminismi", "tappaminen", "rasismi", "ruskea" jne. Virheitä julkistamisen (D) ennusteessa: Tärkein julkistamiseen vaikuttava virhelähde on BOW-mallista saatava pinnallinen merkityskuvaus, vaikka sitä täydennettäisiinkin hansikasvektoreiden antamilla jakauman ominaisuuksilla. Tulkinnan (R) ennustamiseen liittyvät virheet: Monilla käyttäjillä on yleinen käytäntö kysyä suoraan epäillyltä trollilta, trollaako hän vai ei. Vastausstrategiaa (B) koskevat virheet: Joissakin tapauksissa "turhauttamisen" ja "neutralisoinnin" välinen raja on häilyvä. Toinen haastava ongelma on luokkien "Trolli" ja "Osallistu" erottaminen toisistaan.

**Tulos**

Mikä on esimerkki vaikeasti luokiteltavasta tapauksesta?

**Esimerkki 5.804**

Käytämme Kyubyong Parkin ja Edouard Graven osoitteessa al:Kyubyong Park luomia vietnaminkielisiä sanasulkeumia: Hänen projektissaan hän käyttää kahta menetelmää, mukaan lukien fastText ja word2vec, luodakseen sanojen upotuksia wikipedia-tietokannan varmuuskopioista. Edouard Grave et al BIBREF11: He käyttävät fastText-työkalua sanojen upotusten tuottamiseen Wikipediasta.

**Tulos**

Mitä sanasulkeumia käytettiin?

**Esimerkki 5.805**

Arvioijia ohjeistettiin arvioimaan vastausten laatua kolmen kriteerin perusteella: (1) kieliopillinen oikeellisuus - onko vastaus sujuva ja kielioppivirheetön; (2) kontekstuaalinen johdonmukaisuus - onko vastaus kontekstisidonnainen edelliseen dialogihistoriaan nähden; (3) emotionaalinen sopivuus - välittääkö vastaus oikean tunteen ja tuntuuko se siltä, että sen on tuottanut ihminen. Arvioijat antoivat kullekin kriteerille joko 0, 1 tai 2 pistettä, jossa 0 tarkoittaa huonoa, 2 hyvää ja 1 neutraalia.

**Tulos**

Miten ihmisen arviointi suoritetaan?

**Esimerkki 5.806**

Arvioimme menetelmäämme kolmessa merkintätehtävässä: POS-tunnisteiden (Pos), morfologisten tunnisteiden (Morph) ja supertunnisteiden (Stag) tekemistä. Valitsimme nämä tehtävät esimerkkeinä merkintäsovelluksista, koska ne eroavat toisistaan voimakkaasti merkintäjoukkojen koon suhteen. Näiden kolmen tehtävän testitulokset esitetään taulukossa TABREF17 kolmessa ryhmässä. Ensimmäisessä seitsemän sarakkeen ryhmässä ovat tulokset Pos-tehtävässä, jossa sekä LSTM:llä että CNN:llä on kolme variaatiota syöttöominaisuuksista: vain sana ( INLINEFORM0 ), vain merkki ( INLINEFORM1 ) ja molemmat ( INLINEFORM2 ). Morphin ja Stagin osalta käytämme vain INLINEFORM3-asetusta sekä LSTM:ssä että CNN:ssä.

**Tulos**

Selvittävätkö ne yhdessä useita merkintäongelmia?

**Esimerkki 5.807**

Online Retail Data Set koostuu puhtaasta 25873 laskun luettelosta, jossa on yhteensä 541909 riviä ja 8 saraketta. InvoiceNo, CustomerID ja StockCode ovat enimmäkseen 5- tai 6-numeroisia kokonaislukuja, joissa on satunnaisesti kirjaimia. Quantity on enimmäkseen 1-3-numeroisia kokonaislukuja, joista osa on negatiivisia, ja UnitPrice koostuu 1-6-numeroisista liukuluvuista. InvoiceDate on päivämääriä, jotka ovat kaikki samassa muodossa, Country sisältää merkkijonoja, jotka edustavat 38 maata, ja Description on 4224 merkkijonoa, jotka edustavat tuotteiden nimiä. Näistä tiedoista muodostetaan tekstiviestit erottamalla jokainen merkki tyhjällä välilyönnillä ja pinoamalla tietyn laskun rivit, jotka on ryhmitelty InvoiceNo:n mukaan.

**Tulos**

Mikä on taulukoiden lähde?

**Esimerkki 5.808**

Offline-vaiheessa järjestämme kunkin tapauksen sisällön graafiksi, jossa jokainen solmu edustaa lausetta tekstistä ja kysymystä. Tämän jälkeen lisäämme solmujen väliset reunat seuraavan topologian mukaisesti: yhdistämme täysin solmut, jotka edustavat saman tekstikappaleen lauseita (katkoviiva-musta); yhdistämme täysin solmut, jotka edustavat kunkin tekstikappaleen ensimmäistä lausetta (katkoviiva-punainen); lisäämme reunan kysymyksen ja kunkin tekstikappaleen jokaisen solmun välille (katkoviiva-sininen).

**Tulos**

Miten jotkin solmut ovat alun perin yhteydessä toisiinsa tekstin rakenteen perusteella?

**Esimerkki 5.809**

Jotta vertailua aiempaan työhön voitaisiin tehdä, koulutamme perus-muotoisen tekstinmuunnosjärjestelmän BIBREF11:n julkaisemilla tiedoilla.

**Tulos**

Mikä oli lähtötaso?

**Esimerkki 5.810**

Tässä artikkelissa esittelemme työmme, joka on osa ECML PKDD 2019 -konferenssin SociaL Media And Harassment -kilpailua. Käytämme kilpailun tietokokonaisuutta, joka sisältää tekstiä twiiteistä, joilla on edellä mainitut kategoriat.

**Tulos**

Mitä tietokokonaisuuksia tässä asiakirjassa käytettiin?

**Esimerkki 5.811**

Samanlainen suuntaus on nähtävissä myös kahden muun uusimman lähestymistavan, BIBREF9 ja BIBREF8 , suorituskyvyssä.

**Tulos**

Mitkä ovat uusimmat mallit?

**Esimerkki 5.812**

Edellä esitetyllä hyperparametriasetuksella hybridi-NER-malli saavutti F1-pistemäärän $0,995$ syntetisoidussa kyselyssä ja $0,948$ kliinisissä muistiinpanoissa, kun taas i2b2-NER-malli saavutti F1-pistemäärän $0,441$ syntetisoidussa kyselyssä ja $0,927$ kliinisissä muistiinpanoissa (ks. taulukko TABREF23).

**Tulos**

mitkä olivat heidän suorituskykytuloksensa?

**Esimerkki 5.813**

Tutkimme, miten voidaan hyödyntää maalaisjärjellä saatavaa lisätietoa (eli rationaalisia perusteluja) tehtävän syötteenä. Kuten kohdassa SECREF6 mainitsimme, etsimme OMCS-korpuksesta relevantteja lauseita ylimääräisiksi etäisiksi järkiperusteluiksi ja perustotuusperustelujen lauseita dev/testidataa varten. Syötteet eivät enää ole itse käsitejoukkoja, vaan muodossa "[rationales$|$concept-set]" (eli perustelulauseiden ja alkuperäisten käsitejoukkojen merkkijonojen ketjuttaminen).

**Tulos**

Pitääkö mallien tuottaa myös rationaalisia perusteluja?

**Esimerkki 5.814**

Taulukossa TABREF44 esitetään automaattisten ja inhimillisten arviointien keskimääräiset tulokset.

**Tulos**

Kuinka suuri ero ehdotetun mallin ja perusmallin suorituskyvyssä on?

**Esimerkki 5.815**

Porttimekanismi oppii toimialueiden väliset representaatiot. Ne yhdessä ohjaavat tietoa, jonka on kuljettava täysin kytketyn ulostulokerroksen läpi maksimipoolauksen jälkeen. Porttiarkkitehtuurien tehokkuus perustuu ajatukseen, jonka mukaan portti koulutetaan ainoastaan painotuksen määrittämiseksi. Tunneanalyysin tehtävässä tämä painotus vastaa sitä, mitkä painot johtavat lopullisen tappion pienenemiseen tai toisin sanoen tarkimpaan tunteen ennustamiseen. Näin tehdessään porttiarkkitehtuuri oppii, mitkä sanat tai n-grammit vaikuttavat eniten sentimenttiin, ja nämä sanat tai n-grammit ovat usein yhteydessä toimialasta riippumattomiin sanoihin. Toisaalta portti antaa vähemmän painoarvoa n-grammeille, jotka ovat suurelta osin joko alalle ominaisia tai funktionaalisia sanakokonaisuuksia, joiden osuus yleiseen sentimenttiin on vähäinen. Tämä tekee porttiarkkitehtuurista tehokkaan toimialueen mukauttamisessa. Näemme, että porttiarkkitehtuurit ovat lähes aina parempia kuin toistuvat, huomio- ja lineaariset mallit BoW, TFIDF ja PV. Tämä johtuu pitkälti siitä, että kun näitä malleja koulutetaan ja testataan samoilla aloilla, erityisesti toistuvat ja huomiopohjaiset mallit saattavat suoriutua paremmin. Domain Adaptation -mallit eivät kuitenkaan suoriudu kohdealueesta huonosti verrattuna gated-arkkitehtuureihin, koska niistä puuttuu gated-rakenne, jota koulutetaan rinnakkain tärkeyden oppimiseksi. Koska gated-arkkitehtuurit perustuvat konvoluutioihin, ne hyödyntävät rinnakkaistamista ja lisäävät huomattavasti aikakompleksisuutta muihin malleihin verrattuna.

**Tulos**

Onko GCN:n käyttämisessä konseptuaalisia etuja verrattuna monimutkaisempiin arkkitehtuureihin, kuten huomio?

**Esimerkki 5.816**

FBFans-tietokanta sisältää tietoja ydinvoiman vastaisista kiinalaisista Facebook-faniryhmistä syyskuusta 2013 elokuuhun 2014, mukaan lukien viestit sekä niiden kirjoittajan ja tykkääjän tunnukset.

**Tulos**

Mikä aihe sisältyy Kiinan Facebook-tietoihin?

**Esimerkki 5.817**

Tällainen virhe vaikuttaa "trolli"- ja "engage"-luokkien tarkkuuteen ja palautukseen. Ratkaisu tähän ongelmaan voi olla pidempien keskusteluosuuksien sisällyttäminen.

**Tulos**

Mitä mahdollisia ratkaisuja ehdotetaan?

**Esimerkki 5.818**

Otamme tiedot WMT'14:n englannin ja ranskan (En-Fr) sekä englannin ja saksan (En-De) tietokokonaisuuksista.

**Tulos**

Käytetäänkö näitä tekniikoita monikielisten mallien kouluttamiseen ja millä kielillä?

**Esimerkki 5.819**

Tähän työhön on useita mahdollisia laajennuksia. Esimerkiksi kaikkien puhelimelle osoitettujen kehysten käyttäminen sen sijaan, että käytettäisiin vain keskimmäistä kehystä.

**Tulos**

Ehdottavatko he lisäyksiä, joilla voitaisiin parantaa yleistämistä näkymättömiin puhujiin?

**Esimerkki 5.820**

Arvioimme ehdotettuja siirto-oppimistekniikoita kahdessa WMT 2019 -uutiskäännöstehtävän ei-englanninkielisessä kieliparissa: ja saksan$\rightarrow $ tšekin$\rightarrow.

**Tulos**

Suoritetaanko kokeita muilla kielipareilla, miten ehdotettu menetelmä suoriutuu muihin malleihin verrattuna?

**Esimerkki 5.821**

Puhujan kanssa suljetussa tilassa kustakin puhujasta erotettiin kaksi jaksoa kehitys- ja testisarjoiksi. Tämän jälkeen kehitys- ja testijoukkojen kokonaiskoot ovat 1585 IPU:ta, jotka kestävät 2 tuntia 23 minuuttia, ja 1841 IPU:ta, jotka kestävät 2 tuntia ja 48 minuuttia. ASR-malli koulutetaan lopuilla tiedoilla. Koska mallin kouluttaminen olisi vaikeaa, jos kaikki KM:n tai UT:n puhujan tiedot poistettaisiin, kokeita, joissa käytettäisiin niiden puhujia, ei suoritettu.

**Tulos**

Mitä eroa on kaiutin auki ja kaiutin kiinni -asetuksilla?

**Esimerkki 5.822**

Lauseiden esittämisen jälkeen käytämme joitakin luokittelumalleja syötettyjen lauseiden luokitteluun. Nämä mallit kuvataan yksityiskohtaisesti kohdassa SECREF13. Ensimmäinen malli on BIBREF11:ssä ehdotettu TextCNN (kuva FIGREF2). Se sisältää vain CNN-lohkoja, joita seuraa joitakin tiheitä kerroksia. Useiden CNN-lohkojen, joissa on eri kokoiset ytimet, ulostulot on yhdistetty toisiinsa. Toinen malli on VDCNN (kuva KUVA 5), joka on saanut inspiraationsa BIBREF12:ssa tehdystä tutkimuksesta. TextCNN-mallin tavoin se sisältää useita CNN-lohkoja. Kolmas malli on yksinkertainen kaksisuuntainen LSTM-malli (kuva FIGREF15). Se sisältää useita LSTM- kaksisuuntaisia lohkoja pinottuna toisiinsa.Neljäs malli on LSTMCNN (kuva FIGREF24). Ennen CNN-lohkojen läpikäyntiä sanojen upotussarjat muunnetaan LSTM- kaksisuuntaisella lohkolla. Viimeinen malli on järjestelmä nimeltä SARNN (kuva FIGREF25). Se lisää huomiolohkon LTSM-lohkojen väliin. Tässä järjestelmässä käytetään pinoamismenetelmää. Tässä menetelmässä kunkin mallin tuloksena ei ole ainoastaan luokan tunnus vaan myös kunkin luokan todennäköisyys kolmen luokan joukossa. Tästä todennäköisyydestä tulee ensemble-mallin ominaisuus. Tässä pinoava ensemble-malli on yksinkertainen täysyhteysmalli, jossa syötteenä on kaikki todennäköisyydet, jotka tulostuvat alamalleista. Tuloksena on kunkin luokan todennäköisyys.

**Tulos**

Mitä luokittelijaa he käyttävät?

**Esimerkki 5.823**

ISIS:n englanninkielinen verkkolehti nimeltä Dabiq ilmestyi ensimmäisen kerran pimeässä verkossa heinäkuussa 2014 ja jatkoi julkaisemista 15 numeron ajan. Tätä julkaisua seurasi Rumiyah, joka tuotti 13 englanninkielistä numeroa syyskuuhun 2017 asti. Sekä Dabiqia että Rumiyahia tarkasteltaessa monissa lehtien numeroissa on erityisesti naisia käsitteleviä artikkeleita, joiden otsikkoon on yleensä sisällytetty "sisarillemme".

**Tulos**

Raportoidaanko tulokset vain englanninkielisistä tiedoista?

**Esimerkki 5.824**

Tämän artikkelin koeasetukset ja Mboshi-korpuksen arviointiprotokolla (Boundary F-pisteet käyttäen ZRC-puheviitettä) ovat samat kuin BIBREF8:ssa. Taulukossa esitetään tulokset kaksikielisen UWS:n ja monikielisen vivutuksen osalta. Ensin mainitussa tapauksessa saavutamme parhaan tuloksemme käyttämällä yhdenmukaistettuna tietona ranskaa, joka on tämän tietokokonaisuuden alkuperäinen yhdenmukaistettu kieli. Ranskaa lähellä olevat kielet (espanja ja portugali) sijoittuivat paremmiksi, kun taas huonoin tulos oli saksan kieli. Myös englanti suoriutuu kokeissamme erityisen hyvin. Uskomme tämän johtuvan tuloksena olevan tekstin tilastollisista ominaisuuksista. Taulukossa havaitaan, että englannin kieli sisältää kaikista kielistä pienimmän sanaston. Koska koulutamme järjestelmiämme hyvin niukkaresurssisissa ympäristöissä, sanastoon liittyvät piirteet voivat vaikuttaa suuresti järjestelmän kykyyn mallintaa kieltä ja näin ollen myös tuotettujen linjausten lopulliseen laatuun. Jopa runsaasti resursseja sisältävissä ympäristöissä on jo todettu, että jotkin kielet ovat vaikeampia mallintaa kuin toiset BIBREF9.

**Tulos**

Arvioidaanko mallia suhteessa johonkin perustasoon?

**Esimerkki 5.825**

Ensimmäinen on turkkilainen uutisverkkokorpus, joka sisältää 423 miljoonaa sanaa ja 491 miljoonaa merkkiä, eli BOUN Web Corpus BIBREF9 , BIBREF10 . Toinen koostuu 21 miljoonasta turkkilaisesta twiitistä, joissa on 241 miljoonaa sanaa ja 293 miljoonaa merkkiä, ja siihen on yhdistetty 1 miljoona twiittiä TS TweetS:stä, jonka on laatinut Sezer-2013, ja 20 miljoonaa turkkilaista twiittiä, jonka ovat laatineet Bolat ja Amasyalı.

**Tulos**

Mitä tietoja käytettiin sanojen upotusten muodostamiseen?

**Esimerkki 5.826**

Arviointiin käytettiin pientä mutta monipuolista aineistoa, joka koostui 10 englanninkielisestä YouTube-videosta uutiskontekstissa. Valitut videot käsittelevät eri aiheita, kuten teknologiaa, ihmisoikeuksia, terrorismia ja politiikkaa, ja niiden pituus vaihtelee 2-10 minuutin välillä.

**Tulos**

Millaisia Youtube-videoiden transkriptioita he käyttivät?

**Esimerkki 5.827**

Suunnittelemme sanatason tarkentavan dekooderin, koska tämä prosessi on samankaltainen kuin yhteistyöelimen esiharjoitteluprosessissa oleva cloze-tehtävä, joten käyttämällä kontekstuaalisen kielimallin kykyä dekooderi voi tuottaa sujuvampia ja luonnollisempia sekvenssejä.

**Tulos**

Miksi sanojen peittäminen dekooderissa on hyödyllistä?

**Esimerkki 5.828**

Esimerkiksi WSC-kokoelman skeema 23 voidaan kääntää muotoon "Tytöt kiusasivat poikia, joten [rankaisimme/pelastimme] heitä" ja "Pojat kiusasivat tyttöjä, joten [rankaisimme/pelastimme] heitä", jolloin vältetään ennakko-oletukset siitä, kiusaavatko tytöt todennäköisemmin poikia vai päinvastoin.

**Tulos**

Mitä tietoja he tarkastelevat?

**Esimerkki 5.829**

Minimimuutos muutti lauseen merkitystä huomattavasti, mutta silti muunnoksen upotus on hyvin lähellä alkuperäistä lausetta (keskimääräinen samankaltaisuus 0,930).

**Tulos**

Analysoidaanko niissä sitä, miten muutokset muuttivat alkuperäisiä lausekokonaisuuksia?

**Esimerkki 5.830**

Ehdotamme edellä määriteltyä tehtävää varten kuullun ymmärtämisen mallia, Attention-based Multi-hop Recurrent Neural Network (AMRNN) -kehystä, ja osoitamme, että tämä malli pystyy suoriutumaan tehtävästä kohtuullisen hyvin. Ehdotetussa lähestymistavassa tarinoiden ääni transkriboidaan ensin tekstiksi ASR:n avulla, ja ehdotettu malli kehitetään käsittelemään transkriptioita oikean vastauksen valitsemiseksi kysymyksen antamista neljästä vaihtoehdosta.

**Tulos**

Millaista lähestymistapaa tässä työssä ehdotetaan uutta tehtävää varten?

**Esimerkki 5.831**

Sovitamme siis valmiin palkitsemisoppimisalgoritmin BIBREF7 valvottuun ympäristöön automaattista tietojenkäsittelyä varten.

**Tulos**

Mikä RL:n valmis palkkio-oppimisalgoritmi soveltuu tietojenkäsittelyn ja mallin harjoittelun yhteiseen oppimiseen?

**Esimerkki 5.832**

CFQ sisältää 239 357 englanninkielistä kysymys-vastausparia, joihin voidaan vastata julkisten Freebase-tietojen avulla.

**Tulos**

Kuinka suuri on uusi kysymysten vastaustietokanta?

**Esimerkki 5.833**

Arviointitulokset ovat varsin myönteisiä molempien kohteiden osalta ja erityisen hyviä kohde-1:n osalta, kun otetaan huomioon, että ne ovat ensimmäisiä kokeita aineistolla.

**Tulos**

Mikä SVM-menetelmä tuotti parhaan tuloksen?

**Esimerkki 5.834**

Tarkastelemme jatkoanalyysia varten viittä tärkeintä profiilin ominaisuutta. Nämä attribuutit ovat käyttäjänimi, näyttönimi, profiilikuva, sijainti ja kuvaus.

**Tulos**

Mitä profiilin metatietoja käytetään tässä analyysissä?

**Esimerkki 5.835**

Laskimme bag-of-words-pohjaisia vertailuarvoja seuraavilla menetelmillä:Luokittelu TF-IDF + Lineaarinen SVM (TF-IDF + SVM)Luokittelu Depeche++ Emotion-leksikoneilla BIBREF12 + Lineaarinen SVM (Depeche + SVM)Luokittelu NRC Emotion-leksikoneilla BIBREF13, BIBREF14 + Lineaarinen SVM (NRC + SVM)TF-IDF- ja NRC-Emotion-leksikoneiden yhdistelmä (TF-NRC + SVM)Vertailuarvot ::: Doc2Vec + SVMKäytimme myös yksinkertaisia luokittelumalleja, joissa oli opittuja upotuksia. Koulutimme Doc2Vec-mallin BIBREF15 käyttäen tietokokonaisuutta ja käytimme upotettuja asiakirjavektoreita lineaarisen SVM-luokittimen ominaisuuksina.Vertailukohteet ::: Tässä vertailuanalyysissä tarkasteltiin hierarkkista RNN:ää BIBREF16:n mukaisesti. Käytimme kahta BiLSTM:ää BIBREF17, joissa kummassakin on 256 yksikköä lauseiden ja asiakirjojen mallintamiseen. Lauseen merkkejä käsiteltiin muista lauseen merkkeistä riippumatta. Jokaisen token-tason BiLSTM:n suunnan viimeiset tuotokset yhdistettiin ja syötettiin lause-tason BiLSTM:ään syötteinä.BiLSTM:n tuotokset yhdistettiin kahteen tiheään kerrokseen, joissa oli 256 ReLU-yksikköä, ja Softmax-kerrokseen. Merkit alustettiin julkisesti saatavilla olevilla upotuksilla, jotka oli koulutettu GloVe BIBREF18 -ohjelmalla. Lauseiden rajat saatiin SpaCy:ltä. Koulutuksen aikana tiheisiin piilokerroksiin sovellettiin pudotusta: Bi-directional RNN and Self-Attention (BiRNN + Self-Attention)Yksi haaste RNN-pohjaisissa ratkaisuissa tekstiluokittelussa on löytää paras tapa yhdistää sanatason representaatiot korkeamman tason representaatioihin.Self-attention BIBREF19, BIBREF20, BIBREF21 on sovitettu tekstiluokitteluun, mikä on parantanut tulkittavuutta ja suorituskykyä. Käytimme BIBREF20:tä tämän vertailuanalyysin pohjana. vertailuanalyysissä käytettiin kerroksellista kaksisuuntaista RNN:ää (60 yksikköä), jossa oli GRU-soluja ja tiheä kerros. Molemmat itsetarkkailukerrokset olivat kooltaan 60 yksikköä, ja kustannusfunktiona käytettiin risti-entropiaa.Huomaa, että olemme jättäneet pois ortogonaalisen regularisaattoritermin, koska tämä tietokokonaisuus on suhteellisen pieni verrattuna perinteisiin tietokokonaisuuksiin, joita käytetään tällaisen mallin kouluttamiseen. Emme havainneet merkittävää suorituskyvyn paranemista, kun käytimme regularisaattoritermiä kokeissamme: ELMo embedding and Bi-directional RNN (ELMo + BiRNN)Deep Contextualized Word Representations (ELMo) BIBREF22 on viime aikoina osoittanut menestystä useissa NLP-tehtävissä. Kielen mallin valvomaton luonne antaa sille mahdollisuuden hyödyntää suurta määrää saatavilla olevaa merkitsemätöntä dataa oppiakseen parempia representaatioita sanoista.Käytimme tässä vertailuarvossa Tensorhubissa saatavilla olevaa esivalmennettua ELMo-mallia (v2). Syötimme ELMon sanojen upotukset syötteenä yksikerroksiseen kaksisuuntaiseen RNN-malliin (16 yksikköä), jossa on GRU-soluja (dropoutilla) ja tiheä kerros. Kustannusfunktiona käytettiin risti-entropiaa: Hienosäädetty BERTBidirectional Encoder Representations from Transformers (BERT) BIBREF11 on saavuttanut huipputuloksia useissa NLP-tehtävissä, kuten lauseiden luokittelussa.Käytimme alkuperäisessä työssä esitettyä hienosäätömenettelyä sovittaaksemme esivalmennetun koteloimattoman BERT$\_\textrm {{\scriptsize LARGE}}$:n moniluokkaiseen tekstien luokittelutehtävään. Tällä tekniikalla saavutettiin vertailuarvoista paras tulos, ja keskimääräinen mikro-F1-pistemäärä oli 60,4 %.

**Tulos**

Mitkä ovat perustason vertailuarvot?

**Esimerkki 5.836**

Tämä versio ylittää huomattavasti aikaisemman tekniikan tason, mutta siinä on silti vähemmän parametreja kuin aikaisemmissa töissä.

**Tulos**

Vertailevatko ne tuloksia uusimpiin kielimalleihin?

**Esimerkki 5.837**

Oikeudellisesti järkevien vastausten löytämiseksi palkkaamme seitsemän juridisesti koulutettua asiantuntijaa laatimaan vastauksia Turkerin kysymyksiin.

**Tulos**

Keitä asiantuntijoita käytettiin kommentoinnissa?

**Esimerkki 5.838**

Osoitamme, että twiittitekstin ulkopuolisen tiedon (URL-osoitteiden aloitussivuilta) ja käyttäjäominaisuuksien käyttö voi parantaa suorituskykyä merkittävästi.

**Tulos**

Mitä ulkoisia tietolähteitä käytetään?

**Esimerkki 5.839**

Testaamme mgru:ta kahdella tunnetulla tietokannalla, Penn Treebankilla ja Text8:lla.

**Tulos**

Millä tietokokonaisuudella he kouluttavat mallinsa?

**Esimerkki 5.840**

Tämän jälkeen koulutamme kappaleiden vektorimallin käyttämällä Document to Vector (Doc2Vec) -ohjelmistoa BIBREF7 koko esikäsitellyn tekstin (13 miljoonaa) joukolla, vaikka myös pienemmillä joukoilla (1 miljoona) harjoittelu tuottaa hyviä tuloksia.

**Tulos**

Mitä tekstin upottamismenetelmiä käytetään?

**Esimerkki 5.841**

Havaitsimme prosessin aikana joitakin rajoituksia, joita kuvaamme tässä luvussa.Kun päätimme julkaisijan puolueellisuudesta, niiden henkilöiden määrä, joista laskimme pistemäärän, oli pieni. Esimerkiksi de Stentor tavoittaa arviolta 275 000 lukijaa päivittäin virallisella verkkosivustollaan. Yleisöön nojautumisen päättäminen 55 otoksen perusteella altistui otantavirheille. Sitä paitsi pisteet eroavat kustantajien välillä hyvin vähän. Yhdenkään kustantajan absoluuttinen pistemäärä ei ollut suurempi kuin 1, mikä tarkoittaa, että puolueellisin kustantaja oli vain vähän puolueellinen. Sen päättäminen, mitä kustantajia pidämme puolueellisina ja mitä ei, ei siis ole kovin luotettavaa. artikkelitason annotointitehtävä ei ollut yhtä tarkkaan määritelty kuin joukkoistamisalustalla. Sisällytimme kysymykset osaksi olemassa olevaa kyselyä emmekä halunneet aiheuttaa annotoijille suurta taakkaa. Siksi emme antaneet pitkää kuvailevaa tekstiä, jossa selitettiin, miten henkilön tulisi annotoida artikkeli. Näin ollen meillä on annotoijien puolueellisuuden riski. Tämä on yksi syy heikkoon arvioijien väliseen yhteisymmärrykseen.

**Tulos**

Mitä rajoituksia mainitaan?

**Esimerkki 5.842**

Teemme kokeita useilla kielipareilla, kuten saksan ja englannin, ranskan ja englannin sekä japanin ja englannin välillä.

**Tulos**

Mitä kieliä käytetään monikielisessä kuvatekstimallissa?

**Esimerkki 5.843**

Sen sijaan, että lukisimme tokenisoidun syöttötekstin, mallimme lukee raakoja utf-8-tarkkoja tavuja. ASCII-alueen englanninkieliselle tekstille tämä vastaa merkkien käsittelyä yksittäisinä merkkeinä. Muut kuin ASCII-merkit (esim. aksenttimerkit tai muut kuin latinankieliset kirjoitusmerkit) ovat tyypillisesti kaksi tai kolme utf-8-tavua. Käytämme tavallista "muuntajan dekooderia" (pino muuntajakerroksia, joissa on kausaalinen huomiomaski) käsittelemään sekvenssiä $x\_{0:i-1}$ ja ennustamaan seuraavan tavun $x\_i$. Mallin ennuste on arvio todennäköisyysjakaumasta kaikkien mahdollisten 256 tavun arvojen yli. Syötteenä olevan tavun upotusmatriisimme ulottuvuus on 256.

**Tulos**

Kuinka monta merkkiä hyväksytään kielimallin syötteeksi?

**Esimerkki 5.844**

Työntekijöitä pyydetään arvioimaan, voidaanko lausuma johtaa artikkeleista kolmella tasolla, kun heille annetaan lausuma ja artikkelit: Totta, Todennäköistä (eli vastattavissa) tai Epävarmaa (eli vastaamattomia). Jos työntekijä valitsee vaihtoehdon Epävarma, pyydämme työntekijöitä kertomaan kahdesta vaihtoehdosta ("Ei mainita artikkelissa" tai "Muu"), miksi he ovat epävarmoja. Jos työntekijä valitsee arviointitehtävässä vaihtoehdon Totta tai Todennäköistä, kysymme ensin, mitkä annettujen artikkeleiden lauseet ovat perusteluja tietylle väitteelle, samoin kuin HotpotQA BIBREF2:ssa. "Yhteenveto"-tekstilaatikot (eli NLD:t) alustetaan sitten näillä valituilla lauseilla.

**Tulos**

Miten tietokokonaisuutta kommentoitiin?

**Esimerkki 5.845**

Kussakin kokeessa haettujen asiakirjojen tarkkuuden arvioimiseksi käytimme TREC\_Eval-työkalua [3]. TREC\_Eval on vakioväline IR-tehtävien arviointiin, ja sen nimi on lyhenne sanoista Text REtrieval Conference (TREC) Evaluation tool. TREC\_Evalin raportoima keskimääräinen tarkkuus (Mean Average Precision, MAP) oli 27,99 prosenttia ilman kyselyn laajennusta ja 37,10 prosenttia kyselyn laajennuksen kanssa, mikä osoittaa yli 9 prosentin parannusta.

**Tulos**

Mitä arviointimittaria on mitattu?

**Esimerkki 5.846**

Keskitymme tässä Europarl-alueeseen, josta meillä on runsaasti tietoja useilla kielillä, ja käytämme alan sisäisenä harjoitusaineistona Europarl-korpusta BIBREF5 kahta käännössuuntaa varten: Englannin INLINEFORM0 saksa ja englannin INLINEFORM1 ranska.

**Tulos**

mitä kielipareja tutkitaan?

**Esimerkki 5.847**

Arviointimittareina käytimme lebret2016neural-ohjelman mukaisesti BLEU-4-, NIST-4- ja ROUGE-4-mittareita.

**Tulos**

Mitä mittareita käytetään arvioinnissa?

**Esimerkki 5.848**

Autokieli on abstraktia kieltä, joka liittyy auton fyysisiin ominaisuuksiin. Tässä tapauksessa fyysiset ominaisuudet, joihin termi "nopea" viittaa, voivat olla hevosvoimat, tai se voi olla auton muototekijä (miten auto näyttää). Emme kuitenkaan tiedä tarkalleen, mihin ominaisuuksiin termi "nopea" viittaa. Koulutamme sarjan luokittelijoita luokitellaksemme autopuheita. Koulutamme kolme luokittelijaa SECREF8-jaksossa laatimillamme arvosteluvektoreilla. Käyttämämme luokittelijat ovat K Nearest Neighbors (KNN), Random Forest (RF), Support Vector Machine (SVM) ja Multi-layer Perceptron (MLP) BIBREF13.

**Tulos**

Onko autokieli kokoelma abstrakteja piirteitä, joiden perusteella luokittelija koulutetaan myöhemmin?

**Esimerkki 5.849**

Tiedonkeruuprosessi, jossa käytettiin jengiläisten käyttämiä sijaintineutraaleja avainsanoja ja laajennettua hakua heidän uudelleentwiittaus-, ystävä- ja seuraajaverkostoissaan, johti 400 aidon jengiläisen profiilin tunnistamiseen Twitterissä. Tutkimuksessamme havaittiin, että twiittien ja profiilikuvausten teksti, emojien käyttö, profiilikuvat ja musiikkiharrastukset, jotka ilmenevät linkkeinä YouTube-musiikkivideoihin, voivat auttaa luokittelijaa erottamaan toisistaan jengin jäsenten ja muiden kuin jengin jäsenten profiilit.

**Tulos**

Miten jengiin kuulumista koskeva perustotuus määritetään tässä aineistossa?

**Esimerkki 5.850**

Tässä työssä analysoimme tietyn klassisen musiikin radiokanavaan, BBC Radio 3:een, liittyviä twiittejä ja pyrimme havaitsemaan kahdenlaisia musiikillisia nimettyjä entiteettejä, Contributor ja Musical Work.

**Tulos**

Millä kielellä Twitterin sisältö on?

**Esimerkki 5.851**

Tässä käytetään pientä osaa englannin ja saksan välisestä suuresta rinnakkaisesta korpuksesta simulaationa skenaariota varten, jossa meillä ei ole paljon rinnakkaista tietoa: Englanninkielisten tekstien kääntäminen saksaksi.

**Tulos**

Millä kielillä he testaavat aliresursoitua skenaariota varten?

**Esimerkki 5.852**

Kuvassa KUVIO15 esitetään saksan ja japanin testisarjan tarkkuus ilman resursseja suoritettujen vaiheiden määrän funktiona, kun käytössä on vastakohtainen koulutus ja kun sitä ei ole suoritettu. Kuvaajasta käy ilmi, että testitarkkuuden vaihtelu vähenee vastakohtaisen harjoittelun myötä, mikä viittaa siihen, että kieltenvälinen suorituskyky on johdonmukaisempi, kun vastakohtaista harjoittelua käytetään.

**Tulos**

Osoittaako mikään arvioinneista, että vastakkaisoppiminen parantaa suorituskykyä vähintään kahdessa eri kieliperheessä?

**Esimerkki 5.853**

Tässä artikkelissa esitellään uusi toimintamalli, jonka avulla voidaan tuottaa useita toimintoja vuoroa kohti (ns. multi-toiminta), luoda sarja tupleja ja laajentaa agenttien ilmaisuvoimaa. Kukin tuple määritellään muotoon $(\textit {continue}, \textit {act}, \textit {slots})$, jossa continue ilmaisee, jatketaanko vai lopetetaanko uusien tekojen tuottaminen, act on tekotyyppi (esim. inform tai request) ja slots on joukko nykyiseen tekotyyppiin liittyviä slotteja (nimiä). Vastaavasti ehdotetaan uudenlaista dekooderia (kuva KUVA 5) tällaisten sekvenssien tuottamiseksi. Kukin tupletti tuotetaan solulla nimeltä gated Continue Act Slots (gCAS, kuten kuvassa FIGREF7), joka koostuu kolmesta peräkkäin kytketystä gated-yksiköstä, jotka käsittelevät tupletin kolmea komponenttia. Tämä dekooderi voi tuottaa moniaktit kaksinkertaisella rekursiivisella tavalla BIBREF18.

**Tulos**

Mikä on erityistä gCAS-solulle?

**Esimerkki 5.854**

Vertaamme malliamme kahteen perusjoukkoon: MemN2N BIBREF12 on päästä päähän koulutettava versio muistiverkoista BIBREF9 . Attentive and Impatient Readers BIBREF6 käyttää kaksisuuntaisia LSTM-verkkoja kysymyksen ja todisteiden koodaamiseen ja tekee luokittelun laajassa sanastossa näiden kahden koodauksen perusteella.

**Tulos**

Mitkä ovat perustasot?

**Esimerkki 5.855**

Ensimmäisen kerroksen poistaminen käytöstä RTE-tehtävässä parantaa suorituskykyä merkittävästi, sillä absoluuttinen suorituskyky paranee 3,2 prosenttia. Tämän toimenpiteen vaikutukset vaihtelevat kuitenkin tehtävien välillä, ja QNLI- ja MNLI-tehtävissä se aiheuttaa jopa -0,2 prosentin suorituskyvyn laskun.

**Tulos**

Kuinka paljon suorituskyky paranee, jos huomio poistetaan käytöstä tietyissä päissä?

**Esimerkki 5.856**

Arvioimme ehdotetun kielimallin merkkitason muunnelmaa Penn Treebank (PTB) - ja Text8-tietokantojen esikäsitellyn version avulla. Valvomattomassa konstituution jäsentämistehtävässä verrataan mallin päättelemää puurakennetta ihmisasiantuntijoiden kommentoimiin rakenteisiin. Koe suoritetaan WSJ10-tietokannalla.

**Tulos**

Millä tietokokonaisuudella he tekevät kokeita?

**Esimerkki 5.857**

Tämä tietokokonaisuus kerättiin Mechanical Turkin kautta käyttäen 100 simuloitua ympäristöä ja vastaavaa topologista karttaa, ja tietojemme mukaan se on ensimmäinen laatuaan käyttäytymisnavigoinnin osalta. Loimme uuden tietokokonaisuuden navigointiohjeiden seuraamista koskevaa ongelmaa varten BIBREFin5 käyttäytymiseen perustuvan navigointikehyksen mukaisesti. Tämä tietokokonaisuus luotiin käyttämällä Amazon Mechanical Turk -palvelua ja 100 karttaa simuloiduista sisäympäristöistä, joissa kussakin oli 6-65 huonetta. Tietojemme mukaan tämä on ensimmäinen vertailukohde käännösmallien vertailemiseksi käyttäytymiseen perustuvan robottinavigoinnin yhteydessä.Kuten taulukosta TABREF16 käy ilmi, tietokokonaisuus koostuu 8066 parista vapaamuotoisia luonnollisen kielen ohjeita ja navigointisuunnitelmia harjoittelua varten. Tämä harjoitusaineisto kerättiin 88 ainutlaatuisesta simuloidusta ympäristöstä, joissa on yhteensä 6064 erilaista navigointisuunnitelmaa (2002 suunnitelmassa on kussakin kaksi erilaista navigointiohjetta; lopuissa on yksi). Tietokokonaisuus sisältää kaksi testijoukkovaihtoehtoa: Vaikka tietokokonaisuus kerättiin simuloitujen ympäristöjen avulla, navigointiohjeille ei asetettu mitään rakennetta joukkoistamisen aikana. Näin ollen monet aineistossamme olevat ohjeet ovat moniselitteisiä. Lisäksi ohjeissa olevien käyttäytymismallien järjestys ei ole aina sama. Esimerkiksi eräs henkilö sanoi "käänny oikealle ja etene" kuvaamaan osaa reitistä, kun taas toinen henkilö sanoi samankaltaisessa tilanteessa "mene suoraan oikealle kääntymisen jälkeen". Koska aineistomme luonnollisen kielen kuvauksissa on suurta vaihtelua, ohjeiden purkaminen käyttäytymiseksi ei ole aivan yksinkertaista. Katso lisäaineiston liite A, jossa on lisätietoja tiedonkeruupyrkimyksistämme.

**Tulos**

Miten navigointiohjeet kerättiin?

**Esimerkki 5.858**

Vain topologiset upotukset: ( INLINEFORM1 ) Topologian ja tekstin yhteinen upotus: Naivi yhdistelmä, TADW BIBREF5 , CENE BIBREF6 , CANE BIBREF9 , WANE BIBREF10 , DMTE BIBREF34 .

**Tulos**

Mihin muihin sulautumiin niitä verrataan?

**Esimerkki 5.859**

Muiden alojen tutkijat ovat mukauttaneet näitä NLP-yhteisön tehokkaita ideoita muihin kuin sanojen esittämiseen liittyviin tehtäviin, kuten relaatiokokonaisuuksiin BIBREF1 , BIBREF2 , yleisiin tekstipohjaisiin attribuutteihin BIBREF3 , kuvien kuvailevaan tekstiin BIBREF4 , verkkojen graafirakenteiden solmuihin BIBREF5 ja kyselyihin BIBREF6 , muutamia mainitakseni.

**Tulos**

Millaisilla aloilla on näin laajoja sanastoja?

**Esimerkki 5.860**

Tutkimuksessamme katsomme, että twiitistä tuli virus, jos sitä uudelleentwiitattiin yli 1000 kertaa.

**Tulos**

Mikä on heidän määritelmänsä siitä, että twiitit leviävät?

**Tulos**

Mikä on kynnys määrittää, että twiitti on muuttunut viraaliksi?

**Esimerkki 5.861**

Vertailimme useita tiivistämismenetelmiä, jotka voidaan jakaa kolmeen ryhmään: valvomattomat, ei-neuraaliset valvotut ja neuraaliset valvotut menetelmät. Valvomattomien menetelmien osalta testasimme: SumBasic, joka käyttää sanojen frekvenssiä lauseiden sijoittamiseen ja valitsee parhaat lauseet yhteenvedoksi BIBREF13 , BIBREF14 . Lsa, joka käyttää latenttia semanttista analyysia (LSA) asiakirjan termi-lause-matriisin purkamiseen ja poimii lauseet tuloksen perusteella. Kokeilimme näitä kahta lähestymistapaa, joita ehdotettiin BIBREF15:ssä ja BIBREF16:ssä. LexRank, jossa asiakirjasta muodostetaan graafikuvaus, jossa solmut ovat lauseita ja särmät kahden lauseen välistä samankaltaisuutta, ja jossa suoritetaan PageRank-algoritmi kyseiselle graafille ja poimitaan lauseet tuloksena saatujen PageRank-arvojen perusteella BIBREF17 . TextRank, joka on hyvin samankaltainen kuin LexRank, mutta laskee lauseiden samankaltaisuuden yhteisten merkkien määrän perusteella BIBREF19 . Bayes, joka esittää jokaisen lauseen ominaisuusvektorina ja käyttää naiivia Bayes-menetelmää niiden luokitteluun BIBREF5 . Alkuperäisessä artikkelissa TF-IDF-pisteytys lasketaan monisanaisille tunnisteille, jotka tunnistetaan automaattisesti keskinäisen informaation avulla. Me emme tehneet tätä tunnistusta, joten TF-IDF-laskentamme perustuu sanamerkkeihin. Hmm, jossa käytetään piilotettua Markov-mallia, jonka tilat vastaavat sitä, pitäisikö lauseesta poimia BIBREF20 . Alkuperäisessä teoksessa käytetään QR-dekompositiota lauseiden valintaan, mutta meidän toteutuksessamme ei käytetä sitä. Me yksinkertaisesti järjestimme lauseet niiden pistemäärien mukaan ja valitsimme kolme parasta yhteenvedoksi. MaxEnt, joka edustaa kutakin lausetta ominaisuusvektorina ja käyttää maksimi-entropiamallia laskeakseen todennäköisyyden sille, että lause tulisi poimia BIBREF21 . Alkuperäisessä lähestymistavassa merkinnöille asetetaan ennakkojakauma, mutta me asetamme sen sijaan ennakkojakauman painoille. Toteutuksemme on silti alkuperäisen kanssa yhdenmukainen, koska käytimme bias-ominaisuutta, jonka pitäisi pystyä oppimaan etuliitejakauma. Neuraalisen valvotun menetelmän osalta arvioimme NeuralSum BIBREF11 -menetelmää käyttäen tekijöiden alkuperäistä toteutusta. Muutimme heidän toteutustaan hieman, jotta mallia voidaan arvioida ROUGE:n avulla. Huomaa, että kaikki menetelmät ovat ekstraktiivisia. Toteutuskoodimme kaikille edellä mainituille menetelmille on saatavilla verkossa. Käytimme perustasona Lead-N:ää, joka valitsee yhteenvedoksi INLINEFORM0:n johtavat lauseet. Kaikissa menetelmissä poimimme yhteenvedoksi 3 lausetta, koska se on tutkivan analyysin perusteella löytämiemme kultaisten yhteenvetojen lauseiden lukumäärän mediaani.

**Tulos**

Mitä lähestymistapoja he käyttivät?

**Esimerkki 5.862**

Jokainen twiitti on merkitty merkinnällä, jossa ei ole todisteita masennuksesta (esim. "Kansalaiset pelkäävät taloudellista lamaa") tai todisteita masennuksesta (esim. "masentunut pettymyksestä"). Jos twiitti on merkitty todisteeksi masennuksesta, se merkitään lisäksi yhdellä tai useammalla masennusoireella, esimerkiksi masentuneella mielialalla (esim. "masentunut olo"), unihäiriöillä (esim. "taas yksi levoton yö") tai väsymyksellä tai energian menetyksellä (esim. "väsymys on sietämätöntä") BIBREF10 .

**Tulos**

Miten tietokokonaisuus on kommentoitu?

**Esimerkki 5.863**

Käytämme DSTC2 BIBREF20 -tietokokonaisuutta ja Maluuba BIBREF21 -tietokokonaisuutta ehdotetun mallin arvioimiseksi.

**Tulos**

Mitä kahta vertailutietoaineistoa käytetään?

**Esimerkki 5.864**

Tässä tutkimuksessa käsitellään lyhyesti laadunvarmistusjärjestelmän vaiheita ja verrataan seitsemän koneoppimiseen perustuvan luokittelijan (Multi-Layer Perceptron (MLP), Naive Bayes Classifier (NBC), Support Vector Machine (SVM), Gradient Boosting Classifier (GBC), Stochastic Gradient Descent (SGD), K Nearest Neighbour (K-NN) ja Random Forest (RF)) suorituskykyä bengalinkielisten kysymysten luokittelussa luokkiin odotettujen vastausten perusteella.

**Tulos**

mitä ml-pohjaisia lähestymistapoja verrattiin?

**Esimerkki 5.865**

Yksittäisen pään poistamisesta käytöstä saatava hyöty on erilainen eri tehtävissä, ja se vaihtelee STS-B:n pienimmästä absoluuttisesta hyödystä 0,1 %:sta MRPC:n suurimpaan 1,2 %:iin (ks. kuva: disableheadsall). Lisäksi koko kerroksen, eli kaikkien 12 pään poistaminen käytöstä tietyssä kerroksessa, parantaa myös tuloksia. Kuva:disablelayers osoittaa mallin suorituskyvyn GLUE-tavoitetehtävissä, kun eri kerrokset poistetaan käytöstä.

**Tulos**

Missä tietyissä päissä tarkkaavaisuus oli kokeissa pois käytöstä?

**Esimerkki 5.866**

Muihin kohdetehtäviin verrattuna kielioppiin liittyvä CoLA-tehtävä hyötyy huomattavasti ELMon esivalmennuksesta: Paras tulos ilman kielimallin esivalmennusta on alle puolet tuloksesta, joka saavutetaan tällaisen esivalmennuksen avulla. Sitä vastoin merkityssuuntautuneessa tekstin samankaltaisuuden vertailutehtävässä STS saadaan hyviä tuloksia useilla esivalmennuksilla, mutta ELMon käytöstä ei ole merkittävää hyötyä.

**Tulos**

Suoriutuvatko jotkin esivalmennuksen tavoitteet paremmin kuin toiset lausetason ymmärtämistehtävissä?

**Esimerkki 5.867**

Yksinkertaisuuden vuoksi keskitymme analyysissämme New Yorkin Manhattanin Airbnb-ilmoituksiin ajanjaksolla 1. tammikuuta 2016-1. tammikuuta 2017. Meille toimitetut tiedot sisälsivät tietoja noin 40 000:sta Airbnb:ssä julkaistusta Manhattanin majoituspaikasta, jotka julkaistiin Airbnb:ssä tämän määritellyn ajanjakson aikana. Kunkin ilmoituksen osalta meille annettiin tiedot ilmoituksen mukavuuksista (kylpyhuoneiden lukumäärä, makuuhuoneiden lukumäärä ...), ilmoituksen postinumero, isännän kuvaus ilmoituksen sisällöstä, ilmoituksen hinta ja ilmoituksen käyttöaste.

**Tulos**

Mikä on Airbnb:n koko?

**Esimerkki 5.868**

Aiemmissa töissä ei kuitenkaan ole tutkittu loukkaavan kielenkäytön kohdetta, mikä on tärkeää monissa tilanteissa, esimerkiksi tutkittaessa vihapuhetta tietyn kohteen osalta.

**Tulos**

Mitä eroja on tämän tietokokonaisuuden ja jo olemassa olevien tietokokonaisuuksien välillä?

**Esimerkki 5.869**

Kokeelliset tulokset osoittavat, että yhteinen mallimme päihittää pelkän visuaalisen mallin kaikissa tapauksissa ja pelkän tekstin mallin Wikipediassa ja kahdessa arXivin osajoukossa.

**Tulos**

Toimivatko menetelmät, jotka toimivat parhaiten akateemisissa artikkeleissa, parhaiten myös Wikipediassa?

**Esimerkki 5.870**

Vertailun vuoksi koulutamme ja arvioimme näitä järjestelmiä VLSP-korpusten avulla. Erityisesti teemme kokeita Viet Treebankin korpuksella POS-tagaus- ja chunking-tehtäviä varten ja VLSP shared task 2016 -korpuksella NER-tehtävää varten. Kaikki nämä korpukset muunnetaan CoNLL-muotoon. POS-tagitehtävän korpus koostuu kahdesta sarakkeesta, nimittäin sanasta ja POS-tagista. Chunking-tehtävää varten korpuksessa on kolme saraketta, nimittäin sana, POS-tag ja chunk. NER-tehtävän korpus koostuu neljästä sarakkeesta. Näiden sarakkeiden järjestys on sana, POS-tunniste, klusteri ja nimetty entiteetti. NER-korpus on jaettu harjoitus- ja testausosiin, mutta POS-tagaus- ja chunking-tietoaineistoja ei ole aiemmin jaettu. Tästä syystä käytämme näistä aineistoista INLINEFORM0:aa koulutusjoukkona ja loput testausjoukkona. Koska järjestelmässämme käytetään varhaisen lopettamisen menetelmää, käytämme INLINEFORM1:tä näistä koulutusjoukon datajoukoista kehitysjoukkona, kun koulutamme NNVLP-järjestelmää. Taulukoissa TABREF24 ja TABREF25 esitetään kunkin korpuksen tilastot.

**Tulos**

Mitä tietokokonaisuuksia ne käyttävät tehtävissä?

**Esimerkki 5.871**

Tässä asiakirjassa ehdotamme ensimmäistä laajamittaista tietokokonaisuutta sosiaalisen median tietojen laadunvarmistusta varten. Taulukossa TABREF3 on esimerkki TweetQA-aineistosta. Tietokokonaisuus sisältää nyt 10 898 artikkelia, 17 794 twiittiä ja 13 757 joukkoresursoitua kysymys-vastaus -paria.

**Tulos**

Mikä on tämän tietokokonaisuuden koko?

**Esimerkki 5.872**

Kuten kohdassa SECREF2 käsiteltiin, NLP-yhteisössä on saatavilla useita ominaisuuksia, jotka mittaavat tekstin sisältämän entiteetin tärkeyttä. Toteutimme uudelleen Dunietzin ja Gillickin BIBREF11 -julkaisussa esitetyt ominaisuudet. Tämä sisältää erilaisia piirteitä, esimerkiksi sijaintipiirteitä, esiintymistiheyttä sekä entiteetin ja sen sisältämän lauseen sisäistä POS-rakennetta.

**Tulos**

Mitä ominaisuuksia käytetään kuvaamaan entiteettien tärkeyttä ja suhteellista arvovaltaa?

**Esimerkki 5.873**

Esittelimme uuden lähestymistavan, jolla muunnetaan maskuliinisesti taivutettujen ja feminiinisesti taivutettujen substantiivilauseiden välillä morfologisesti rikkaissa kielissä. Tätä varten otimme käyttöön Markovin satunnaiskentän, jossa on valinnainen neuraalinen parametrisointi, joka päättelee, millä tavalla lauseen on muututtava, jotta morfosyntaktinen yhteisymmärrys säilyy, kun tiettyjen substantiivien kieliopillista sukupuolta muutetaan.

**Tulos**

Minkä mallin avulla he vaihtavat maskuliinisesti taivutetut ja feminiinisesti taivutetut lauseet?

**Esimerkki 5.874**

Toinen suuntaus on se, että käyttämällä alle 50 prosenttia käytettävissä olevista tiedoista malli pyrkii sovittamaan tiedot liian hyvin, kuten validointitappion johdonmukainen kasvu osoittaa noin 15 epookin jälkeen (katso katkoviivat kuvan FIGREF28 oikealla puolella).

**Tulos**

Mitkä kokeelliset tulokset viittaavat siihen, että alle 50 prosentin käyttö käytettävissä olevista harjoitusesimerkeistä saattaa johtaa ylisovittamiseen?

**Esimerkki 5.875**

Tässä vaiheessa lasketaan kaikkien sanojen esiintymät harjoitusjoukossa ja lajitellaan ne esiintymien vähenemisjärjestykseen. Kuten edellä mainittiin, sanavarastoksi katsotaan INLINEFORM0 useimmin esiintyvät sanat, jotka esiintyvät vähintään INLINEFORM1 kertaa. Se on toteutettu Sparkissa suoraviivaisena map-reduce-tehtävänä.

**Tulos**

Suorittavatko ne mitään morfologista tokenisointia?

**Esimerkki 5.876**

Ensimmäisenä tietokokonaisuutena käytimme turkkilaisia elokuva-arvosteluja, jotka on kerätty suositulta verkkosivustolta. Tässä elokuvakorpuksessa on 20 244 elokuva-arvostelua, ja arvostelujen keskimääräinen sanamäärä on 39 sanaa. Jokaisella arvostelulla on tähtiarvosana, joka kertoo elokuvan tunnetilasta. Nämä polariteettipisteet ovat arvojen 0,5 ja 5 välillä 0,5:n välein. Katsomme, että arvostelu on negatiivinen, jos sen pistemäärä on 2,5 tai pienempi. Toisaalta, jos se on yhtä suuri tai suurempi kuin 4, sen oletetaan olevan positiivinen. Olemme valinneet satunnaisesti 7 020 negatiivista ja 7 020 positiivista arviota ja käsitelleet vain ne.

**Tulos**

Mitä tietoja elokuvatietokannasta annetaan?

**Esimerkki 5.877**

Malli koostuu koodaajasta (§SECREF5) ja dekooderista, jossa on huomiomekanismi (§SECREF7), jotka molemmat on toteutettu käyttämällä rekursiivisia neuroverkkoja (RNN); koodaaja muuntaa lähdesanat vektoreiden sekvenssiksi ja dekooderi tuottaa kohdekielen sanat yksi kerrallaan huomiomekanismin avulla yhtälössä DISPLAY\_FORM2 ja DISPLAY\_FORM3 esitetyn ehdollisen todennäköisyyden perusteella.

**Tulos**

Mitä malliarkkitehtuuria he käyttävät mallin rakentamiseen?

**Esimerkki 5.878**

Erityisesti NLG-tehtävässä yksittäinen mallimme päihitti kilpailevat mallit sekä Rouge-L:n että Bleu-1:n osalta.

**Tulos**

Miten tiivistelmien laatua mitataan?

**Esimerkki 5.879**

Käytämme BIBREF5:n katastrofitietoja. Se sisältää erilaisia tietokokonaisuuksia, mukaan lukien CrisiLexT6 -tietokokonaisuus, joka sisältää kuusi englanninkielisiin twiitteihin liittyvää kriisitapahtumaa vuosina 2012 ja 2013, jotka on merkitty kunkin kriisin aiheeseen liittyvyyden (on-topic ja off-topic) mukaan. Kukin kriisitapahtuman twiitti sisältää lähes 10 000 leimattua twiittiä, mutta keskitymme vain tulviin liittyviin twiitteihin, joten kokeilimme vain kahta tulvatapahtumaa eli Queenslandin tulvaa Queenslandissa, Australiassa ja Albertan tulvaa Albertassa, Kanadassa, ja leimasimme kaikki aiheeseen liittyvät twiitit aiheeseen liittyviksi ja aiheeseen liittymättömät twiitit aiheeseen liittymättömiksi, jotta ymmärrettäisiin implisiittiset luokkatunnisteet tässä tapauksessa.

**Tulos**

Mitä tietokokonaisuutta he käyttivät?

**Esimerkki 5.880**

Neuraalinen konekäännös koulutettiin Nematus-ohjelmalla. Sekä NMT- että PreMT-järjestelmässä käytettiin oletuskonfiguraatiota.

**Tulos**

Mitä NMT-arkkitehtuuria ne käyttävät?

**Esimerkki 5.881**

BioNLP 2009 Shared Task BIBREF195 perustui GENIA-korpukseen BIBREF196 , joka sisältää PubMed-artikkelien tiivistelmiä ihmisen verisoluissa esiintyvistä transkriptiotekijöistä.

**Tulos**

Mitä tietokokonaisuuksia tässä työssä käytetään?

**Esimerkki 5.882**

Huomioon perustuva käännösmalli Käytämme BIBREF6 -järjestelmää, joka on konvolutiomalli sekvenssistä sekvenssiin.

**Tulos**

Mitä käännösjärjestelmää he käyttävät kääntääkseen englanniksi?

**Esimerkki 5.883**

Tietokokonaisuuksien keräämisessä käytämme positiivisina näytteinä otsikoita, jotka on kerätty qin2018automatic, lin2019learning -ohjelmassa Tencent Newsilta, joka on yksi suosituimmista kiinalaisista uutissivustoista. Noudatamme samaa datan jakoa kuin alkuperäisessä artikkelissa. Koska osa linkeistä ei ole enää saatavilla, saamme 170 754 harjoitusnäytettä ja 4511 validointinäytettä. Negatiivisten harjoitusnäytteiden keräämistä varten valitsemme satunnaisesti luotuja otsikoita osoitingeneraattori BIBREF0-mallista, joka on koulutettu LCSTS:n tietokokonaisuudella BIBREF5, ja luomme tasapainoisen harjoituskorpuksen, joka sisältää 351 508 harjoitusnäytettä ja 9 022 validointinäytettä. Koulutetun luokittelijan arvioimiseksi muodostamme testijoukon ottamalla satunnaisesti 100 otsikkoa LCSTS-tietokannan testiosasta, ja merkinnät saadaan 11 ihmisen annotoijalta. Käytämme LCSTS BIBREF5 -tietokokonaisuutta tiivistelmämallin kouluttamiseen. Aineisto on kerätty kiinalaiselta mikroblogisivustolta Sina Weibo. Se sisältää yli 2 miljoonaa kiinankielistä lyhyttä tekstiä, joiden otsikot on antanut kunkin tekstin kirjoittaja. Tietokokonaisuus on jaettu 2 400 591 näytteeseen koulutusta varten, 10 666 näytteeseen validointia varten ja 725 näytteeseen testausta varten. Jokainen lause merkitään Jieba-ohjelmalla, ja sanaston koko on 50000.

**Tulos**

Käytettiinkö arvioinnissa toisen alan tietokokonaisuuksia?

**Esimerkki 5.884**

Lisäksi QA:lla ja QG:llä on todennäköisyyskorrelaatio, koska molemmat tehtävät liittyvät $q$:n ja $a$:n yhteiseen todennäköisyyteen. Kun kysymys-vastaus -parin $\kysymys q, a \kysymys $ annetaan, yhteistodennäköisyys $P(q, a)$ voidaan laskea kahdella vastaavalla tavalla.$$P(q, a) = P(a) P(q|a) = P(q)P(a|q)$$ (Yht. 1)Ehdollinen jakauma $P(q|a)$ vastaa täsmälleen QG-mallia, ja ehdollinen jakauma $P(a|q)$ on läheisesti yhteydessä QA-malliin. Olemassa olevat tutkimukset oppivat tyypillisesti QA-mallin ja QG-mallin erikseen minimoimalla omia häviöfunktioitaan ja jättämällä huomiotta niiden välisen todennäköisyyskorrelaation.Näiden näkökohtien perusteella esittelemme koulutuskehyksen, joka hyödyntää QA:n ja QG:n kaksinaisuutta molempien tehtävien parantamiseksi. QA:n ja QG:n kaksinaisuutta voidaan hyödyntää eri tavoin. Tässä työssä hyödynnämme QA:n ja QG:n välistä todennäköisyyskorrelaatiota regularisointiterminä vaikuttaaksemme molempien tehtävien koulutusprosessiin. Tarkemmin sanottuna kehyksemme koulutustavoitteena on oppia yhdessä $\theta \_{qa}$:lla parametrisoitu QA-malli ja $\theta \_{qg}$:lla parametrisoitu QG-malli minimoimalla niiden häviöfunktiot seuraavan rajoituksen mukaisesti.$$P\_a(a) P(q|a;\theta \_{qg}) = P\_q(q)P(a|q;\theta \_{qa})$$ (Yhtälö 3)$P\_a(a)$ ja $P\_q(q)$ ovat vastauslauseiden ja kysymyslauseiden kielimallit. Kaiken kaikkiaan kehys sisältää kolme komponenttia, nimittäin QA-mallin, QG-mallin ja regularisointitermin, joka kuvastaa QA:n ja QG:n kaksinaisuutta. QA-kohtaisen tavoitteen tavoitteena on minimoida häviöfunktio $l\_{qa}(f\_{qa}(a,q;\theta \_{qa}), label)$ , jossa $label$ on 0 tai 1, joka ilmaisee, onko $a$ oikea vastaus kohtaan $q$ vai ei. Jokaisen oikean kysymys-vastausparin osalta QG-kohtainen tavoite on minimoida seuraava häviöfunktio,$$l\_{qg}(q, a) = -log P\_{qg}(q|a;\theta \_{qg})$$ (Yht. 6)$$, jossa $a$ on oikea vastaus kysymykseen $q$ . Kolmas tavoite on regularisointitermi, joka täyttää yhtälössä 3 esitetyt todennäköisyysduraliteettirajoitukset . Tarkemmin sanottuna, kun on annettu oikea $\kulma q, a \kulma $ pari, haluamme minimoida seuraavan häviöfunktion,$$ \nonumber l\_{dual}(a,q;\theta \_{qa}, \theta \_{qg}) &= [logP\_a(a) + log P(q|a;\theta \_{qg}) \\\ & - logP\_q(q) - logP(a|q;\theta \_{qa})]^2$$ (Yht. 9)missä $P\_a(a)$ ja $P\_q(q)$ ovat marginaalijakaumia, jotka saadaan helposti kielimallin avulla.

**Tulos**

Mitä tarkoittaa "hyödyntää nimenomaisesti niiden todennäköisyyskorrelaatiota molempien mallien koulutusprosessin ohjaamiseksi"?

**Esimerkki 5.885**

Viittaamme lukijaa BIBREF10:een, jossa on lisätietoja propagandatekniikoista; alla on luettelo tekniikoista: Propagandatekniikat ::: 1. Kuormitettu kieli.Sanojen/lauseiden käyttäminen, joilla on voimakas emotionaalinen merkitys (positiivinen tai negatiivinen) yleisöön vaikuttamiseksi BIBREF11.Propagandatekniikat ::: 2. Nimittely tai leimaaminen.Propagandan kohteen leimaaminen joksikin, mitä kohdeyleisö pelkää, vihaa, pitää epätoivottavana tai muuten rakastaa tai ylistää BIBREF12.Propagandatekniikat ::: 3. Toistaminen.Saman viestin toistaminen yhä uudelleen ja uudelleen, jotta yleisö lopulta hyväksyisi sen BIBREF13, BIBREF12.Propagandatekniikat ::: 4. Liioittelu tai vähättely.Joko jonkin asian esittäminen liiallisella tavalla: asioiden tekeminen suuremmiksi, paremmiksi tai huonommiksi tai jonkin asian esittäminen vähemmän tärkeältä tai pienemmältä kuin se todellisuudessa on BIBREF14, esim. sanomalla, että loukkaus oli vain vitsi.Propagandatekniikat ::: 5. Epäilys.Jonkun tai jonkin asian uskottavuuden kyseenalaistaminen.Propagandatekniikat ::: 6. Pelkoon/ennakkoluuloihin vetoaminen.Pyritään saamaan tukea ajatukselle herättämällä väestössä ahdistusta ja/tai paniikkia vaihtoehtoa kohtaan, mahdollisesti ennakkokäsityksiin perustuen.Propagandatekniikat :::: 7. Lippujen heiluttaminen.Vahvoilla kansallisilla tunteilla (tai ryhmään, esim. rotuun, sukupuoleen tai poliittisiin mieltymyksiin liittyvillä tunteilla) leikittely toiminnan tai ajatuksen oikeuttamiseksi tai edistämiseksi BIBREF15.Propagandatekniikat :::: 8. Syy-seuraussuhteiden liiallinen yksinkertaistaminen.Oletetaan yksi syy, vaikka asian taustalla on useita syitä. Siihen kuuluu myös syntipukkius: syyllisyyden siirtäminen yhteen henkilöön tai ihmisryhmään tutkimatta asian monimutkaisuutta.Propagandatekniikat ::: 9. Iskulauseet: Lyhyt ja näyttävä lause, joka voi sisältää leimaamista ja stereotypisointia. Iskulauseilla on taipumus toimia emotionaalisina vetoomuksina BIBREF16.Propagandatekniikat ::: 10. Auktoriteettiin vetoaminen: Väite, jonka mukaan väite on totta vain siksi, että asiaa käsittelevä auktoriteetti/asiantuntija tukee sitä, ilman muita tukevia todisteita BIBREF17. Sisällytämme mukaan erikoistapauksen, jossa viittaus ei ole auktoriteetti/asiantuntija, vaikka kirjallisuudessa sitä kutsutaankin todistukseksi BIBREF14.Propagandatekniikat ::: 11. Mustavalkoinen harhaluulo, diktatuuri.Kahden vaihtoehtoisen vaihtoehdon esittäminen ainoina mahdollisuuksina, vaikka todellisuudessa vaihtoehtoja on useampia BIBREF13. Ääritapauksessa yleisölle kerrotaan täsmälleen, mihin toimiin sen tulisi ryhtyä, jolloin kaikki muut mahdolliset vaihtoehdot suljetaan pois (diktatuuri).Propagandatekniikat :::: 12. Ajatuksen katkaiseva klisee.Sanoja tai lauseita, jotka estävät kriittisen ajattelun ja mielekkään keskustelun tietystä aiheesta. Ne ovat tyypillisesti lyhyitä ja yleisiä lauseita, jotka tarjoavat näennäisen yksinkertaisia vastauksia monimutkaisiin kysymyksiin tai jotka vievät huomion pois muilta ajatuslinjoilta BIBREF18.Propagandatekniikat ::: 13. Whataboutism.Discredit an opponent's position by charging them with hypocrisy without directly disproving their argument BIBREF19.Propaganda Techniques ::: 14. Reductio ad Hitlerum.Yleisön suostuttelu paheksumaan jotakin toimintaa tai ajatusta vihjaamalla, että kyseinen ajatus on suosittu kohdeyleisön halveksimissa ryhmissä. Se voi viitata mihin tahansa henkilöön tai käsitteeseen, jolla on negatiivinen konnotaatio BIBREF20.Propagandatekniikat :::: 15. Punahilkka.Epäolennaisen materiaalin tuominen käsiteltävään asiaan, jotta kaikkien huomio harhautuu pois esitetyistä asioista BIBREF11. Punaisen harhaanjohtavan väitteen kohteeksi joutuneet johdatellaan pois keskustelun keskiössä olleesta asiasta ja kehotetaan seuraamaan havaintoa tai väitettä, joka saattaa liittyä alkuperäiseen väitteeseen, mutta joka ei ole kovinkaan olennainen kiistanalaisen asian kannalta BIBREF20.Propagandatekniikat ::: 16. Bandwagon.Yritetään saada kohdeyleisö liittymään mukaan ja ryhtymään toimintatapaan, koska "kaikki muutkin ryhtyvät samaan toimintaan" BIBREF15.Propagandatekniikat ::: 17. Hämärtäminen, tarkoituksellinen epämääräisyys, hämmennys.Käytetään tarkoituksellisesti epäselviä sanoja, jotta yleisö saisi oman tulkintansa BIBREF21, BIBREF11. Esimerkiksi silloin, kun argumentissa käytetään epäselvää lausetta, jolla on useita mahdollisia merkityksiä, eikä se näin ollen oikeastaan tue johtopäätöstä.Propagandatekniikat ::: 18. Straw man: Kun vastustajan väite korvataan samankaltaisella, joka sitten kumotaan alkuperäisen BIBREF22 sijasta.

**Tulos**

Mitkä ovat 18 propagandatekniikkaa?

**Esimerkki 5.886**

Kokeemme tehdään todellisella uhanalaisella kielellä, Mboshi (Bantu C25), jota puhutaan Kongo-Brazzavillessa, käyttäen BIBREF17:n kaksikielistä ranska-mboshi 5K -korpusta.

**Tulos**

Mihin kieliperheeseen Mboshi kuuluu?

**Esimerkki 5.887**

Corpus utilizados ::: Corpus 5KLEste corpus fue constituido con aproximadamente 5 000 documentos (en su mayor parte libros) en español. Los documentos originales, en formatos heterogéneos, fueron procesados para crear un único documento codificado en utf8. Las frases fueron segmentadas automáticamente, usando un programa en PERL 5.0 y expresiones regulares, para obtener una frase por línea.Las características del corpus 5KL se encuentran en la Tabla TABREF4. Este corpus es empleado para el entrenamiento de los modelos de aprendizaje profundo (Deep Learning, Sección SECREF4).El corpus literario 5KL posee la ventaja de ser muy extenso y adecuado para el aprendizaje automático. Tiene sin embargo, la desventaja de que no todas las frases son necesariamente "frases literarias". Muchas de ellas son frases de lengua general: estas frases a menudo otorgan una fluidez a la lectura y proporcionan los enlaces necesarios a las ideas expresadas en las frases literarias.Otra desventaja de este corpus es el ruido que contiene. El proceso de segmentación puede producir errors en la detección de fronteras de frases. También los números de página, capítulos, secciones o índices producen errores. No se realizó ningún proceso manual de verificación, por lo que a veces se introducen informaciones indeseables: copyrights, datos de la edición u otros. Estas son, sin embargo, las condiciones que presenta un corpus literario real.Corpus utilizados ::: Corpus 8KFUn corpus heterogéneo de casi 8 000 frases literarias fue constituido manualmente a partir de poemas, discursos, citas, cuentos y otras obras. Se evitaron cuidadosamente las frases de lengua general, y también aquellas demasiado cortas ($N \le 3$ palabras) or demasiado largas ($N \ge 30$ palabras). El vocabulario empleado es complejo y estético, además que el uso de ciertas figuras literarias como la rima, la anáfora, la metáfora y otras pueden ser observadas en estas frases.Las características del corpus 8KF se muestran en la Tabla TABREF6. Este corpus fue utilizado principalmente en los dos modelos generativos: modelo basado en cadenas de Markov (Sección SECREF13) y modelo basado en la generación de Texto enlatado (Canned Text, Sección SECREF15).

**Tulos**

Mitä tietokokonaisuuksia käytetään?

**Esimerkki 5.888**

Toinen lähestymistapa kyselyjen ja kuvien esitystapojen luomiseen on kuvien käsitteleminen mustana laatikkona. Kun otetaan huomioon tietokokonaisuutemme tilastot (3B kysely- ja kuvaparia, joissa on 220 miljoonaa yksilöllistä kyselyä ja 900 miljoonaa yksilöllistä kuvaa), tiedämme, että eri kyselyt esiintyvät samojen kuvien kanssa. Jos kysely $q\_1$ esiintyy useiden samojen kuvien kanssa kuin kysely $q\_2$ , niin $q\_1$ ja $q\_2$ ovat todennäköisesti semanttisesti samankaltaisia riippumatta yhteisten kuvien visuaalisesta sisällöstä. Näin ollen voimme käyttää menetelmää, joka käyttää vain esiintymistilastoja, ymmärtääkseen paremmin, kuinka hyvin pystymme kuvaamaan kyselyiden välisiä suhteita.

**Tulos**

Voisitko oppia tällaisen sulauttamisen pelkästään kuvamerkintöjen perusteella ja ilman visuaalista informaatiota?

**Esimerkki 5.889**

Ensinnäkin se tarjoaa kommentointikäyttöliittymän, jonka avulla käyttäjät voivat määritellä sisältöelementtejä, ladata asiakirjoja ja kommentoida asiakirjoja. Alustamme pystyy ottamaan vastaan asiakirjoja eri muodoissa, kuten PDF- ja Microsoft Word -muodoissa, ja muuntaa nämä muodot tavalliseksi tekstiksi ennen niiden esittämistä kommentoijille.

**Tulos**

Minkä tyyppisiä asiakirjoja annotaatioalusta tukee?

**Esimerkki 5.890**

Arvioimme ehdotettua malliamme Stanford Network Analysis Projectin (SNAP) Amazonin arvostelutietokannassa BIBREF8, joka sisältää arvostelujen ja arvosanojen lisäksi myös kultaisia tiivistelmiä. Vertailemme empiirisesti eri menetelmiä käyttämällä Amazon SNAP Review Dataset BIBREF20 -aineistoa, joka on osa Stanford Network Analysis Project -hanketta. Jotta voimme tehdä oikeudenmukaisen vertailun aiempiin töihin, otamme käyttöön samat osiot kuin aiemmissa töissä BIBREF6, BIBREF7, eli jokaisella alueella ensimmäiset 1000 näytettä otetaan kehitysjoukoksi, seuraavat 1000 näytettä testijoukoksi ja loput harjoitusjoukoksi.

**Tulos**

Mitä tarkistustietokokonaisuutta he käyttävät?

**Esimerkki 5.891**

Keskitymme ensisijaisesti pienempään 645 käsin merkityn artikkelin aineistoon, joka annettiin tehtävään osallistujille sekä harjoittelua että validointia varten. Esikoulutimme BERT-basea 600 000:lla artikkelilla ilman merkintöjä käyttämällä samaa Cloze-tehtävää BIBREF5 , jota BERT oli alun perin käyttänyt esikoulutuksessa.

**Tulos**

Kuinka pitkä tietokokonaisuus on?

**Esimerkki 5.892**

Arvioimme menetelmäämme IWSLT16:n saksan ja englannin (DE-EN) välisessä käännöksessä molempiin suuntiin, WMT15:n englannin ja saksan (EN-DE) välisessä käännöksessä molempiin suuntiin ja NIST:n kiinan ja englannin (ZH$\rightarrow $EN) välisessä käännöksessä.

**Tulos**

Millä kielillä he tekevät kokeita?

**Esimerkki 5.893**

Koska tässä tutkimuksessa keskitytään espanjankielisiin twiitteihin, kaikki englanninkielisten tietokokonaisuuksien twiitit käännettiin espanjaksi.

**Tulos**

Miltä muilta kieliltä he käänsivät tiedot?

**Esimerkki 5.894**

mcdonald:11 totesi, että kun kohdekielellä ei ole saatavilla puupankki-annotaatioita, harjoittelu useilla lähdekielillä päihittää harjoittelun yhdellä kielellä (eli usean lähteen mallinsiirto päihittää yhden lähteen mallinsiirron). Guo:16:n mukaisesti kunkin kohdekielen osalta jäsennin koulutetaan kuudella muulla kielellä Googlen universaalissa riippuvuuspankissa versiossa 2.0 (de, en, es, fr, it, pt, sv, lukuun ottamatta kohdekieltä) ja käytetään kultaisia karkean tason POS-tunnisteita.

**Tulos**

Miten malli toimii, jos puupankkia ei ole saatavilla?

**Esimerkki 5.895**

Testataksemme, voivatko opitut representaatiot erottaa äänneasukategoriat toisistaan, käytämme minimaalisen parin ABX-erottelutehtävää BIBREF19 , BIBREF20 . Puhujan sisäisessä tehtävässä kaikki puhelintripletit kuuluvat samalle puhujalle (esim. $x$3 ). Lopuksi jokaisen keskeisen puhujaparin pisteet keskiarvoistetaan ja vähennetään 1:stä, jotta saadaan raportoitu puhujan sisäinen ABX-virheprosentti. Puhujien välisessä tehtävässä $x$4 ja $x$5 kuuluvat samaan puhujaan ja $x$6 eri puhujaan (esim. $x$7 ). Tietyn minimiparin pisteet lasketaan ensin keskiarvona kaikista niistä puhujapareista, joille tämä kontrasti voidaan tehdä. Kuten edellä, tuloksena saadut pisteet keskiarvoistetaan kaikkien kontekstien ja kaikkien keskuspuhelinparien osalta ja muunnetaan virheprosentiksi.

**Tulos**

Mikä on se mittari, jota tässä asiakirjassa mitataan?

**Esimerkki 5.896**

Tässä työssä käytämme perusluokittelijana MLP:tä (joka on muutettu käsittelemään tietojemme esitystapaa).

**Tulos**

Mitä luokittelualgoritmia he käyttävät s2sL:ssä?

**Esimerkki 5.897**

BERT-ADA BIBREF33 on APC-tehtävää varten ehdotettu toimialaan mukautettu yhteistyöhön perustuva yhteistyömalli, jossa BERT-BASE-mallia on hienosäädetty tehtävään liittyvällä korpuksella. Tämä malli saavutti huipputarkkuuden Laptops-tietokannassa. Rakennamme BERT-BASE-malliin perustuvan yhteisen mallin ATE:n ja APC:n monitehtävää varten. Kun malliparametrit oli optimoitu empiiristen tulosten perusteella, BERT-BASE-malliin perustuva yhteinen malli saavutti toivoa herättävän suorituskyvyn kaikissa kolmessa tietokokonaisuudessa ja jopa ylitti muut ehdotetut BERT-pohjaiset parannetut mallit joissakin tietokokonaisuuksissa, kuten BERT-PT, AEN-BERT, SDGCN-BERT jne.

**Tulos**

Mikä oli SemEval-2014 task4 -tehtävän ravintola- ja kannettava tietokone -tietokannan nykytila?

**Esimerkki 5.898**

Ensin vertaamme malleja BiLSTMBIBREF14, BiLSTM+CNNBIBREF20, BiLSTM+CRFBIBREF1, BiLSTM+CNN+CRFBIBREF2 CNN-malliinBIBREF0 ja Stanfordin CRF-malliinBIBREF21.

**Tulos**

Mikä on lähtötaso?

**Tulos**

Mitä malleja käytetään Nepalin kielen NER:n ratkaisemiseen?

**Esimerkki 5.899**

Kuten taulukosta TABREF39 käy ilmi, ehdotettu järjestelmä saa jatkuvasti parhaat tulokset kaikissa neljässä tietokokonaisuudessa. SemEval ABSA -kilpailun voittaneisiin järjestelmiin verrattuna kehyksemme saavuttaa 5,0 %, 1,6 %, 1,4 % ja 1,3 % absoluuttisen parannuksen INLINEFORM0-, INLINEFORM1-, INLINEFORM2- ja INLINEFORM3-tietokannoissa.Kehyksemme voi päihittää riippuvuuksien jäsennykseen perustuvan huippumallin RNCRF:n kaikissa tietokannoissa. Huomaamme myös, että RNCRF ei suoriudu hyvin INLINEFORM0- ja INLINEFORM1-tietokannoissa (3,7 % ja 3,9 % huonommin kuin meidän mallimme). Huomaamme, että INLINEFORM2 ja INLINEFORM3 sisältävät paljon epävirallisia arvosteluja, joten RNCRF:n suorituskyvyn heikkeneminen johtuu luultavasti riippuvuuksien jäsentäjän virheistä, kun se käsittelee tällaisia epävirallisia tekstejä.

**Tulos**

Kuinka paljon ne ovat tehokkaampia kuin uusimmat menetelmät?

**Esimerkki 5.900**

CIC on sopiva täydentävä mittari ROUGE:lle, koska se ottaa huomioon tärkeimmät tiedot kullakin dialogialueella. CIC:tä voidaan soveltaa mihin tahansa tiivistämistehtävään, jossa on ennalta määriteltyjä olennaisia kokonaisuuksia.

**Tulos**

Onko uusi arviointimittari ROGUEn laajennus?

**Esimerkki 5.901**

Käytimme Pasokh-tietokokonaisuutta BIBREF42 , joka sisältää 100 persialaista uutisasiakirjaa, joihin kuhunkin liittyy 5 tiivistelmää. Kukin tiivistelmä koostuu useista lauseista alkuperäisestä tekstistä, jotka on valinnut ihmisasiantuntija. Joitakin lauseita on hieman muutettu, eivätkä ne näin ollen ole täsmällinen kopio alkuperäisistä lauseista. Asiakirjat on luokiteltu kuuteen luokkaan, kuten poliittisiin, taloudellisiin ja muihin. Asiakirjojen pituus vaihtelee 4:stä 156:een virkkeeseen. Kaiken kaikkiaan siinä on noin 2 500 lausetta.

**Tulos**

Mitä tietokokonaisuutta käytetään tässä tehtävässä?

**Esimerkki 5.902**

Otimme tehtävän käyttöön Amazon Mechanical Turkissa (AMT). Nähdäksemme, miten päättelytaidot vaihtelevat työntekijöiden välillä, palkkaamme 3 väkijoukkotyöntekijää yhtä instanssia kohden. Palkkaamme luotettavia työntekijöitä, joilla on $\ge 5 000$ HIT-kokemusta ja hyväksyntäprosentti $\ge $ 99.0 %, ja maksamme ¢20 palkkiota instanssia kohden.

**Tulos**

Käyttivätkö he jotain joukkoistamisalustaa?

**Esimerkki 5.903**

Seuraavan kysymyksemme ratkaisemiseksi arvioimme putkistomme toimintaa, kun suhteita opitaan itsenäisesti (eli erikseen) verrattuna suhteiden yhteiseen oppimiseen RDN:ssä, mikä näkyy taulukossa TABREF22 .

**Tulos**

Mitä he oppivat yhdessä?

**Esimerkki 5.904**

Syöttöominaisuudet ovat 40 mel-skaalattua log-suodatinpankki-egregaattia (FBanks), jotka lasketaan 10 ms:n välein 25 ms:n ikkunassa ja jotka on yhdistetty deltoihin ja delta-deltoihin (120 ominaisuutta vektorissa). Yritimme myös käyttää spektrogrammeja ja kokeilimme erilaisia normalisointitekniikoita.

**Tulos**

Mitä ominaisuuksia he kokeilevat?

**Esimerkki 5.905**

Nyt on selvää, että jotkin joissakin arkkitehtuureissa käytetyt konfiguraatiot (esim. ylimääräiset RNN-kerrokset) ovat itse asiassa merkityksettömiä ja ne voidaan poistaa kokonaan ilman, että tarkkuus kärsii.

**Tulos**

Mitkä ovat vähiten tärkeät osatekijät, jotka on tunnistettu VQA-mallien koulutuksessa?

**Esimerkki 5.906**

Kuvion 1 kaavioissa esitetään päällekkäisten ja ei-yllekkäisten kysymysten väliset etäisyydet eri sulautusjärjestelmillä. Large saavutti tässä tehtävässä F1-pistemäärän 0,71 ilman erityistä harjoittelua yksinkertaisesti valitsemalla kynnysarvon, jonka alapuolella kaikki lauseparit katsotaan kaksoiskappaleiksi. Testataksemme, voidaanko nämä tulokset yleistää meidän alallamme, suunnittelimme testin, jossa hyödynnettäisiin sitä vähäistä dataa, joka meillä oli arvioitavana.

**Tulos**

Miten järjestelmän tarkkuus mitataan?

**Esimerkki 5.907**

Tässä kokeessa laskemme ehdotetun NED-mittarin korrelaation potilaan havaitsemien tunnesidearvioiden kanssa. Koska ehdotettu mittari on luonteeltaan epäsymmetrinen, laskemme mittarit sekä potilaan ja terapeutin väliselle että terapeutin ja potilaan väliselle sitoutumiselle. Ilmoitamme Pearsonin korrelaatiokertoimet ( INLINEFORM0 ) tätä koetta varten taulukossa TABREF26 yhdessä niiden INLINEFORM1 -arvojen kanssa.

**Tulos**

Miten he korreloivat NED:n ja tunnesiteen tasojen välillä?

**Esimerkki 5.908**

Vertaamme malliamme useisiin perusmalleihin, jotka kuvataan jäljempänä. Huomaamme, että kahden ensimmäisen tulokset on aiemmin raportoitu BIBREF2.Hasty Student BIBREF2 on heuristiikkapohjainen yksinkertainen malli, joka jättää reseptin huomiotta ja antaa vastauksen tarkastelemalla vain kysymystä ja vastausjoukkoa käyttäen etäisyyksiä visuaalisessa ominaisuusavaruudessa.Impatient Reader BIBREF19 on yksinkertainen neuraalinen malli, joka on saanut nimensä siitä, että se laskee toistuvasti huomion reseptin yli sen jälkeen, kun se on tarkastellut jokaista kyselyssä olevaa kuvaa. biDAF BIBREF14 on vahva luetun ymmärtämisen malli, joka käyttää kaksisuuntaista huomionvirtausmekanismia saadakseen kysymystietoisen representaation ja perustaa ennusteensa tähän representaatioon. Alun perin se on span-valintamalli syötekontekstista. BiDAF w/ static memory on laajennettu versio BiDAF-mallista, joka muistuttaa ehdottamaamme PRN-mallia siten, että se sisältää muistin yksikön olioille.

**Tulos**

Mitä ovat aiemmin raportoidut mallit?

**Esimerkki 5.909**

Tunnetunnistusta tekstistä varten transkriboimme manuaalisesti kaikki AMMER-tutkimuksen lausumat. Kehitämme siirto-oppimismenetelmän hyödyntääkseen olemassa olevia ja saatavilla olevia tietokokonaisuuksia, jotka ovat suurempia kuin AMMER-tietokokonaisuus. Käytämme neuroverkkoa, jossa on sulauttamiskerros (jäädytetyt painot, esivalmennettu Common Crawl ja Wikipedia BIBREF36), kaksisuuntainen LSTM BIBREF37 ja kaksi tiheää kerrosta, joita seuraa pehmeä maksimilähtökerros. Tämä asetelma on BIBREF38:n innoittama. Käytämme kaikissa kerroksissa 0,3:n pudotusprosenttia ja optimoimme Adam BIBREF39:llä oppimisnopeudella $10^{-5}$ (nämä parametrit ovat samat kaikissa myöhemmissä kokeissa). Rakennamme Keras-kirjaston päälle TensorFlow-backendillä. Pidämme tätä asetelmaa perusmallina.

**Tulos**

Mikä on tehtävän perusmenetelmä?

**Esimerkki 5.910**

Jotta voimme suunnitella tulkittavissa olevan sanojen upotusmallin pienille korpuksille, tunnistamme uusia yhteyksiä sanojen upotusten ja aihepiirimallien välillä ja mukautamme aihepiirimallinnuksen edistysaskeleita. Jakaumahypoteesin BIBREF23 mukaisesti skip-grammin sanojen upotukset parametrisoivat diskreettejä todennäköisyysjakaumia sanoille INLINEFORM0, joilla on taipumus esiintyä yhdessä ja jotka ovat yleensä semanttisesti yhtenäisiä - ominaisuus, jota BIBREFin Gaussin LDA-malli hyödyntää21. Tulkitsemme siis skip-grammin uudelleen tietyn valvotun naivin Bayesin aihepiirimallin parametrisoinniksi (taulukko TABREF2 , ylhäällä oikealla). Kuten LDA:ssa, tätä mallia voidaan parantaa korvaamalla naiivi Bayes-oletus sekajäsenyysoletuksella. Soveltamalla mixed membership -esitystä tähän skip-grammin aihepiirimalliversioon saadaan taulukon TABREF2 oikeassa alakulmassa oleva malli. Kun tämä malli on jälleen kerran parametrisoitu sanojen upotusten avulla, saadaan lopullinen mallimme, mixed membership skip-gram (MMSG).

**Tulos**

Mitä tekniikoita käytetään sanojen upotuksiin ja aihepiirimalleihin?

**Esimerkki 5.911**

Havainnollistamme konseptia käsittelemällä joitakin hiljattain kokeilemiamme instansseja. Suunnitelmassa on myös selkeä "relevanssipaikka", jossa on mahdollisuus semanttiseen neuvotteluun, nimittäin ennen lopullisen päätöksen tekemistä. Tästä on esimerkki alla olevassa esimerkissä. Kuten SECREF12 osoittaa, jopa tapauksissa, joissa päätös voidaan tehdä nopeasti, voi olla selkeä keskinäinen vahvistusvaihe ennen (hiljaisen) päätöksentekosignaalin lähettämistä. Kolmas tutkimamme asetelma BIBREF19 tuo käsitteellisen neuvottelun selvemmin etualalle. Kyseisessä pelissä pelaajille esitetään kuvia tiettyyn lajiin kuuluvista linnuista, ja heidän tehtävänään on keksiä kuvaus yhteisistä ominaisuuksista. Tässäkin tapauksessa molempien osallistujien on hyväksyttävä lopullinen vastaus. Kuten SECREF13 osoittaa, tämä voi johtaa käsitteellistä sisältöä koskevaan eksplisiittiseen neuvotteluun.

**Tulos**

Suorittavatko kirjoittajat kokeita ehdottamallaan menetelmällä?

**Esimerkki 5.912**

Stanford Sentiment Treebank Stanford Sentiment Treebank (SST) BIBREF14 . Kyseessä on elokuva-arvostelujen tunteen ennustaminen. Tästä korpuksesta on johdettu kaksi tietokokonaisuutta: (1) SST-1, joka sisältää viisi luokkaa: erittäin negatiivinen, negatiivinen, neutraali, positiivinen ja erittäin positiivinen. (2) SST-2, jossa on vain kaksi luokkaa: negatiivinen ja positiivinen. Molemmista poistetaan harjoitusjoukosta alle 4 pituiset lauseet. Subj BIBREF15 . Tässä pyritään luokittelemaan lauseet joko subjektiivisiksi tai objektiivisiksi. Kokonaisuuteen kuuluu 5000 kappaletta kumpaakin. TREC BIBREF16 . Kysymysten luokittelua koskeva tietokokonaisuus, joka sisältää kuusi luokkaa: lyhenne, entiteetti, kuvaus, ihminen, sijainti ja numeerinen. Harjoitustapauksia on 5500 ja testitapauksia 500. Ironia BIBREF17 . Tämä tietokokonaisuus sisältää 16 006 lausetta redditistä, jotka on merkitty ironisiksi (tai ei ironisiksi). Tietokanta on epätasapainoinen (suhteellisen harvat lauseet ovat ironisia). Ennen harjoittelua negatiiviset tapaukset otettiin liian pienestä näytteestä, jotta luokkien koot olisivat yhtä suuret. Huomaa, että tästä tietokokonaisuudesta raportoimme tarkkuuden sijasta AUC-arvon (Area Under Curve), koska se on epätasapainoinen.

**Tulos**

Minkä tietokokonaisuuden/korpuksen perusteella tätä arvioidaan?

**Esimerkki 5.913**

Toinen ongelma on se, että Googlen julkaisema BERT-malli on koulutettu vain arabian kielen Wikipediaan, joka on lähes yksinomaan modernia arabiaa (MSA). Jaetun tehtävän tietokokonaisuus sisältää 5 030 twiittiä, jotka liittyvät Lähi-idän eri poliittisiin kysymyksiin ja tapahtumiin vuosina 2011-2018. Twiitit sisältävät sekä MSA:ta että eriasteisia murteita, kuten egyptiläistä, Persianlahden murretta ja Levantin murretta.

**Tulos**

Mitä murretta käytetään Googlen yhteistyömallissa ja mitä käytetään tehtävätiedoissa?

**Esimerkki 5.914**

CodeInternational: Google Translate -työkalu, jolla voidaan kääntää koodia eri kielten välillä.

**Tulos**

Perustuuko tämä automaattinen käännöstyökalu neuroverkkoihin?

**Esimerkki 5.915**

Pyysimme lääkäreitä, joilla on kokemusta lääketieteellisiin kokonaisuuksiin liittyvän tiedon poimimisesta teksteistä, kommentoimaan edellä kuvattuja kokonaisuuksia. Aluksi pyysimme neljää annotoijaa testaamaan ohjeistustamme kahdella tekstillä. Tämän jälkeen havaituista ongelmista keskusteltiin ja ne ratkaistiin. Tämän pilottiannotointivaiheen jälkeen pyysimme kahta eri annotoijaa annotoimaan kaksi tapausraporttia ohjeidemme mukaisesti. Samat kommentoijat kommentoivat 53 tapausselostuksen kokoelman. Annotoinnissa käytettiin WebAnno BIBREF7 -nettityökalua, joka on web-pohjainen työkalu kielellistä annotointia varten. Annotaattorit saattoivat valita kunkin tekstin valmiiksi annotoidun version tai tyhjän version. Ennalta annotoidut versiot sisälsivät ehdotettuja entiteettikokonaisuuksia, jotka perustuivat ehtoluetteloiden ja löydösten synonyymiluetteloiden merkkijonotuloksiin.

**Tulos**

Miten merkinnät tehtiin?

**Esimerkki 5.916**

Parannamme aiempaa huipputason BIBREF35 -menetelmää VQA-aineistossa noin 6 % BLEU-pisteissä ja 10 % METEOR-pisteissä. VQG-COCO-tietokannan osalta parannamme BIBREF5:tä 3,7 prosenttia ja BIBREF36:ta 3,5 prosenttia METEOR-pisteiden osalta.

**Tulos**

Mitkä olivat aiemmat huipputason vertailuarvot?

**Esimerkki 5.917**

Harvennamme tämän graafin minimitilannepuun ja k-naapurien (MST-kNN) graafin BIBREF14 yhdistelmäksi. Tämä geometrinen rakenne poistaa vähemmän tärkeät samankaltaisuudet, mutta säilyttää graafin ja siten myös tietokokonaisuuden globaalin liitettävyyden. Tämän jälkeen MST-kNN-graafia analysoidaan Markov Stability BIBREF15, BIBREF16, BIBREF17, BIBREF18 -menetelmällä, joka on moniresoluutioinen graafin jakomenetelmä, jolla tunnistetaan merkitykselliset aliraafit (eli asiakirjojen klusterit) eri rakeisuustasoilla.

**Tulos**

Mitä klusterin tunnistamismenetelmää käytetään tässä asiakirjassa?

**Esimerkki 5.918**

Päädymme kahteen merkittävään rooliin, joita kutsutaan ankkuripuhujiksi ja täsmällisiksi puhujiksi: ankkuripuhujat (A) ovat yli 1 prosentin kynnyksen molemmissa kriteereissä, mikä tarkoittaa, että he puhuvat usein ja pitkään ja ovat siten tärkeässä asemassa vuorovaikutuksessa; täsmälliset puhujat (PS) sitä vastoin ovat alle 1 prosentin kynnyksen sekä vuorojen kokonaismäärän että kokonaispuheajan osalta.

**Tulos**

Kuinka monta luokkaa kirjoittajat määrittelevät puhujan roolille?

**Esimerkki 5.919**

HASOC 2019 -tapahtuman tietokokonaisuudet annettiin kolmella kielellä: Hindi, englanti ja saksa. Hindi- ja englanninkielisessä datasetissä oli kummassakin kolme osatehtävää, kun taas saksankielisessä vain kaksi osatehtävää. Osallistuimme kaikkiin järjestäjien antamiin tehtäviin ja päätimme kehittää yhden ainoan mallin, joka olisi kieliagnostinen. Käytimme samaa malliarkkitehtuuria kaikilla kolmella kielellä. Hindi-tietokannan harjoitustietokanta oli tasapainoisempi kuin englannin- tai saksankielinen tietokanta. Näin ollen tulokset olivat noin 0,78. Koska saksan kielen tietokanta oli erittäin epätasapainoinen, tulokset putosivat 0,62:een.

**Tulos**

Mitä kieliä käytetään mallin testaamiseen?

**Esimerkki 5.920**

Kartoitamme jokaisen relaatiotyypin $R(x,y)$ vähintään yhdeksi parametrisoiduksi luonnollisen kielen kysymykseksi $q\_x$ , jonka vastaus on $y$ .

**Tulos**

Miten syötekolmio muunnetaan aukkotäytteiseksi tehtäväksi?

**Esimerkki 5.921**

Kiillekirja on muinainen kirjoitusjärjestelmä, jonka sumerilaiset keksivät yli kolmen vuosituhannen ajan.

**Tulos**

Mikä on yksi maailman ensimmäisistä kirjoitusjärjestelmistä?

**Esimerkki 5.922**

Kilpailu on jaettu viiteen osatehtävään, joihin kuuluvat standardiluokittelu, ordinaaliluokittelu ja jakauman estimointi. Yksityiskohtaisempi kuvaus löytyy osoitteesta BIBREF0 .

**Tulos**

Mitkä olivat viisi englanninkielistä osatehtävää?

**Esimerkki 5.923**

Tämän jälkeen LSTM-dekooderi tuottaa merkit sanamuodossa käyttäen kooderin tiloja ja huomiomekanismia.

**Tulos**

Mikä on dekooderin arkkitehtuuri?

**Tulos**

Onko malli tarkkaavainen?

**Esimerkki 5.924**

Tämän työn tulokset ovat:

**Tulos**

Etsivätkö he epäjohdonmukaisuuksia eri UD-pankkien välillä?

**Esimerkki 5.925**

Arvioimme lähestymistapaamme kahdella hyvin vertaillulla sekvenssien merkintätehtävällä, CoNLL 2003 NER-tehtävällä BIBREF13 ja CoNLL 2000 Chunking-tehtävällä BIBREF14 .

**Tulos**

mitkä ovat arviointitietokannat?

**Esimerkki 5.926**

Operationalisoimme kysymyksen lähettämällä valitut testidatat samalle toimittajapohjaiselle transkriptioputkelle, jota Microsoft käyttää tuotantodataa varten (mallien kouluttamista ja sisäistä arviointia varten), ja vertaamalla tuloksia ASR-järjestelmän tuloksiin NIST:n pisteytysprotokollan mukaisesti.

**Tulos**

mitä standardoitua puheen transkriptioputkea käytettiin?

**Esimerkki 5.927**

OLID-tietokannassa käytämme hierarkkista annotaatiomallia, joka on jaettu kolmeen tasoon, jotta voidaan erottaa toisistaan, onko kieli loukkaavaa vai ei (A), sekä loukkaavan kielen tyyppi (B) ja kohde (C). Kutakin tasoa kuvataan yksityiskohtaisemmin seuraavissa alaluvuissa, ja esimerkkejä esitetään taulukossa TABREF10 . Taso A: Loukkaavan kielen havaitseminenTasolla A erotetaan toisistaan loukkaavat (OFF) ja ei-loukkaavat (NOT) twiitit. Taso B: Loukkaavan kielen luokitteluTasolla B luokitellaan loukkauksen tyyppi, ja käytetään kahta merkintää: kohdistetut (TIN) ja kohdistamattomat (INT) loukkaukset ja uhkaukset. Taso C: Loukkaavan kielen kohteiden tunnistaminenTasolla C luokitellaan loukkausten ja uhkausten kohteet yksilöihin (IND), ryhmiin (GRP) ja muihin (OTH).

**Tulos**

Mitkä ovat huomautusjärjestelmän kolme kerrosta?

**Esimerkki 5.928**

LastStateRNN on klassinen RNN-malli, jossa viimeinen tila kulkee MLP:n läpi ja LR-kerros arvioi vastaavan todennäköisyyden. AvgRNN-mallissa sen sijaan tarkastellaan kaikkien soluista lähtevien tilojen keskimääräistä vektoria. AttentionRNN-malli on se, joka on esitetty BIBREF9:ssä.

**Tulos**

Mikä oli lähtötaso?

**Esimerkki 5.929**

Tammikuussa 2020 uusin CoVo-versio 2019-06-12 sisältää 29 kieltä. CoVoST perustuu tällä hetkellä kyseiseen julkaisuun ja kattaa seuraavat 11 kieltä: Ranska, saksa, hollanti, venäjä, espanja, italia, turkki, persia, ruotsi, mongoli ja kiina. Common Voice BIBREF10 on joukkoistettu puheentunnistuskorpus, jolla on avoin CC0-lisenssi. Osallistujat nauhoittavat äänileikkeitä lukemalla lahjoitettujen lauseiden pankista.

**Tulos**

Miten aineisto kerättiin?

**Esimerkki 5.930**

Kaiken kaikkiaan tämä vahvistaa alkuperäistä intuitiotamme, jonka mukaan parhaiten toimivan suosioon perustuvan lähestymistavan ja parhaan samankaltaisuuteen perustuvan lähestymistavan yhdistämisen pitäisi johtaa korkeimpaan tarkkuuteen (eli INLINEFORM7 INLINEFORM8 -menetelmään).

**Tulos**

mikä algoritmi oli parhaiten menestynyt?

**Esimerkki 5.931**

SnapCaptions-tietokanta koostuu 10 000 käyttäjän luoman kuvan (snap) ja tekstin kuvatekstipareista, joissa kuvatekstien nimetyt entiteetit on merkitty manuaalisesti asiantuntevilla ihmisannotoijilla (entiteettityypit: PER, LOC, ORG, MISC).

**Tulos**

Minkä tyyppisiä nimettyjä entiteettejä ne tunnistavat?

**Tulos**

Kuinka suuri on heidän MNER SnapCaptions -tietokantansa?

**Esimerkki 5.932**

Ainoa vaatimus on, että malli hyväksyy syötteenä (entiteettien ja relaatioiden) upotuskerroksen. Jos malli täyttää tämän vaatimuksen (ja monet tietämysgraafeihin perustuvat neuraaliset mallit täyttävät sen), voimme käyttää Doloresin upotuksia korvaavana vaihtoehtona. Aloitamme vain vastaavan upotuskerroksen Doloresin upotuksilla.

**Tulos**

Tarvitaanko hienosäätöä, jotta nämä sulautukset voidaan sisällyttää nykyisiin malleihin?

**Esimerkki 5.933**

Tässä jaksossa kuvataan kokeellinen tutkimus. Sovitamme sefe-mallin kolmeen tietokokonaisuuteen ja vertaamme sitä efe BIBREF10 -malliin. Kvantitatiiviset tulokset osoittavat, että kontekstivektoreiden jakaminen tuottaa parempia tuloksia ja että amortisointi ja hierarkkinen rakenne parantavat tuloksia entisestään. Aineisto. Sovellamme sefe-mallia kolmeen tietokokonaisuuteen: ArXiv-papereihin, Yhdysvaltain senaatin puheisiin ja ostoksiin supermarketin päivittäistavarakaupan ostosdatassa. Kuvaamme nämä aineistot jäljempänä, ja taulukossa TABREF17 on yhteenveto aineistoista. ArXiv-paperit: Tämä tietokokonaisuus sisältää huhtikuun 2007 ja kesäkuun 2015 välisenä aikana ArXivissa 19 eri tunnisteella julkaistujen papereiden tiivistelmät. Käsittelemme kutakin tagia ryhmänä ja sovitamme sefejä tavoitteenamme paljastaa, millä sanoilla on voimakkain muutos käytössä. Jaamme tiivistelmät harjoittelu-, validointi- ja testijoukkoihin, joiden osuudet ovat vastaavasti INLINEFORM0 , INLINEFORM1 ja INLINEFORM2. Senaattipuheet: Tämä tietokokonaisuus sisältää Yhdysvaltain senaatin puheita vuodesta 1994 vuoden 2009 puoliväliin. Toisin kuin ArXiv-kokoelma, se on transkriptio puhutusta kielestä. Ryhmittelemme aineiston puhujan kotiosavaltion ja puoluekannatuksen mukaan. Ainoastaan republikaaniseen ja demokraattiseen puolueeseen kuuluminen otetaan huomioon. Tuloksena on 83 ryhmää (republikaanit Alabamasta, demokraatit Alabamasta, republikaanit Arkansasista jne.). Joitakin osavaltio/puolue-yhdistelmiä ei ole saatavilla tiedoissa, koska joissakin 50 osavaltiosta on ollut vain senaattoreita, joilla on sama puoluekanta. Jaoimme puheet harjoitteluun ( INLINEFORM0 ), validointiin ( INLINEFORM1 ) ja testaukseen ( INLINEFORM2 ). päivittäistavarakauppa-aineisto: Tämä aineisto sisältää INLINEFORM0-asiakkaiden ostokset. Aineisto kattaa 97 viikon ajanjakson. Kun matalataajuiset nimikkeet on poistettu, aineisto sisältää INLINEFORM1:n yksilöllisiä nimikkeitä 1.10upc-tasolla (Universal Product Code). Jaamme aineiston harjoitus-, testi- ja validointijoukkoihin, joiden osuudet ovat INLINEFORM2 , INLINEFORM3 ja INLINEFORM4 . Koulutusdata sisältää yhteensä INLINEFORM5 ostosmatkoja ja INLINEFORM6 ostoksia. Tekstikorpusten osalta kiinnitämme sanaston 15k yleisimpiin termeihin ja poistamme kaikki sanastoon kuulumattomat sanat. BIBREF2:n mukaisesti poistamme lisäksi jokaisen sanan, jonka todennäköisyys on INLINEFORM0 , jossa INLINEFORM1 on sanan taajuus. Tämä pienentää näytteitä erityisesti usein esiintyvistä sanoista ja nopeuttaa harjoittelua. (Taulukossa TABREF17 ilmoitetut koot ovat esikäsittelyn jälkeen jäljelle jäävien sanojen lukumäärä). Mallit. Tavoitteenamme on sovittaa sefe-malli näihin tietokokonaisuuksiin. Käytämme tekstidatassa Bernoulli-jakaumaa ehdollisena eksponenttiperheenä, kun taas ostosdatassa käytämme Poisson-jakaumaa, joka sopii paremmin laskentadataan. Kussakin tietokokonaisuudessa verrataan neljää sefe-malliin perustuvaa lähestymistapaa kahteen efe BIBREF10 -perusasetukseen. Kaikki on sovitettu sgd BIBREF34 -ohjelmalla. Vertaamme erityisesti seuraavia menetelmiä:

**Tulos**

Mitä kokeita käytetään tämän lähestymistavan hyötyjen osoittamiseksi?

**Esimerkki 5.934**

29 794 Wikipedia-aineisto koostuu englanninkielisen Wikipedian artikkeleista, joille Wikipedia-yhteisö on antanut laatuluokkamerkinnät. Wikipedia-artikkelit on merkitty johonkin kuudesta laatuluokasta, jotka ovat laadun mukaan alenevassa järjestyksessä: Hyvä artikkeli ("GA"), B-luokan artikkeli ("B"), C-luokan artikkeli ("C"), aloitusartikkeli ("Start") ja tynkäartikkeli ("Stub"). Otimme satunnaisotannalla 5 000 artikkelia kustakin laatuluokasta ja poistimme kaikki uudelleenohjaussivut, jolloin saimme 29 794 artikkelin tietokokonaisuuden.

**Tulos**

Kuinka suuri on heidän aineistonsa?

**Esimerkki 5.935**

Kuten yhtälöstä ( EQREF6 ) käy ilmi, äänensävyn ennustamisen aliverkko ( INLINEFORM0 ) ottaa videon ja pinyin-sekvenssin syötteenä ja ennustaa vastaavan äänensävysekvenssin. Videokontekstivektorit INLINEFORM0 ja pinyin-kontekstivektorit INLINEFORM1 fuusioidaan, kun äänimerkkiä ennustetaan kussakin dekooderivaiheessa. Videokooderi on sama kuin kohdassa SECREF7 ja pinyin-kooderi on: DISPLAYFORM0 Syötetty videosekvenssi syötetään ensin VGG-malliin BIBREF9 visuaalisten ominaisuuksien poimimiseksi. VGG:n conv5:n ulostuloon liitetään globaali keskiarvopoolaus BIBREF10, jotta saadaan 512-ulotteinen piirrevektori. Tämän jälkeen 512-dim-muotoinen ominaisvektori syötetään videokooderiin.

**Tulos**

Mikä visuaalinen informaatio luonnehtii sävyjä?

**Esimerkki 5.936**

Käytämme leimattua dataa IEMOCAP-tietokannasta, jossa on leimatut aktivointi- ja valenssiarvot, jotka molemmat on mitattu viiden pisteen Likert-asteikolla kolmen eri kommentoijan toimesta.

**Tulos**

Mitä monitehtäväannotoitua korpusta käytetään?

**Esimerkki 5.937**

Perinteiset tekstistä puheeksi -järjestelmät (TTS) koostuvat monimutkaisista putkistoista BIBREF0 , joihin kuuluu usein akustisia etuosia, kestomalli, akustinen ennustemalli ja vokooderimalleja. Neuraaliset tekstistä puheeksi -järjestelmät ovat herättäneet suurta tutkimusintoa viimeisten kahden vuoden aikana. Ensimmäinen, joka tutki tätä tutkimusalaa perusteellisesti, oli Googlen tacotron BIBREF1 -järjestelmä. Mallimme arkkitehtuurissa käytetään RNN-pohjaista Seq2Seq-mallia mel-spektrogrammin tuottamiseen tekstistä. Arkkitehtuuri on samankaltainen kuin Tacotron 2:ssa BIBREF4 Alkuperäisessä Tacotron 2:ssa huomiomekanismina käytettiin sijaintiherkkää huomiota BIBREF12 yhdistettynä alkuperäiseen additiiviseen Seq2Seq BIBREF7 Bahdanau-huomioon. ehdotamme tämän huomion korvaamista yksinkertaisemmalla query-key-huomioinnilla muuntajamallista Edellä esitetyn logiikan mukaisesti käytämme BIBREF6:sta tuttua samankaltaista menetelmää, joka lisää kokonaistappiotavoitteeseen ylimääräisen ohjatun huomion häviön, joka toimii siten, että huomiomekanismista tulee mahdollisimman varhaisessa vaiheessa monotoinen. Kuten FIGREF24:stä nähdään, luodaan huomiohäviömaski, INLINEFORM0 , soveltaa häviötä, joka pakottaa huomiokohdistuksen, INLINEFORM1 , olemaan lähes diagonaalinen.

**Tulos**

Mitä muutoksia ne tekevät vakiintuneisiin Seq2seq-arkkitehtuureihin?

**Esimerkki 5.938**

Yleisesti ottaen vastakohtainen harjoittelu parantaa tarkkuutta kaikilla kielillä, ja parannus on toisinaan dramaattinen verrattuna BERT:n ei-vastakohtaiseen perustasoon. Kun BERT-yhteistyöhön yhdistetään adversaalinen oppiminen, kieltenväliset F1-pisteet paranevat saksan osalta verrattuna ei-adversaaliseen perustasoon, ja espanjan ja hollannin osalta tulokset pysyvät pitkälti samoina.

**Tulos**

Onko adversio-oppimisella saavutettu vahvempi suorituskyky tekstiluokittelussa vai NER:ssä?

**Esimerkki 5.939**

Arvioimme Bertramia BIBREF0:n WNLaMPro-tietokannassa.

**Tulos**

Mikä on sanojen kartoitustehtävän tietokanta?

**Esimerkki 5.940**

Tässä jaksossa ehdotamme uusia yksinkertaisia disentanglementtimalleja, jotka toimivat aiempia menetelmiä paremmin, ja tarkastelemme uudelleen aiempaa työtä.

**Tulos**

Kokeilivatko he korpusta?

**Esimerkki 5.941**

Tässä jaksossa keskitytään äskettäin julkaistuihin tietokokonaisuuksiin ja LID-tutkimuksiin, joita voidaan soveltaa Etelä-Afrikan tilanteeseen. Syvällinen katsaus algoritmeihin, ominaisuuksiin, tietokokonaisuuksiin, jaettuihin tehtäviin ja arviointimenetelmiin on BIBREF0.DSL 2015 & DSL 2017 - jaettujen tehtävien tietokokonaisuuksia BIBREF1 käytetään usein LID-vertailumittauksissa, ja ne ovat saatavilla myös Kaggle-verkkopalvelussa. Hiljattain julkaistu rinnakkainen JW300-korpus BIBREF2 kattaa yli 300 kieltä, ja siinä on keskimäärin noin 100 tuhatta rinnakkaista lausetta kieliparia kohti. Yksikieliseen kirjoitetun luonnollisen kielen tunnistamiseen tarkoitettu WiLI-2018-vertailuaineisto BIBREF4 sisältää noin 1000 kappaletta 235 kielestä. NCHLT-tekstikorporaatio BIBREF7 on todennäköisesti hyvä lähtökohta eteläafrikkalaisten kielten BIBREF8 yhteiselle LID-tehtävätietokannalle.

**Tulos**

Mitä tietokokonaisuuksia käytetään Etelä-Afrikan kielten LID-tutkimuksessa?

**Esimerkki 5.942**

BIBREF0-, BIBREF1- ja BIBREF2-ohjelmissa Pasca et al. poimivat ensin mahdollisia luokka-attribuuttipareja käyttämällä kielellisesti motivoituja malleja strukturoimattomasta tekstistä, mukaan lukien kyselylokit ja kyselyistunnot, ja pisteyttävät sitten attribuutit Bayesin mallin avulla. Rahul Rai ehdotti julkaisussa BIBREF3 , että tuoteattribuutteja tunnistetaan asiakkaiden online-arvosteluista käyttämällä part-of-speech(POS)-merkintämalleja ja että niiden tärkeyttä arvioidaan useilla eri frekvenssimittareilla. BIBREF4 -julkaisussa Lee et al. kehittivät järjestelmän käsite-attribuuttiparien poimimiseksi useista tietolähteistä, kuten Probase-tietokannasta, yleisistä verkkodokumenteista, kyselylokeista ja ulkoisesta tietopohjasta, ja eri lähteistä saatujen painotusten yhdistämiseksi yhdeksi johdonmukaiseksi tyypillisyyspisteytykseksi käyttämällä Ranking SVM -mallia. BIBREF5 -julkaisussa Li et al. esittelivät OntoRank-algoritmin semanttisten web-objektien tärkeysjärjestykseen asettamista varten kolmella rakeisuustasolla: dokumentti, termit ja RDF-grafit. Algoritmi perustuu rational surfer -malliin, jota käytetään menestyksekkäästi Swoogle-semanttisessa web-hakukoneessa.

**Tulos**

Mitkä ovat perinteiset menetelmät tärkeiden ominaisuuksien tunnistamiseksi?

**Esimerkki 5.943**

Käytämme NIST:n DUC-05, DUC-06 ja DUC-07 jaettujen tehtävien BIBREF7, BIBREF19 ja BIBREF20 tietokokonaisuuksia. Kun kilpailijoille annettiin kysymys ja joukko uutisdokumentteja, heitä pyydettiin laatimaan 250 sanan tiivistelmä, joka vastaa kysymykseen. DUC-05 sisältää 1 600 tiivistelmää (50 kysymystä x 32 järjestelmää), DUC-06 sisältää 1 750 tiivistelmää (50 kysymystä x 35 järjestelmää) ja DUC-07 sisältää 1 440 tiivistelmää (45 kysymystä x 32 järjestelmää).

**Tulos**

Mitä tietokokonaisuutta he käyttävät?

**Esimerkki 5.944**

Sama BUTD-malli, joka saavutti 73,02 %:n tuloksen korjaamattomassa SNLI-VE-testijoukossa, saavuttaa 73,18 %:n tasapainotetun tarkkuuden, kun sitä testataan SNLI-VE-2.0:n korjatulla testijoukolla. Tämän mallin suorituskyvyssä ei siis ole merkittävää eroa. Tämä saattaa johtua satunnaisuudesta.

**Tulos**

Kuinka suuri on nykyisen mallin suorituskyvyn ero alkuperäisen ja korjatun korpuksen välillä?

**Esimerkki 5.945**

Kaksoisoppimisen ydinajatus on hyödyntää primaaritehtävän (kartoitus alueesta $\mathcal {X}$ alueeseen $\mathcal {Y}$) ja kaksoistehtävän (kartoitus alueesta $\mathcal {Y}$ alueeseen $\mathcal {X}$ ) välistä kaksoistavoitetta molempien tehtävien suorituskyvyn parantamiseksi. MADL BIBREF0 laajentaa kaksoisoppimisen BIBREF1- ja BIBREF2-kehystä ottamalla käyttöön useita primaari- ja duaalimalleja.

**Tulos**

Miten moniagenttinen kaksoisoppiminen toimii?

**Esimerkki 5.946**

Mallissamme otetaan käyttöön monikäänteinen päättelymekanismi, jolla käsitellään moninäkökulmaisia yhteensovitusominaisuuksia.

**Tulos**

Mitä yhteensovitusominaisuuksia ne käyttävät?

**Esimerkki 5.947**

syötteiden fuusiokerros, joka mahdollistaa vuorovaikutuksen syötetietojen ja uudenlaisen huomioperusteisen GRU:n välillä, joka mahdollistaa loogisen päättelyn järjestettyjen syötteiden perusteella.

**Tulos**

Miten malli kiertää tukevien faktojen puutteen koulutuksen aikana?

**Esimerkki 5.948**

Supertaggauksen osalta havaitsemme, että perusmallin ristiinentropian avulla koulutettu malli parantaa ennusteitaan palkkihakupohjaisen dekoodauksen avulla 2 tarkkuuspistettä ahneeseen dekoodaukseen verrattuna, mikä viittaa siihen, että palkkihaku on jo hyödyllinen tässä tehtävässä, vaikka hakutietoista koulutusta ei tarvittaisikaan.

**Tulos**

Kuinka paljon ne ovat parempia kuin ahneen ja ristiinentrooppisen säteen perusdekoodaus?

**Esimerkki 5.949**

Validoimme lähestymistapamme Wikipedian myrkyllisiä kommentteja koskevalla tietokokonaisuudella BIBREF18 .

**Tulos**

Mitä tietokokonaisuuksia ne käyttävät?

**Esimerkki 5.950**

Koulutusaineistona käytetään kahden monikielisen käännöskorpuksen sekoitusta: MultiUN BIBREF20 ja OpenSubtitles BIBREF21. MultiUN koostuu 463 406 virallisesta asiakirjasta kuudella kielellä ja sisältää noin 300 miljoonaa sanaa kullakin kielellä. OpenSubtitles on elokuvien ja television tekstityksistä koostuva korpus, joka sisältää 2,6 miljardia lausetta 60 kielellä. Valitsimme näistä kahdesta korpuksesta neljä yhteistä kieltä: englanti, espanja, venäjä ja kiina. Harjoituskorpuksen tilastot esitetään taulukossa TABREF14.

**Tulos**

Mitä monikielistä rinnakkaista aineistoa käytetään ehdotetun mallin kouluttamiseen?

**Esimerkki 5.951**

Lopuksi dekooderitehtävä, joka ennustaa kohdejonon todennäköisyyden ajankohtana INLINEFORM3 aikaisempien tulosteiden ja kontekstitietojen INLINEFORM4 perusteella, voidaan muotoilla seuraavasti: DISPLAYFORM0

**Tulos**

Miten heidän dekooderinsa tuottaa tekstiä?

**Esimerkki 5.952**

Käyttämällä joukon semanttisesti suuntautuneita tehtäviä, jotka edellyttävät eksplisiittisiä semanttisia ristiinkielisiä representaatioita, osoitimme, että mBERT-kontekstin upotukset eivät representoi samankaltaisia semanttisia ilmiöitä samalla tavalla, eivätkä ne siksi ole suoraan käyttökelpoisia ristiinkielisiin nollakohtaisiin tehtäviin.

**Tulos**

Millaisia haasteita tämä työ asettaa, jotka on ratkaistava, jotta voidaan rakentaa parempia kielineutraaleja esityksiä?

**Esimerkki 5.953**

Kuvassa FIGREF32 havainnollistetaan jäsennysverkon arvioima syntaktinen etäisyys, kun luetaan kolmea eri sekvenssiä PTB-testisarjasta. Havaitaan, että syntaktinen etäisyys on yleensä suurempi sanan viimeisen merkin ja välilyönnin välillä, joka on järkevä taitekohta sanojen erottamiseksi toisistaan. Malli havaitsi itsenäisesti välttämään sanojen välisen huomiokytkennän ja käyttämään välilyönnin (erotinmerkkien) piilotettuja tiloja tiivistämään aiempaa tietoa. Tämä on vahva todiste siitä, että malli pystyy ymmärtämään datan piilevän rakenteen.

**Tulos**

Miten he osoittavat mallinsa paljastavan taustalla olevan syntaktisen rakenteen?

**Esimerkki 5.954**

Luvussa "Materiaalit ja menetelmät" käsittelemme aineistomme ja menetelmämme, mukaan lukien tutkimamme aineisto, miten esikäsittelimme aineiston ja loimme "kausaalisen" korpuksen ja vastaavan "kontrolli"-korpuksen, sekä yksityiskohtaiset tiedot tilastollisista ja kielianalyysityökaluista, joilla tutkimme näitä korpuksia.

**Tulos**

Käytetäänkö korporaatioiden analysoinnissa asiantuntijoiden annotaatioita, joukkoistamista vai pelkästään automaattisia menetelmiä?

**Esimerkki 5.955**

Tämä algoritmi löytää optimaalisen lukumäärän minimoimalla kustannusfunktion, joka perustuu sanojen samankaltaisuusmatriisin ominaisvektorirakenteeseen. Tarkemmat yksityiskohdat löytyvät asiaa koskevasta kirjallisuudesta.Tulokset ja keskustelu

**Tulos**

Mitkä ovat kuusi kohdekieltä?

**Esimerkki 5.956**

Lausuma ketjutetaan erityisellä symbolilla, joka merkitsee sekvenssin loppua. Alustamme sanojen upotukset käyttämällä 300-ulotteista GloVe BIBREF30 -sanastoa ja hienosäädämme niitä harjoittelun aikana.

**Tulos**

Käyttävätkö ne esivalmisteltuja sanavektoreita dialogin kontekstin upottamiseen?

**Esimerkki 5.957**

FastTextBIBREF4: Se käyttää tekstin luokittelussa ominaisuuksina sanapusseja ja n-grammipusseja, jotka keräävät tehokkaasti osittaista tietoa paikallisesta sanajärjestyksestä. BiLSTM: Toisin kuin feedforward-neuraaliverkot, BiLSTM:n kaltaiset rekursiiviset neuroverkot käyttävät historiatietoon perustuvaa muistia oppiakseen pitkän matkan piirteitä ja ennustaakseen sitten tuloksen. Käytämme kaksikerroksista BiLSTM-arkkitehtuuria, jossa on GloVe-sanojen upotukset, vahvana RNN-perustasona. BERT BIBREF5: Se on kontekstisidonnainen sanojen esitysmalli, joka käyttää kaksisuuntaisia muuntajia ja joka on esivalmennettu suurella 3,3 miljardin dollarin sanakorpuksella. Käytämme $BERT\_{large}$-mallia, joka on hienosäädetty harjoitusaineistoon.

**Tulos**

Mikä on kokeiden lähtötaso?

**Esimerkki 5.958**

Yksi suosituimmista QG:n mittareista, BLEU BIBREF21 , tarjoaa joukon mittareita, joilla voidaan verrata automaattisesti luotuja tekstejä yhteen tai useampaan viitteeseen. Erityisesti BLEU-N perustuu ehdokkaan ja sitä vastaavan viitteen (vastaavien viitteiden) välisten päällekkäisten n-grammien lukumäärään. Sen vuoksi käytämme BIBREFissä33 alun perin ehdotettua Self-BLEU:ta luotujen tekstisekvenssien monimuotoisuuden mittarina. Näin ollen, kun kysymyksen ja kontekstin pari annetaan QA-mallin syötteeksi, voidaan laskea kahdenlaisia mittareita: n-grammiin perustuva pistemäärä: mitataan haetun vastauksen ja perustotuuden keskimääräistä päällekkäisyyttä.todennäköisyyspistemäärä: QA-mallin luottamus haettuun vastaukseen; tämä vastaa todennäköisyyttä, jolla QA-malli antaa oikean vastauksen haetulle vastaukselle.

**Tulos**

Mitä automatisoituja mittareita kirjoittajat tutkivat?

**Esimerkki 5.959**

On mielenkiintoista havaita, että perusmalli vahvistaa harjoitteluaineiston harhaa INLINEFORM0- ja INLINEFORM1-arvoilla mitattuna. Kuvattuja harhaisuusmittareita käyttävien mittausten perusteella menetelmämme lieventää tehokkaasti harhaisuutta kielen mallintamisessa ilman, että perpleksisyys lisääntyy merkittävästi.

**Tulos**

miten sukupuoleen perustuvien ennakkoluulojen lieventämistä arvioidaan?

**Esimerkki 5.960**

Taulukossa TABREF9 on yhteenveto keskimääräisistä luokittelutarkkuustuloksista.

**Tulos**

Mitä arviointimittaria käytetään?

**Esimerkki 5.961**

Analysoimme ensin koodien sekoittamissa sanaleikeissä esiintyviä rakennetyyppejä ja luokittelimme ne kahteen luokkaan, jotka ovat peräkkäiset ja sanansisäiset.

**Tulos**

Mihin luokkiin koodisekoitetut sanaleikit kuuluvat?

**Esimerkki 5.962**

Taulukossa TABREF14 esitetään ehdotetun s2sL-lähestymistavan tulokset verrattuna MLP:n tuloksiin puhe/musiikki ja neutraali/sad-luokittelutehtävissä, kun otetaan huomioon eri osuudet harjoitusdatasta.

**Tulos**

Mihin malleihin/kehyksiin niitä verrataan?

**Esimerkki 5.963**

Lisäksi ehdotamme tieliikennetutkimusta uudeksi mittariksi, jolla voidaan osoittaa mallin kestävyys altistumisharhoja vastaan.

**Tulos**

Mikä on tiekoemittari?

**Esimerkki 5.964**

Haasteeseen vastataan seuraavasti: kun annetaan luonnollisen kielen syötesekvenssi, joka kuvaa kohtausta, kuten osa tarinasta, joka on peräisin transkriptiosta, tavoitteena on päätellä, mikä toiminto tapahtuu todennäköisimmin seuraavaksi.

**Tulos**

Käsitelläänkö tätä kirjaimellisesti vain "ennustaa seuraava tekstissä esiintyvä loitsu"?

**Esimerkki 5.965**

Toisaalta MHD:n tapauksessa huomiotason integroinnin sijasta jokaiselle päähän määritetään useita dekoodereita ja integroidaan niiden tuotokset lopullisen tuotoksen saamiseksi.

**Tulos**

Laskeeko jokainen dekooderin tarkkaavaisuuspää saman ulostulon?

**Esimerkki 5.966**

Huomaamme, että ketjuttaminen on jatkuvasti parempi kuin gated-attention-mekanismi sekä harjoittelu- että testausohjeiden osalta. Epäilemme, että gated-attention on hyödyllinen skenaarioissa, joissa objekteja kuvataan useiden attribuuttien avulla, mutta sillä ei ole haitallista vaikutusta, kun kyseessä ovat järjestysliittimet.

**Tulos**

Miksi he päättelevät, että Gated-Attentionin käyttö ei tarjoa kilpailuetua ketjuttamiseen verrattuna tässä tilanteessa?

**Esimerkki 5.967**

Kuten näemme, kaikki CRU-mallimme variantit pystyvät antamaan huomattavia parannuksia perinteiseen GRU-malliin verrattuna, sillä kolmessa tietokokonaisuudessa voidaan havaita 2,7 %:n, 1,0 %:n ja 1,9 %:n enimmäishyöty.

**Tulos**

Osoittavatko koetulokset uuden lähestymistavan johdonmukaisen merkittävän parannuksen perinteisiin CNN- ja RNN-malleihin verrattuna?

**Esimerkki 5.968**

Tehtäviä on siis yhteensä 44 (22 INLINEFORM10 2). Lisäksi jokaisesta tehtävästä on 2 inhimillistä yhteenvetoa. Valitsimme kolme kilpailevaa järjestelmää (SumBasic, ILP ja ILP+MC), joten meillä on kolme järjestelmä-järjestelmä -paria (ILP+MC vs. ILP, ILP+MC vs. SumBasic ja ILP vs. SumBasic) kutakin tehtävää ja kutakin inhimillistä yhteenvetoa varten.

**Tulos**

Rakentavatko he yhden mallin kutakin aihetta vai kaikkia aiheita varten?

**Esimerkki 5.969**

Luomme kolme uutta harjoitustietokokonaisuutta lisäämällä alkuperäiseen Wikipedian korpukseen (jakso lm) lauseita, jotka sisältävät objektin RC-merkkejä. Tätä varten poimimme satunnaisesti 30 miljoonaa lausetta Wikipediasta (jotka eivät ole päällekkäisiä alkuperäisen korpuksen lauseiden kanssa), analysoimme ne samalla jäsentimellä ja suodatamme niistä 680 000 lausetta, jotka sisältävät objektin RC:n. Näin saamme 680 000 lausetta.

**Tulos**

Miten ne suorittavat tietojen lisäämisen?

**Esimerkki 5.970**

Jos otamme CBT:llä koulutetun parhaan psr-kokoonpanon perustasoksi, malliarkkitehtuurin parantaminen BIBREF6 , BIBREF7 , BIBREF8 , BIBREF9 , BIBREF10 , BIBREF11 -mallien mukaisesti ja alkuperäisen CBT-koulutusdatan käyttämistä jatkamalla johtaa INLINEFORM0- ja INLINEFORM1-luokkien absoluuttisiin parannuksiin nimettyjen entiteettien ja tavallisten substantiivien osalta. Sitä vastoin koulutustietokannan paisuttaminen paransi INLINEFORM2:n arvoa, kun käytettiin samaa mallia.

**Tulos**

Kuinka suuria ovat Attention-Sum Reader -mallin parannukset, kun käytetään BookTest-tietokokonaisuutta?

**Esimerkki 5.971**

Stanford - Twitter Sentiment Corpus (STS Corpus): STS-korpus sisältää 1 600 000 harjoitustwiittiä, jotka on kerätty BIBREF0-tietokannasta. BIBREF0 rakensi manuaalisesti testijoukon, jossa oli 177 negatiivista ja 182 positiivista twiittiä. Stanfordin testijoukko on pieni. Sitä on kuitenkin käytetty laajalti eri arviointitehtävissä BIBREF0 BIBREF5 BIBREF13 .Sanders - Twitter Sentiment Corpus: Tämä tietokokonaisuus koostuu käsin luokitelluista twiiteistä, jotka on kerätty hakusanojen avulla: INLINEFORM0 , #google, #microsoft ja #twitter. Rakennamme tietokokonaisuuden BIBREF14 binääriluokittelua varten.Terveydenhuollon uudistus (HCR): Tämä tietokokonaisuus rakennettiin indeksoimalla twiitit, jotka sisälsivät hashtagin #hcr BIBREF15 . Tehtävänä on ennustaa positiiviset/negatiiviset twiitit BIBREF14 .

**Tulos**

Mitä tietokokonaisuuksia he käyttivät?

**Tulos**

Mitä kolmea Twitterin tunnetilaluokitusdataa käytetään kokeiluissa?

**Esimerkki 5.972**

Toiseksi, twiittien poistaminen suoraan Twitterin verkkosivuilta. Toisen vaihtoehdon avulla ladattiin kiinnostavien osakkeiden päivittäiset twiitit vuodesta 2015 tammikuusta 2017 kesäkuuhun. Tästä syystä tutkitaan kahta saman toimialan yritystä, Teslaa ja Fordia, siitä, miten Twitter-tunnelmat voisivat vaikuttaa osakekurssiin. Kunkin twiitin kääntämiseksi sentimenttipisteytykseksi käytettiin Stanfordin coreNLP-ohjelmistoa.

**Tulos**

Mitä twiittejä käytetään päivittäisen sentimenttisignaalin tuottamiseen?

**Esimerkki 5.973**

Suuntaviivat uskollisuuden arvioimiseksi Ehdotamme seuraavia suuntaviivoja selitysten uskollisuuden arvioimiseksi. Näissä ohjeissa käsitellään kirjallisuudessa havaitsemiamme yleisiä sudenkuoppia ja epäoptimaalisia käytäntöjä.Guidelines for Evaluating Faithfulness ::: Ole selkeä sen suhteen, mitä arvioit.Uskottavuuden ja uskottavuuden sekoittaminen on haitallista. Sinun tulisi olla selvillä siitä, mitä niistä arvioit, ja käyttää sopivia menetelmiä kummallekin. Sama pätee tietysti myös tulkintatekniikoita suunniteltaessa - ole selvillä siitä, mitä ominaisuuksia priorisoidaan.Guidelines for Evaluating Faithfulness :::: Uskollisuuden arviointiin ei pitäisi sisältyä inhimillistä arviota tulkinnan laadusta. huomioimme, että: (1) ihmiset eivät voi arvioida, onko tulkinta uskollinen vai ei: jos he ymmärtäisivät mallin, tulkinta olisi tarpeetonta; (2) samankaltaisista syistä emme voi myöskään saada valvontaa tähän ongelmaan. Sen vuoksi ihmisen arvostelukykyä ei pitäisi ottaa mukaan uskollisuuden arviointiin, koska ihmisen arvostelukyky mittaa uskottavuutta. guidelines for Evaluating Faithfulness ::: Meidän pitäisi pystyä tulkitsemaan mallin virheellisiä ennusteita aivan samoin kuin oikeita ennusteita. Arviointimenetelmiin, jotka perustuvat kultaisiin merkintöihin, vaikuttavat ihmisen ennakko-odotukset siitä, mitä mallin pitäisi tehdä, ja ne taas työntävät arviointia uskottavuuden suuntaan.Guidelines for Evaluating Faithfulness ::: Älä luota "luontaista tulkittavuutta" koskeviin väitteisiin.Luontainen tulkittavuus on väite, kunnes toisin todistetaan. "Luontaisesti tulkittavissa olevien" mallien antamiin selityksiin on sovellettava samoja standardeja kuin post hoc -tulkintamenetelmiin, ja niiden uskottavuutta on arvioitava käyttämällä samoja arviointitekniikoita.Guidelines for Evaluating Faithfulness ::: Lopputehtävän käyttäjän suorituskyky HCI-ympäristöissä on vain osoitus uskottavuuden ja mallin suorituskyvyn välisestä korrelaatiosta, vaikka tämä korrelaatio olisi kuinka pieni tahansa. Vaikka se on tärkeää tulkintojen hyödyllisyyden arvioimiseksi joissakin käyttötapauksissa, se ei liity uskollisuuteen.

**Tulos**

Mitkä ovat keskeisiä kohtia uskollisuuden arviointia koskevissa suuntaviivoissa?

**Esimerkki 5.974**

Klusterointi suoritettiin erikseen kunkin lääkärin erikoisalan osalta. Lisäksi tarkasteltiin lääkärien tunnusten jakautumista saatuihin klustereihin. Kävi ilmi, että jotkin klusterit kattoivat lähes täsmälleen yhden lääkärin kirjoittamat kuvaukset. Näin kävi erikoisaloilla, joilla klusterit erottuvat toisistaan suurilla marginaaleilla (esim. psykiatria, pediatria, kardiologia).

**Tulos**

Tutkivatko he tekstien samankaltaisuutta eri lääkäreiden välillä?

**Esimerkki 5.975**

Kokeet ::: Vertailua varten valitsemme useita julkisia malleja, mukaan lukien semanttisen jäsennyksen mallit: BiDAF BIBREF3, MRC-malli, joka käyttää kaksisuuntaista huomiovirtaverkkoa kysymyksen ja tekstin koodaamiseen; QANet BIBREF12, joka käyttää konvoluutioita ja itsehuomautuksia koodaajien rakennuspalikoina kysymyksen ja tekstin esittämiseen; BERT BIBREF23, valmiiksi harjoittelemamme kaksisuuntainen Transformer-pohjainen kielimalli, joka on viime aikoina saavuttanut huippusuorituskyvyn monissa julkisissa MRC-tietoaineistoissa; ja numeerisia MRC-malleja: NAQANet+, itse toteuttamamme parannettu versio NAQANetista, jossa otetaan lisäksi huomioon reaaliluku (esim. "2,5"), monipuolisempi aritmeettinen ilmaisu, datan lisääminen jne. Syn Dep BIBREF6, neuraalinen semanttinen jäsennysmalli (KDG) BIBREF22 Stanfordin riippuvuuksiin perustuvilla lause-esityksillä;OpenIE BIBREF6, KDG avoimeen tiedonlouhintaan perustuvilla lause-esityksillä;SRL BIBREF6, KDG semanttiseen roolimerkintään perustuvilla lause-esityksillä;ja perinteiset MRC-mallit:

**Tulos**

mihin nykyisiin malleihin niitä verrataan?

**Esimerkki 5.976**

On myös huomattava, että käytännön ongelmien vuoksi tämän työn tavoitteena on luoda reseptiehdokkaita lääkemääräysmenettelyn helpottamiseksi sen sijaan, että korvattaisiin kokonaan ihmislääkärit.

**Tulos**

Miksi he pitivät tätä hyvänä ajatuksena?

**Esimerkki 5.977**

BIBREF4 sisältää monenlaista uutisiin liittyvää tietoa, kuten kuvia, kuvatekstejä, maantieteellisiä sijaintitietoja ja kommentteja, joita voidaan käyttää artikkelin suosion mittarina. Tämän tietokokonaisuuden artikkelit kerättiin tammikuun ja joulukuun 2014 välisenä aikana. Vaikka yritimme hakea koko tietokokonaisuuden, pystyimme lataamaan vain 38 182 artikkelia, koska tietokokonaisuudessa julkaistiin kuolleita linkkejä. Haetut artikkelit julkaistiin tärkeimmissä uutiskanavissa, kuten Yahoo News, The Guardian tai The Washington Post. NowThisNews-tietokannan osalta normalisoimme tiedot ryhmittelemällä artikkelit kustantajakohtaisesti ja luokittelemalla ne suosituiksi, kun kommenttien määrä ylittää kyseisen kustantajan mediaaniarvon.

**Tulos**

Mikä on uutisartikkelien lähde?

**Esimerkki 5.978**

Käytämme hyperopt-kirjastoa hyperparametrien valitsemiseen seuraavilla arvoilla: LSTM-kerroksen koko (16, 32, 64), pudotusarvo ($0.0-0.9$), aktivointifunktio ($relu$, $selu$, $tanh$), optimoija ($sgd$, $adam$, $rmsprop$), oppimisnopeuden (1e-1,...,-5) ja erän koon (4, 8, 16) vaihtelulla.

**Tulos**

Mitä aktivointifunktiota he käyttävät mallissaan?

**Esimerkki 5.979**

Käytämme BIBREF6 aristo2016:combiningin tekstikorporaatiota (S) BIBREF6 aristo2016:combiningin tuple KB:n rakentamiseen. Tarkemmin sanottuna käytämme jokaisen Q\_\mathit {tr}$:n monivalintakysymyksen $(q,A) \ ja jokaisen A$:n vaihtoehdon $a \ osalta kaikkia $q$:n ja $a$:n sisältämiä muita kuin pysähtymättömiä sanamerkkejä ElasticSearch-kyselynä S:lle. Otamme 200 parasta osumaa, suoritamme Open IE v4:n ja yhdistämme tuloksena saadut tuplat kaikkien $a \in A$:n ja kaikkien Q\_\mathit {tr}$:n sisältämien kysymysten osalta luodaksemme tuple KB (T). Lause-korpus (S) koostuu 80 000 lauseesta ja 280 GB tavallisesta tekstistä, jotka on poimittu BIBREF6 aristo2016:combiningin käyttämiltä verkkosivuilta.

**Tulos**

Mistä korpuksesta OpenIE-uutteet on otettu?

**Esimerkki 5.980**

Käytämme word2vec BIBREF0:ta sanojen upotusten kouluttamiseen.

**Tulos**

Millaisen mallin he rakentavat lyhenteiden laajentamiseksi?

**Esimerkki 5.981**

Päästä päähän -menetelmämme lukee tekstiä ja kirjoittaa molempiin muistipaikkoihin ja niiden välisiin reunoihin. Intuitiivisesti muistipaikat vastaavat entiteettejä ja reunat entiteettien välisiä suhteita, jotka kukin esitetään vektorina.

**Tulos**

Miten tieto palautuu muistiin?

**Esimerkki 5.982**

Semanttinen haku: Käsittelimme neuraalista semanttista hakua sekä kappale- että lausetasolla binäärisenä luokitusongelmana, jossa mallien parametrit päivitetään minimoimalla binäärinen ristiinentropiahäviö. Tarkemmin sanottuna syötimme kyselyn ja kontekstin BERT:iin seuraavasti: Sovelsimme affiinikerrosta ja sigmoidiaktivaatiota viimeisen kerroksen ulostuloon [$\mathit {CLS}$]-merkkiä, joka on skalaarinen arvo. Parametrit päivitettiin tavoitefunktiolla: jossa $\hat{p}\_i$ on mallin ulostulo, $\mathbf {T}^{p/s}\_{pos}$ on positiivinen joukko ja $\mathbf {T}^{p/s}\_{neg}$ on negatiivinen joukko. Kuten kuvasta FIGREF2 käy ilmi, lausetasolla positiivisina esimerkkeinä käytettiin perustotuuden lauseita, kun taas negatiivisina esimerkkeinä käytettiin muita lauseita, jotka oli haettu ylävirran joukosta. Vastaavasti kappaletasolla positiivisina esimerkkeinä käytettiin kappaleita, joissa oli jokin perustotuuslause, ja negatiivisina esimerkkeinä käytettiin muita kappaleita, jotka olivat peräisin termipohjaisista hakuprosesseista.

**Tulos**

Miten he kouluttavat hakumoduulit?

**Esimerkki 5.983**

Teemme myös hienojakoisen syntaktisen analyysin, joka osoittaa, miten menetelmämme avulla voidaan hallita ensimmäisen ja toisen persoonan pronominien sekä niihin liittyvien verbien ja adjektiivien morfologista toteutumista.

**Tulos**

Mitä johtopäätöksiä syntaktisesta analyysistä tehdään?

**Esimerkki 5.984**

Kehyksemme sisältää kolme keskeistä moduulia: Retrieve, Fast Rerank ja BiSET. Retrieve pyrkii palauttamaan jokaiselle lähdeartikkelille muutamia ehdokkaita malleiksi harjoituskorpuksesta. Tämän jälkeen Fast Rerank -moduuli tunnistaa nopeasti parhaan mallin ehdokkaiden joukosta. Lopuksi BiSET valitsee vastavuoroisesti tärkeät tiedot lähdeartikkelista ja mallista luodakseen parannetun artikkeliesityksen tiivistämistä varten. Tämä moduuli alkaa tavallisella tiedonhakukirjastolla, jolla haetaan pieni joukko ehdokkaita hienojakoista suodatusta varten cao2018retrievenä. Tätä varten kaikki muut kuin alfabeettiset merkit (esim. päivämäärät) poistetaan, jotta ne eivät vaikuttaisi artikkelien vastaavuuteen. Hakuprosessi aloitetaan tekemällä lähdeartikkelin avulla kysely harjoituskorpuksesta, jotta löydetään muutama (5-30) aiheeseen liittyvä artikkeli, joiden tiivistelmiä käsitellään ehdokasmalleina. Edellä kuvattu hakuprosessi perustuu lähinnä pinnalliseen sanojen yhteensovittamiseen, eikä sillä voida mitata kahden artikkelin välistä syvää semanttista suhdetta. Sen vuoksi kehitetään Fast Rerank -moduuli, joka tunnistaa ehdokkaista parhaan mallin sen perusteella, mikä on niiden syvällinen semanttinen yhteys lähdeartikkeliin. Mallina pidetään ehdokasta, jolla on suurin relevanssi. Kuten kuvassa FIGREF6 on esitetty, tämä moduuli koostuu konvoluutiokooderilohkosta, samankaltaisuusmatriisista ja poolointikerroksesta.

**Tulos**

Miten mallit löydetään harjoitusdatasta?

**Esimerkki 5.985**

Aloitimme vastaamalla aina KYLLÄ (erissä 2 ja 3) saadaksemme perustason suorituskyvyn. Erässä 4 käytimme päättelyä.

**Tulos**

Mikä oli perusmalli?

**Esimerkki 5.986**

Huomattakoon, että tässä tarkasteltavat kysymyksiin vastaamisen tehtävät ovat multimodaalisia, sillä vaikka konteksti on menettelytapatekstiä, kysymys ja monivalintatehtävät koostuvat kuvista.

**Tulos**

Mitä multimodaalisuutta aineistossa on käytettävissä?

**Esimerkki 5.987**

Tulevaisuudessa aiomme kiinnittää huomiota koodisekoitettujen tekstien yleistettyyn kielimalliin, jolla voidaan käsitellä myös hindi-koodisekoitettuja ja muita monikielisiä koodisekoitettuja tietokokonaisuuksia (eli pyrimme vähentämään riippuvuutta kielikohtaisista koodisekoitetuista resursseista). Järjestelmämme päihittää kaikki aiemmat aggressioiden tunnistamiseen käytetyt uusimmat lähestymistavat englanninkielisessä koodisekoitetussa TRAC-aineistossa, ja koska järjestelmä on koulutettu vain Facebook-kommenttien perusteella, se päihittää muut lähestymistavat myös Twitter-testijoukossa. Aggressiivisuuden/aggression tunnistamisen hienojakoinen määritelmä on TRAC-2018 BIBREF0, BIBREF2 -tapahtuman järjestäjien toimittama. He ovat luokitelleet aggressiivisuuden kolmeen merkintään (Overtly aggressive(OAG), Covertly aggressive(CAG), Non-aggressive(NAG)). Yksityiskohtainen kuvaus kustakin kolmesta leimasta on kuvattu seuraavasti: Overtly Aggressive(OAG) - Tämäntyyppinen aggressiivisuus osoittaa suoraa sanallista hyökkäystä, joka osoittaa tiettyyn yksilöön tai ryhmään. Esimerkiksi: "Hyvin sanottu sonu..sinulla on rohkeutta vastustaa muslimien dadagiria".Covertly Aggressive(CAG) - Tässä aggressiotyypissä hyökkäys ei ole suora vaan piilotettu, hienovarainen ja epäsuorempi, vaikka se useimmiten ilmaistaan kohteliaasti. Esimerkiksi: "Rakas Intia, lakkaa leikkimästä kansasi tunteilla äänten saamiseksi." Non-Aggressive(NAG) - Yleensä tämäntyyppisissä teksteissä ei ole minkäänlaista aggressiota, vaan niitä käytetään pääasiassa tosiasioiden toteamiseen, tilaisuuksien toivomiseen, kohteliaisuuteen ja tukemiseen.

**Tulos**

Mihin TRAC-tietokannassa on sekoitettu englantia?

**Esimerkki 5.988**

Tulevassa työssä aiomme myös lisätä tämäntyyppisiä tutkimuksia ERP-ennusteisiin.Keskustelu

**Tulos**

Mitä tietokokonaisuuksia käytetään?

**Tulos**

Mitkä kaksi kirjallisuudessa esiintyvää toiminnanohjausjärjestelmäparia hyötyvät yhteisestä harjoittelusta?

**Esimerkki 5.989**

Julkaisimme Redditissä kyselyn, jossa pyysimme tanskankielisiä käyttäjiä ehdottamaan loukkaavia, seksistisiä ja rasistisia termejä sanastoa varten. Kieli ja käyttäjäkäyttäytyminen vaihtelevat eri alustoilla, joten tavoitteena on kerätä alustakohtaisia termejä. Näin saatiin 113 loukkaavaa ja vihamielistä termiä, joita käytettiin loukkaavien kommenttien löytämiseen. Loput korpuksen kommenteista sekoitettiin, ja osajoukkoa tästä korpuksesta käytettiin lopun lopullisen tietokokonaisuuden täyttämiseen. Tuloksena saatu tietokokonaisuus sisältää 3600 käyttäjien tuottamaa kommenttia, joista 800 Ekstra Bladetista Facebookissa, 1400 r/DANMAGista ja 1400 r/Denmarkista.

**Tulos**

Kuinka suuri tanskalaisten kommenttien aineisto oli?

**Esimerkki 5.990**

Korektorin jälkeen COSTRA 1.0 -tietokannan muodostavat 4262 yksilöllistä lausetta (mukaan lukien 150 siemenlausetta).

**Tulos**

Kuinka monta lausemuunnosta on keskimäärin käytettävissä jokaista aineiston yksittäistä lausetta kohti?

**Esimerkki 5.991**

Ensimmäinen haaste on hankkia erittäin suuri vietnaminkielinen korpus ja käyttää sitä luokittelijan rakentamiseen, mikä voisi parantaa tarkkuutta entisestään. Toinen haaste on vietnamilaisten asiakirjojen suuren tietovaraston ja analyyttisen kehyksen suunnittelu ja kehittäminen, mikä vastaa Web 2.0 -sovelluksista, kuten Facebookista, Twitteristä ja niin edelleen, peräisin olevien artikkelien ja/tai asiakirjojen jättimäisen määrän nopeaa ja jatkuvaa kasvua. Viimeinen haaste liittyy sellaisen järjestelmän rakentamiseen, joka pystyy oppimaan asteittain uusia korporaatioita ja käsittelemään palautetta vuorovaikutteisesti.

**Tulos**

Miksi sanojen segmentointi vietnamin kielessä aiheuttaa haasteita?

**Esimerkki 5.992**

Taulukon TABREF13 tuloksissa on kymmenen koetinta. ELMo-transformerin ja mSynC:n suorituskyky on jälleen samankaltainen, mutta mSynC suoriutuu hieman huonommin seitsemästä yhdeksästä tehtävästä.

**Tulos**

Kuinka monessa koetehtävässä matala-syntaksitietoinen kontekstisidonnainen sulautus toimii paremmin kuin ELMon sulautus?

**Esimerkki 5.993**

Perusdekooderimme on vakiomuotoinen säteenhakupdekooderi BIBREF5 , johon on tehty useita suoraviivaisia suorituskykyoptimointeja.

**Tulos**

Mitä perusviivapurkulaitetta he käyttävät?

**Esimerkki 5.994**

Aineistossamme on 3 209 arviota 553 eri autosta 49 eri autonvalmistajalta.

**Tulos**

Kuinka suuri on autokielen kielitietokanta?

**Esimerkki 5.995**

Sen jälkeen pyydämme erillistä turkkilaisryhmää arvioimaan tarinoiden yleistä laatua ja kolmea parannusaluetta. Kaikki arvosanat annetaan viisiportaisella asteikolla. Keräämme kaksi arviota kutakin juttua kohden ja hylkäämme arviot, jotka eroavat toisistaan yli kaksi pistettä. Kaikkiaan 11 prosenttia arvioinneista hylättiin, joten analysoitavaksi jäi neljä mittaria 241 jutun osalta.

**Tulos**

Miten he arvioivat tuotettuja tarinoita?

**Esimerkki 5.996**

Ensimmäinen luokittelumalli toimii suodattimena luokittelun toisessa vaiheessa. Käytämme sekä SVM:ää että NB:tä tulosten vertailuun ja valitsemme SVM:n myöhemmin ensimmäisen vaiheen luokittelumalliksi paremman F-tuloksen vuoksi. Koulutus suoritetaan tweeteillä, jotka on merkitty luokilla , ja jotka perustuvat unigrammeihin ominaisuuksina. Luomme twiitissä olevista merkkijonoista sanavektorit käyttämällä WEKA API BIBREF9 -ohjelmassa saatavilla olevaa suodatinta ja suoritamme ristiinvalidoinnin tavanomaisilla luokittelutekniikoilla.

**Tulos**

Mitä luokittelijaa käytetään hätätilanteiden havaitsemiseen?

**Esimerkki 5.997**

Niissä tapauksissa, joissa mallin mielestä kaikkia historiallisia näytteitä olisi pidettävä yhtä tärkeinä peräkkäisessä analyysitehtävässä, meidän on etsittävä muualta laskennallisesti edullinen keino ymmärtää, mitä pysäytyskohdassa tapahtui.

**Tulos**

Voidaanko heidän tapaansa luoda informatiivisempaa visuaalista kuvaa soveltaa muihinkin tehtäviin kuin keskustelujen vuorotteluun?

**Esimerkki 5.998**

Lisäksi niitä käsitellään niiden meluisan luonteen vuoksi käyttämällä joitakin Twitter-kohtaisia tekniikoita, kuten URL-osoitteiden, käyttäjien mainintojen, hashtagien ja hymiöiden korvaamista/poistamista, oikeinkirjoituksen korjaamista, pidennysten normalisointia, lyhenteiden etsimistä, välimerkkien poistamista, vahvistimien ja vähentäjien havaitsemista, negaatioiden tunnistamista jne. Kyseinen kieli osoittautui varsin haastavaksi luovan oikeinkirjoituksen ja välimerkkien käytön, kirjoitusvirheiden, slangin, uusien sanojen, URL-osoitteiden sekä genrekohtaisen terminologian ja lyhenteiden, esim. re-tweetin RT ja #hashtagien, käytön vuoksi. Genre-eron lisäksi myös pituusero: sosiaalisen median viestit ovat yleensä lyhyitä, usein pituudeltaan rajattuja, kuten Twitterissä, eli pikemminkin lause tai otsikko kuin kokonainen dokumentti.

**Tulos**

Millaisia vaikeuksia Twitterin tunteiden analysoinnissa on verrattuna tunteiden analysointiin muilla aloilla?

**Esimerkki 5.999**

jiant tarjoaa tukea huippuluokan lauseenkoodausmalleille, mukaan lukien tuki Huggingface Transformersille. Tuettuja malleja ovat mm: BERT BIBREF17, RoBERTa BIBREF27, XLNet BIBREF28, XLM BIBREF29, GPT BIBREF30, GPT-2 BIBREF31, ALBERT BIBREF32 ja ELMo BIBREF33. jiant tukee myös (kaksisuuntaisten) LSTM-mallien BIBREF34 ja syvien sanasäkkimallien BIBREF35 sekä syntaksitietoisten mallien, kuten PRPN BIBREF36 ja ON-LSTM BIBREF37, koulutusta alusta alkaen. jiant tukee myös sanojen upotuksia, kuten GloVe BIBREF38.

**Tulos**

Onko jiant yhteensopiva minkä tahansa ohjelmointikielen mallien kanssa?

**Esimerkki 5.1000**

Tässä jaksossa esittelemme kaksi kokeissamme käytettyä tietokokonaisuutta: Se sisältää 4090 viestiä ja niihin liittyviä videoita NowThisNews Facebook-sivulta, jotka on kerätty 07/2015 ja 07/2016 välisenä aikana. Kunkin viestin osalta keräsimme sen otsikon ja siihen liittyvän videon katselukertojen määrän, jota pidämme suosion mittarina. Melko pitkän tiedonkeruuprosessin vuoksi päätimme normalisoida aineistomme ryhmittelemällä ensin viestit niiden julkaisukuukauden mukaan ja merkitsemällä sitten viestit, joiden suosiometriikka ylittää kuukausittaisen mediaaniarvon, suosituiksi ja loput epäsuosituiksi.

**Tulos**

Mistä he saavat uutisvideot?

**Esimerkki 5.1001**

Käytämme luokittelukokeissamme BIBREF4:n käyttöön ottamaa MH17 Twitter -tietokokonaisuutta, joka on kerätty tutkimaan MH17-lento-onnettomuutta koskevan (epä)tiedon kulkua Twitterissä. Se sisältää avainsanahaun perusteella kerättyjä twiittejä, jotka lähetettiin 17. heinäkuuta 2014 (lento-onnettomuuspäivä) ja 9. joulukuuta 2016 välisenä aikana.BIBREF4 tarjoaa annotaatioita osajoukolle tietokannan sisältämiä englanninkielisiä twiittejä.

**Tulos**

Mitkä kielet sisältyvät tietokantaan?

**Esimerkki 5.1002**

Käytämme Waseemin ja Hovyn BIBREF5-, Davidsonin ym. BIBREF9- ja Waseemin ym. BIBREF10-malleja perustasona ja vertailemme tuloksia eri hienosäätöstrategioillamme käyttäen esivalmistettua BERTbase-mallia.

**Tulos**

Mihin olemassa oleviin lähestymistapoihin niitä verrataan?

**Tulos**

Mitä perustasoa käytetään?

**Esimerkki 5.1003**

Luokkiin II ja III kuuluvia virheitä on vaikeampi välttää, ja niitä voitaisiin parantaa käyttämällä päättelyä BIBREF36 tai ironian havaitsemismenetelmiä BIBREF37.

**Tulos**

Mitä suosituksia annetaan suorituskyvyn parantamiseksi tulevaisuudessa?

**Esimerkki 5.1004**

Ihmisten merkitsemien harjoitusaineistojen luominen koneoppimista varten muistuttaa usein sisällönanalyysia, joka on vakiintunut menetelmä humanistisissa ja yhteiskuntatieteissä (erityisesti kirjallisuudessa, viestinnän tutkimuksessa ja kielitieteissä), mutta josta on myös versioita, joita käytetään biotieteissä, ekologiassa ja lääketieteessä. Sisällönanalyysi on saanut viime vuosisadan aikana monia muotoja, positivistisemmista menetelmistä, joissa vahvistetaan muodollisesti rakenteellisia tapoja arvioida sisältöä, tulkinnallisempiin menetelmiin, joissa hyväksytään monitulkintaisuus ja moninaiset tulkinnat, kuten grounded theory BIBREF16. Nykyään strukturoitua sisällönanalyysia (jota kutsutaan myös "suljetuksi koodaukseksi") käytetään kaikenlaisten laadullisten tai jäsentymättömien tietojen muuttamiseen strukturoiduiksi ja/tai määrällisiksi tiedoiksi, mukaan lukien mediatekstit, vapaamuotoiset kyselyvastaukset, haastattelupöytäkirjat ja videotallenteet. Hankkeisiin osallistuu yleensä koodaajien (joita kutsutaan myös annotaattoreiksi, merkitsijöiksi tai tarkastajiksi) ryhmiä, joissa tarvitaan ihmistyövoimaa koodaamaan, merkitsemään tai merkitsemään aineistokokonaisuuksia.

**Tulos**

Missä mielessä tietojen annotointi muistuttaa strukturoitua sisällönanalyysia?

**Esimerkki 5.1005**

Testataksemme, onko horisontaalinen vai vertikaalinen johdonmukaisuus hyödyllisempi, harjoittelemme ja arvioimme M-Bertiä tietokokonaisuuden muunnoksella, jossa kaikki kysymykset ovat englanninkielisiä.

**Tulos**

Suoritettiinko kieltenvälinen vs. yksikielinen arviointi?

**Esimerkki 5.1006**

Tietokannassamme tallennamme täydellisen twiittiobjektin, joka sisältää twiitin tunnisteen, käyttäjänimen, hashtagit ja twiitin maantieteellisen sijainnin. Loimme luettelon yleisimmistä COVID-19:ään liittyvistä arabian kielen avainsanoista. Twitterin streaming API:n avulla etsimme kaikki twiitit, jotka sisälsivät avainsanan (avainsanat) twiitin tekstissä. Taulukossa TABREF1 on luettelo käytetyistä avainsanoista sekä kunkin avainsanan seurannan alkamispäivä. Lisäksi taulukossa TABREF2 on luettelo seuraamistamme hashtageista sekä kustakin hashtagista kerättyjen twiittien määrä.

**Tulos**

Mitä lisätietoa aineistosta löytyy?

**Esimerkki 5.1007**

Ominaisuudet. Väitämme, että erityyppiset ominaisuudet, kuten tekstin tunnetila, moraali ja muut tekstipohjaiset ominaisuudet, ovat kriittisiä tunnistettaessa Twitter-tilejä, jotka eivät ole tosiseikkoja, hyödyntämällä niiden esiintymistä uutisten raportoinnin aikana tilin aikajanalla. Käytämme runsaasti ominaisuuksia, jotka on lainattu aiemmista väärennettyjen uutisten, puolueellisuuden ja huhujen havaitsemiseen liittyvistä töistä BIBREF0, BIBREF1, BIBREF8, BIBREF9.[leftmargin=4mm]Tunne: Muodostamme tunnevektorin käyttämällä 15 tunnetyypin sanaesiintymiä kahdesta saatavilla olevasta tunnelexikoniasta. Käytämme NRC:n sanakirjaa BIBREF10, joka sisältää $\sim $14K sanoja, jotka on merkitty käyttämällä Plutchikin kahdeksaa tunnetta BIBREF11. Toinen leksikko on SentiSense BIBREF12, joka on käsitepohjainen affektiivinen leksikko, joka liittää tunnemerkityksiä WordNet-tietokannan käsitteisiin. Siinä on $\sim 5,5 dollaria sanoja, jotka on merkitty tunteilla 14 tunnekategorian joukosta Käytämme kategorioita, joita ei ole NRC-leksikossa.Sentiment: Poimimme twiittien sentimentin käyttämällä EffectWordNet BIBREF13, SenticNet BIBREF14, NRC BIBREF10 ja subj\_lexicon BIBREF15, joissa kussakin on kaksi sentimenttiluokkaa, positiivinen ja negatiivinen: BIBREF16, jossa sanat merkitään johonkin seuraavista 10 luokasta (huolenpito, vahinko, oikeudenmukaisuus, huijaaminen, lojaalisuus, petos, auktoriteetti, kumouksellisuus, pyhyys ja alennustila).Style: Käytämme kanonisia tyylipiirteitä, kuten kysymysmerkkien, huutomerkkien, peräkkäisten merkkien ja kirjainten, linkkien, hashtagien ja käyttäjien mainintojen lukumäärää. Lisäksi poimimme isojen kirjainten osuuden ja twiitin pituuden.Words embeddings: Poimimme twiitin sanojen upotukset käyttämällä $Glove\-840B-300d$ BIBREF17 -esivalmennettua mallia. Twiitin lopullinen esitys saadaan keskiarvoistamalla sen sanojen upotukset.

**Tulos**

Mitä ominaisuuksia uutetaan?

**Esimerkki 5.1008**

Arviointitietokantana käytetään CoNLL-2012 shared task BIBREF21 -korpusta, joka on valittu Ontonotes 5.0 -tietokannasta.

**Tulos**

Millä tietokokonaisuudella he arvioivat malliaan?

**Esimerkki 5.1009**

Arvioidaksemme hypersphere-ominaisuutemme vaikutusta valmiisiin NER-järjestelmiin, suoritamme NE-tunnistuksen kahdella vakiomuotoisella NER-vertailuaineistolla, CoNLL2003 ja ONTONOTES 5.0.

**Tulos**

Arvioidaanko ne NER-tietoaineistoja?

**Esimerkki 5.1010**

Käytämme mandariinikiinan ja kantonin kielen tietoja. Kummankin kielen aineisto koostuu saman puhujan nauhoittamista puhutuista sanoista. Mandariinikiinan kielen aineisto on peräisin naispuoliselta puhujalta, ja sen on toimittanut Shtooka, ja kantonin kielen aineisto on peräisin miespuoliselta puhujalta, ja se on ladattu Forvo-verkkopalvelusta, joka on joukkorahoitteinen ääntämyssanakirja.

**Tulos**

Mitä tietokokonaisuutta käytetään harjoittelussa?

**Esimerkki 5.1011**

BERT: Toteutamme kaksi BERT-pohjaista peruslinjaa BIBREF51 todisteiden tunnistamista varten. Ensiksi koulutamme BERT:n jokaiselle kysely-käytäntö-lauseparille binäärisenä luokittelutehtävänä tunnistamaan, onko lause todiste kysymyksen kannalta vai ei (Bert). Kokeilemme myös kaksivaiheista luokittelua, jossa koulutamme mallin erikseen vain kysymyksille, jotta voimme ennustaa vastaamiskelpoisuuden. Jos päättelyhetkellä vastauskelpoinen luokittelija ennustaa, että kysymys on vastauskelpoinen, todisteiden tunnistamisen luokittelija tuottaa joukon ehdokaslauseita (Bert + vastauskelpoinen).

**Tulos**

Minkä tyyppistä neuraalista mallia käytettiin?

**Esimerkki 5.1012**

jiant System Overview ::: jiant ComponentsTasks: Tehtävät: Tehtävissä on viittauksia tehtävätietoihin, menetelmiä tietojen käsittelyyn, viittauksia luokittelupäätteisiin sekä menetelmiä suorituskykymittareiden laskemiseen ja ennusteiden tekemiseen.

**Tulos**

Sisältyykö jiant-ohjelmaan tietokokonaisuuksia 50 NLU-tehtävää varten?

**Esimerkki 5.1013**

Swissmetro-tietokanta koostuu tutkimustiedoista, jotka kerättiin Sveitsin St. Gallenin ja Geneven välisissä junissa maaliskuussa 1998.

**Tulos**

Mitä tietokokonaisuuksia käytetään arvioinnissa?

**Esimerkki 5.1014**

Harjoitusaineistomme koostuu 2,09 miljoonasta lauseparista, jotka on poimittu LDC-korpuksesta.

**Tulos**

Mistä sanasto tulee?

**Esimerkki 5.1015**

Tässä työssä arvioimamme aineisto on WMT English-French NewsTest2014, jossa on 380 miljoonaa sanaa rinnakkaista harjoitusdataa ja 3003 lauseen testijoukko. Validointiin käytetään NewsTest2013-joukkoa.

**Tulos**

Testaavatko he vain yhdellä tietokokonaisuudella?

**Esimerkki 5.1016**

Kokosimme kolme kiinankielistä tekstikorpusta verkkoaineistosta kolmelle alalle, nimittäin "hotelli", "matkapuhelin (mobiili)" ja "matkailu". Kaikki tekstit koskevat käyttäjien arvosteluja.

**Tulos**

Mitkä ovat tietolähteet?

**Esimerkki 5.1017**

Optimaalisen kontekstin aktiivisesti valitsevien tarkkaavaisuusmallien parempi suorituskyky verrattuna malliin, jonka kontekstina on koko säie (hLSTM), osoittaa, että kontekstin päättely parantaa intervention ennustamista verrattuna oletusarvoisen täydellisen kontekstin käyttöön.

**Tulos**

Mitkä keskustelun näkökohdat ovat huomiomekanismin mukaan merkityksellisiä ohjaajan intervention kannalta?

**Esimerkki 5.1018**

BIBREF16 on ottanut käyttöön BIBREF-tietokannan, jossa syötekysely on Wikipedia-artikkelin otsikon ja yhden sen jakson otsikon ketjutus. Relevantit kohdat ovat kyseisen osion sisältämiä kappaleita. Korpus koostuu kaikista englanninkielisen Wikipedian kappaleista lukuun ottamatta tiivistelmiä. Julkaistussa tietokokonaisuudessa on viisi ennalta määriteltyä taitosta, ja käytämme neljää ensimmäistä harjoitusjoukkona (noin 3 miljoonaa kyselyä) ja loput validointijoukkona (noin 700 000 kyselyä). Testijoukko on sama, jota käytettiin TREC-CAR 2017 -tapahtumaan lähetettyjen aineistojen arvioinnissa (noin 1800 kyselyä).

**Tulos**

Mikä on TREC-CAR-tietokanta?

**Esimerkki 5.1019**

Sentimenttianalyysissä käytetty korpus on BIBREF11:n elokuva-arvosteluja sisältävä IMDb-tietokanta, kun taas NER-korpus on BIBREF12:n Groningenin merkityspankki (GMB), joka sisältää 47 959 lausenäytettä.

**Tulos**

Mitä Named Entity Recognition -tietokokonaisuutta käytetään?

**Esimerkki 5.1020**

Viime aikoina WMT 2018 APE -yhteistehtävässä on esitetty useita muuntajaarkkitehtuurin mukautuksia monilähteistä APE:tä varten. pal-EtAl:2018:WMT ehdotti APE-mallia, jossa käytetään kolmea itsetarkkailuun perustuvaa koodaajaa. He esittelevät ylimääräisen yhteisen koodaajan, joka kiinnittää huomiota kahden koodatun sekvenssin yhdistelmään $mt$:stä ja $src$:stä. tebbifakhr-EtAl:2018:WMT, WMT 2018:n NMT-alatehtävän voittaja ($wmt18^{nmt}\_{best}$), käyttää sekvenssitason häviöfunktioita, jotta vältetään altistumisharhat harjoittelun aikana ja jotta se on johdonmukainen automaattisten arviointimetriikoiden kanssa. shin-lee:2018:WMT ehdottaa, että kullakin koodaajalla on oma itsehuomautus- ja feed-forward-kerroksensa jokaisen syötteen käsittelemiseksi erikseen. Dekooderin puolelle he lisäävät kaksi ylimääräistä monipäistä huomiokerrosta, yhden $src \rightarrow mt$:lle ja toisen $src \rightarrow pe$:lle. Tämän jälkeen toinen monipäähuomio näiden huomiokerrosten ulostulon välissä auttaa dekooderia ottamaan talteen $mt$:ssä olevat yhteiset sanat, joiden pitäisi pysyä $pe$:ssä. WMT 2018 -tapahtuman ($wmt18^{smt}\_{best}$) APE PBSMT-osatehtävän voittaja BIBREF11 esitteli myös toisen muuntajapohjaisen monilähteisen APE:n, jossa käytetään kahta koodaajaa ja pinotaan ylimääräinen ristikkäistarkkailukomponentti $src \rightarrow pe$:lle edellisen ristikkäistarkkailukomponentin päälle $mt \rightarrow pe$:lle. Kun shin-lee:2018:WMT:n lähestymistapaa verrataan voittajajärjestelmään, arkkitehtuurissa on vain kaksi eroa: i) $src \rightarrow mt$:n ja $src \rightarrow pe$:n ristiinhuomaamisjärjestys dekooderissa ja ii) $wmt18^{smt}\_{best}$ jakaa lisäksi parametreja kahden kooderin välillä. Äskettäin WMT 2018 APE shared -tehtävässä on esitetty useita muuntajaarkkitehtuurin mukautuksia monilähteistä APE:tä varten. pal-EtAl:2018:WMT ehdotti APE-mallia, jossa käytetään kolmea itsetarkkailuun perustuvaa kooderia. He esittelevät ylimääräisen yhteisen koodaajan, joka kiinnittää huomiota kahden koodatun sekvenssin yhdistelmään $mt$:stä ja $src$:stä. tebbifakhr-EtAl:2018:WMT, WMT 2018:n NMT-alatehtävän voittaja ($wmt18^{nmt}\_{best}$), käyttää sekvenssitason häviöfunktioita, jotta vältetään altistumisharhat harjoittelun aikana ja jotta se on johdonmukainen automaattisten arviointimetriikoiden kanssa. shin-lee:2018:WMT ehdottaa, että kullakin koodaajalla on oma itsehuomautus- ja feed-forward-kerroksensa jokaisen syötteen käsittelemiseksi erikseen. Dekooderin puolelle he lisäävät kaksi ylimääräistä monipäistä huomiokerrosta, yhden $src \rightarrow mt$:lle ja toisen $src \rightarrow pe$:lle. Tämän jälkeen toinen monipäähuomio näiden huomiokerrosten ulostulon välissä auttaa dekooderia ottamaan talteen $mt$:ssä olevat yhteiset sanat, joiden pitäisi pysyä $pe$:ssä. WMT 2018 -tapahtuman ($wmt18^{smt}\_{best}$) APE PBSMT-osatehtävän voittaja BIBREF11 esitteli myös toisen muuntajapohjaisen monilähteisen APE:n, jossa käytetään kahta koodaajaa ja pinotaan ylimääräinen ristikkäistarkkailukomponentti $src \rightarrow pe$:lle edellisen ristikkäistarkkailukomponentin päälle $mt \rightarrow pe$:lle. Kun shin-lee:2018:WMT:n lähestymistapaa verrataan voittajajärjestelmään, arkkitehtuurissa on vain kaksi eroa: i) $src \rightarrow mt$:n ja $src \rightarrow pe$:n ristiinhuomaamisjärjestys dekooderissa ja ii) $wmt18^{smt}\_{best}$ jakaa lisäksi parametreja kahden kooderin välillä.

**Tulos**

Mikä oli automaattisen jälkikäsittelyn aiempi nykytilamalli?

**Esimerkki 5.1021**

Arvioimme sukupolvia ROUGE BIBREF29- ja METEOR BIBREF30 -mittareilla käyttäen kohteina todellisia lauseita.

**Tulos**

Mitä mittareita käytetään arvioinnissa?

**Esimerkki 5.1022**

Käytämme useita standardimittareita, joita käytetään laajalti olemassa olevissa teoksissa mittaamaan vuoropuhelun luomismallien suorituskykyä, mukaan lukien BLEU BIBREF24, sulauttamiseen perustuvat mittarit (Average, Extrema, Greedy ja Coherence) BIBREF25, BIBREF26, entropiaan perustuvat mittarit (Ent-{1,2}) BIBREF0 ja erilliset mittarit (Dist-{1,2,3} ja Intra-{1,2,3}) BIBREF1, BIBREF6.

**Tulos**

Mitä automaattisia arviointimittareita käytetään?

**Esimerkki 5.1023**

Siksi käytämme mallin signaalin voimakkuuden mittarina oikean suhteen softmax-todennäköisyyden vähenemistä. Kutsumme tätä metriikkaa ${\Delta }\_s$ (delta-softmax), joka voidaan kirjoittaa seuraavasti: missä $rel$ on EDU-parin oikea relaatio, $t\_i$ edustaa merkkiä indeksillä $i$ $N$ merkistä ja $X\_{mask=i}$ edustaa syötesekvenssiä, jossa on maskeerattu kohta $i$ (kun $i \ in 1 \ldots N$ ilman erottimia tai $\phi $, tyhjä joukko).

**Tulos**

Miten delta-softmax lasketaan?

**Esimerkki 5.1024**

MielialaluokitusKokeilemme ensin mielialaluokitusta monitehtäväkokeella.Käytämme 16 eri tietokokonaisuutta useista suosituista BIBREFissä20 käytetyistä arvostelukorpuksista. Nämä tietokokonaisuudet koostuvat 14 tuotearvostelutietokokonaisuudesta ja kahdesta elokuva-arvostelutietokokonaisuudesta.Kaikki kunkin tehtävän tietokokonaisuudet jaetaan satunnaisesti koulutusjoukkoon, kehitysjoukkoon ja testausjoukkoon, joiden osuudet ovat 70 %, 10 % ja 20 %. Jaetun lauseen edustuksen siirrettävyysHuomiointimekanismin avulla ehdotetuissa malleissamme oleva jaetun lauseen koodaaja voi tuottaa yleisempiä tehtävävariantteja edustuksia, joita voidaan pitää valmiiksi varastoituna tietämyksenä ja käyttää sitten uusissa tehtävissä, joita ei ole vielä nähty. Jaksojen merkitseminen aputehtävänäHyvän lauseen esityksen tulisi sisältää sen kielellistä tietoa. Siksi sisällytämme sekvenssien merkitsemistehtävän (kuten POS-taggaus ja pilkkominen) aputehtäväksi monitehtäväoppimisjärjestelmään, joka koulutetaan yhdessä ensisijaisten tehtävien kanssa (edellä mainitut 16 tunteiden luokittelutehtävää).

**Tulos**

Mitä tehtäviä he kokeilivat?

**Esimerkki 5.1025**

Tässä asiakirjassa korostetaan ensin televisio- ja radiolähetysten merkitystä sivutoimisen tunnistamisen tietolähteenä ja niiden mahdollisia vaikutuksia. Sen jälkeen tehdään tilastollinen analyysi sukupuolen edustuksesta aineistossa, joka koostuu neljästä uusimmasta ranskankielisen lähetyksen korpuksesta, joita käytetään laajasti puheyhteisössä. Lopuksi pohdimme ASR-järjestelmän näkökulmasta, miten tällainen edustus vaikuttaa tähän aineistoon perustuviin järjestelmiin.

**Tulos**

Millaisia tehtäviä he käyttivät mies- ja naispuhujien suoritusten arvioinnissa?

**Esimerkki 5.1026**

Taulukossa TABREF21 esitetään korean INLINEFORM0-englannin tulokset käyttäen samoja kokoonpanoja (1, 2 ja 8) kuin saksanenglannin kohdalla. Tuloksemme vahvistavat, että soveltamamme tekniikat ovat menestyksekkäitä kaikissa tietokokonaisuuksissa, ja ne johtavat vahvempiin järjestelmiin kuin aiemmin on raportoitu tässä tietokokonaisuudessa, sillä ne saavuttavat 10,37 BLEU:ta verrattuna gu-EtAl:2018:EMNLP1:n raportoimaan 5,97 BLEU:hun.

**Tulos**

Mihin menetelmiin niitä verrataan koreanenglantilaisessa tietokokonaisuudessa?

**Tulos**

Mitkä olivat heidän kokeelliset tuloksensa vähäisten resurssien aineistossa?

**Esimerkki 5.1027**

Tunteiden luokittelutehtävässä kokeilimme malliamme seuraavilla julkisilla tietokokonaisuuksilla. MR Elokuva-arvostelut, joissa kussakin on yksi lause. IMDB Elokuva-arvostelut IMDB-sivustolta, jossa jokainen elokuva-arvostelu on merkitty binääriluokilla, joko positiivinen tai negatiivinen BIBREF19. SUBJ$^1$ Elokuva-arvostelu, joka on merkitty subjektiivisella tai objektiivisella BIBREF20. Testasimme CRU-malliamme myös cloze-tyylisessä luetun ymmärtämistehtävässä. Suoritimme kokeita julkisilla tietokokonaisuuksilla: CBT NE/CN BIBREF25.

**Tulos**

Mitä tietokokonaisuuksia käytetään tunnetilaluokituksen ja luetun ymmärtämisen testaamiseen?

**Esimerkki 5.1028**

Koulutimme ja arvioimme algoritmimme Microsoft COCO (MS-COCO) 2014 Captions -tietokannassa BIBREF21 . Raportoimme tulokset Karpathyn validointi- ja testijaoilla BIBREF8 , joita käytetään yleisesti muissa kuvatekstijulkaisuissa. Tietokokonaisuus sisältää 113 000 harjoituskuvaa, joihin kuhunkin kuvaan on liitetty 5 ihmisen kommentoimaa kuvatekstiä. Karpathyn testi- ja validointijoukot sisältävät kumpikin 5K kuvaa. Arvioimme mallejamme käyttämällä CIDEr-D BIBREF22 , SPICE BIBREF23 , BLEU BIBREF24 , METEOR BIBREF25 ja ROUGE-L BIBREF26 -mittareita. Vaikka kokeellisesti on osoitettu, että BLEU:lla ja ROUGE:lla on alhaisempi korrelaatio ihmisten tekemien arvioiden kanssa kuin muilla mittareilla BIBREF23 , BIBREF22 , yleinen käytäntö kuvatekstejä käsittelevässä kirjallisuudessa on ilmoittaa kaikki mainitut mittarit.

**Tulos**

Mitkä ovat yleiset tekstitysmittarit?

**Esimerkki 5.1029**

Open IE BIBREF7 -tehtävästä tehtiin ulkoinen arviointi.

**Tulos**

Käytetäänkö semanttista hierarkiaesitystä mihinkään tehtävään?

**Esimerkki 5.1030**

Arvioimme menetelmäämme kääntämällä englantia kolmelle morfologisesti rikkaalle kielelle, joilla kullakin on oma morfologinen typologiansa: Osoitamme, että mallimme pystyy saavuttamaan paremman käännöstarkkuuden ja yleistettävyyden kuin perinteiset avoimen sanaston NMT-menetelmät.

**Tulos**

Mitä kolmea kieltä tutkielmassa tarkastellaan?

**Esimerkki 5.1031**

Sosiaalisen median käyttö on viime vuosina kasvanut nopeasti. Ihmiset julkaisevat säännöllisesti päivittäisiä tapahtumia. BIBREF0 ehdottaa neljää tehtävää lääkkeiden nimien havaitsemiseksi, lääkkeiden käytön luokittelemiseksi, lääkkeiden haittavaikutusten luokittelemiseksi ja rokotuskäyttäytymisen havaitsemiseksi twiiteistä. Osallistuimme tehtäviin Task2 ja Task4.

**Tulos**

Arvioitiinko järjestelmä vain toisen jaetun tehtävän osalta?

**Esimerkki 5.1032**

Tässä luvussa kuvaamme lähestymistapaamme, jonka avulla luokittelemme käyttäjien reaktiot johonkin kahdeksasta diskurssityypistä: yhteisymmärrys, vastaus, arvostus, erimielisyys, tarkennus, huumori, kielteinen reaktio tai kysymys tai ei mikään annetuista merkinnöistä, jota kutsumme "muuksi", käyttämällä kielellisesti toimivia neuroverkkomalleja.

**Tulos**

Mitkä ovat yhdeksän tyyppiä?

**Esimerkki 5.1033**

Käytimme CNN-verkkoa (Convolutional Neural Network), jonka rakenne on samankaltainen kuin BIBREF29:n ehdottama.

**Tulos**

Mitä neuraalisia arkkitehtuureja käytetään?

**Esimerkki 5.1034**

Käytämme asiantuntijoiden antamia merkintöjä abstraktia varten, jos ne ovat olemassa, muuten käytämme yleisön antamia merkintöjä.

**Tulos**

Miten he sovittavat kommentoijat ja instanssit yhteen?

**Esimerkki 5.1035**

Koulutamme MaLOPaa kaikkien seitsemän kielen harjoitusjaksojen rinnakkaistiedostojen avulla.

**Tulos**

Kuinka monella kielellä tätä jäsentäjää on kokeiltu?

**Esimerkki 5.1036**

Harjoittelussa käytettiin Kaldi Babel -reseptiä, jossa käytettiin 198 tuntia dataa kymmenellä kielellä (bulgaria, tšekki, ranska, saksa, korea, puola, portugali, venäjä, thai, vietnam) GlobalPhone-ohjelmasta.

**Tulos**

Mitä kieliä otetaan huomioon?

**Esimerkki 5.1037**

TABREF5 näyttää sellaisten mallien tulostuskäyttäytymisen, joita vahingoitimme siten, että ne muistuttavat afasiaa aiheuttavaa vahinkoa. Koska kaikkien toimialueidemme tulostuskielet käyttävät monissa tapauksissa merkkejä merkitysten esittämiseen, on odotettavissa, että Wernicken alueen analogi on vastuussa korkean tarkkuuden ylläpitämisestä.

**Tulos**

Suorittavatko he kvantitatiivisen analyysin tietämyksen vääristymiä osoittavasta mallistaan?

**Esimerkki 5.1038**

Käytämme tunnetunnistukseen valmista työkalua (valmistajaa ei voida paljastaa lisenssirajoitusten vuoksi). Se antaa kehyskohtaiset pisteet ($\in [0;100]$) erillisille tunnetiloille, joita ovat ilo, viha ja pelko. Poimimme äänisignaalin samasta jaksosta kuin kasvonilmeiden kohdalla ja käytämme valmiiksi saatavilla olevaa tunnetunnistustyökalua. Ohjelmisto tuottaa yksittäiset luokituspisteet 24 erillisen tunnetilan joukolle koko lausahduksen osalta.

**Tulos**

Mitä tunteiden tunnistustyökaluja käytetään äänen ja kasvojen syöttämiseen?

**Esimerkki 5.1039**

Monitehtäväoppimisessa opitaan joukko toisiinsa liittyviä tehtäviä (esim. tunneaktivointi) sekä ensisijainen tehtävä (esim. tunnevalenssi); molemmat tehtävät jakavat osia verkon topologiasta ja siten ne koulutetaan yhdessä, kuten kuvassa KUVA4 on esitetty.

**Tulos**

Mitkä ovat tehtävät monitehtäväoppimisasetelmassa?

**Esimerkki 5.1040**

Menetelmät ::: BIBREF11:n innoittamana käytämme pituuskoodausta antaaksemme verkolle tietoa jäljellä olevasta lauseen pituudesta dekoodauksen aikana. Ehdotamme kahdenlaista pituuden koodausta: absoluuttista ja suhteellista. Olkoon pos ja len merkin sijainti ja jakson loppu, jotka molemmat ilmaistaan numeromerkkeinä. Tällöin absoluuttinen lähestymistapa koodaa jäljellä olevan pituuden:missä $i=1,\ldots ,d/2$.Vastaavasti suhteellinen ero koodaa suhteellisen sijainnin loppuun nähden. Tämä esitys saadaan yhdenmukaiseksi absoluuttisen koodauksen kanssa kvantisoimalla suhteellisten paikkojen avaruus äärelliseen $N$ kokonaislukujen joukkoon:missä $q\_N: [0, 1] \rightarrow \lbrace 0, 1, ..., N\rbrace $ määritellään yksinkertaisesti seuraavasti: $q\_N(x) = \lfloor {x \times N}\rfloor $. Koska olemme kiinnostuneita kohdesekvenssin merkin pituudesta, len ja pos annetaan merkkeinä, mutta esitämme sekvenssin BPE-segmentoitujen osasanojen sekvenssinä BIBREF17. Epäselvyyden ratkaisemiseksi len on merkkijonon pituus, kun taas pos on kaikkien edeltävien merkkien merkkimäärä.

**Tulos**

Miten ne rikastuttavat sijaintitietoa pituustiedolla?

**Esimerkki 5.1041**

Päätimme tutkia, miten CBT:tä voitaisiin vielä parantaa, testaamalla ihmisiä satunnaisella 50:llä nimetyn entiteetin ja 50:n yleisen substantiivin validointikysymyksen osajoukolla, joihin psr-kokoonpano ei pystynyt vastaamaan oikein.

**Tulos**

Miten ne osoittavat, että parannuksille on vielä tilaa?

**Esimerkki 5.1042**

Tämä menetelmä, josta käytämme nimitystä progressiiviset dynaamiset esteet (progressive dynamic hurdles, PDH), antaa jatkuvasti hyvin suoriutuville malleille mahdollisuuden harjoitella useampia vaiheita. Se alkaa tavallisena turnausvalinnan evoluutioarkkitehtuurihakuna, jossa on varhainen pysäytys, jolloin jokainen lapsimalli harjoittelee suhteellisen pienen $s\_0$ määrän askelia ennen kuin sen soveltuvuus arvioidaan. Kun ennalta määrätty määrä lapsimalleja, $m$ , on kuitenkin arvioitu, luodaan este, $h\_0$ , laskemalla nykyisen populaation keskimääräinen kunto. Seuraaville $m$ tuotetuille lapsimalleille, joiden kunto on suurempi kuin $h\_0$ $s\_0$ koulutusvaiheen jälkeen, annetaan $s\_1$ lisäkoulutusvaihetta, minkä jälkeen ne arvioidaan uudelleen lopullisen kuntonsa määrittämiseksi. Kun toiset $m$ mallia on tarkasteltu tällä tavalla, muodostetaan toinen este, $h\_1$ , laskemalla kaikkien niiden nykyisen populaation jäsenten keskimääräinen kunto, joita on koulutettu maksimimäärän askelia. Seuraavien $m$$ lapsimallien kohdalla koulutus ja arviointi jatkuu samalla tavalla, paitsi että malleille, joiden kunto on suurempi kuin $m$0 $m$1 koulutusaskeleen jälkeen, annetaan $m$2 lisää koulutusaskelia, ennen kuin niiden lopullinen kunto arvioidaan. Tätä prosessia toistetaan, kunnes saavutetaan tyydyttävä määrä maksimikoulutusaskeleita.

**Tulos**

Mikä on ehdotettu Progressive Dynamic Hurdles -menetelmä?

**Esimerkki 5.1043**

Lopuksi huomiomalli MDREA ylittää myös parhaat olemassa olevat tutkimustulokset (WAP 0,690-0,688) BIBREF19 .

**Tulos**

Kuinka paljon heidän mallinsa on parempi kuin uusimmat tulokset?

**Esimerkki 5.1044**

Neuroverkkomallimme koostuu pääasiassa kolmesta rinnakkaisesta LSTM BIBREF21 -kerroksesta.

**Tulos**

Millainen arkkitehtuuri neuroverkolla on?

**Esimerkki 5.1045**

Validoimme ehdotetun s2sL:n suorituskyvyn esittämällä alustavia tuloksia kahdesta eri tehtävästä, nimittäin puheen ja musiikin erottelusta ja tunteiden luokittelusta. Puhe- ja musiikkiluokittelutehtävää varten käytimme GTZAN Music-Speech -tietokokonaisuutta [17], joka koostuu 120 äänitiedostosta (60 puhetta ja 60 musiikkia). Kukin äänitiedosto (kestoltaan 2 sekuntia) esitetään 13-ulotteisella MFCC-vektorilla (mel-frequency cepstral coefficient), jossa kukin MFCC-vektori on kaikkien kehystason (kehyksen koko 30 sekuntia ja päällekkäisyys 10 sekuntia) MFCC-vektoreiden keskiarvo. On huomattava, että päätavoitteemme tässä tehtävässä ei ole parempi ominaisuuksien valinta vaan lähestymistapamme tehokkuuden osoittaminen erityisesti vähäisen datamäärän skenaarioissa. Tunteiden luokittelutehtävässä käytetään Berliinin puheen tunnetietokantaa (EMO-DB) [18], joka koostuu 535 lausumasta, jotka vastaavat seitsemää eri tunnetta.

**Tulos**

Kuinka monella näytteellä he tekevät kokeita?

**Esimerkki 5.1046**

Käyttämämme arvostelut on poimittu TripAdvisorista, ja niitä on alun perin ehdotettu BIBREF10- ja BIBREF11 -julkaisuissa.

**Tulos**

Mistä hotelliarvostelut ovat peräisin?

**Esimerkki 5.1047**

Tarkemmin sanottuna mallissamme käytämme sanan kontekstia ennustamaan sen merkinnän, ja näin mallimme oppii merkintätietoisen kontekstin jokaiselle lauseen sanalle. Parantaaksemme sanaesityksen ja sen kontekstin välistä vuorovaikutteisuutta lisäämme sanaesityksen ja sen kontekstin keskinäistä informaatiota.

**Tulos**

Miten heidän mallissaan hyödynnetään asiayhteystietoa kunkin teoksen osalta annetussa lauseessa monitehtävässä ympäristössä? ympäristössä?

**Esimerkki 5.1048**

Peruslinjat ovat: (a) koulutettu S-SQuAD:lla, (b) koulutettu T-SQuAD:lla ja sen jälkeen hienosäädetty S-SQuAD:lla ja (c) edellinen paras malli, joka on koulutettu S-SQuAD BIBREF5:llä käyttäen Dr.QA BIBREF20:tä. Ehdotettu lähestymistapa (rivi (f)) ylittää edellisen parhaan mallin (rivi (c)) 2 %:lla EM-tuloksella ja yli 1,5 %:lla F1-tuloksella.

**Tulos**

Mikä oli edellinen paras malli?

**Esimerkki 5.1049**

Tunnistamisen poistotehtävissä kolme mittaria, joita käytämme arkkitehtuurimme suorituskyvyn arvioimiseen, ovat tarkkuus, palautus ja INLINEFORM0-pisteet, jotka määritellään jäljempänä.

**Tulos**

Mitä arviointimittareita ne käyttävät?

**Esimerkki 5.1050**

Ensiksi vertaamme sen suorituskykyä muihin BERT-malleihin ja uusimpiin tunneanalyysin järjestelmiin osoittaaksemme sen suorituskyvyn luokittelutehtävissä. Toiseksi vertaamme sen suorituskykyä hiljattain toteutetussa hollannin kielen tehtävässä, nimittäin demonstratiivipronominien erottelussa, jolloin voimme lisäksi verrata meidän ja muiden BERT-mallien suorituskykyä nollapisteessä eli käyttämällä vain esivalmennettua mallia ilman mitään hienosäätöä.

**Tulos**

Mitä kielellisiä tehtäviä he kokeilivat?

**Esimerkki 5.1051**

Merkitsemme jokaisen arvosteluvektorin sillä autolla, jota se arvostelee.

**Tulos**

Mitä ovat autojen puhekielitietokannan merkinnät?

**Esimerkki 5.1052**

Flickr30K-tietokanta BIBREF0 on kokoelma yli 30 000 kuvaa, joista jokaisessa on 5 joukkoistettua kuvausta. Sitä käytetään yleisesti kouluttamaan ja arvioimaan neuroverkkomalleja, jotka tuottavat kuvien kuvauksia (esim. BIBREF2 ). Tämän artikkelin tavoitteena on antaa yleiskatsaus kielellisistä ennakkoluuloista ja stereotypioista ja ennakkoluuloista johtuvista perusteettomista johtopäätöksistä. Rakennan aiempaan työhön, joka koskee kielellistä harhaa yleisesti BIBREF3 , annan esimerkkejä Flickr30K-aineistosta ja esittelen perusteettomien johtopäätösten taksonomian. Lopuksi käsittelen useita menetelmiä, joilla aineistoa voidaan analysoida ennakkoluulojen havaitsemiseksi.

**Tulos**

Mikä on tietokokonaisuuden koko?

**Esimerkki 5.1053**

Puitteena käytämme FLAIRin BIBREF46 lauseenluokittimen konfiguraatiota, jossa on biLSTM-kooderi/luokitinarkkitehtuuri, jota syötetään merkki- ja sanatason representaatioilla, jotka koostuvat kiinteistä 300-ulotteisista GloVe-editoinneista BIBREF47 , valmiiksi koulutetuista kontekstualisoiduista FLAIRin sanaeditoinneista ja valmiiksi koulutetuista kontekstualisoiduista merkkien etumerkkieditoinneista, jotka on saatu AllenNLP:stä BIBREF48 ja joissa on FLAIRin oletusarvoisilla hyperparametreilla.

**Tulos**

Onko joitakin malleja arvioitu tällä mittarilla, ja mitkä ovat tulokset?

**Esimerkki 5.1054**

Käytämme UltraSuitea: lasten puheterapiaistunnoista kerättyjen ultraääni- ja akustisten tietojen arkisto BIBREF15 . Käytimme kaikkia kolmea tietokokonaisuutta arkistosta: UXTD (tallennettu tyypillisesti kehittyvien lasten kanssa) sekä UXSSD ja UPX (tallennettu lasten kanssa, joilla on puheäänihäiriöitä). Kaikkiaan tietokokonaisuus sisältää 13 815 puhuttua lausetta 86 puhujalta, mikä vastaa 35,9 tuntia äänityksiä. Lausumat on luokiteltu sen mukaan, minkä tyyppinen tehtävä lapselle annettiin, ja ne on merkitty seuraavasti: Sanat (A), ei-sanat (B), lause (C), artikulaatio (D), ei-puhe (E) tai keskustelu (F).

**Tulos**

Annotoivatko he oman tietokokonaisuutensa vai käyttävätkö he olemassa olevaa tietokokonaisuutta?

**Esimerkki 5.1055**

Python-pakettia twitterscraper käytetään twiittien keräämiseen twitteristä. Ohjelmasta poimittiin 10 478 twiittiä kahdelta viime vuodelta aloilta kuten "urheilu", "politiikka" ja "viihde".

**Tulos**

Mistä korpuksen tekstit ovat peräisin?

**Esimerkki 5.1056**

Vaikka ehdokkaat antoivat suostumuksensa haastattelujensa käyttöön, aineistoa ei julkaista julkisesti tämän tutkimuksen ulkopuolella, koska videot ovat henkilötietoja, joihin sovelletaan tiukkoja yksityisyyden suojaa koskevia rajoituksia.

**Tulos**

Ovatko ehdokkaat antaneet suostumuksensa siihen, että heidän videoitaan käytetään tutkimuksessa?

**Esimerkki 5.1057**

Lisäksi vertasimme mallejamme muihin olemassa oleviin teoksiin, kuten OpATT BIBREF6 ja Neural Content Planning with conditional copy (NCP+CC) BIBREF4.

**Tulos**

Mikä on tehtävään soveltuva uusin malli?

**Esimerkki 5.1058**

Ihmisten arvioinnitSeuraamme BIBREF11 , BIBREF12 ja valtavaa määrää aiempaa semanttista samankaltaisuutta koskevaa työtä, ja pyydämme yhdeksää opiskelijaa arvioimaan 360 relaatioparin samankaltaisuutta Wikidata BIBREF8 -aineiston osajoukosta, jotka on valittu niin, että ne kattavat korkeasta matalaan samankaltaisuustasoon. Kokeessamme koehenkilöitä pyydettiin antamaan jokaiselle parille kokonaislukuinen samankaltaisuuspistemäärä 0:sta (ei yhtäläisyyttä) 4:ään (täysin samanlainen).

**Tulos**

Miten ne keräävät inhimillisiä arvioita suhteiden samankaltaisuudesta?

**Esimerkki 5.1059**

Esittelemme tässä askeleen tähän suuntaan: probabilistinen semanttinen jäsentäjä, joka käyttää suurta tietopohjaa (NELL) muodostaakseen ennakkotodennäköisyysjakauman jäsentämiensä lauseiden merkityksistä ja joka "ymmärtää" jokaisen lauseen joko tunnistamalla olemassa olevat uskomukset, jotka vastaavat lauseen merkitystä, tai luomalla uusia uskomuksia.

**Tulos**

Mitä tietopohjia ne käyttävät?

**Esimerkki 5.1060**

Sekä itse distributiohypoteesi että Tugendhatin tulkinta Fregen teoksesta ovat esimerkkejä holistisesta lähestymistavasta merkitykseen, jossa kokonaisuuden merkitys määrittää osien merkityksen.

**Tulos**

Miten Fregen holistinen ja funktionaalinen lähestymistapa merkitykseen liittyy yleiseen distribuutiohypoteesiin?

**Esimerkki 5.1061**

Pinotut LSTM:t Solutietoiset pinotut LSTM:tJatketaan edellä määriteltyä pinotun LSTM:n muotoilua edellisessä alaluvussa mainitun ongelman ratkaisemiseksi. Lauseen koodaajatKokeissamme käyttämämme lauseen koodausverkko ottaa syötteenä INLINEFORM0-sanoja (joiden oletetaan olevan yhden pisteen vektoreita). Ylimmän kerroksen luokittelijat Luonnollisen kielen päättelykokeissa käytämme ominaisuuksien louhinnassa seuraavaa BIBREF36:n ehdottamaa heuristista funktiota: DISPLAYFORM0, jossa INLINEFORM0 tarkoittaa vektorien ketjuttamista ja INLINEFORM1 ja INLINEFORM2 sovelletaan elementtiviisaasti.

**Tulos**

Mitä malleja he kokeilivat?

**Esimerkki 5.1062**

Arkkitehtuuripohjaiset menetelmät pyrkivät joko luomaan tehtäväkohtaisen arkkitehtuurin esiharjoittelun aikana (tehtäväkohtaiset menetelmät) tai pyrkivät rakentamaan yleisen esiharjoitteluarkkitehtuurin, joka soveltuu kaikkiin myöhempiin tehtäviin (tehtäväkohtaiset menetelmät).

**Tulos**

Miten arkkitehtuuripohjainen menetelmä käsittelee NLG:n esteitä?

**Esimerkki 5.1063**

Käytämme videotekstien teksteihin VATEX-tietokokonaisuutta, joka sisältää yli 41 250 videota ja 825 000 englannin- ja kiinankielistä kuvatekstiä. Kuvatekstien joukossa on yli 206 000 englannin ja kiinan välistä rinnakkaiskäännösparia. Se kattaa 600 ihmisen toimintaa ja monenlaista videosisältöä. Jokaisen videon parina on 10 englanninkielistä ja 10 kiinankielistä erilaista kuvatekstiä. Noudatamme virallista jakoa, jossa on 25 991 videota harjoittelua varten, 3 000 videota validointia varten ja 6 000 julkista testivideota lopullista testausta varten.

**Tulos**

Kuinka suuri on käytetty tietokokonaisuus?

**Esimerkki 5.1064**

Tavoitteenamme oli rakentaa yleistettävissä oleva tunneanalyysimalli, joten käytimme kolmea erilaista harjoitusjoukkoa, jotka on esitetty taulukossa TABREF5 . Yksi näistä kolmesta tietokokonaisuudesta (Amazonin arvostelut BIBREF23 , BIBREF24 ) on laajempi, ja siinä on tuotearvosteluja useista eri luokista, kuten kirja-arvosteluja, elektroniikkatuotteiden arvosteluja ja sovellusarvosteluja. Kahden muun tietokokonaisuuden tarkoituksena on tehdä mallista toimialueeseen erikoistuneempi. Tässä asiakirjassa keskitymme toimialueena ravintola-arvosteluihin ja käytämme Yelp-ravintola-arvostelujen tietokokonaisuutta, joka on poimittu Yelp Dataset Challenge BIBREF25 -kilpailusta, ja ravintola-arvostelujen tietokokonaisuutta, joka on osa Kaggle-kilpailua BIBREF26 .

**Tulos**

mitä tietokokonaisuutta käytettiin harjoittelussa?

**Esimerkki 5.1065**

Käytämme kolmea luovaa englanninkielistä tietokokonaisuutta, joilla on erilaiset kielelliset ominaisuudet: (1) 740 klassisen ja nykyaikaisen englanninkielisen runon korpus, (2) 14950 metaforalauseen korpus, joka on haettu metaforatietokannan verkkosivulta, ja (3) 1500 laulun sanoituksen korpus, joka koostuu eri genreistä.

**Tulos**

Raportoidaanko tulokset vain englanninkielisistä tiedoista?

**Esimerkki 5.1066**

Näihin jäsentymättömiin ominaisuuksiin kuuluvat: Nämä ovat: 1) kunkin riskitekijän osa-alueeseen liittyvien lauseiden suhteellinen määrä vastaanottotiedotteissa (vastaanottotiedotteessa olevien lauseiden kokonaismäärästä) ja 2) kliiniset sentimenttipisteet kullekin riskitekijäalueelle, eli sentimenttipisteet, jotka arvioivat potilaan psykososiaalisen toimintakyvyn tasoa (positiivinen, negatiivinen tai neutraali) suhteessa kuhunkin riskitekijäalueeseen.Nämä sentimenttipisteet saatiin automaattisesti aiemmassa työssämme BIBREF15 esitellyn aiheen poiminta- ja sentimenttianalyysiputken avulla, ja ne harjoitteltiin ennalta sisäisellä psykiatrisella EHR-tekstillä. Takaisinottoriskiluokittimen syötteeksi poimittiin 45 kliinisesti tulkittavissa olevaa piirrettä sisäänottoa kohti.

**Tulos**

Miten ne sisällyttävät tunneanalyysin?

**Esimerkki 5.1067**

Heidän tuloksensa paransivat BIBREFin8 antamaa perustasoa. Toinen vaikutusvaltainen työ oli BIBREF12 , jossa käytettiin Double Propagation -nimistä valvomatonta algoritmia, joka karkeasti ottaen koostuu siemenjoukon lisäämisestä asteittain riippuvuuksien jäsentämisen avulla. n tästä huolimatta taulukko TABREF20 osoittaa, että järjestelmämme on parempi kuin parhaat aiemmat lähestymistavat kaikilla viidellä kielellä. Joissakin tapauksissa, kuten turkki ja venäjä, parhaat aiemmat tulokset olivat ABSA:n järjestäjien antamia perustuloksia, mutta hollannin, ranskan ja espanjan kielissä järjestelmämme on huomattavasti parempi kuin nykyinen huipputekniikka. Huolimatta siitä, että käytämme samaa järjestelmää kaikilla kielillä, olemme paremmat kuin GTI:n toimittamassa ehdotuksessa, jossa käytettiin CRF-järjestelmää, jossa oli espanjankieliselle BIBREF-tietokannalle28 ominaisia kielellisiä piirteitä.

**Tulos**

Mikä oli lähtötaso?

**Esimerkki 5.1068**

BIBREF12:n innoittamana integroimme tässä asiakirjassa rajojen kokoamisvaiheen uusimpaan LSTM-malliin kiinan sanojen segmentointia varten ja syötämme tuloksen CRF-malliin NER:ää varten. Tuloksena on 2 prosentin absoluuttinen parannus F1-kokonaispistemäärään verrattuna nykyisiin uusimpiin menetelmiin.

**Tulos**

Mitä huipputason syvää neuroverkkoa käytetään?

**Esimerkki 5.1069**

Chowdhury BIBREF14 ja Thomas et al. BIBREF11 ehdottivat menetelmiä, joissa käytetään kielellisiä ilmiöitä ja kaksivaiheista SVM-menetelmää DDI:iden luokitteluun. FBK-irst BIBREF10 on jatkotyö, jossa kernel-menetelmää sovelletaan olemassa olevaan malliin ja se on parempi kuin se. Neuroverkkoihin perustuvia lähestymistapoja on ehdotettu useissa töissä. Liu et al. BIBREF9 käyttävät ensimmäistä kertaa CNN:ää DDI:iden louhintaan, mikä päihittää perinteiset koneoppimiseen perustuvat menetelmät. Sahu et al. BIBREF12 ehdottivat LSTM:ään perustuvaa DDI:n louhintamenetelmää, joka on parempi kuin CNN:ään perustuva menetelmä, koska LSTM käsittelee lausetta sekvenssinä diaikkunoiden sijasta.

**Tulos**

Mitkä ovat asiakirjassa mainitut nykyiset menetelmät?

**Esimerkki 5.1070**

Lisäksi FSDM:ssä on uusi moduuli nimeltä vastauspaikkojen binääriluokittelija, joka lisää lisävalvontaa vastauksessa esiintyvien paikkojen tuottamiseksi tarkemmin ennen lopullisen tekstimuotoisen agenttivastauksen tuottamista (katso lisätietoja kohdasta "Menetelmät").

**Tulos**

Miten rakobinääriluokittelijat parantavat suorituskykyä?

**Esimerkki 5.1071**

Esittelemme ja arvioimme tätä varten kehittämämme mallin nimeltä QuaSP+Zero, joka muuttaa QuaSP+-parseria seuraavasti: Dekoodauksen aikana kohdissa, joissa jäsentäjä valitsee, mikä ominaisuus sisällytetään LF:ään (esim. kuva FIGREF31 ), se ei ota huomioon vain kysymysmerkkejä vaan myös näiden merkkien ja laadullisessa mallissa käytettyjen ominaisuuksien INLINEFORM0 välisen suhteen. Tämä lähestymistapa noudattaa BIBREF11 Krishnamurthy2017NeuralSP:n käyttämää entiteettien linkittämistä koskevaa lähestymistapaa, jossa kysymysmerkkien ja (entiteetteihin liittyvien) sanojen välinen samankaltaisuus - jota kutsutaan entiteettien linkittämispisteytykseksi - auttaa päättämään, mitkä entiteetit sisällytetään LF:ään jäsennyksen aikana.

**Tulos**

Miten QuaSP+Zero-malli toimii?

**Esimerkki 5.1072**

Tein myös useita tapaustutkimuksia. Sain asiakirjojen upotukset samassa latentissa avaruudessa kuin aiheen upotukset laskemalla yhteen kunkin merkin jälkikeskiarvovektorit INLINEFORM0 ja visualisoin ne kaksiulotteisesti käyttämällä INLINEFORM1 -SNE BIBREF24 (kaikki vektorit normalisoitiin yksikköpituisiksi). Unionin tilaa koskevat puheet (kuva KUVIO FIGREF27 ) upotetaan lähes lineaarisesti vuosien mukaan, ja New Deal -aikakauden (1930-luku) ympärillä on suuri hyppäys, ja ne ovat hyvin erillään puolueittain millä tahansa aikajaksolla.

**Tulos**

Mikä on esimerkki laskennallisesta yhteiskuntatieteellisestä NLP-tehtävästä?

**Esimerkki 5.1073**

BIBREF25 ehdotti myös ajatusta käyttää sanojen upotusten keskiarvoa asiakirjan yleisen kontekstin esittämiseen. Heidän työstään poiketen me päätämme korruptoida alkuperäisen asiakirjan poistamalla satunnaisesti merkittävän osan sanoista ja esitämme asiakirjan käyttämällä vain jäljelle jääneiden sanojen upotuksia.

**Tulos**

Onko heidän lähestymistapansa samanlainen kuin sanavektoreiden painotetun keskiarvosumman muodostaminen, jossa painot heijastavat sanojen frekvenssejä?

**Esimerkki 5.1074**

Koska yhtäkään aiempien töiden tietokokonaisuuksista ei ole julkaistu, päätämme rakentaa uuden tietokokonaisuuden. Keräämme satunnaisesti INLINEFORM0-tapauksia China Judgments Online -verkkopalvelusta, joista INLINEFORM1-tapaukset ovat harjoittelua varten ja INLINEFORM2-tapaukset validointia ja testausta varten.

**Tulos**

mitä tietokokonaisuuksia kokeessa käytetään?

**Esimerkki 5.1075**

Alustan käyttäjät antavat vaikuttavuusääniä arvioidakseen, kuinka vaikuttava tietty väite on. Käyttäjät voivat valita yhden viidestä mahdollisesta väitteen vaikutusmerkinnästä: ei vaikutusta, vähäinen vaikutus, keskisuuri vaikutus, suuri vaikutus ja erittäin suuri vaikutus. Arvioidessaan väitteen vaikutusta käyttäjät voivat tutustua väitteen koko asiayhteyteen, joten he voivat arvioida, kuinka vaikuttava väite on väitteen tietyssä asiayhteydessä. Mielenkiintoinen havainto on, että tässä aineistossa sama väite voi saada eri vaikutusmerkinnät sen mukaan, missä yhteydessä se esitetään.

**Tulos**

Miten pargmatiivinen ja diskurssikonteksti lisätään tietokokonaisuuteen?

**Esimerkki 5.1076**

Tämän vastapainoksi käytämme vasemmalta oikealle suuntautuvaa huomiomaskia, joka on samanlainen kuin alkuperäisessä Transformer-dekooderissa BIBREF1. Sovellamme $X$:n syötemerkkien osalta tällaista maskia kaikkiin $X$:n kanssa ketjutettuihin kohdemerkkeihin $Y$, jotta syötemerkit voivat kiinnittää huomiota vain muihin syötemerkkeihin. Vastaavasti kohdemerkkien $y\_t$ kohdalla asetamme huomiomaskin kaikkiin merkkeihin $y\_{>t}$, jolloin kohdemerkit $y\_t$ voivat huomioida vain syötemerkit ja jo muodostetut kohdemerkit.

**Tulos**

Mitä eroa on BERT-genissä ja tavallisessa BERT:ssä?

**Esimerkki 5.1077**

CBOW-arkkitehtuurissa tehtävänä on ennustaa sana kontekstin perusteella ja SG-arkkitehtuurissa sanan perusteella. Rakensimme 16 mallia sanojen upotuksista käyttäen CBOW- ja Skip-gram-menetelmien toteutusta FastText-työkalussa BIBREF9 .

**Tulos**

Mitä erityistä on tietyissä upotuksissa?

**Esimerkki 5.1078**

Tutkimme seuraavia kuutta idiomiresurssia: Wiktionary, Oxford Dictionary of English Idioms (ODEI, BIBREF31), UsingEnglish.com (UE), Sporleder-korpus BIBREF10, VNC-tietokanta BIBREF9 ja SemEval-2013 Task 5 -tietokanta BIBREF15.

**Tulos**

Mitä sanakirjoja käytetään PIE:iden automaattiseen poimintaan?

**Esimerkki 5.1079**

Järjestelmämme, mukaan lukien ohjelmisto ja korpus, on saatavilla avoimen lähdekoodin projektina vapaaseen tutkimustarkoitukseen, ja uskomme, että se on hyvä lähtökohta tulevien vietnamilaisten SRL-järjestelmien kehittämiselle ja vertailulle.

**Tulos**

Ovatko niiden korpus ja ohjelmistot julkisia?

**Esimerkki 5.1080**

Mallimme BiLSTM+CNN(grafeemitaso) ja BiLSTM+CNN(G)+POS päihittää kaikki muut mallit, joita kokeiltiin OurNepali- ja ILPRL-tietokannoissa.

**Tulos**

Mikä on paras malli?

**Esimerkki 5.1081**

Kokeet ::: Tulokset ::: Natural Language Inference: (81.2 vs. 80.2 XLM:lle) ja käyttää alle puolet vähemmän parametreja (110M vs. 250M). Sen suorituskyky jää kuitenkin edelleen jälkeen alkuperäisellä englanninkielisellä harjoitusjoukolla koulutetuista malleista TRANSLATE-TEST-asetuksessa (81,2 vs. 82,91 RoBERTa). On huomattava, että CamemBERT käyttää paljon vähemmän parametreja kuin RoBERTa (110M vs. 355M parametria).

**Tulos**

Missä tehtävissä CamemBERT ei parane?

**Esimerkki 5.1082**

Sitä vastoin tässä asiakirjassa ehdotetaan jaksotietojen hyödyntämiseksi sekä globaalin (koko asiakirja) että paikallisen kontekstin (esim. jakso/aihe) hajautettua esittämistä, kun päätetään, pitäisikö lause sisällyttää tiivistelmään.

**Tulos**

Mitä he tarkoittavat globaalilla ja paikallisella kontekstilla?

**Esimerkki 5.1083**

Sanaintruusiotestiä on kallista soveltaa, koska se edellyttää ihmisen suorittamaa manuaalista arviointia erikseen kunkin upotusulottuvuuden osalta. Lisäksi sanaintruusiotesti ei määrittele sulautusulottuvuuksien tulkittavuuden tasoa, vaan se antaa binäärisen päätöksen siitä, onko ulottuvuus tulkittavissa vai ei. Jatkuvien arvojen käyttäminen on kuitenkin tarkoituksenmukaisempaa kuin binääriarviointien tekeminen, koska tulkittavuuden tasot voivat vaihdella asteittain eri ulottuvuuksien välillä.

**Tulos**

Mitä etuja heidän ehdottamallaan tulkinnanvaraisuuden määrällisellä menetelmällä on verrattuna ihmislähtöiseen arviointiin, johon he vertaavat sitä?

**Esimerkki 5.1084**

Arvioimme QRNN-arkkitehtuuria suositussa dokumenttitason tunteiden luokittelun vertailukohteessa, IMDb:n elokuva-arvostelutietokannassa BIBREF17 . Tietokokonaisuus koostuu tasapainoisesta otoksesta, joka koostuu 25 000 positiivisesta ja 25 000 negatiivisesta arvostelusta, jotka on jaettu samankokoisiin train- ja testijoukkoihin ja joiden asiakirjojen keskimääräinen pituus on 231 sanaa BIBREF18 . Vertaamme vain muihin tuloksiin, joissa ei käytetä ylimääräistä merkitsemätöntä dataa (pois lukien esimerkiksi BIBREF19 ).

**Tulos**

Mitä tunnetilaluokittelua koskevaa tietokokonaisuutta käytetään?

**Esimerkki 5.1085**

Kun validoinnin aikana on poistettu 135 kysymystä, lopullinen tietokokonaisuus sisältää 13 939 kysymystä, joista 3 827 edellyttää maalaisjärjen tietämystä (eli 27,4 %).

**Tulos**

Mikä on heidän aineistonsa koko?

**Esimerkki 5.1086**

Tässä esitellään arviointiasetelma ja analysoidaan artikkelin ja objektin (AEP) sijoittelutehtävän tuloksia. Raportoimme arviointimittarit vain "merkityksellisten" uutis-artikkeliparien osalta. Perusasetukset. Tarkastelemme seuraavia perustasoja tätä tehtävää varten.B1. Ensimmäisessä perustasossa käytetään ainoastaan Dunietzin ja Gillickin BIBREF11 .B2. Toisessa perustasossa parille INLINEFORM0 annetaan arvo relevantti, jos ja vain jos INLINEFORM1 esiintyy INLINEFORM2:n otsikossa. Seuraavassa esitellään ASP-tehtävän arviointiasetelma ja keskustellaan tuloksista keskittyen kolmeen päänäkökohtaan: (i) kokonaissuorituskyky eri vuosina, (ii) entiteettiluokkakohtainen suorituskyky ja (iii) vaikutus entiteettiprofiilin laajentamiseen ehdottamalla puuttuvia osioita entiteeteille valmiiksi laskettujen mallien perusteella. Perustasot. Tietojemme mukaan emme ole tietoisia mistään vertailukelpoisesta lähestymistavasta tähän tehtävään. Sen vuoksi tarkastelemamme peruslinjat ovat seuraavat:S1: Valitaan malli INLINEFORM0:sta se jakso, jolla on suurin leksikaalinen samankaltaisuus INLINEFORM1:n kanssa : S1 INLINEFORM2S2: Sijoitetaan uutinen INLINEFORM0:n yleisimpään jaksoon.

**Tulos**

Mitä perusmallia käytetään?

**Esimerkki 5.1087**

Tässä työssä ehdotamme, että kokonaislukuiseen lineaariseen ohjelmointiin (integer linear programming, ILP) perustuvaa yhteenvetokehystä täydennetään yhteisesiintymismatriisin matalan sijan approksimaatiolla, ja arvioimme lähestymistapaa laajalla joukolla tietokokonaisuuksia, joissa on suuri leksikaalinen monimuotoisuus.

**Tulos**

Mitä ne rajoittavat käyttämällä kokonaislukuista lineaarista ohjelmointia?

**Esimerkki 5.1088**

Arvioimme CAHANin eri versioiden ja HAN-perustason avulla opittujen asiakirjojen upotusten laatua kolmella laajamittaisella asiakirjaluokitustietokannalla, jotka esiteltiin BIBREF14:ssä ja joita käytettiin alkuperäisessä HAN-julkaisussa BIBREF5. Ne jakautuvat kahteen luokkaan: aihepiiriluokittelu (Yahoo) ja hienojakoinen tunneanalyysi (Amazon, Yelp).

**Tulos**

Mitä tietokokonaisuuksia käytetään

**Esimerkki 5.1089**

Annotoijien joukko valitaan useiden lyhyiden harjoittelukierrosten jälkeen, joissa he saivat laajan henkilökohtaisen palautteen ja joissa oli enintään 15 predikaattia kierrosta kohden.

**Tulos**

Miten työntekijät koulutetaan?

**Esimerkki 5.1090**

Ehdotetun BiLSTM-mallin arvioimiseksi huomion kanssa (BiLSTM+att) sitä verrataan kolmeen sen omaan vaihtoehtoon: BiLSTM ilman huomiota (BiLSTM) sekä yksi eteenpäin suuntautuva LSTM-kerros huomion kanssa (LSTM+att) ja ilman huomiota (LSTM). Muita perusmalleja ovat BIBREF32 , joka on jo ehdottanut LSTM-pohjaista arkkitehtuuria, joka käyttää vain muita kuin ajallisia piirteitä, sekä SVM-pohjainen estimointimalli, jota BIBREF24 on alun perin käyttänyt palkkion estimointiin.

**Tulos**

Mitä mallia he käyttävät perustasona tyytyväisyyden arvioimiseksi?

**Esimerkki 5.1091**

Vaikka tällaista lähestymistapaa on käytetty eri tutkimuksissa ominaisuuksien suunnittelussa, sanavektoreiden valinta ja klusterien lukumäärä ovat edelleen kokeilu- ja virhemenettelyä.

**Tulos**

Mitä muita hyperparametreja kuin klusterien lukumäärää arvioidaan tyypillisesti tämäntyyppisessä tutkimuksessa?

**Esimerkki 5.1092**

Huomasimme, että valvomaton morfologinen analysaattori, joka pystyy käyttämään ristikkoja, paransi foneemien tunnistuksen ja sanojen segmentoinnin tarkkuutta. Näin ollen tämä tulos viittaa siihen, että tämä sanojen segmentointimenetelmä ottaa huomioon puheentunnistuksen moninaiset hypoteesit kokonaisuutena ja vähentää epävarmuutta, kuten tunnistuksen vaihtelua, käyttämällä tavujen tunnistustuloksia ristikkomuodossa.

**Tulos**

Miten he osoittavat, että paikkojen nimien hankkiminen auttaa itsensä paikallistamisessa?

**Esimerkki 5.1093**

He kommentoivat 7,8 000 verbiä ja raportoivat keskimäärin 2,4 QA-paria predikaattia kohti. Vaikka useamman annotaattorin osoitettiin tuottavan suuremman kattavuuden, heidän julkaisemansa tietokokonaisuus tuotettiin käyttämällä vain yhtä annotaattoria verbiä kohden.

**Tulos**

Miten kattavuutta mitattiin?

**Esimerkki 5.1094**

Seuraavaksi analysoimme kaikki loukkaaviksi merkityt twiitit, jotta ymmärtäisimme paremmin, miten arabian puhujat käyttävät loukkaavaa kieltä. Seuraavassa on käytön jakautuminen: Suora nimittely: Yleisin hyökkäys on kutsua henkilöä eläimellä, ja käytetyimmät eläinten nimet olivat كلب> ("klb" - "koira"), حمار> ("HmAr" - "aasi") ja بهيم> ("bhym" - "peto"). Toiseksi yleisintä oli henkisten kykyjen loukkaaminen käyttämällä sanoja kuten غبي> ("gby" - "tyhmä") ja عبيط> ("EbyT" - "idiootti"). Joitakin kulttuurikohtaisia eroja on otettava huomioon. Kaikkia eläinten nimiä ei käytetä loukkauksina. Esimerkiksi sellaisia eläinten nimiä kuin أسد> ("Asd" - "leijona"), صقر> ("Sqr" - "haukka") ja غزال> ("gzAl" - "gaselli") käytetään tyypillisesti ylistykseen. Muihin loukkauksiin ihmiset käyttävät: joitakin lintujen nimiä, kuten دجاجة> ("djAjp" - "kana"), بومة> ("bwmp" - "pöllö") ja غراب>. ("grAb" - "varis"); hyönteiset, kuten ذبابة> ("\*bAbp" - "kärpänen"), صرصور> ("SrSwr" - "torakka") ja حشرة>. ("H$rp" - "hyönteinen"); mikro-organismit, kuten جرثومة> ("jrvwmp" - "mikrobi") ja طحالب> ("THAlb" - "levä"); elottomat esineet, kuten جزمة> ("jzmp" - "kengät") ja سطل> ("sTl" - "ämpäri") muiden käyttötapojen ohella.Vertaus ja metafora: Käyttäjät käyttävät vertausta ja metaforaa, jos he vertaisivat henkilöä johonkin: eläimeen, kuten زي الثور> ("zy Alvwr" - "kuin härkä"), سمعني نهيقك> ("smEny nhyqk" - "anna minun kuulla karjuntaasi") ja هز ديلك>. ("hz dylk" - "heiluta häntääsi"); henkilö, jolla on henkinen tai fyysinen vamma, kuten منغولي> ("mngwly" - "mongolialainen (down-syndrooma)"), معوق> ("mEwq" - "vammainen") ja قزم> ("qzm" - "kääpiö"); ja vastakkaiseen sukupuoleen kuten جيش نوال>. ("jy$ nwAl" - "Nawalin armeija (Nawal on naispuolinen nimi)") ja نادي زيزي> ("nAdy zyzy" - "Zizin kerho (Zizi on naispuolinen lemmikkinimi)").Epäsuora puhe: Tällaista loukkaavaa kieltä ovat mm: sarkasmia, kuten أذكى إخواتك> ("A\*kY AxwAtk" - "sisaruksistasi fiksuin") ja فيلسوف الحمير> ("fylswf AlHmyr" - "aasien filosofi"); kysymykset, kuten ايه كل الغباء ده>. ("Ayh kl AlgbA dh" - "mitä tämä kaikki typeryys on"); ja epäsuora puhe, kuten النقاش مع البهايم غير مثمر> ("AlnqA$ mE AlbhAym gyr mvmr" - "turha puhua karjalle").Pahan toivominen: Tämä tarkoittaa kuoleman tai suuren vahingon toivomista jollekin, kuten ربنا ياخدك> ("rbnA yAxdk" - "Jumala vieköön (tappakoon) sinut"), الله يلعنك>. ("Allh ylEnk" - "Jumala/ Allah kiroaisi sinut"), ja روح في داهية> ("rwH fy dAhyp" - vastaa "mene helvettiin").Nimenmuutos: Yksi yleinen tapa loukata muita on muuttaa kirjain tai kaksi heidän nimissään, jotta saadaan uusia loukkaavia sanoja, jotka rimmaavat alkuperäisten nimien kanssa. Esimerkkejä tällaisesta ovat الجزيرة> ("Aljzyrp" - "Aljazeera (kanava)") muuttaminen الخنزيرة>:ksi. ("Alxnzyrp" - "sika") ja خلفان> ("xlfAn" - "Khalfan (henkilön nimi)") خرفان> ("xrfAn" - "hullu").Yhteiskunnallinen kerrostuneisuus: Jotkut loukkaukset liittyvät: tiettyihin ammatteihin, kuten بواب> ("bwAb" - "ovimies") tai خادم> ("xAdm" - "palvelija"); ja tiettyihin yhteiskunnan osiin, kuten بدوي> ("bdwy" - "beduiini") ja فلاح> ("flAH" - "maanviljelijä").Moraaliton käyttäytyminen: Nämä loukkaukset liittyvät negatiivisiin moraalisiin piirteisiin tai käyttäytymiseen, kuten حقير> ("Hqyr" - "ilkeä"), خاين> ("xAyn" - "petturi") ja منافق> ("mnAfq" - "tekopyhä").Sukupuoleen liittyvät: Niihin kuuluvat sellaiset ilmaukset kuin خول> ("xwl" - "homo"), وسخة> ("wsxp" - "prostituoitu") ja عرص> ("ErS" - "parittaja").

**Tulos**

Mitkä ovat arabian puhujien loukkaavan kielenkäytön erityispiirteet?

**Esimerkki 5.1095**

Ensimmäinen raportoimamme mittari on reaktiotyyppi. Toinen raportoimamme mittari on reaktionopeus. Jin et al. jin2013epidemiologisessa tutkimuksessa todettiin, että luotettavat uutistarinat leviävät nopeammin kuin väärät tiedot tai huhut; Zeng et al. zeng2016rumors havaitsivat, että huhut kiistävillä twiiteillä oli lyhyemmät viiveet kuin niitä tukevilla twiiteillä. Toinen tavoitteemme on selvittää, säilyvätkö nämä suuntaukset erityyppisten uutislähteiden osalta Twitterissä ja Redditissä. Tutkiaksemme, reagoivatko käyttäjät luotettavista lähteistä peräisin olevaan sisältöön eri tavalla kuin harhaanjohtavista lähteistä peräisin olevaan sisältöön, mittaamme reaktioviiveen, jonka määrittelemme linkin tai sisällön lähettämisen/twiittaamisen ja reaktiokommentin tai -twiittaamisen välillä kuluneeksi ajaksi. Raportoimme kumulatiiviset jakaumafunktiot (CDF) kullekin lähdetyypille ja käytämme Mann Whitney U (MWU) -testejä vertaillaksemme, reagoivatko käyttäjät tietyllä reaktiotyypillä merkittävästi erilaisilla viiveillä uutislähteisiin, joiden uskottavuus on erilainen.

**Tulos**

Miten nopeus mitataan?

**Esimerkki 5.1096**

Opetussuunnitelman uskottavuus ::: Keskusteluominaisuudet ::: SpesifisyysNeuraalisen dialogin generointimallin tunnettu ongelma on, että malli on altis tuottamaan geneerisiä vastauksia. Opetussuunnitelman uskottavuus ::: Conversational Attributes :::: ToistuvuusToistuvia vastauksia on helppo tuottaa nykyisessä auto-regressiivisessä vastausten dekoodauksessa, jossa vastausten tuottaminen kiertää usein, kun taas monipuoliset ja informatiiviset vastaukset ovat paljon monimutkaisempia neuraalisessa dialogin tuottamisessa. Opetussuunnitelman uskottavuus ::: Keskusteluattribuutit :::: JatkuvuusYhteensopiva vastaus ei ainoastaan vastaa annettuun kyselyyn, vaan myös käynnistää seuraavan lausahduksen. Opetussuunnitelman uskottavuus ::: Conversational Attributes :::: Model ConfidenceHeuristisista keskusteluattribuuteista huolimatta otamme lisäksi käyttöön mallin luottamuksen attribuuttina, joka erottaa helposti opittavat näytteet heikosti opittavista näytteistä mallin oppimiskyvyn kannalta. Opetussuunnitelman uskottavuus ::: Keskusteluattribuutit ::: Query-relatednessKeskustelua pidetään johdonmukaisena, jos vastaus korreloi hyvin annetun kyselyn kanssa.

**Tulos**

Mitä viittä vuoropuhelun ominaisuutta analysoitiin?

**Esimerkki 5.1097**

On osoitettu, että NER-järjestelmän sisältämää semanttista informaatiota voidaan lisätä merkittävästi, kun syväoppimismenetelmällä saadut entiteetit yhdistetään onnistuneesti tietopohjan BIBREF26 , BIBREF27 vastaaviin entiteetteihin. Uudelleenohjaus: Wikidata-linkityselementin osalta tunnistamme, että hakua rajoittaa kunkin entiteetin yleisin hakunimi.

**Tulos**

Miten he yhdistävät syväoppimismallin ja tietopohjan?

**Esimerkki 5.1098**

Tässä työssä tarkastelemme suunnattuja graafeja. Olkoon INLINEFORM0 graafi, joka koostuu joukosta kärkipisteitä INLINEFORM1 ja joukosta reunoja ( INLINEFORM2 ), jotka ovat järjestettyjä pareja. Lisäksi kullekin reunalle voidaan määrittää reaaliarvoinen paino. Olkoon INLINEFORM3 edustaa asiakirjaa, joka koostuu merkkeistä INLINEFORM4 . Tekstin merkkien esiintymisjärjestys tiedetään, joten INLINEFORM5 on täysin järjestetty joukko. Mahdollinen tapa rakentaa graafi asiakirjasta on yksinkertaisesti tarkkailla sanojen yhteisesiintymiä. Kun kaksi sanaa esiintyy yhdessä, niitä käytetään reunana. Tällaisissa lähestymistavoissa ei kuitenkaan oteta huomioon sanojen järjestysluonnetta, jolloin järjestys katoaa. Pyrimme ottamaan tämän näkökohdan huomioon seuraavasti. Annettu korpus käydään läpi, ja jokaiselle elementille INLINEFORM6 muodostetaan sen seuraaja INLINEFORM7 yhdessä tietyn elementin kanssa suunnattu reuna INLINEFORM8 . Lopuksi tällaisia reunoja painotetaan sen mukaan, kuinka monta kertaa ne esiintyvät tietyssä korpuksessa. Näin ollen graafi, joka on muodostettu tietyn korpuksen läpikäynnin jälkeen, koostuu kaikista paikallisista naapurustoista (järjestys yksi), jotka on yhdistetty yhdeksi yhteiseksi rakenteeksi. Globaalit kontekstuaaliset tiedot säilyvät mahdollisesti koskemattomina (painojen avulla), vaikka ne on havaittava verkkoanalyysin avulla, kuten seuraavassa ehdotetaan.

**Tulos**

Miten kuvaajat johdetaan tietystä tekstistä?

**Esimerkki 5.1099**

MIMIC-III on vapaasti saatavilla oleva, tunnistamaton tietokanta, joka sisältää Beth Israel Deaconess Medical Centerin teho-osastolle vuosina 2001-2012 otettujen potilaiden sähköiset terveystiedot.

**Tulos**

mitä tietokokonaisuuksia käytettiin?

**Esimerkki 5.1100**

Tällä hetkellä puolustaja voi hyödyntää DNN:ien kestävyyden parantamisessa kuva-alalta tekstiin sovellettavia menetelmiä, kuten vastakohtaista koulutusta BIBREF107 , lisäkerroksen lisäämistä BIBREF113 , risti-entropiafunktion optimointia BIBREF114 , BIBREF115 tai vastakohtaisten esimerkkien siirrettävyyden heikentämistä.

**Tulos**

Mitkä strategiat ovat lupaavimpia näiden hyökkäysten estämiseksi?

**Esimerkki 5.1101**

Keräämme lisää ihmisten kirjoittamia kohtauksia jokaiselle dev- ja testisarjan käsitejoukolle Amazon Mechanical Turk -alustan kautta. Jokaista syötettyä käsitejoukkoa kommentoi vähintään kolme eri ihmistä. Annotoijien edellytetään myös antavan perusteluina lauseita, jotka kannustavat heitä käyttämään tervettä järkeä kohtauksia luodessaan.

**Tulos**

Ovatko perustelut syntyneet lauseiden kirjoittamisen jälkeen?

**Esimerkki 5.1102**

Tulevaisuuden työssä aiomme ratkaista kolmikoita, joissa toisena entiteettinä on useita entiteettejä, jotka eivät kuulu tämän asiakirjan ongelman piiriin. QA4IE:n syötteenä on dokumentti $D$ ja olemassa oleva tietopohja $K$, ja tuloksena on joukko relaatiokolmioita $R = \lbrace e\_i, r\_{ij}, e\_j\rbrace $ $:ssa $D$, jossa $e\_i$ ja $e\_j$ ovat kaksi yksittäistä entiteettiä ja $r\_{ij}$ on niiden välinen suhde.

**Tulos**

Voidaanko tällä lähestymistavalla mallintaa n-alkuisia suhteita?

**Esimerkki 5.1103**

Bi-LSTM BIBREF11 on neuromallien perustaso. Bi-LSTM$\_{+ att. + LEX + POS}$ BIBREF10 on monitehtäväinen oppimiskehys WSD:lle, POS-tunnisteille ja LEX:lle, jossa on itsehuomiointimekanismi, joka muuntaa WSD:n sekvenssioppimistehtäväksi. GAS$\_{ext}$ BIBREF12 on GAS:n muunnos, joka on kiiltoa laajentava muistiverkon muunnos laajentamalla kiiltotietoa. CAN$^s$ ja HCAN BIBREF13 ovat lausetason ja hierarkkisen yhteishuomion neuroverkkomalleja, jotka hyödyntävät kiiltotietoa.

**Tulos**

Miten neuroverkkoarkkitehtuuri mukautuu tuntemattomaan määrään aisteja sanaa kohti?

**Esimerkki 5.1104**

Kuten luvussa 2 mainittiin, CAEVO on nykyisin uusin järjestelmä ominaisuuksiin perustuvaan ajallisten tapahtumasuhteiden louhintaan BIBREF10 . Sitä käytetään laajalti perustasona TB-Dense-tietojen arvioinnissa. Otamme sen käyttöön CaTeRS- ja RED-tietokokonaisuuksien arvioinnin perustana.

**Tulos**

Mitkä olivat perinteiset kielellisiin ominaisuuksiin perustuvat mallit?

**Tulos**

Minkälainen lähtötaso kahdelle tietokokonaisuudelle luodaan?

**Esimerkki 5.1105**

Se kerättiin joukkoistamalla BIBREF4. Tiedonkeruu tehtiin Android-sovelluksella.

**Tulos**

Mikä on tietojen lähde?

**Esimerkki 5.1106**

Mallissamme on viisi riippumatonta dekooderia, yksi kullekin sarjan kuvalle. Näin kukin dekooderi voi oppia tietyn kielimallin kuhunkin sekvenssin kohtaan.

**Tulos**

Onko kaikilla dekooderin LSTM-muisteilla samat painot?

**Esimerkki 5.1107**

Viimeksi GANeja käytettiin alkuperäisen maalauksen luomiseen ilman valvontaa BIBREF24.

**Tulos**

Koulutetaanko teksti-kuvasynteesiä valvotulla vai valvomattomalla tavalla?

**Esimerkki 5.1108**

Tunnepisteiden jakauman mukaan Teslan tunnelmat ovat testausjakson aikana hieman vinoutuneet positiivisiksi.

**Tulos**

Antavatko kirjoittajat esimerkkejä suurista tapahtumista, jotka herättävät yleisön huomion, ja niiden vaikutuksesta osakekurssiin?

**Esimerkki 5.1109**

Kuten taulukosta TABREF24 käy ilmi, MUSE päihittää kaikki aiemmat En-De- ja En-Fr-käännösmallit, mukaan lukien sekä uusimmat itsenäisen tarkkaavaisuuden mallit BIBREF0, BIBREF13 että konvoluutiomallit BIBREF11, BIBREF15 ja BIBREF10. Tämä tulos osoittaa, että pelkkä itsetarkkailu tai konvoluutio ei riitä sekvenssistä sekvenssiin oppimiseen. Ehdotettu rinnakkainen moniulotteinen huomio parantaa niitä sekä En-De- että En-Fr-käännöksessä.Verrattuna Evolved Transformer BIBREF19:ään, joka on rakennettu NAS:n avulla ja jossa sekoitetaan myös eri ytimen kokoisia konvoluutioita, MUSE saavuttaa 2,2 BLEU:n parannuksen En-Fr-käännöksessä.

**Tulos**

Kuinka suuri on suorituskyvyn parannus Transformersiin verrattuna?

**Esimerkki 5.1110**

Kneser-Ney-sujutus Vertaamme erityisesti Kneser-Ney-sujutusta, joka on yleisesti hyväksytty tekniikan tasoksi ennen NLM:ää, nykyisiin parhaisiin NLM:iin.

**Tulos**

mitä klassisia kielimalleja artikkelissa mainitaan?

**Esimerkki 5.1111**

Käytämme kokeissamme julkisesti saatavilla olevaa tietokokonaisuutta KVRET BIBREF5. Harjoittelua varten on 2425 dialogia, validointia varten 302 ja testausta varten 302 dialogia, kuten taulukon TABREF12 yläosasta käy ilmi.

**Tulos**

Mikä on tietokokonaisuuden koko?

**Esimerkki 5.1112**

Edellä mainittujen haasteiden ratkaisemiseksi ehdotamme paikallisesti lineaarista meta-kuvioinnin oppimismenetelmää, joka a) vaatii vain kunkin lähdekuvion sanaston sanat, eikä sen tarvitse ennustaa puuttuvien sanojen kuvioita, b) voi kuvioida eri ulottuvuuksisia lähdekuvioita ja c) on herkkä lähdekuvioiden naapurustojen monimuotoisuudelle.Ehdotetussa menetelmässä on kaksi vaihetta: naapuruston rekonstruointivaihe (kohta "Lähimmän naapurin rekonstruointi" ) ja projektiovaihe (kohta "Projektio meta-sulautuma-avaruuteen" ). Rekonstruktiovaiheessa sanan upotus esitetään lineaarisesti painotettuna yhdistelmänä sen lähimpien naapureiden upotuksista kussakin lähteen upotusavaruudessa.

**Tulos**

Mikä on tässä asiakirjassa esitelty meta-embedding-menetelmä?

**Esimerkki 5.1113**

Kokeelliset tutkimukset ::: Koska BERT on jo saavuttanut kysymysten vastaamisen huipputason suorituskyvyn, tässä jaksossa vertaamme ehdotettua malliamme uusimpiin kysymysten vastaamisen malleihin (eli QANet BIBREF39 ja BERT-Base BIBREF26). BERT-mallilla on kaksi versiota: BERT-Base ja BERT-Large, mutta laskentaresurssien puutteen vuoksi voimme verrata vain BERT-Base-mallia BERT-Large-mallin sijaan.

**Tulos**

Mihin vertailukohtia ehdotettua mallia verrataan?

**Esimerkki 5.1114**

$D\_a$ sisälsi kaikki twiitit, jotka kerättiin hyökkäyspäivänä kohdassa 4.2 mainituista viidestä hyökkäyksestä. Ja $D\_b$ sisälsi kaikki ennen viittä hyökkäystä kerätyt twiitit. $D\_a$:ssa on 1180 twiittiä ja $D\_b$:ssa 7979 twiittiä. Hyökkäyspäivien twiitit ($D\_a$) on kommentoitu manuaalisesti, ja vain 50 prosenttia näistä twiiteistä koskee todellisuudessa DDoS-hyökkäystä.

**Tulos**

Onko Twitter-käyttäjillä tapana twiitata DOS-hyökkäyksestä, kun se tapahtuu? Kuinka paljon tietoja tukee tätä oletusta?

**Esimerkki 5.1115**

Luomme kaksi mallia, jotka molemmat koostuvat kolmesta pääosasta: koodaaja, vuorovaikutus ja luokittelija, ja otamme syötteenä kaksi sekvenssiä. Kooderi on jaettu sekvenssien kesken ja käyttää yksinkertaisesti kahta pinottua GRU-kerrosta. Vuorovaikutusosa koostuu toisessa mallissa vain huomiosta, kun taas toisessa mallissa se koostuu huomion ja konfliktin yhdistelmästä, kuten (yhtälö 11) osoittaa. Luokittelijaosa on yksinkertaisesti pinottuja täysin kytkettyjä kerroksia.

**Tulos**

Mitä neuraalista arkkitehtuuria ne käyttävät huomiokonfliktimekanismiensa perustana?

**Esimerkki 5.1116**

Tässä koesarjassa BERT-toteutusta verrataan useisiin MEDDOCAN-haasteeseen osallistuneisiin järjestelmiin: CRF-luokittelijaan BIBREF18, spaCy-entiteettitunnistimeen BIBREF18 ja NLNDE BIBREF12:een, joka voitti jaetun tehtävän ja on espanjankielisen kliinisen tekstin arkaluonteisen tiedon tunnistamisen ja luokittelun nykytekniikan taso. Saatujen tulosten perusteella yhteistyöelin jää kuitenkin vain 0,3 F1-pistettä jälkeen ja olisi saavuttanut toisen sijan kaikkien MEDDOCANin jaetun tehtävän kilpailijoiden joukossa. Kun otetaan huomioon, että vain 3 prosenttia kultaisista merkinnöistä on edelleen virheellisesti merkitty, tehtävää voidaan pitää lähes ratkaistuna, eikä ole selvää, ovatko järjestelmien väliset erot todella merkittäviä vai johtuvatko ne vähäisistä vaihteluista alustuksessa tai vähäisten merkintäepävarmuuksien pitkästä sarjasta.

**Tulos**

Saavuttaako yhteistyöhön perustuva yhteistyö parhaan suorituskyvyn kaikista vertailluista algoritmeista?

**Esimerkki 5.1117**

Lopuksi yhdistämme CNN- ja RNN-mallit äänestysprosessin avulla. Sovellamme kuhunkin testisarjan lauseeseen useita CNN- ja RNN-malleja, jotka on esitetty taulukoissa TABREF12 ja TABREF14, ja ennustamme eniten ääniä saaneen luokan. Jos äänet menevät tasan, valitsemme satunnaisesti yhden yleisimmistä luokista.

**Tulos**

Miten heidän yksinkertainen äänestysjärjestelmänsä toimii?

**Esimerkki 5.1118**

Vertaamme parasta F1-pistemääräämme muihin uusimpiin lähestymistapoihin taulukossa TABREF39 , mikä osoittaa, että mallillamme on kilpailuetua lääkkeiden ja lääkkeiden välisestä vuorovaikutuksesta.

**Tulos**

Mikä on heidän mallinsa suorituskyky?

**Tulos**

Kuinka paljon heidän mallinsa on parempi kuin nykyiset menetelmät?

**Esimerkki 5.1119**

Käytämme siis Krishnamurthyn ja Mitchellin esittelemää krishnamurthy-2015-semparse-open-vocabulary -tietokokonaisuutta, joka koostuu ClueWeb09-verkkokorpuksesta sekä Googlen FACC-entiteettilinkistä, joka yhdistää kyseisen korpuksen Freebase BIBREFiin9 .

**Tulos**

Mitä tietopohjaa he käyttävät?

**Esimerkki 5.1120**

Valitsimme SVMhmm BIBREF111-toteutuksen rakenteellisista tukivektorikoneista sekvenssien merkitsemistä varten. SVM-toteutus

**Tulos**

Mitä koneoppimismenetelmiä kokeissa käytetään?

**Esimerkki 5.1121**

Pysyäksemme johdonmukaisina LSTM:llä tehtyjen kokeiden kanssa käytämme morfessoria suomen kielen alasanojen tokenisointiin.

**Tulos**

Onko LSTM:n perustaso alasanamalli?

**Esimerkki 5.1122**

Testaamme malliamme ensin yhden toimialueen tietokokonaisuudella, WoZ2.0 BIBREF19 . Se koostuu 1 200 dialogista ravintolavarauksen alueelta, jossa on kolme ennalta määriteltyä kohtaa: ruoka, hintaluokka ja alue. Koska nimikenttä esiintyy aineistossa harvoin, sitä ei ole sisällytetty kokeiluihimme aikaisemman kirjallisuuden BIBREF3 , BIBREF20 mukaisesti. Malliamme testataan myös monialuetietoaineistolla MultiWoZ BIBREF9 . Siinä on monimutkaisempi ontologia, jossa on 7 toimialuetta ja 25 ennalta määriteltyä paikkaa. Koska uskomustilojen yhdistettyä slot-arvoparien esitystä on sovellettava malliin, jossa on $O(n)$ ITC, slottien kokonaismäärä on 35.

**Tulos**

Mitä tietokokonaisuuksia käytetään suorituskyvyn arvioinnissa?

**Esimerkki 5.1123**

Loimme kahdeksan erilaista luokittelijaa, joista kukin käytti jotakin seuraavista kahdeksasta ominaisuudesta, jotka ovat saatavilla Twitterin API:n virrasta haetusta twiitistä: Käyttäjän sijainti (uloc): Tämä on käyttäjän profiilissaan ilmoittama sijainti. Vaikka tämä ominaisuus saattaa vaikuttaa a priori hyödylliseltä, se on kuitenkin jossain määrin rajallinen, sillä kyseessä on vapaa tekstikenttä, jonka käyttäjät voivat jättää tyhjäksi, syöttää sijainnin nimen, joka on moniselitteinen tai jossa on kirjoitusvirheitä, tai merkkijonon, joka ei täsmää mihinkään tiettyyn sijaintiin (esim. "kotona"). Tarkasteltaessa käyttäjien itse ilmoittamia sijainteja Hecht et al. BIBREF49 havaitsivat, että 66 prosenttia ilmoitti tietoja, jotka voidaan kääntää, tarkasti tai epätarkasti, maantieteelliseksi sijainniksi, ja loput 34 prosenttia oli joko tyhjiä tai ei ollut maantieteellisesti paikannettavissa: Tämä on käyttäjän itse ilmoittama käyttöliittymän kieli. Käyttöliittymän kieli voi olla osoitus käyttäjän alkuperämaasta, mutta käyttäjä on voinut myös määrittää käyttöliittymän eri kielelle, kuten englanniksi, koska se oli oletuskieli, kun hän rekisteröityi tai koska hänen valitsemansa kieli ei ole käytettävissä.Aikavyöhyke (tz): Tämä ilmaisee aikavyöhykkeen, jonka käyttäjä on määritellyt asetuksissaan, esim. "Pacific Time (US & Canada)". Kun käyttäjä on määritellyt asetuksissaan tarkan aikavyöhykkeen, se voi olla osoitus hänen alkuperämaastaan; joidenkin käyttäjien asetuksissa voi kuitenkin olla oletusaikavyöhyke tai he voivat käyttää vastaavaa aikavyöhykettä, joka kuuluu eri paikkaan (esim. "Eurooppa/Lontoo" Portugalissa asuvalle käyttäjälle). Twitterin aikavyöhykeluettelo ei myöskään sisällä kaikkia maita.Tweet language (tlang): Twitter tunnistaa automaattisesti kielen, jolla twiitin uskotaan olevan kirjoitettu. Sen on todettu olevan tarkka tärkeimpien kielten osalta, mutta se jättää toivomisen varaa vähemmän käytettyjen kielten osalta. Twitterin kielitunnisteen on myös todettu olevan vaikeuksissa monikielisten twiittien kanssa, joissa twiitin osat on kirjoitettu eri kielillä BIBREF50 .Offset (offset): Tämä on käyttäjän asetuksissaan määrittelemä siirtymä UTC/GMT:hen nähden. Se on samanlainen kuin aikavyöhyke, vaikkakin rajoitetumpi, koska se on jaettu useiden maiden kanssa.käyttäjän nimi (name): Tämä on käyttäjän asetuksissaan määrittelemä nimi, joka voi olla hänen oikea nimensä tai hänen valitsemansa vaihtoehtoinen nimi. Käyttäjän nimi voi joissakin tapauksissa paljastaa hänen alkuperämaansa.Käyttäjän kuvaus (description): Tämä on vapaa teksti, jossa käyttäjä voi kuvailla itseään, kiinnostuksen kohteitaan jne: Teksti, joka muodostaa twiitin varsinaisen sisällön. Sisällön käyttöön liittyy useita varoituksia. Yksi niistä on se, että sisältö saattaa muuttua ajan myötä, ja näin ollen uudet twiitit saattavat käsitellä uusia aiheita, joita luokittelijat eivät ole nähneet aiemmin. Toinen varoitus on se, että twiitin sisältö ei välttämättä ole paikkakuntakohtaista; aiemmassa tutkimuksessa Rakesh et al. BIBREF51 havaitsivat, että 10 000 twiitistä vain 289:n sisältö oli paikkakuntakohtaista.

**Tulos**

Mitkä ovat kahdeksan mainittua ominaisuutta?

**Esimerkki 5.1124**

Suodatuksen jälkeen jätettiin 27 534 viestiä. Tätä aineistoa käytetään kokeiluissa.

**Tulos**

Mikä on merkityn tietokokonaisuuden koko?

**Esimerkki 5.1125**

Plackett-Luce-mallin log-likelihood ei ole BLEU-pistemäärän tiukka yläraja, mutta se korreloi kuitenkin hyvin BLEU-pistemäärän kanssa, kun kyseessä ovat rikkaat ominaisuudet. Käsite "rikas" on itse asiassa kvalitatiivinen, ja sen määrittely on epäselvää eri sovelluksissa. Tarjoamme empiirisesti kaavan rikkauden mittaamiseksi konekääntämisen skenaariossa. Mitä suurempi, sitä rikkaampi. Käytännössä olemme havainneet, että r:n karkea kynnysarvo on 5 %.

**Tulos**

Miten he mittaavat kestävyyttä kokeissa?

**Esimerkki 5.1126**

Teimme kokeita CMU:n ARCTIC-tietokannalla BIBREF33, joka sisältää rinnakkaisia äänitteitä ammattimaisista yhdysvaltalaisista englannin puhujista 16 kHz:n taajuudella. Yksi naispuolinen puhuja (slt) valittiin kohdepuhujaksi ja yksi miespuolinen puhuja (bdl) ja yksi naispuolinen puhuja (clb) valittiin lähteiksi. Valitsimme kumpikin 100 lausumaa validointia ja arviointia varten, ja muita 932 lausumaa käytettiin harjoitusaineistona. TTS-korpusta valitsimme yhdysvaltalaisen naispuolisen englanninkielisen puhujan (judy bieber) M-AILABS-puhetietokannasta BIBREF34 yhden puhujan Transformer-TTS-mallin kouluttamista varten. Kun näytteenottotaajuus oli myös 16 kHz, harjoitusjoukko sisälsi 15 200 lausetta, jotka olivat noin 32 tuntia pitkiä.

**Tulos**

Mitä tietokokonaisuuksia kokeillaan?

**Esimerkki 5.1127**

Aiemmissa tutkimuksissa on sovellettu reduktionistista lähestymistapaa, jossa näitä kahta ongelmaa, "mitä" ja "miten", on tarkasteltu erikseen sisällön valinnan ja kysymysten rakentamisen avulla. Neuraaliset mallit sen sijaan motivoivat kokonaisarkkitehtuuria. Syväoppimisen kehykset poikkeavat reduktionistisesta lähestymistavasta ja sallivat lähestymistavat, jotka yhdessä optimoivat sekä "mitä" että "miten" yhtenäisessä kehyksessä.

**Tulos**

Mitä oppimisparadigmoja tässä tutkimuksessa käsitellään?

**Esimerkki 5.1128**

BIBREF17:ssä VAE:tä koskevassa työssä akustisten upotusten oppimiseksi tehtiin kokeita TIMIT-tietoaineistolla. BIBREF17:ssä tehtyjen VAE:hen perustuvien puhelinten luokittelukokeiden perustason suorituskyky raportoi 72,2 prosentin tarkkuuden. Työmme perustana oleva uudelleen toteutettu versio antoi 72,0 prosentin tarkkuuden, ja tämän tuloksen katsottiin tarjoavan uskottavan pohjan jatkotyölle.

**Tulos**

Mitä luokittelun perustasoja käytetään vertailussa?

**Esimerkki 5.1129**

Noudatamme samankaltaista arviointiprotokollaa kuin BIBREF6 , BIBREF8 , BIBREF9 -julkaisuissa, eli käytämme oppimiamme representaatioita ominaisuuksina matalan kompleksisuuden luokittimelle (tyypillisesti lineaariselle) uudessa valvotussa tehtävässä/alueella, jota ei ole nähty harjoittelun aikana, päivittämättä lauseen representaatiomallimme parametreja. Tarkastelemme myös tällaista siirto-oppimisen arviointia keinotekoisesti rakennetussa vähäresurssisessa ympäristössä. Lisäksi arvioimme myös oppimiemme yksittäisten sanarepresentaatioiden laatua käyttämällä vakiomuotoisia vertailukohtia BIBREF36 , BIBREF37 .

**Tulos**

Miten he arvioivat lause-esityksiään?

**Esimerkki 5.1130**

Vaikka sekoitusstrategia korvaa suurimman osan Fr-De\*:n ja Fr\*-De:n välisestä erosta (3,01 $\rightarrow $ 0,17) De $\rightarrow $ Fr -tapauksessa, tuloksena saatu PSEUDOmix osoittaa edelleen alhaisempaa BLEU-arvoa kuin kohde-alkuperää oleva Fr-De\*-korpus. Parannamme siis lähdeaineistosta peräisin olevan Fr\*-De-korpuksen synteettisten esimerkkien laatua kouluttamalla edelleen sen käännösmallia (En $\rightarrow $ Fr). Kuten kuvasta 2 käy ilmi, kun kohteena oleva Fr-De\*-korpus on kiinteä, lähteenä olevalla Fr\*-De-aineistolla ja PSEUDOmixilla koulutettujen mallien laatu kasvaa suhteessa Fr\*-De-korpuksen emomallin laatuun. Lopulta PSEUDOmix osoittaa korkeinta BLEU-arvoa, ja se on parempi kuin sekä Fr\*-De- että Fr-De\*-aineisto. Kuten taulukossa 6 esitetään, havaitaan, että hienosäätö, jossa käytetään perustotuuden rinnakkaisdataa, parantaa huomattavasti kaikkien NMT-mallien käännöslaatua. Kaikista hienosäädetyistä malleista PSEUDOmix osoittaa parasta suorituskykyä kaikissa kokeissa. Tämä on erityisen rohkaisevaa De $\rightarrow $ Fr -mallissa, jossa PSEUDOmixin BLEU-arvo oli alhaisempi kuin Fr-De\*-datan ennen hienosäätöä. Jopa siinä tapauksessa, että PSEUDOmix osoittaa vertailukelpoisia tuloksia muiden synteettisten korporaatioiden kanssa Pseudo Only -skenaariossa, se osoittaa suurempia parannuksia käännöksen laadussa, kun sitä hienosäädetään oikean rinnakkaisen datan kanssa.

**Tulos**

Kuinka monta parannusta ranskan ja saksan käännösten vertailuarvoon?

**Esimerkki 5.1131**

Suoritamme sitten seuraavan analyysin.Puhujan sukupuolen vaikutukset: Etsimme yksikön ensimmäisen persoonan pronomineja, joilla on subjektin tapaus (ani, sukupuolen osalta merkitsemätön, vastaa englannin I:tä), ja otamme huomioon sitä hallitsevan verbin sukupuolen (tai adjektiivit kopulaarirakenteissa, kuten `I am nice'). Mahdolliset sukupuolet ovat `maskuliininen', `feminiininen' ja `kumpikin', joista jälkimmäinen tarkoittaa tapausta, jossa diakritisoimaton kirjoitusmuoto sallii sekä maskuliinisen että feminiinisen lukutavan. Odotamme, että sukupuoli vastaa etuliitteessä pyydettyä sukupuolta.Interlocutors' Gender and Number Effects: Etsimme toisen persoonan pronomineja ja tarkastelemme niiden sukupuolta ja lukumäärää. Subjektin asemassa olevien pronominien osalta otamme huomioon myös niitä hallitsevien verbien (tai adjektiivien kopulaarirakenteissa) sukupuolen ja lukumäärän. Kun kyseessä on yksikössä oleva yleisö, odotamme, että sukupuoli ja lukumäärä vastaavat pyydettyjä. Monikossa odotamme maskuliinisen ja monikollisen muodon vastaavuutta.

**Tulos**

Millainen syntaktinen analyysi suoritetaan?

**Esimerkki 5.1132**

Koska harjoitusaineisto koostuu vain lausuma-denotaatiopareista, ranker koulutetaan maksimoimaan oikean vastauksen $z$ log-likelihoodia käsittelemällä loogisia muotoja latenttina muuttujana. Rankerin tehtävänä on pisteyttää jäsentimen tuottamat loogiset muotokandidaatit; testiaikana korkeimman pistemäärän saanutta loogista muotoa käytetään suoritukseen. Ranker on diskriminoiva log-lineaarinen malli loogiselle muodolle $y$, kun kyseessä on lausuma $x$ :

**Tulos**

Miten malli laskee todennäköisyyden suorittaa korjaus semanttiseen denotaatioon?

**Esimerkki 5.1133**

Vertailimme malliamme MLE-, RL- ja GAN-perusmalleihin. Koska COCO:lla ja EMNLP2017 WMT:llä ei ole syötettä, kun taas WeiboDial pitää viestejä syötteenä, valitsimme seuraavat perusmallit: MLE: RNN-malli, joka on koulutettu MLE-tavoitteella BIBREF4 . Sen laajennus, Seq2Seq, voi toimia dialogitietokannassa BIBREF2 .SeqGAN: Ensimmäinen teksti-GAN-malli, joka päivittää generaattoria politiikkagradientilla, joka perustuu erottelijan palkintoihin BIBREF7 .LeakGAN: SeqGANin muunnos, joka tarjoaa palkintoja erottelijan vuotaneen tiedon perusteella generaattorille BIBREF11 .MaliGAN: SeqGANin muunnos, joka optimoi generaattorin normalisoidulla maksimiluotettavuustavoitteella BIBREF8 .IRL: Tämä käänteinen vahvistusoppimismenetelmä korvaa diskriminaattorin palkitsemisapproksimaattorilla tiheiden palkkioiden tarjoamiseksi BIBREF12 .RAML: RL-lähestymistapa, jolla MLE-tavoite sisällytetään RL-koulutuskehykseen ja jossa BLEU:ta pidetään palkkiona BIBREF17 .DialogGAN: DialogGAN:n laajennus, joka on viritetty dialogin generointitehtävään ja jossa MLE-tavoite on lisätty vastakkaiseen tavoitteeseen BIBREF16 .DPGAN: DialogGAN:n muunnos, jossa käytetään kielimalliin perustuvaa diskriminaattoria ja jossa ristiinentropiaa pidetään palkkiona BIBREF13 .

**Tulos**

Mitä GAN-malleja käytettiin vertailukohtana?

**Esimerkki 5.1134**

Arkkitehtuurimme on suunniteltu osoittamaan, että käyttämällä mallia, joka tallentaa paremmin tietoa tekstin kontekstista ja peräkkäisestä luonteesta, voidaan päihittää kirjallisuudessa yleisesti käytetyt leksikonipohjaiset menetelmät.

**Tulos**

Mitä merkityksellistä tietoa GRU-malli tallentaa, mitä perinteiset ML-mallit eivät?

**Esimerkki 5.1135**

Kuva-avusteisen mallin (W+C+V; ylempi rivi kuvassa FIGREF19 ) osalta vahvistamme, että modaliteettihuomio vaimentaa onnistuneesti epäolennaisia signaaleja (selfiet jne.) ja vahvistaa merkityksellisiä modaliteettipohjaisia konteksteja tietyn merkin ennustamisessa.

**Tulos**

Tarkastavatko he malliaan nähdäkseen, oppiiko heidän mallinsa yhdistämään kuvan osia kokonaisuuksiin liittyviin sanoihin?

**Esimerkki 5.1136**

Yhden puhujan vuoropuheluvuoro voi olla paitsi suora vastaus toisen puhujan kysymykseen, myös todennäköisesti jatkoa hänen omalle aiemmalle lausumalleen. Kun mallinnetaan dialogin vuoroa $k$, ehdotamme, että vuoron $k-2$ viimeinen RNN-tila liitetään suoraan vuoron $k$ alkutilaan sen sijaan, että sen annettaisiin levitä vuoron $k-1$ RNN:n kautta.

**Tulos**

Kuinka pitkä dialogihistoria on tallennettu?

**Esimerkki 5.1137**

U.S. Newsin tilastoista poimitaan 200 korkeakoulua koskevat neljä tosiasioihin liittyvää ominaisuutta, jotka koostuvat seuraavista: Undergraduate Enrollment, Male/Female Ratio, Private/Public ja Region (Northeast, South, West ja Midwest).

**Tulos**

Mitä suuria maantieteellisiä alueita tutkitaan?

**Esimerkki 5.1138**

Mallissamme käytetään hiljattain käyttöön otettua biaffiinista huomiota BIBREF14 parantamaan roolien pisteytystä.

**Tulos**

Mikä on biaffinen maalintekijä?

**Esimerkki 5.1139**

Olemme käyttäneet kokeiluissamme 14 TD:n tietokokonaisuutta.

**Tulos**

Mitä tietokokonaisuuksia ne käyttävät?

**Esimerkki 5.1140**

Koska menetelmämme syötteenä on tekstimuotoinen tieto, noudatamme BIBREF15:n lähestymistapaa ja kuvaamme tekstin kiinteän kokoiseksi vektoriesitykseksi. Tätä varten käytämme sanojen upotuksia, joita on sovellettu menestyksekkäästi muilla aloilla. Noudatamme BIBREF5:tä ja käytämme valmiiksi koulutettuja GloVe-sanavektoreita BIBREF16 upotuskerroksen alustamiseen (tunnetaan myös nimellä look-up table). Jaksossa SECREF18 käsitellään upotuskerrosta yksityiskohtaisemmin.

**Tulos**

Mitä esivalmennettuja sanavektoreita he käyttivät?

**Esimerkki 5.1141**

Otamme käyttöön seuraavat kaksi tietokokonaisuutta: Quora BIBREF1: Quora Question Pairs -tietokanta sisältää kysymyspareja, jotka on merkitty merkinnöillä, jotka osoittavat, ovatko nämä kaksi kysymystä parafraaseja. Käytämme samaa tietokokonaisuuden osiota kuin BIBREF5:ssä, jossa on 384 348/10 000/10 000 paria harjoittelu-/kehitys- ja testijoukossa.MRPC BIBREF34: Microsoft Research Paraphrase Corpus koostuu lausepareista, jotka on kerätty online-uutisista. Jokainen pari on merkitty merkinnällä, joka osoittaa, ovatko kaksi lausetta semanttisesti samanarvoisia. Harjoitus-/testijoukossa on 4 076/1 725 paria.

**Tulos**

Mitkä ovat benhmark-tietoaineistot parafraasien tunnistamista varten?

**Esimerkki 5.1142**

Twiitit tunnukset 6, 8 ja 10 ovat joitakin näytteitä, jotka sisältävät loukkaavia sanoja ja solvauksia, jotka eivät kaikissa tapauksissa ole vihamielisiä tai loukkaavia ja joiden kirjoittajat käyttivät tällaista kieltä päivittäisessä viestinnässään. Kun otetaan huomioon nämä todisteet ja tarkastelemalla twiittien sisältöä, voimme nähdä twiittien ID:t 3, 4 ja 9 kohdalla, että BERT-pohjainen luokittimemme pystyy erottamaan twiitit, joissa ei ole vihamielistä eikä implisiittistä vihasisältöä.

**Tulos**

Mitä todisteita kirjoittajat esittävät siitä, että malli voi ottaa huomioon joitakin harhoja tietojen merkinnöissä ja keräämisessä?

**Esimerkki 5.1143**

Käytämme seuraavia kolmea arviointimittaria: Foneemivirheprosentti (PER) on ennustettujen foneemisekvenssien ja kultaisen standardin foneemisekvenssien välinen Levenshteinin etäisyys jaettuna kultaisen standardin foneemisekvenssien pituudella.Sanavirheprosentti (Word Error Rate, WER) on niiden sanojen prosenttiosuus, joissa ennustettu foneemisekvenssi ei täsmälleen vastaa kultaista standardia.Sanavirheprosentti 100 (Word Error Rate 100, WER 100) on niiden sanojen prosenttiosuus testijoukosta, joissa oikea arvaus ei ole järjestelmän sadan ensimmäisen arvausarvion joukossa.

**Tulos**

mitä arviointimittareita käytettiin?

**Esimerkki 5.1144**

Päästä päähän -järjestelmässä (etuliite E2E) käytetään DNN-topologiaa, joka on kuvattu kuvassa KUVA8 . Esitämme tulokset kolmella eri kokokokoonpanolla (etuliitteet 700K, 318K ja 40K), joista kukin edustaa likimääräisten parametrien määrää, ja kahdella erityyppisellä harjoitusreseptillä (etuliitteet 1stage ja 2stage), jotka vastaavat vastaavasti end-to-end- ja encoder+decoder-järjestelmiä, kuten UID7:ssä on kuvattu.

**Tulos**

Kuinka monta parametria esitetyssä mallissa on?

**Esimerkki 5.1145**

KGR10, joka tunnetaan myös nimellä plWordNet Corpus 10.0 (PLWNC 10.0), on tulosta työstä, joka koskee verkkosivujen sisällön automaattiseen hankintaan ja uuttamiseen tarkoitettua työkaluketjua nimeltä CorpoGrabber BIBREF19 . Se on työkalupaketti, jonka avulla saadaan verkkosivuston olennaisin sisältö, mukaan lukien kaikki alasivustot (käyttäjän määrittelemään syvyyteen asti). Ehdotettua työkaluketjua voidaan käyttää suuren tekstidokumenttien web-korpuksen rakentamiseen. Se tarvitsee syötteenä luettelon juurisivustoista. CorpoGrabberin muodostavat työkalut on mukautettu puolankielelle, mutta useimmat osatehtävät ovat kielestä riippumattomia. Koko prosessi voidaan suorittaa rinnakkain yhdellä koneella, ja se sisältää seuraavat tehtävät: Kunkin syöttösivun URL-osoitteen HTML-alasivujen lataaminen HTTrackin avulla, pelkän tekstin poimiminen jokaiselta alasivulta poistamalla "boilerplate"-sisältö (kuten navigointilinkit, otsikot, alatunnisteet, mainokset HTML-sivuilta) BIBREF20 , pelkän tekstin deduplikointi BIBREF20 , huonolaatuisten asiakirjojen poistaminen käyttämällä Morphological Analysis Converter and Aggregatoria (MACA) BIBREF21 , asiakirjojen merkitseminen Wrocław CRF Taggerilla (WCRFT) BIBREF22 . Kaksi viimeistä vaihetta ovat käytettävissä vain puolan kielellä.

**Tulos**

Miten KGR10-korpus luotiin?

**Esimerkki 5.1146**

BIBREF14 esittelee EMPATHETICDIALOGUES-tietokannan, joka on uusi tietokanta, joka sisältää 25 000 keskustelua ja sisältää emotionaalisia kontekstitietoja, jotka helpottavat tekstimuotoisen keskustelujärjestelmän harjoittelua ja arviointia. BIBREF2:n työ tuottaa tietokokonaisuuden, joka sisältää 1,5 miljoonaa Twitter-keskustelua, jotka on kerätty Twitter API:n avulla 62 tuotemerkin asiakaspalvelutililtä useilta eri toimialoilta. Tätä tietokokonaisuutta käytettiin sävytietoisen asiakaspalvelukeskustelurobotin rakentamiseen. Lopuksi BIBREF29 yritti parantaa SEMAINE-korpusta BIBREF30 käyttämällä crowdsourcing-skenaariota, jotta saataisiin inhimillistä arviota siitä, mikä vastaus herättää positiivisia tunteita.

**Tulos**

Mitkä ovat tällä hetkellä saatavilla olevat EAC:n tietokokonaisuudet?

**Esimerkki 5.1147**

Ensin esivalmennamme generaattorimme Gutenbergin tietokokonaisuudella BIBREF24 20 epookin ajan ja sitten hienosäädämme BIBREF19 -tietokantoja kohdetietokantojamme varten kielen mallintamistavoitteella.

**Tulos**

Mitä tavoitefunktiota käytetään GAN:ssa?

**Esimerkki 5.1148**

Tässä työssä rakennamme uuden regressiomallin, joka perustuu kielellisiin, sisällöllisiin, käyttäytymiseen liittyviin ja aihepiirteisiin ja jonka avulla voimme havaita arabiankieliset Twitter-robotit ja ymmärtää niiden vaikutusta uskonnollisen vihan levittämiseen arabiankielisessä Twitter-avaruudessa.

**Tulos**

Ehdottavatko he uutta mallia, jolla voitaisiin havaita paremmin nimenomaan arabialaiset botit?

**Esimerkki 5.1149**

Tämän prosessin tuloksena saadaan yhteensä INLINEFORM0 10700 harjoitusesimerkkiä, joista noin 2000-3000 esimerkkiä luokkaa kohti, ja jokaisella puhujalla on keskimäärin 185 esimerkkiä.

**Tulos**

Kuinka monta tapausta heidän tietokokonaisuudessaan on?

**Esimerkki 5.1150**

convertible.pl: DCG-sääntöjen toteuttaminen kolmivaiheisen muunnoksen 1. ja 3. vaihetta varten sekä muita sääntöjä, mukaan lukien sanasto.

**Tulos**

Mitä DCG:tä käytetään?

**Esimerkki 5.1151**

Tuloksena syntynyt mallimme tuottaa huippuluokan tuloksia sekä VQA-tietokannassa että bAbI-10k-tekstikysymys-vastaus -tietokannassa, mikä osoittaa, että malli voidaan yleistää useilla eri aloilla.

**Tulos**

Toteutuuko DMN+-mallissa uusin tekniikka?

**Esimerkki 5.1152**

Järjestelmämme sijoittui kilpailussa toiseksi vain 0,3 BLEU-pistettä voittajatiimi UPC-TALP:n jälkeen. Kaikkien joukkueiden saamat suhteellisen alhaiset BLEU- ja korkeat TER-pisteet johtuvat kilpailussa annetuista toimialueen ulkopuolisista tiedoista, jotka tekivät tehtävästä yhtä haastavan kaikille osallistujille. Se, että järjestäjät tarjosivat toimialueen ulkopuolista dataa, johti haastavaan mutta mielenkiintoiseen skenaarioon kaikille osallistujille.

**Tulos**

Parantaako toimialueen ulkopuolisten tietojen käyttö menetelmän suorituskykyä?

**Esimerkki 5.1153**

Tässä artikkelissa esittelen tulkittavissa olevan sanojen upotusmallin ja siihen liittyvän aihepiirimallin, jotka on suunniteltu toimimaan hyvin, kun ne koulutetaan pienellä tai keskikokoisella kiinnostavalla korpuksella. Testasin representaatioiden suorituskykyä ominaisuuksina asiakirjojen luokittelu- ja regressiotehtävissä. Tulokset on esitetty taulukossa TABREF26 . Asiakirjojen luokittelussa käytin kolmea vakiomuotoista vertailuaineistoa: 20 uutisryhmää (19 997 uutisryhmän viestiä), Reuters-150-uutislehden artikkeleita (15 500 artikkelia ja 150 luokkaa) ja Ohsumed-tietokannan lääketieteellisiä tiivistelmiä 23 sydän- ja verisuonitaudista (20 000 artikkelia). I Analysoin myös regressiotehtävää, jossa ennustetaan unionin tilaa käsittelevän puheen vuosi sen tekstitietojen perusteella.

**Tulos**

Millaisia valvottuja oppimistehtäviä näillä esityksillä yritetään suorittaa?

**Esimerkki 5.1154**

Arviointimittareina käytimme BLEU BIBREF8- ja RIBES BIBREF9 -mittareita käännöstarkkuuden mittaamiseen ja token-tason viivettä viiveen mittaamiseen. Japanin käännöstarkkuuden arvioinnissa käytimme Kytea BIBREF10 -menetelmää tokenize-menetelmänä.

**Tulos**

Mitä mittareita ne käyttävät simultaanikääntämisen arviointiin?

**Esimerkki 5.1155**

Vaikka lähestymistapamme on selvästi parempi kuin satunnaiset vertailut, suhteellisen alhaiset F1-pisteet osoittavat, että sen ennustaminen, mikä sana kaikui selityksissä, on hyvin haastava tehtävä. Tästä seuraa, että pystymme vain rajallisesti ymmärtämään, miten ihmiset valitsevat sanojen kaikuisuuden selityksissä. Se, missä määrin selitysten rakentaminen on pohjimmiltaan satunnaista BIBREF47 vai onko olemassa muita tunnistamattomia malleja, on tietenkin avoin kysymys. Toivomme, että tutkimuksemme ja julkaisemamme resurssit kannustavat jatkamaan työtä selitysten pragmatiikan ymmärtämiseksi.

**Tulos**

Antavatko kirjoittajat mitään selitystä kiehtoville sanojen kaikuille?

**Esimerkki 5.1156**

Yhdistetystä tulosjoukosta poimitaan yksilölliset käyttäjät ja twiitit, jolloin saadaan noin 60 000 yksilöllisen twiitin tietokokonaisuus, joka liittyy 51 104 yksilölliseen käyttäjään.

**Tulos**

Kuinka monta seuraajaa he analysoivat?

**Tulos**

Kuinka monta twiittiä tässä paperissa tutkitaan?

**Esimerkki 5.1157**

Ehdotamme uutta valvomatonta mallia, Sent2Vec, universaalien lauseen upotusten oppimiseen. Muodollisesti opimme lähde- (tai konteksti-) upotuksen INLINEFORM0 ja kohde- upotuksen INLINEFORM1 jokaiselle sanaston sanalle INLINEFORM2, ja upotusulottuvuudet INLINEFORM3 ja INLINEFORM4 ovat samat kuin ( EQREF6 ). Lauseen upotus määritellään sen muodostavien sanojen lähdesanojen upotusten keskiarvona, kuten ( EQREF8 ). Täydennämme tätä mallia lisäksi oppimalla myös lähdesanojen upotukset kussakin lauseessa esiintyvien unigrammien lisäksi myös n-grammien osalta ja keskiarvoistamalla n-grammien upotukset yhdessä sanojen kanssa, eli lauseen INLINEFORM5:n upotus INLINEFORM6:lle mallinnetaan seuraavasti: DISPLAYFORM0

**Tulos**

Miten n-grammiominaisuuksissa otetaan huomioon kompositionaalisuus?

**Esimerkki 5.1158**

Siksi ehdotamme huomiomalleja, joiden avulla voidaan päätellä piilevä konteksti eli viestisarja, joka käynnistää intervention.

**Tulos**

Minkälaista latenttia kontekstia käytetään ohjaajan intervention ennustamiseen?

**Esimerkki 5.1159**

Arab-tweet. Iän ja sukupuolen mallintamiseen käytämme Arap-Tweet BIBREF19 , josta käytetään jatkossa nimitystä Arab-Tweet. UBC:n Twitterin sukupuolitietoaineisto. Kehitämme myös oman Twitter-tietokannan sukupuolta varten. Merkitsimme manuaalisesti 1 989 käyttäjää kustakin 21 arabimaasta. AraNetin murteiden tunnistamismalli perustuu MADARin jaetussa tehtävässä 2 BIBREF20 voittaneeseen järjestelmäämme, joka on kuvattu BIBREF12:ssa. Korpus on jaettu train-, dev- ja testikorpukseen, ja järjestäjät peittävät testisarjan merkinnät. Hyödynnämme kahta tietokokonaisuutta, BIBREF22:n LAMA-DINA-tietokokonaisuutta, Twitter-tietokokonaisuutta, jossa on yhdistelmä BIBREF23:n kultaisia merkintöjä ja etävalvonnan merkintöjä. Loput tietokokonaisuudesta on merkitty vain etänä valvottuna (LAMA-DIST) ($182 605$ twiittejä) . Lisätietoja tietokokonaisuudesta on BIBREF22:ssa. Käytämme tietokokonaisuutta IDAT@FIRE2019 jaetun tehtävän BIBREF24 julkaisemien arabiankielisten twiittien ironian tunnistamiseen. Keräämme 15 arabian kielen sentimenttianalyysiin liittyvää tietokokonaisuutta, mukaan lukien MSA ja murteet BIBREF25, BIBREF26, BIBREF27, BIBREF1, BIBREF28, BIBREF29, BIBREF30, BIBREF31, BIBREF32, BIBREF33, BIBREF34. Taulukossa TABREF28 esitetään kaikki käyttämämme korporaatiot.

**Tulos**

Mitä tietokokonaisuuksia käytetään koulutuksessa?

**Esimerkki 5.1160**

Käytämme hienosäätöä varten esivalmennettua karsimatonta BERT$\_\mathrm {BASE}$-mallia, koska havaitsimme, että BERT$\_\mathrm {LARGE}$-malli suoriutuu tässä tehtävässä hieman huonommin kuin BERT$\_\mathrm {BASE}$.

**Tulos**

Käytetäänkö suurta vai pientä BERT:iä?

**Esimerkki 5.1161**

DeepMine on julkisesti kaikkien saatavilla, ja eri käyttäjille on tarjolla erilaisia lisenssejä. Se kerättiin joukkoistamalla BIBREF4. Tiedonkeruu tehtiin Android-sovelluksella. Kukin vastaaja asensi sovelluksen henkilökohtaiseen laitteeseensa ja nauhoitti useita lauseita eri istunnoissa.

**Tulos**

miten puhe kerättiin?

**Esimerkki 5.1162**

TF-IDF: Tukivektorikone-luokittelija (SVM), logistinen regressioluokittelija (LR), Naive Bayes -luokittelija (NB) ja satunnaismetsä (RF); CNN: Käytimme TF-IDF-ominaisuuksia neljän luokittelijan syötteenä: LSTM: Sovelsimme LSTM-mallia BIBREF20 lauseiden luokitteluun valmiiksi koulutettujen sanojen upotusten avulla; LSTM-soft: Tämän jälkeen lisäsimme LSTM-mallin päälle pehmeän tarkkaavaisuuskerroksen BIBREF21 , jossa laskimme pehmeät kohdistamispisteet jokaiselle piilotetulle tilalle; LSTM-self: Sovelsimme LSTM-malliin BIBREF22-kerroksen.

**Tulos**

Mihin perusmalleihin ehdotettua mallia verrataan?

**Esimerkki 5.1163**

Davidson et al. BIBREF2 osoitti, että viidessä Twitterin vertailutietoaineistossa, jotka oli kommentoitu loukkaavan kielenkäytön havaitsemista varten, oli systemaattisia ja huomattavia rasistisia vääristymiä. Tarkastelemalla useampia näytteitä ja ottaen huomioon hiljattain tehdyt tutkimukset BIBREF2, BIBREF24 ja BIBREF19 on selvää, että monet virheet johtuvat tiedonkeruusta BIBREF19 ja annotaatiosäännöistä BIBREF24 johtuvista harhoista eivätkä itse luokittelijasta.

**Tulos**

Mitkä ovat olemassa olevat ennakkoluulot?

**Esimerkki 5.1164**

Käytämme ensin uusimpia PDTB-tunnisteiden merkitsimiä perustasoa BIBREF13 varten, BIBREF12:ta mallimme kausaalisuusennusteen arvioimiseksi (BIBREF12 vaatii syötteenä tekstistä poimittuja lauseita, joten käytimme jäsentäjäämme poimimaan lauseita viestistä).

**Tulos**

Mitä lähtökohtia he ottivat huomioon?

**Esimerkki 5.1165**

Havaitseminen: Hypernymian havaitsemisessa tehtävänä on luokitella, ovatko sanaparit hypernymiasuhteessa. Suunta: Suunnan ennustamisessa tehtävänä on tunnistaa, mikä termi on laajempi tietyssä sanaparissa. Graded Entailment: Graded entailment -toiminnossa tehtävänä on määrittää, missä määrin hypernyymisuhde on voimassa.

**Tulos**

Mitä hypernymian tehtäviä he tutkivat?

**Esimerkki 5.1166**

Näiden menetelmien mukaisesti arvioimme NCEL:ää seuraavilla viidellä tietokokonaisuudella: (1) CoNLL-YAGO BIBREF22 : CoNLL 2003:n jaettu tehtävä, joka sisältää 4791 mainintaa sisältävän testa-aineiston 216 asiakirjassa ja 4485 mainintaa sisältävän testb-aineiston 213 asiakirjassa. (2) TAC2010 BIBREF39 : muodostettu Text Analysis Conference -konferenssia varten ja sisältää 676 mainintaa 352 asiakirjassa testausta varten. (3) ACE2004 BIBREF23 : osajoukko ACE2004:n yhteisviiteasiakirjoja, joka sisältää 248 mainintaa 35 asiakirjassa ja joka on annotoitu Amazon Mechanical Turkilla. (4) AQUAINT BIBREF40 : 50 uutisartikkelia, joihin sisältyy 699 mainintaa kolmesta eri uutistoimistosta. (5) WW BIBREF19 : uusi vertailukohde, jossa mainintojen ennakkojakaumat ovat tasapainossa, mikä johtaa vaikeaan disambiguointitapaukseen. Siinä on 6374 mainintaa 310 asiakirjassa, jotka on poimittu automaattisesti Wikipediasta.

**Tulos**

Mitä tietokokonaisuuksia ne käyttävät?

**Esimerkki 5.1167**

Taulukon TABREF38 tulokset vahvistavat BIBREF13:n tulokset ja viittaavat siihen, että olemme onnistuneet toistamaan suuren osan niiden ominaisuuksista. Tulokset kaikille kolmelle ennusteasetukselle (yksi lähtevä reuna: INLINEFORM0 , tuki/hyökkäys: INLINEFORM1 ja tuki/hyökkäys/kumpikaan: INLINEFORM2 ) kaikkien tyyppimuuttujien ( INLINEFORM3 , INLINEFORM4 ja INLINEFORM5 ) osalta esitetään taulukossa TABREF39 . Kaikki mallit ylittävät merkittävästi enemmistön perusmallin makro F1:n osalta.

**Tulos**

Mitä perusviivoja ja luokittelujärjestelmiä kokeissa käytetään?

**Esimerkki 5.1168**

Tätä varten rakensimme uuden tietokokonaisuuden, jossa on 23 700 kyselyä, jotka ovat lyhyitä ja jäsentymättömiä ja jotka ovat samantyylisiä kuin tehtäväkeskeisten järjestelmien todellisten käyttäjien tekemät kyselyt.

**Tulos**

Mikä on tämän tietokokonaisuuden koko?

**Esimerkki 5.1169**

Näin ollen jatkuva relaksaatio top-k-argmax-operaatioon voidaan yksinkertaisesti toteuttaa käyttämällä iteratiivisesti max-operaatiota, joka on jatkuva ja mahdollistaa gradienttivirran takaisinkulkeutumisen aikana.

**Tulos**

Mitä häviämismittareita he kokeilevat uudessa koulutusmenetelmässään, jota arvioidaan palkkihaun tuloksen perusteella?

**Esimerkki 5.1170**

Vaikka tässä korpuksessa on virheellisiä lauseita ja niiden tunnemerkintöjä, niistä puuttuvat vastaavat korjatut lauseet, joita tarvitaan mallimme harjoitteluun. Saadaksemme nämä puuttuvat tiedot, käytämme natiivien englanninkielisten puhujien apuna puolueetonta ja anonyymiä alustaa nimeltä Amazon Mechanical Turk (MTurk) BIBREF19, joka on maksullinen markkinapaikka ihmisälytehtäville (Human Intelligence Tasks, HIT). Käytämme tätä alustaa luodaksemme natiivien englanninkielisten puhujien tehtäviä alkuperäisten virheellisten twiittien muotoilemiseksi oikeiksi lauseiksi. Taulukossa TABREF12 on joitakin esimerkkejä. Mallien suorituskyvyn arvioinnissa käytetty tietokokonaisuus on Braun et al. BIBREF20 esittämä Chatbot Natural Language Unerstanding (NLU) Evaluation Corpus NLU-palvelujen testaamiseksi. Se on julkisesti saatavilla oleva vertailukohde, ja se koostuu lauseista, jotka on saatu saksalaisesta Telegram-keskustelurobotista, jota käytetään vastaamaan kysymyksiin julkisista liikenneyhteyksistä. Aineistossa on kaksi tarkoitusta, nimittäin lähtöaika ja yhteyden etsiminen, ja siinä on 100 harjoitus- ja 106 testinäytettä, jotka esitetään taulukossa TABREF18. Vaikka vertailukohteen pääkieli on englanti, tämä tietokokonaisuus sisältää muutamia saksankielisiä asemien ja katujen nimiä.

**Tulos**

Raportoidaanko tulokset vain englanninkielisistä tietokokonaisuuksista?

**Esimerkki 5.1171**

TOEFL-kuullunymmärtämiskoe on tarkoitettu englannin kielen oppijoille, joiden äidinkieli ei ole englanti. Tässä artikkelissa kerrotaan, miten nykypäivän kone pystyy suoriutumaan tällaisesta testistä.Tässä tarkasteltava kuullun ymmärtämistehtävä liittyy läheisesti Spoken Question Answering (SQA) BIBREF0 , BIBREF1 .

**Tulos**

Mikä on tässä työssä ehdotettu uusi tehtävä?

**Esimerkki 5.1172**

Valitsimme lauseiden koodausmalleiksi yksinkertaisen yksikerroksisen kaksisuuntaisen LSTM:n, jossa on maksimipoolaus (BiLSTM-max) ja jonka piilokoko on 600D suuntaa kohti ja jota käytettiin esimerkiksi InferSent BIBREF17 ja HBMP BIBREF18 -malleissa. Muihin malleihin olemme valinneet ESIM BIBREF19 , joka sisältää lauseiden välisen huomion, ja KIM BIBREF2 , jossa on lauseiden välinen huomio ja jossa hyödynnetään ulkoista tietoa. Valitsimme myös kaksi mallia, joissa on mukana valmiiksi koulutettu kielimalli, nimittäin ESIM + ELMo BIBREF20 ja BERT BIBREF0 .

**Tulos**

Mitä malleja vertailtiin?

**Esimerkki 5.1173**

Ennen kaikkea otamme käyttöön kaksi yhteistä peruslinjaa. Ensimmäisessä valitaan vain johtavat lauseet yhteenvedon muodostamiseksi. Sitä käytetään usein DUC:n virallisena perustasona, ja nimemme sen "LEAD". Toisen järjestelmän nimi on "QUERY\_SIM", joka asettaa lauseet suoraan paremmuusjärjestykseen sen TF-IDF-kosiinimuunnoksen samankaltaisuuden mukaan. Ennen kaikkea esittelemme kaksi yleistä perusviivaa. Ensimmäisessä valitaan vain johtavat lauseet yhteenvedon muodostamiseksi. Sitä käytetään usein DUC:n virallisena perustasona, ja se on nimeltään "LEAD". Toisen järjestelmän nimi on "QUERY\_SIM", joka asettaa lauseet suoraan paremmuusjärjestykseen sen mukaan, mikä on niiden TF-IDF-kosiinimuunnos samankaltaisuus kyselyn kanssa. Lisäksi toteutamme kaksi suosittua kyselyyn keskittyvää tiivistämismenetelmää, jotka ovat MultiMR BIBREF2 ja SVR BIBREF20 . Koska mallimme on täysin datalähtöinen, otamme käyttöön tuoreen tiivistämisjärjestelmän DocEmb BIBREF9 , joka myös käyttää vain syviä neuroverkko-ominaisuuksia lauseiden sijoittamiseen. Yhteisen mallin tehokkuuden todentamiseksi suunnittelemme ISOLATION-nimisen perustason, joka suorittaa erottelun ja merkityksellisyyden luokittelun erikseen.

**Tulos**

Mihin malleihin niitä verrataan?

**Esimerkki 5.1174**

NMT-kokeessa käytimme Torch-toteutusta kouluttaaksemme 2-kerroksisen LSTM-mallin, jossa on 500 piilotettua yksikköä sekä koodaimessa että dekoodaimessa, 12 epookin aikana.

**Tulos**

Mitä NMT-malleja he kokeilivat?

**Esimerkki 5.1175**

Käytämme tässä artikkelissa kahta eri upotusavaruutta. Ensimmäinen on laajalti käytetty esitys, joka perustuu GoogleNewsin BIBREF-tietokantaan8 . Toinen on otettu BIBREF2 -tietokannasta, ja se on koulutettu Reddit-tietokannalla BIBREF9 .

**Tulos**

Missä sulautumissa ne havaitsevat vääristymiä?

**Esimerkki 5.1176**

LDA:n tuottamien aiheiden johdonmukaisuus on kuitenkin odotettua heikompi.Tämän johdonmukaisuuden puutteen korjaamiseksi käytimme ei-negatiivista matriisifaktorointia (NMF).

**Tulos**

Miten Dabiqissa ja Rumiyahissa tunnistetaan näkyvät aiheet?

**Esimerkki 5.1177**

Perustaso on hierarkkinen fraasipohjainen järjestelmä BIBREF29 , jossa on 4-grammainen kielimalli, jonka ominaisuuksien painotukset on viritetty MIRA BIBREF30 -ohjelmalla. Siksi sopiva ylimääräinen perustaso olisi merkitä käännössäännöt näillä indikaattorifunktioilla, mutta ilman pisteitä, mikä vastaa sellaisten sääntöjen tunnistamista, joissa on fraaseja (Baseline + SegOn).

**Tulos**

Mihin käännösjärjestelmiin niitä verrataan?

**Esimerkki 5.1178**

Järjestelmämme ovat huomiokooderi-dekooderiverkkoja BIBREF0 . Perustamme toteutuksemme dl4mt-oppaaseen, jota olemme parantaneet uusilla ominaisuuksilla, kuten ensemble-dekoodauksella ja läpäisevällä pudotuksella.

**Tulos**

Mitkä ovat perusjärjestelmät?

**Esimerkki 5.1179**

Käytämme BIBREF-tietokannassa46 esiteltyä valtavaa 45 000 masentuneesta käyttäjästä koostuvaa tietokantaa, jossa on 1500 masennukseen viittaavasta termistä koostuva masennusoireiden sanasto, joka luotiin psykologin avulla ja jota käytettiin masentuneiksi itsensä ilmoittaneiden henkilöiden profiilien keräämiseen. Kaksi ihmistuomaria tarkisti BIBREF46:ssa 8 770 käyttäjän (24 miljoonaa aikaleimattua twiittiä) osajoukon, joka sisälsi 3981 masentunutta ja 4789 vertailukäyttäjää (jotka eivät osoita masennusta).

**Tulos**

Miten tiedot on kommentoitu?

**Esimerkki 5.1180**

Arvioimme vastakohtaisesti koulutetut mallit, kuten taulukossa TABREF18 on esitetty.Vastakohtaisen koulutuksen jälkeen kaikkien kohdemallien suorituskyky paranee merkittävästi, kun taas alkuperäisten esimerkkien suorituskyky pysyy vertailukelpoisena.

**Tulos**

Kuinka paljon suorituskyky paranee kokeissa, kun mallit koulutetaan luotujen vastakkaisten esimerkkien avulla?

**Esimerkki 5.1181**

$Coherence@k$:n on osoitettu korreloivan hyvin eri aihepiirien mallinnusmenetelmillä opittujen aiheiden inhimillisen tulkinnanvaraisuuden kanssa BIBREF7 . Näin ollen voimme odottaa tulkinnanvaraisia upotuksia maksimoimalla sen. Tulkittavuuden arvioimiseksi käytämme $Coherence@k$ (yhtälö 6 ) , automatisoituja ja manuaalisia sanojen tunkeutumistestejä.

**Tulos**

Miten ne arvioivat tulkittavuutta?

**Esimerkki 5.1182**

Esimerkiksi viestiä voidaan pitää sellaisenaan harmittomana, mutta kun otetaan huomioon aiemmat viestiketjut, sitä voidaan pitää loukkaavana ja päinvastoin. Tämä tekee loukkaavan kielenkäytön havaitsemisesta erittäin työlästä jopa ihmisannotoijille, minkä vuoksi on vaikea rakentaa suurta ja luotettavaa tietokokonaisuutta BIBREF10 .

**Tulos**

Millaisia esimerkkejä kirjoittajat antavat verkkoväkivallan kontekstiriippuvaisen luonteen aiheuttamista vaikeuksista?

**Esimerkki 5.1183**

Alan ehkä yllättävin tulos on kuitenkin Zipfin laki, jonka mukaan jos sanat asetetaan järjestykseen niiden esiintymistiheyden mukaan laajassa tekstissä, saadaan noin potenssilakiin perustuva jakauma kaikille kielille BIBREF0, BIBREF11. Toinen mielenkiintoinen tekstien luonnehdinta on Heaps-Herdanin laki, joka kuvaa sitä, miten sanavarasto eli erilaisten sanojen joukko kasvaa tekstin koon myötä, ja jonka määrän on empiirisesti havaittu kasvavan tekstin koon potenssina BIBREF18, BIBREF19. On syytä huomata, että on väitetty, että tämä laki on seurausta Zipfin laista. BIBREF20, BIBREF21

**Tulos**

Miten Zipfin ja Herdan-Heapin lait eroavat toisistaan?

**Esimerkki 5.1184**

Kuvaamme WikiSQL:ää koskevat sääntömme tässä. Sääntömme KBQA:ta varten on yksinkertainen ilman kuratoitua kartoitussanakirjaa. SequentialQA:n sääntöputki on samanlainen kuin WikiSQL:ssä.

**Tulos**

Ovatko säännöt tietokokonaisuuskohtaisia?

**Esimerkki 5.1185**

Vertailua varten luomme kaksi perusjärjestelmää kumpaankin suuntaan: toisessa käytetään perinteistä fraasipohjaista tilastollista konekääntämistä (SMT) ja toisessa NMT-järjestelmää.

**Tulos**

Mikä oli lähtötaso?

**Esimerkki 5.1186**

Koulutustietokannassa on 1018 väärennettyä uutista, joissa on kaksoiskappale, kun taas koulutuksessa on 322 aitoa sisältöä sisältävää artikkelia, joissa on kaksoiskappale.

**Tulos**

Mikä on tietokokonaisuuden koko?

**Esimerkki 5.1187**

Kehittämisprosessissa kehitettiin annotaatio-ohjeita ja koulutettiin annotaattoreita, jotka eivät olleet tietoturva-asiantuntijoita. Darkoden ja Hack Forumsin kehitys- ja testisarjat kommentoitiin muiden tiimin jäsenten toimesta (viisi Darkoden ja yksi Hack Forumsin osalta), minkä jälkeen kaikista erimielisyyksistä keskusteltiin ja ne ratkaistiin lopullisen annotaation laatimiseksi. Kirjoittajat, jotka ovat joko NLP:n tai tietoturvan tutkijoita, tekivät kaikki annotaatiot.

**Tulos**

Kuka kommentoi tietoja?

**Esimerkki 5.1188**

Kuten kuviossa KUVIO 1 havainnollistetaan, keskeinen ajatuksemme on, että voimme hyödyntää diskurssisuhteita BIBREF4 levittääksemme polariteettia tehokkaasti siemenpredikaateista, jotka ilmoittavat suoraan henkilön tunteista (esim. "olla iloinen" on positiivinen). Oletetaan, että tapahtumat $x\_1$ ja $x\_2$ ovat diskurssirelaatiossa Cause (eli $x\_1$ aiheuttaa $x\_2$). Jos siemenleksikon mukaan $x\_2$ on positiivinen, myös $x\_1$ on todennäköisesti positiivinen, koska se laukaisee positiivisen tunteen. Se, että $x\_2$:n tiedetään olevan negatiivinen, viittaa $x\_1$:n negatiiviseen napaisuuteen. Vastaavasti, jos $x\_1$ ja $x\_2$ ovat myönnytysdiskurssisuhteessa (eli $x\_2$ on $x\_1$:sta huolimatta), $x\_2$:n polariteetin käänteisyys voi levitä $x\_1$:een. Vaikka $x\_2$:n polariteettia ei tiedettäisikään etukäteen, voimme käyttää hyväksi sitä, että $x\_1$ ja $x\_2$ ovat joko samapolarisia (syy) tai käänteispolarisia (myönnytys), vaikka heuristiikka ei olekaan vapaa vastakohdista. Muunnamme tämän ajatuksen objektiivisiksi funktioiksi ja koulutamme neuroverkkomalleja, jotka ennustavat tietyn tapahtuman polariteetin.

**Tulos**

Miten suhteita käytetään polariteetin levittämiseen?

**Esimerkki 5.1189**

Ehdotettu multimodaalinen differentiaaliverkko (MDN) koostuu esitysmoduulista ja yhteisestä sekoitusmoduulista. Esitysmoduulissa käytetään triplettiverkkoa BIBREF41 , BIBREF42 . Triplettiverkko koostuu kolmesta osa-alueesta: kohde-, tuki- ja kontrastiverkko. Kaikilla kolmella verkolla on samat parametrit. Mixture-moduuli tuo kuvan ja kuvatekstin upotukset yhteiseen ominaisuuksien upotusavaruuteen.

**Tulos**

Miten kirjoittajat määrittelevät differentiaalisen verkon?

**Esimerkki 5.1190**

Tietokanta (joka koostuu 2,6 gigatavusta raakatekstiä) sisältää 421\ 829\ 960$ sanaa, jotka on jaettu 17\ 305\ 401$ lauseeseen.

**Tulos**

Kuinka suuri on aineisto, jota käytetään italian kielen Word2Vec-koulutukseen?

**Esimerkki 5.1191**

SVMRank on SVM:n muunnos, jossa jokaiselle datapisteelle annetaan pisteet ja tulokset voidaan asettaa paremmuusjärjestykseen ( BIBREF26 ). Käytämme SVMRankia seuraavissa kokeissa.

**Tulos**

mikä on heidän kehittämänsä valvottu malli?

**Esimerkki 5.1192**

Rokotuksiin suhtautuminen luokiteltiin seuraavasti: "kielteinen", "neutraali", "myönteinen" ja "ei selvä".

**Tulos**

Sallivatko ne sen, että rokotuksiin liittyviä avainsanoja sisältävät viestit ovat neutraaleja?

**Esimerkki 5.1193**

Morfologisen analyysin tieto-ohjatut mallit rakennetaan käyttäen INLINEFORM0-koulutusdataa, joka koostuu INLINEFORM1-koulutusesimerkeistä. Vertailemamme perusmalli BIBREF5 pitää mallin lähtöavaruutta osajoukkona INLINEFORM2, jossa INLINEFORM3 on kaikkien tässä harjoitusaineistossa nähtyjen tunnistejoukkojen joukko. Tarkemmin sanottuna ne ratkaisevat tehtävän moniluokkaisena luokitusongelmana, jossa luokat ovat yksittäisiä tunnistejoukkoja. Vähäisten resurssien skenaarioissa tämä osoittaa, että INLINEFORM4 ja jopa niiden INLINEFORM5:ssä esiintyvien tagijoukkojen osalta olemme saattaneet nähdä hyvin vähän harjoitusesimerkkejä. Tunnistejoukkojen sekvenssin ehdollinen todennäköisyys lauseen perusteella muotoillaan 0. kertaluvun CRF:nä. DISPLAYFORM0

**Tulos**

Mihin muihin monikielisiin lähestymistapoihin mallia verrataan?

**Esimerkki 5.1194**

Kytean käytön jälkeen japaninkielisiin teksteihin sovelletaan LSW-algoritmia OOV-sanojen korvaamiseksi niiden synonyymeillä.

**Tulos**

Otetaanko japanilais-vietnamilaisessa tehtävässä huomioon synonyymiset suhteet?

**Esimerkki 5.1195**

Mallin syötteenä on sanan upotuksen ja toisen upotuksen yhdistelmä, joka osoittaa, onko tämä sana predikaatti: $ \mathbf {x}\_t = [\mathbf {W}\_{\text{emb}}(w\_t), \mathbf {W}\_{\text{mask}}(w\_t = v)]. $

**Tulos**

Mikä on OpenIE-tuplien syöttöesitys malliin?

**Esimerkki 5.1196**

Tässä artikkelissa käsittelemme affektiivisen tapahtuman polariteetin tunnistamista, jota edustaa pistemäärä, joka vaihtelee arvosta $-1$ (negatiivinen) arvoon 1 (positiivinen).

**Tulos**

Mitä merkintöjä on saatavilla tietokannassa valvontaa varten?

**Esimerkki 5.1197**

GlossBERT-ohjelmassa GlossBERT-ohjelmalla voidaan hyödyntää glossitietoa täysimääräisesti ja muodostaa konteksti-glossi-parit kaikista WordNetin mahdollisista kohdesanan merkityksistä, jolloin WSD-tehtävää voidaan käsitellä lauseparien luokittelun ongelmana.

**Tulos**

Sisällytetäänkö WordNet malliin?

**Esimerkki 5.1198**

Tällöin dekoodausfunktio on lineaarinen projektio, joka on $= f\_{\text{de}}()=+ $ , jossa $\in ^{d\_\times d\_}$ on koulutettava painomatriisi ja $\in ^{d\_\times 1}$ on harhautustermi. NICE BIBREF -ohjelmassa17 suunniteltiin bijektiivisten muunnosten perhe, ja yksinkertaisin jatkuva bijektiivinen funktio $f:^D\rightarrow ^D$ ja sen käänteisfunktio $f^{-1}$ määritellään seuraavasti:$$h: \hspace{14.22636pt} \_1 &= \_1, & \_2 &= \_2+m(\_1) \luku \\\ h^{-1}: \hpace{14.22636pt} \_1 &= \_1, & \_2 &= \_2-m(\_1) \nonumber $$ (Yht. 15)missä $\_1$ on $d$ -ulotteinen osio syötteestä $\in ^D$ ja $m:^d\rightarrow ^{D-d}$ on mielivaltainen jatkuva funktio, joka voi olla koulutettava monikerroksinen feedforward-neuraaliverkko, jossa on epälineaarisia aktivointifunktioita. Sitä kutsutaan `additiiviseksi kytkentäkerrokseksi' BIBREF17 , jolla on yksikkö-Jacobin determinantti. Jotta oppimisjärjestelmä voisi tutkia tehokkaampia muunnoksia, noudatamme `affiinisen kytkentäkerroksen' BIBREF24 :$$h: \hspace{5.69046pt} \_1 &= \_1, & \_2 &= \_2 \odot \text{exp}(s(\_1)) + t(\_1) \nonumber \\\ h^{-1}: \hspace{5.69046pt} \_1 &= \_1, & \_2 &= (\_2-t(\_1)) \odot \text{exp}(-s(\_1)) \nonumber $$ (Yht. 16)missä $s:^d\rightarrow ^{D-d}$ ja $t:^d\rightarrow ^{D-d}}$ ovat molemmat neuroverkkoja, joilla on lineaariset ulostuloyksiköt.Jatkuvan bijektiivisen muunnoksen vaatimuksena on, että syötteen $$ ja ulostulon $$ dimensioiden on täsmälleen vastattava toisiaan.

**Tulos**

Mitkä ovat nämä kaksi dekoodaustoimintoa?

**Esimerkki 5.1199**

Havaitsemme mielenkiintoisia piilokorrelaatioita tiedoissa. Kuvassa FIGREF24 on valittu aihe 2. Aihealue 2 sisältää neljä eniten esiintyvää avainsanaa: "vegaani", "jooga", "työ", "every\_woman", joilla on suurin termifrekvenssi. Aiheesta voidaan päätellä erilaisia asioita, kuten "naiset harrastavat yleensä joogaa enemmän kuin miehet", "naiset opettavat joogaa ja pitävät sitä työnä", "joogi noudattaa vegaanista ruokavaliota". Sanoisimme, että tiedoissa on havaittavissa korrelaatioita, eli `jooga-veganismi', `naiset-jooga'. Naiset-jooga

**Tulos**

Mitä muita mielenkiintoisia korrelaatioita on havaittu?

**Esimerkki 5.1200**

Korpus sisältää 2000 lausetta.

**Tulos**

Kuinka pitkä tietokokonaisuus on?

**Esimerkki 5.1201**

Suorituskyvyn heikkenemisen syyn ymmärtämiseksi on ensin ymmärrettävä, miten yhteistyöelimen on käsiteltävä syötettyjä tekstitietoja. Yhteistyöelin käyttää WordPiece-tokenisaattoria tekstin tokenisointiin. WordPiece-tokenizer-lausekkeet perustuvat pisimmän etuliitteen vastaavuusalgoritmiin, jolla luodaan merkkejä . Näin saadut merkit syötetään BERT-mallin syötteeksi. WordPiece tokenizer käyttäytyy erittäin mielenkiintoisesti, kun kyseessä on meluisan datan merkitseminen. Kirjoitusvirheiden vuoksi näitä sanoja ei löydy suoraan BERTin sanakirjasta. Näin ollen WordPiece-tokenizer jakaa meluiset sanat alasanoiksi. Se päätyy kuitenkin hajottamaan ne osasanoiksi, joiden merkitys voi poiketa hyvin paljon alkuperäisen sanan merkityksestä. Usein tämä muuttaa lauseen merkityksen kokonaan, mikä johtaa huomattavaan suorituskyvyn heikkenemiseen.

**Tulos**

Mikä on syy suorituskyvyn heikkenemiseen, kun käytetään BERT:iä joissakin suosituissa tehtävissä?

**Esimerkki 5.1202**

Samalla arkkitehtuurilla RCRN:n suorituskyky on parempi kuin ablatiivisten perusratkaisujen BiLSTM by INLINEFORM2 ja 3L-BiLSTM by INLINEFORM3 keskimäärin 16 tietokokonaisuudessa.

**Tulos**

Kuinka paljon paremmat ne ovat kuin BiLSTM:t tunneanalyysissä?

**Esimerkki 5.1203**

Varmistaaksemme, että menetelmämme (eli monilinjainen huomio) korvaa Transformerissa monipäähuomion, suoritamme kaksi NLP-tehtävää, jotka ovat kielten mallintaminen (LM) ja neuraalinen konekääntäminen (NMT). Sen jälkeen testaamme eri mallikonfiguraatioita PTB BIBREF25 , WikiText-103 BIBREF26 ja One-Billion Word benchmark BIBREF27 -tietokannoissa ja raportoimme tulokset taulukoissa 1 ja 2. Tässä tehtävässä olemme kouluttaneet Transformer-mallin BIBREF2 WMT 2016 English-German datasetillä BIBREF36 .

**Tulos**

Millaisilla tietokokonaisuuksilla tai tehtävillä he tekevät kokeita?

**Esimerkki 5.1204**

Käytämme Dialogue State Tracking Competition 2 (DSTC2) -tietokokonaisuutta BIBREF27 , joka on laajimmin käytetty tietokokonaisuus tehtäväkeskeisten chatbottien tutkimuksessa. Käytimme myös kahta muuta Google Research BIBREF28 -tietokokonaisuutta, jotka Google Research on hiljattain avannut. Nämä ovat M2M-sim-M (elokuva-alan tietokokonaisuus) ja M2M-sim-R (ravintola-alan tietokokonaisuus).

**Tulos**

Mitä kolmea tietokokonaisuutta käytetään?

**Esimerkki 5.1205**

Ensimmäisessä skenaariossa dekooderi varustetaan ylimääräisellä morfologiataulukolla, joka sisältää kohdepuolen affiksit. Asetamme taulukon päälle huomiomoduulin, jota dekooderi ohjaa. Kun dekooderi ottaa näytteen merkistä, se etsii jokaisessa vaiheessa taulukosta olennaisimmat tiedot, jotka voivat rikastuttaa sen tilaa. Taulukosta lähetettyjä signaaleja voidaan tulkita lisärajoituksina. Toisessa skenaariossa dekooderi jaetaan kahden lähtökanavan kesken. Ensimmäinen ottaa näytteen kohdemerkistä ja toinen ennustaa merkin morfologisen merkinnän. Tämä monitehtäväinen lähestymistapa pakottaa dekooderin lähettämään morfologiatietoista tietoa viimeiselle kerrokselle, mikä johtaa parempiin ennusteisiin. Kolmannessa skenaariossa yhdistämme nämä kaksi mallia.

**Tulos**

Miten morfologiataulukon apusignaalit sisällytetään dekooderiin?

**Esimerkki 5.1206**

Viimeisenä mutta ei vähäisimpänä eettisyys ja oikeudenmukaisuus ovat tärkeitä näkökohtia, joita on syytä tutkia. Tässä mielessä yksilöllisten ja yleisten vääristymien havaitseminen olisi asetettava etusijalle, jotta käytännön toimijoille voidaan antaa hyödyllistä palautetta. Lisäksi harkitsemme vastakkaisoppimisen käyttöä, kuten BIBREF:ssä33 , jotta voidaan varmistaa oikeudenmukaisuus koulutusprosessin aikana.

**Tulos**

Onko tutkimuksessa otettu huomioon eettisiä näkökohtia?

**Tulos**

Analysoidaanko he, onko heidän järjestelmässään minkäänlaista puolueellisuutta?

**Esimerkki 5.1207**

Mallien kouluttamiseen on käytetty kolmea eri tietokokonaisuutta: Toronton kirjakorpusta, Wikipedian lauseita ja twiittejä. Sent2Vec-mallimme ovat myös keskimäärin parempia tai samantasoisia kuin C-PHRASE-malli, vaikka ne jäävätkin huomattavasti jälkeen STS 2014 WordNet- ja News-osatehtävissä. Tämä havainto johtuu siitä, että suuri osa C-PHRASE-mallin koulutettavasta datasta on peräisin englanninkielisestä Wikipediasta, mikä auttaa sitä suoriutumaan hyvin määritelmiä ja uutisia sisältävissä aineistoissa.

**Tulos**

Raportoidaanko tulokset vain englanninkielisistä tiedoista?

**Esimerkki 5.1208**

Osoitimme BIBREF22:ssa, että Pearsonin korrelaatio soveltuu huonosti STS:ään. Sen sijaan laskemme Spearmanin korrelaation lauseiden upotusten ja kultaisten etikettien kosinus-similaarisuuden välillä.

**Tulos**

Mitä mittareita käytetään STS-tehtävissä?

**Esimerkki 5.1209**

Arvioimme hiljattain ehdottamiamme malleja ja niihin liittyviä perusmalleja useissa eri kielten vähäresurssiympäristöissä todellisilla, etäisesti valvotuilla aineistoilla, joissa ei ole synteettistä kohinaa. Meluisten merkintöjen edistynyt mallintaminen parantaa suorituskykyä huomattavasti, jopa 36 prosenttia verrattuna menetelmiin, joissa ei käsitellä kohinaa, ja jopa 9 prosenttia verrattuna kaikkiin muihin kohinan käsittelyä käyttäviin perusmenetelmiin.

**Tulos**

Arvioivatko he lähtötasoa vasten?

**Esimerkki 5.1210**

Etusijalle asetetaan merkinnät, jotka ovat täydellisesti yhteensopivia kaikkien neuraalisten annotaattoreiden kanssa. Taulukosta TABREF11 nähdään, että molemmissa tietokokonaisuuksissa on noin 40 prosenttia täsmälleen yhteensopivia merkintöjä kaikissa malleissa (AM). Tämän jälkeen etusijalle asetetaan kontekstipohjaiset mallit, jotta voidaan tarkistaa, täsmäävätkö merkinnät täydellisesti kaikissa kontekstimalleissa. Jos kaksi kolmesta kontekstimallista on oikein, tarkistetaan, onko kyseinen etiketti tuotettu myös vähintään yhdellä kontekstittomalla mallilla. Sitten sallitaan, että etiketit luottavat näihin vähintään kahteen kontekstimalliin. Tämän tuloksena noin 47 prosenttia etiketeistä otetaan kontekstimallien (CM) perusteella. Kun huomaamme, että mikään kontekstimalli ei tuota samoja tuloksia, asetamme etiketit paremmuusjärjestykseen niiden luottamusarvojen avulla, jotka on tuotettu todennäköisyysjakaumana käyttäen $softmax$-funktiota. Merkinnät lajitellaan luottamusarvojen mukaan laskevaan järjestykseen. Tämän jälkeen tarkistetaan, että kolme ensimmäistä (kun yksi kontekstimalli ja molemmat muut kuin kontekstimallit tuottavat saman merkinnän) tai vähintään kaksi merkintää vastaavat toisiaan, ja valitaan kyseinen merkintä. IEMOCAP:ssa niitä on noin 3 % ja MELD:ssä (BM) 5 %. Jos mikään edellä mainituista ehdoista ei täyty, jätetään tuntemattoman kategorian sisältävä etiketti pois. Lopullisissa merkinnöissä tämä tuntematon kategoria merkitään merkinnällä `xx', ja niitä on noin 7 % IEMOCAPissa ja 11 % MELDissä (NM).

**Tulos**

Miten ensemble-merkitsijä poimii lopullisen merkinnän?

**Esimerkki 5.1211**

Kaivannaisteollisuuden rakentaminen CNN/Daily Mail

**Tulos**

Mikä on nykyisten mittareiden ongelma, johon he yrittävät puuttua?

**Esimerkki 5.1212**

Viime aikoina SMT-järjestelmiin on sisällytetty sanojen ja lausekkeiden jatkuvatoimisia avaruusesityksiä neuroverkkojen avulla. Yksikielisten neuroverkkojen kielimallien BIBREF13 , BIBREF14 , neuroverkkojen yhteismallien (neural network joint models, NNJM) BIBREF4 ja neuroverkkojen globaalien leksikonimallien (neural network global lexicon models, NNGLM) BIBREF3 lisääminen on osoittautunut käyttökelpoiseksi SMT:ssä Kielioppivirheiden korjaus (GEC) on haastava tehtävä virheiden tyypin vaihtelevuuden ja virheiden syntaktisten ja semanttisten riippuvuuksien vuoksi. Useimmissa kielioppivirheiden korjausjärjestelmissä käytetään luokittelua ja sääntöpohjaisia lähestymistapoja tiettyjen virhetyyppien korjaamiseen. Nämä järjestelmät käyttävät kuitenkin useita kielellisiä vihjeitä ominaisuuksina. Tavanomaiset kielianalyysityökalut, kuten POS-taggerit (part-of-speech) ja jäsentimet, on usein koulutettu hyvin muodostetulle tekstille, ja ne toimivat huonosti epäkieliopillisessa tekstissä. Tämä aiheuttaa lisää virheitä ja rajoittaa sääntöpohjaisten ja luokitteluun perustuvien lähestymistapojen suorituskykyä GEC:ssä. Tämän seurauksena lausepohjainen tilastollinen konekääntäminen (SMT) on saavuttanut suosiota, koska se pystyy oppimaan tekstin muunnoksia virheellisestä tekstistä oikeaksi tekstiksi virheiden korjaamista rinnakkaisista korpuksista ilman kielellistä lisätietoa. Teemme kokeita sisällyttämällä NNGLM:n ja NNJM:n sekä itsenäisesti että yhdessä perusjärjestelmään Edellä kuvatun perusjärjestelmän lisäksi sisällytämme ominaisuuksina kaksi neuroverkkokomponenttia, neuroverkon globaalin leksikonin mallin (NNGLM) ja neuroverkon yhteisen mallin (NNJM). Sekä NNGLM että NNJM koulutetaan käyttämällä perusjärjestelmämme käännösmallin kouluttamiseen käytettyä rinnakkaista dataa.

**Tulos**

Miten ne yhdistävät kaksi ehdotettua neuroverkkomallia?

**Esimerkki 5.1213**

Havaitsemme, että yhteisöissä, joille on ominaista erikoistunut, jatkuvasti päivittyvä sisältö, on korkeampi käyttäjien pysyvyysaste, mutta niissä on myös suurempia kielellisiä eroja, jotka erottavat uudet tulokkaat vakiintuneista jäsenistä. Kun tarkastellaan tarkemmin tekijöitä, jotka voivat vaikuttaa tähän kielelliseen eroon, havaitaan, että erityisesti omaleimaisissa yhteisöissä vakiintuneilla käyttäjillä on suurempi taipumus sitoutua yhteisön erikoissisältöön kuin uusilla tulokkailla (SECREF5 ).

**Tulos**

Millaisia malleja he havaitsevat siitä, miten käyttäjien sitoutuminen vaihtelee yhteisön ominaisuuksien mukaan?

**Esimerkki 5.1214**

Koulutamme kolme järjestelmää (S1, S2 ja S3) taulukossa TABREF5 esitetyillä korpuksilla. Kaksi ensimmäistä järjestelmää ovat muunnosmalleja, jotka on koulutettu erilaisilla tietomäärillä (6M vs. 18M rinnakkaista lausetta, kuten taulukosta näkyy). Kolmas järjestelmä sisältää muutoksen, jossa otetaan huomioon koko asiakirjan täydelliset ydinviittausketjut ja täydennetään käännettävää lausetta tällä tiedolla, ja se on koulutettu samalla määrällä lausepareja kuin S1.

**Tulos**

Mitkä kolme neuraalista konekäännösjärjestelmää analysoidaan?

**Esimerkki 5.1215**

Jotkin kirjoittajat käyttävät esivalmisteltuja upotuksia (erityisesti silloin, kun heidän aineistokokonaisuutensa on liian pieni omien upotusten kouluttamiseen) tai yrittävät muuttaa näitä upotuksia ja mukauttaa ne omaan aineistokokonaisuuteensa. Näiden lähestymistapojen suurin haittapuoli on kuitenkin se, että upotusten kouluttamiseen käytettävä korpus ei välttämättä liity tiettyyn tehtävään, jossa upotuksia käytetään. Monet lääketieteelliset käsitteet eivät sisälly tunnettuihin sulautumispohjiin. Lisäksi sanojen samankaltaisuus voi vaihdella eri yhteyksissä. Tämän jälkeen lasketaan käsitteiden upotukset (GloVe-menetelmällä) erikseen haastattelukuvauksille ja tutkimuskuvauksille.

**Tulos**

Hienosäätävätkö he lääketieteellisissä teksteissään käyttämiään sanojen upotuksia?

**Esimerkki 5.1216**

Meidän määritelmämme puolueellisuudesta on nyt: [Tulkintaharha episteemisessä ME-pelissä on todennäköisyysjakauma tyyppeihin, jonka keskustelijoiden tai pelaajien tai tuomariston uskomusfunktio antaa. Huomaa, että ME-pelissä on tyypillisesti useita tulkinnallisia ennakkoluuloja: jokaisella pelaajalla on oma ennakkoluulonsa, samoin kuin tuomaristolla.

**Tulos**

Mitä tulkinnallisia ennakkoluuloja tässä asiakirjassa analysoidaan?

**Esimerkki 5.1217**

DDI-korpus sisältää tuhansia XML-tiedostoja, joista jokainen koostuu useista tietueista. Lauseessa, joka sisältää INLINEFORM0-lääkkeitä, on INLINEFORM1-lääkepareja.

**Tulos**

Kuinka suuri on arvioitava tietokokonaisuus?

**Esimerkki 5.1218**

Koska poimimme vain viittaukset muihin oikeuden päätöksiin, saimme 471 319 viittausta korkeimman oikeuden päätöksistä, 167 237 viittausta korkeimman hallinto-oikeuden päätöksistä ja 264 463 viittausta perustuslakituomioistuimen päätöksistä. Nämä ovat niiden tekstikatkelmien lukumäärät, jotka tunnistettiin viittauksiksi ennen SECREF3-jaksossa kuvattua jatkokäsittelyä.

**Tulos**

Kuinka suuri tietokokonaisuus on?

**Esimerkki 5.1219**

Tulokset esitetään taulukossa TABREF11 . Tulokset osoittavat, että itsetarkkailumekanismin sisällyttäminen koodaajaan on hyödyllistä useimmissa tehtävissä. Alkuperäiset mallit olivat kuitenkin parempia joissakin tehtävissä (CR, MPQA, MRPC), mikä viittaa siihen, että itsetarkkailumekanismi voi joskus aiheuttaa kohinaa lauseiden ominaisuuksiin. Kaiken kaikkiaan itsehuomioivan lause-esityksen hyödyntäminen parantaa suorituskykyä viidessä tehtävässä kahdeksasta.

**Tulos**

Kuinka paljon parempi suorituskyky ehdotetulla menetelmällä on verrattuna perusmenetelmiin?

**Esimerkki 5.1220**

Asiakirjojen esikäsittelyssä kokeillaan kolmea vaiheittaista tasoa: raakateksti, tekstin puhdistus asiakirjan loogisen rakenteen tunnistamisen avulla ja avainsanojen harvojen osien poistaminen asiakirjasta. Näin esitämme ensimmäisen johdonmukaisen vertailun eri avainsanojen poimintamalleista ja tutkimme niiden kestävyyttä kohinaisten tekstien käsittelyssä.

**Tulos**

mitä asiakirjojen esikäsittelyn tasoja tarkastellaan?

**Esimerkki 5.1221**

Tätä mittaria käytetään arvioitaessa vuoropuhelun uskomustenseurantajärjestelmän BIBREF1 tarkkuutta.APRA: APRA-mittarilla (Action Per-Response Accuracy) arvioidaan vuoropuhelupolitiikan laatijan tuottamien vuoropuhelutoimien tarkkuutta kierrosta kohden. Perusasetusten osalta APRA:lla arvioidaan dialogipolitiikan laatijan luokittelutarkkuutta. Mutta meidän mallimme tuottaa itse asiassa jokaisen yksittäisen toimintamerkin, ja pidämme ennustusta oikeana vain, jos mallin tuotoksen jokainen merkki vastaa vastaavaa merkkiä perustotuudessa.BLEU BIBREF19: Mittari arvioi luonnollisen kielen generaattorin tuottaman lopullisen vastauksen laatua. Mittaria käytetään yleensä tehtäväkeskeisen dialogijärjestelmän suorituskyvyn mittaamiseen.

**Tulos**

Mitä mittareita käytetään mallien suorituskyvyn mittaamiseen?

**Esimerkki 5.1222**

Kuvaamme ensin mallit yksitellen ja sitten käyttämämme yhdistämistekniikan. Seuraavassa MN tarkoittaa Memory Networks -verkkoja keskusteluhistorian koodaamiseksi, RCNN tarkoittaa R-CNN:ää kuvan objektitason esityksiä varten, Wt tarkoittaa lineaarista lisäkerrosta dekooderissa ja LF myöhäistä fuusiomekanismia, joka on määritelty BIBREF0.Models ::: LF-RCNN Myöhäinen fuusiokooderi BIBREF0, jossa on yhdistetty historia. Käytämme kaksikerroksisia LSTM-muisteja, joissa on 512 piilotettua yksikköä kysymysten ja historian upottamiseen. Objektitason piirteet punnitaan käyttämällä vain kysymysten upotuksia. Glove-vektoreista saadut sanojen upotukset jäädytetään, eikä niitä hienosäädetä. Kuvassa FIGREF6 on yleiskuva arkkitehtuurista: MN-RCNNMuistiverkkokooderi BIBREF0, jossa on kaksisuuntaiset GRU:t ja sanojen upotukset hienosäädetty. Objektitason piirteitä punnitaan kysymysten ja kuvatekstien upotuksilla. Muu osa järjestelmästä on sama kuin edellä. (Kuva FIGREF6)Mallit ::: MN-RCNN-WtSama kuin edellä, mutta ylimääräinen lineaarinen kerros, jota sovelletaan vastausehdokkaan ja koodaimen tuloksen pistetuotokseen, ja joka ohjataan tanh-funktion avulla. Vertaa kuvaa FIGREF6 kuvaan FIGREF6.

**Tulos**

Mitä kolmea erottelumallia he käyttivät?

**Esimerkki 5.1223**

Tätä haastetta silmällä pitäen esittelemme Torch-Structin, jolla on kolme erityistä panosta:Modulaarisuus: mallit esitetään jakaumina, joilla on syväoppimiskehykseen integroitu joustava standardikäyttöliittymä.Täydellisyys: laaja valikoima klassisia algoritmeja on toteutettu, ja uusia malleja voidaan helposti lisätä Pythonissa.Tehokkuus: toteutukset on suunnattu laskennalliseen/muistitehokkuuteen grafiikkasuorittimille, ja taustapohjaan sisältyy laajennuksia optimointia varten.

**Tulos**

Onko tämä kirjasto toteutettu Torchiin vai onko se kehysriippumaton?

**Esimerkki 5.1224**

Teimme kaksi inhimillistä arviointia Mechanical Turk -palvelun avulla vertaillaksemme mallimme ja perustason suorituskykyä.

**Tulos**

Liittyykö tämän perhetyön arviointiin mitään inhimillistä arviointia?

**Esimerkki 5.1225**

Kieliopin matemaattisen rakenteen tutkiminen on osoittanut, että lauseiden perustana eivät ole sanat vaan jotkin atomiset kieliopilliset tyypit, kuten substantiivi- ja lausetyyppi BIBREF23 , BIBREF24 , BIBREF25 . Transitiiviverbityyppi ei ole atomaattinen kieliopillinen tyyppi, vaan kahdesta substantiivityypistä ja yhdestä lausetyypistä koostuva komposiitti. Näin ollen erityisen mielenkiintoista tässä on se, että atomisella ei todellakaan tarkoiteta pienintä... Toisaalta, aivan kuten hiukkasfysiikassa, jossa meillä on hiukkasia ja antihiukkasia, atomisiin tyyppeihin kuuluu sekä tyyppejä että anti-tyyppejä. Mutta toisin kuin hiukkasfysiikassa, antityyppejä on kahdenlaisia, nimittäin vasemmanpuoleisia ja oikeanpuoleisia. Tämä tekee kielestä vieläkin ei-kommutatiivisemman kuin kvanttiteoria!

**Tulos**

Jaetaanko sanojen merkitykset alkeishiukkasiksi kuten kvanttiteorian standardimallissa?

**Esimerkki 5.1226**

Toiseksi laskemme kolme metriikkaa poimitulle tiedolle: $\bullet $ Relation Generation (RG) arvioi, kuinka hyvin järjestelmä pystyy tuottamaan tekstiä, joka sisältää faktatietoja (eli oikeita tietueita). Mittaamme niiden $\hat{y}\_{1:T}$:sta poimittujen ainutlaatuisten relaatioiden $r$ tarkkuutta ja absoluuttista lukumäärää (RG-P% ja RG-#), jotka esiintyvät myös $s$:s$:ssä.$\bullet $ Sisällönvalinta (CS) mittaa, kuinka hyvin luotu dokumentti vastaa kultaista dokumenttia mainittujen tietueiden osalta. Mitataan niiden $\hat{y}\_{1:T}$:sta poimittujen ainutlaatuisten suhteiden $r$ tarkkuus ja palautus (merkitään vastaavasti CS-P% ja CS-R%), jotka on poimittu myös $y\_{1:T}$:sta.$\bullet $ Content Ordering (CO) analysoi, kuinka hyvin järjestelmä järjestää kuvauksessa käsitellyt tietueet. Mittaamme normalisoidun Damerau-Levenshteinin etäisyyden BIBREF36 niiden $\hat{y}\_{1:T}$:sta poimittujen tietueiden jaksojen välillä, jotka on poimittu myös $y\_{1:T}$:sta.

**Tulos**

Mitä laadullisia mittareita käytetään arvioinnissa?

**Esimerkki 5.1227**

Ensimmäisenä perustasona käytämme Bag-of-Wordsia, joka on tunnettu ja vankka tekstin esitysmuoto, jota käytetään eri aloilla BIBREF21 , yhdistettynä tavalliseen matalaan luokittimeen eli lineaarisella ytimellä varustettuun tukivektorikoneeseen. Käytimme SVM:n LIBSVM-toteutusta.

**Tulos**

Mitä matalia lähestymistapoja he kokeilivat?

**Esimerkki 5.1228**

Vaikka luomme uusia malleja esimerkiksi tunneanalyysin ja sukupuolen tunnistamisen kaltaisiin tehtäviin osana AraNetiä, keskitymme enemmänkin itse työkalupaketin kokoamiseen ja vahvojen vertailukelpoisten perusmallien luomiseen. Vaikka tarjoammekin joitakin perusmalleja joihinkin tehtäviin, emme vertaile niitä nimenomaisesti aiempaan tutkimukseen, koska useimmissa nykyisissä töissä joko hyödynnetään pienempiä aineistoja (jolloin vertailu ei ole oikeudenmukaista) tai käytetään BERT:iä edeltäviä menetelmiä (jolloin meidän mallimme todennäköisesti päihittävät ne). Monille mallintamillemme tehtäville ei ole ollut vakiomuotoisia vertailuarvoja, joiden avulla malleja voitaisiin vertailla keskenään. Tämä vaikeuttaa edistymisen mittaamista ja sellaisten alojen tunnistamista, joille kannattaa kohdentaa ponnisteluja ja määrärahoja.

**Tulos**

Mihin malleihin niitä verrattiin?

**Esimerkki 5.1229**

Hyödynsimme olemassa olevaa, annotoitua Twitter-tietokokonaisuutta, joka on rakennettu masennukseen liittyvien oireiden hierarkkisen mallin BIBREF12 , BIBREF13 perusteella. Tietokanta sisältää 9 473 annotaatiota 9 300 twiitille.

**Tulos**

Mitä tietokokonaisuutta tässä tutkimuksessa käytetään?

**Esimerkki 5.1230**

Koulutustietokanta sisältää 2 815 esimerkkiä, joista 1 940 (eli 69 %) on valeuutisia ja 1 968 (eli 70 %) klikkaushuijauksia; lisäksi meillä on 761 testausesimerkkiä.

**Tulos**

mitä tietokokonaisuuksia käytettiin?

**Esimerkki 5.1231**

Teemme kokeita itse keräämillämme CAIS-tietokannoilla arvioidaksemme niiden yleistettävyyttä eri kielillä. Vertailussa käytetään kahta perusmallia, joista toinen on suosittu BiLSTM + CRF-arkkitehtuuri BIBREF36 sekvenssien merkintätehtävään ja toinen tehokkaampi sententce-state LSTM BIBREF21. Taulukossa TABREF50 luetellut tulokset osoittavat CM-verkkomme yleistettävyyden ja tehokkuuden eri alojen ja eri kielten käsittelyssä.

**Tulos**

Mitkä olivat perusmallit?

**Esimerkki 5.1232**

Tutkimme lisäksi INLINEFORM0 -grammin pituuden vaikutusta malliimme (eli huomion jäsentämismoduulin käyttämien covolutional-kerrosten suodatinkokoa). Kuvassa FIGREF39 esitetään linkkien ennustamisen AUC-pisteet Cora-tietokannassa INLINEFORM1 -grammin pituuden vaihtelun suhteen.

**Tulos**

Mittaavatko ne erityisesti sitä, miten hyvin ne suoriutuvat pidemmistä jaksoista?

**Esimerkki 5.1233**

Lopuksi suoritimme hyperparametrin virityksen säätämällä yhteistarkkailulohkojen lukumäärää, eräkokoa ja koulutettujen epookkien lukumäärää ja yhdistimme kolme parasta verkkoamme.

**Tulos**

Mitä hyperparametreja on viritetty?

**Esimerkki 5.1234**

Vertasimme mallejamme seuraaviin uusimpiin perusmalleihin: Sequence to Sequence (Seq2Seq): Hierarkkinen LSTM (HLSTM): Yksinkertainen koodaaja-dekooderimalli, joka yhdistää neljä lausetta pitkäksi lauseeksi huomiomekanismin avulla BIBREF31: Tarinan kontekstiä edustaa hierarkkinen LSTM: sanatason LSTM kullekin lauseelle ja lausetason LSTM, joka yhdistää neljä lausetta BIBREF29 . Sovelletaan hierarkkista huomiomekanismia, joka huolehtii vastaavasti kahden LSTM:n tiloista. HLSTM+Copy: Hierarkkisiin tiloihin sovelletaan kopiointimekanismia BIBREF32 , jolla kopioidaan tarinan kontekstissa olevat sanat generointia varten.HLSTM+Graph Attention(GA): Sovellamme monilähteistä HLSTM-huomiota, jossa commonsense-tieto koodataan graafihuomion avulla.HLSTM+Contextual Attention(CA): Sovelletaan kontekstuaalista huomiota commonsense-tiedon esittämiseen.

**Tulos**

Vertasivatko ne Transformer-pohjaisiin suuriin kielimalleihin?

**Tulos**

Mitä perusviivoja he käyttävät?

**Esimerkki 5.1235**

Käytämme vakiomittareita ROUGE-1, ROUGE-2 ja ROUGE-L BIBREF29 kaikkien yhteenvetomallien arvioimiseen. BIBREF22:n ja BIBREF23:n mukaisesti käytämme F-mittausta ROUGE XSUM:n ja CNN/DailyMailin osalta ja rajoitetun pituuden recall-mittausta ROUGE:n osalta NYT:n ja DUC:n osalta.

**Tulos**

Mitä mittaria käytettiin arviointivaiheessa?

**Esimerkki 5.1236**

Epäilemme, että tämän epäjohdonmukaisuuden syynä on dekoodauksen puuttuminen maksimiluotettavuusestimoinnista, ja ehdotamme, että tulevaisuudessa vaihtoehtona tutkitaan sekvenssitason oppimista. Epäjohdonmukaisuus voi johtua dekoodauksen puuttumisesta tämän optimointiongelman ratkaisemisessa. Maximum likelihood -oppiminen sovittaa mallia $p\_{\theta }$ käyttäen datan jakaumaa, kun taas koulutetusta mallista dekoodattu sekvenssi noudattaa dekoodausalgoritmin aiheuttamaa jakaumaa $q\_{\mathcal {F}}$. Tämän ristiriidan perusteella esitämme vahvan olettamuksen: emme voi taata, että saamme hyvän johdonmukaisen sekvenssigeneraattorin käyttämällä maksimaalisen todennäköisyyden oppimista ja ahnetta dekoodausta.

**Tulos**

Onko äärettömän pituisten sekvenssien tuottaminen seurausta maksimaalisen todennäköisyyden harjoittelusta?

**Esimerkki 5.1237**

Lähestymistapamme koostuu muuntajaverkon syötteen jäsentämisestä muuntajien omatoimisen huomion käyttämiseksi ja ohjaamiseksi, jolloin se ehdollistuu kokonaisuuteen. Pääasiallinen tapamme koodata syötettä, eli entiteetti-ensimmäinen menetelmä, on esitetty kuvassa KUVA 4. Meillä on myös entiteetti-viimeinen muunnos, jossa entiteetti havaitaan ensisijaisesti juuri ennen luokittelumerkkiä, jotta [CLS]-merkki ehdollistuu sen mukaisesti [CLS]-merkin itsehuomioon. Lisävariaationa voimme joko ajaa muuntajan kerran asiakirjaa kohti useilla [CLS]-merkkeillä (asiakirjatason malli, kuten kuvassa KUVIO 4) tai erikoistaa ennusteen yhteen aikavaiheeseen (lausetason malli).

**Tulos**

Millä tavoin syötteet on järjestetty uudelleen?

**Esimerkki 5.1238**

Ottaen huomioon etiikan ja moraalin monimutkaisuuden ja koskemattomuuden rajoitumme tämän rakenteen melko yksinkertaiseen toteutukseen deontologisen etiikan teorioiden mukaisesti. Niissä kysytään, mitkä valinnat ovat moraalisesti välttämättömiä, kiellettyjä tai sallittuja sen sijaan, että kysyttäisiin, millainen ihmisen tulisi olla tai mitä tekojemme seurauksia tulisi suosia. Siten normit ymmärretään universaaleiksi säännöiksi siitä, mitä tulee tehdä ja mitä ei tule tehdä. Siksi keskitymme sosiaalisen hyväksynnän arvostukseen yksittäisissä verbeissä ja yksittäisissä verbeissä, joihin liittyy ympäröivää kontekstitietoa - esim. luota ystävääni tai luota koneeseen - selvittääksemme, mitkä niistä edustavat tekemistä ja mitkä ovat yleensä tekemättä jättämistä.

**Tulos**

Miten kirjoittajat määrittelevät deontologisen eettisen ajattelun?

**Esimerkki 5.1239**

Fenerbahçe Olemme päättäneet pitää suosituista urheiluseuroista kertovia twiittejä verkkotunnistuksemme kohteena. Näin ollen olemme määritelleet kohteiksemme Galatasarayn (eli Kohde-1) ja Fenerbahçen (eli Kohde-2), jotka ovat kaksi Turkin suosituinta jalkapalloseuraa. Sitten olemme merkinneet näiden kohteiden twiittien kannanottotiedot joko kannattajiksi tai vastustajiksi. Tätä tutkimusta varten emme ole merkinneet twiittejä luokkaan Neither.

**Tulos**

Miten twiitit kommentoitiin?

**Esimerkki 5.1240**

Kaikkien käsitteiden keskiarvon laskemisen jälkeen menetämme tiedon kunkin käsitteen leksikaalisesta vaihtelusta, mutta toisaalta voimme nyt tutkia, millä alueilla on samanlaista geolektistä vaihtelua, mikä johtaa hyvin määriteltyihin kielellisiin lajikkeisiin. Ne solut, joilla on samanlaiset värit joko kuviossa FIGREF16 tai kuviossa FIGREF17, voidaan olettaa kuuluvan samaan murrealueeseen. Näin ollen kartoissa voidaan erottaa kaksi pääaluetta tai klusteria. Violetti tausta peittää suurimman osan kartasta ja edustaa maaseutualueita, joilla on pieni, hajallaan oleva väestö. Analyysimme osoittaa, että tällä soluryhmällä on leksikossaan enemmän erityisiä sanoja. Vihreät ja keltaiset solut sen sijaan muodostavat toisen klusterin, joka on suurelta osin keskittynyt keskustaan ja rannikolle, jotka vastaavat suuria kaupunkeja ja teollisuusalueita. Näissä soluissa espanjan kielen vakiokielen käyttö on yleistä, mikä johtuu todennäköisesti kouluopetuksesta, tiedotusvälineistä, matkailijoista jne. Sen sanaston luonne on yhtenäisempi verrattuna violettiin ryhmään. Siinä missä violetti rypäs suosii tiettyjä lauseita, kaupunkiryhmän sanasto sisältää suurimman osan avainsanoista.

**Tulos**

Mitkä ovat kaupungin murteen ominaispiirteet?

**Esimerkki 5.1241**

Kun otamme LM-sulautumat mukaan järjestelmäämme, kokonaissuorituskyky kasvaa 90,87 prosentista 91,93 prosenttiin INLINEFORM0 CoNLL 2003 NER -tehtävässä, mikä merkitsee yli 1 prosentin absoluuttista F1-parannusta ja huomattavaa parannusta aiempaan tekniikan tasoon verrattuna. Saavutamme myös uuden huipputuloksen (96,37 % INLINEFORM1 ) CoNLL 2000 Chunking -tehtävässä.

**Tulos**

millaisia tuloksia niillä saavutetaan?

**Esimerkki 5.1242**

Tämän jakson tavoitteena on validoida teoreettisten tulostemme soveltuvuus - joiden mukaan itsetarkkaavaisuus voi suorittaa konvoluution - ja tutkia, oppivatko itsetarkkaavaisuuskerrokset käytännössä toimimaan konvoluutiokerrosten tavoin, kun niitä koulutetaan tavanomaisiin kuvanluokittelutehtäviin. Tutkimme erityisesti itsetarkkailun ja konvoluution välistä suhdetta kvadraattisten ja opittujen suhteellisten sijaintikoodausten kanssa. Huomaamme, että molemmissa tapauksissa opitut huomiotodennäköisyydet noudattavat yleensä lemman UNKREF15 ehtoja, mikä vahvistaa hypoteesimme. Varmistaaksemme, että mallimme oppii mielekkään luokittelijan, vertaamme sitä kuitenkin vakiomalliin ResNet18 BIBREF14 CIFAR-10-tietokannassa BIBREF15.

**Tulos**

Mitä numeerisia kokeita he tekevät?

**Esimerkki 5.1243**

VQABQ-mallissa on kaksi päämoduulia, peruskysymysten luontimoduuli (moduuli 1) ja visuaalisten kysymysten yhteistarkkailuun perustuva vastausmoduuli (moduuli 2).

**Tulos**

Mistä kahdesta päämoduulista heidän lähestymistapansa koostuu?

**Esimerkki 5.1244**

Verrattuna kontekstin sisäiseen jäsennykseen perustuvaan järjestelmään, yhdistelmä, jossa käytetään tarkkaa merkkijonojen täsmäytystä, tuottaa yli 6 %:n parannuksen palautukseen, ja yhdistelmä, jossa käytetään taivutusmerkkijonojen täsmäytystä, tuottaa vieläkin suuremman, lähes 8 %:n parannuksen, kun taas tarkkuuden menetys on 0,6 % ja 0,8 %.

**Tulos**

Millaisia täydentäviä PIE-uuttomenetelmiä käytetään luotettavuuden lisäämiseksi entisestään?

**Esimerkki 5.1245**

Käytämme 300-ulotteista Glove Common Crawl Embeddings (840B Token) BIBREF11 ja hienosäädämme ne tehtävään sopiviksi.

**Tulos**

Mitä sulautusalgoritmia ja ulottuvuuden kokoa käytetään?

**Esimerkki 5.1246**

Mallimme koostuu kahdesta neuroverkkomoduulista eli uuttimesta ja abstrahoijasta. Extractor koodaa lähdeasiakirjan ja valitsee asiakirjasta lauseita, minkä jälkeen abstractor parafrasioi tiivistelmäehdokkaat. Uuttolaite perustuu koodaaja-dekooderi-kehykseen. Sovitamme BERT:n kooderia varten, jotta voimme hyödyntää kontekstisidonnaisia esitystapoja, jotka on saatu valmiiksi koulutetuista muunnoksista. Käytämme dekooderina LSTM-pointteriverkkoa BIBREF22 valitaksemme uutetut lauseet edellä mainittujen lause-esitysten perusteella. Abstraktorimme on käytännössä identtinen BIBREF8:ssa ehdotetun abstraktorin kanssa.

**Tulos**

Mitä menetelmää tässä käytetään?

**Esimerkki 5.1247**

Arvioimme MPAD:n avulla opittujen asiakirjojen upotusten laatua 10 asiakirjaluokittelutietokannassa, jotka kattavat aiheen tunnistamisen, karkean ja hienojakoisen tunneanalyysin ja mielipiteiden louhinnan sekä subjektiivisuuden havaitsemisen. Seuraavaksi esittelemme lyhyesti tietokokonaisuudet. Niiden tilastotiedot esitetään taulukossa TABREF21.(1) Reuters. Tämä tietokokonaisuus sisältää uutistoimisto Reutersin vuonna 1987 keräämiä juttuja. Yleisen käytännön mukaisesti käytimme ModApte-jakoa ja otimme huomioon vain ne 10 luokkaa, joissa oli eniten positiivisia harjoitusesimerkkejä. Poistimme myös asiakirjat, jotka kuuluivat useampaan kuin yhteen luokkaan, ja sitten luokat, joihin ei jäänyt yhtään asiakirjaa (2 luokkaa). 2) BBCSport BIBREF30 sisältää BBC Sport -sivuston asiakirjoja, jotka vastaavat urheilu-uutisartikkeleita vuosilta 2004-2005. 3) Polarity BIBREF31 sisältää positiivisesti ja negatiivisesti leimattuja pätkiä Rotten Tomatoesista.(4) Subjektiivisuus BIBREF32 sisältää elokuva-arvostelukatkelmia Rotten Tomatoesista (subjektiiviset lauseet) ja Internet Movie Database -elokuvatietokannan juonitiivistelmiä (objektiiviset lauseet).(5) MPQA BIBREF33 koostuu positiivisista ja negatiivisista lauseista, jotka on annotoitu osana kesällä 2002 järjestettyä NRRC:n moninäkökulmaista kysymysten vastaamista käsittelevää työpajakokoelmaa (NRRC Workshop on Multi-Perspective Question Answering).(6) IMDB BIBREF34 on kokoelma IMDB:stä löytyviä vahvasti polarisoituja elokuva-arvosteluja (positiivisia ja negatiivisia). Kustakin elokuvasta on enintään 30 arvostelua.(7) TREC BIBREF35 koostuu kysymyksistä, jotka on luokiteltu kuuteen eri kategoriaan.(8) SST-1 BIBREF36 sisältää samoja pätkiä kuin Polarity. Kirjoittajat käyttivät Stanford Parser -ohjelmaa pätkien jäsentämiseen ja niiden jakamiseen useisiin lauseisiin. Sen jälkeen he käyttivät Amazon Mechanical Turk -palvelua annotoidakseen tuloksena saadut lauseet polariteetin mukaan (erittäin negatiivinen, negatiivinen, neutraali, positiivinen, erittäin positiivinen).(9) SST-2 BIBREF36 on sama kuin SST-1, mutta siitä on poistettu neutraalit arvostelut ja pätkät on luokiteltu positiivisiksi tai negatiivisiksi.(10) Yelp2013 BIBREF26 sisältää arvosteluja, jotka on saatu vuoden 2013 Yelp Dataset Challenge -kilpailusta.

**Tulos**

Mitä tietokokonaisuuksia käytetään?

**Esimerkki 5.1248**

Kaupallisista NLP-työkalupaketeista (esim. BIBREF14, BIBREF15, BIBREF16) valitsimme kokeiluihimme BIBREF17 ja BIBREF18, jotka ovat tietojemme mukaan ainoat julkisesti saatavilla olevat kaupalliset sovellusliittymät, jotka on tarkoitettu tekstialasta riippumattomaan entiteettitason tunneanalyysiin. Raportoimme myös TensiStrength BIBREF13:n, TwitterNLP BIBREF6:n, BIBREF19:n, CogComp-NLP BIBREF20:n ja Stanford NLP NER BIBREF21:n tulokset.

**Tulos**

Mitä työkalupaketteja he käyttävät?

**Esimerkki 5.1249**

Taulukossa TABREF15 mitataan BLEU BIBREF19-, NIST BIBREF20-, METEOR BIBREF21-, ROUGE-L BIBREF22- ja CIDEr BIBREF23 -mittarit vuoden 2018 E2E NLG Challenge -testidatalla käyttäen järjestäjien toimittamaa arviointiskriptiä. Käytämme kultaisen standardin valintaa tekstinmuodostusmallin harjoittelussa ja validoinnissa sekä automaattisessa arvioinnissa. Kun otamme tekstinmuodostusmallimme käyttöön manuaalista arviointia varten, käytämme CRF-mallia (Conditional Random Field, ehdollinen satunnaiskenttä) ennustamaan, mitkä tapahtumat mainitaan. Toisessa inhimillisessä arvioinnissa pyrittiin arvioimaan tuotoksen hyväksyttävyyttä uutistoimiston tuotantokäyttöön. Minimimuokkausarvioinnissa, jonka suoritti uutiskorpuksen luonut kommentoija, korjataan vain asiavirheet ja kielioppivirheet, jolloin teksti voi jäädä kömpelöksi tai epäsuoraksi. Luodun tekstin sanavirheprosentti (WER) verrattuna sen korjattuun varianttiin on 5,6 prosenttia (6,2 prosenttia, jos välimerkkejä ei oteta huomioon). Taulukossa TABREF23 on yhteenveto asiavirheistä ja niiden tyypeistä. Järjestelmän tuottamista 510 pelitapahtumasta 78:ssa oli asiavirhe, eli 84,7 prosenttia oli tuotettu ilman asiavirheitä. Suurin osa sujuvuusongelmista liittyy raportin yleiseen sujuvuuteen ja rakenteeseen. Näiden ongelmien ratkaiseminen edellyttäisi, että malli ottaisi huomioon useita pelitapahtumia ja yhdistelisi tietoja joustavammin toiston välttämiseksi. Toisessa inhimillisessä arvioinnissa pyrittiin arvioimaan tuotoksen hyväksyttävyyttä uutistoimiston tuotantokäyttöön. Tuotosta arvioidaan sen käyttökelpoisuuden kannalta uutiskanavalla, joka on merkitty koneellisesti tuotetuksi, eli sen ei ole tarkoitus vastata ihmistoimittajan tasoa, joka on varustettu huomattavilla taustatiedoilla. Arvioinnin suoritti kaksi STT-toimiston toimittajaa, jotka jakoivat 59 peliä keskenään suunnilleen tasan.

**Tulos**

Mitä arviointikriteerejä ja -mittareita käytettiin tuotetun tekstin arvioinnissa?

**Esimerkki 5.1250**

Tutkitaan systemaattisesti merkkitason vastahyökkäysten vaikutusta tunnetilaluokitukseen kahdella arkkitehtuurilla ja neljällä eri syötemuodolla. Tarkastelemme myös parafraasien havaitsemista.

**Tulos**

Minkä lopputehtävien perusteella ne arvioidaan?

**Esimerkki 5.1251**

Kuten taulukosta TABREF48 käy ilmi, kaksi malliamme osoittavat kilpailukykyisiä tuloksia verrattuna uusimpaan tekniikkaan Visual Dialog challenge 2018 -haasteessa (DL-61 oli Visual Dialog challenge 2018 -haasteen voittaja).

**Tulos**

Mikä malli voitti Visual Dialog -haasteen 2018?

**Esimerkki 5.1252**

Vaikka LangID-High ei anna tarkempaa tulosta, se antaa kuitenkin kompaktimman tuloksen: LangID-High:n koko on 15,4 Mt, kun taas wFST:n yhdistetyt korkean resurssin mallit ovat 197,5 Mt.

**Tulos**

miten mallin tiiviyttä mitataan?

**Esimerkki 5.1253**

NUS:lla koulutettujen toimintatapojen keskimääräinen onnistumisprosentti (SR) oli 94,0 %, kun niitä testattiin ABUS:lla, ja 96,6 %, kun niitä testattiin NUS:lla.

**Tulos**

kuinka paljon nus päihitti abusin?

**Esimerkki 5.1254**

Poikkeamien havaitseminen. Outlier Detection -tehtävässä BIBREF0 määritetään, mikä sana INLINEFORM1-sanojen luettelossa INLINEFORM0 ei liity muihin INLINEFORM2-sanoihin, jotka valittiin liittyviksi. Kullekin INLINEFORM3 -sanalle voidaan laskea sen tiiviyspisteet INLINEFORM4 , joka on INLINEFORM5 -sanan tiiviys. Tunneanalyysi. Tarkastelemme myös BIBREF31 :ssä kuvattua sentimenttianalyysia.

**Tulos**

Testaavatko he sanojen upotuksiaan myöhemmissä tehtävissä?

**Esimerkki 5.1255**

Tietokannat Englannin ja indin NMT-järjestelmien kouluttamiseen käytämme IITB:n englannin ja indin rinnakkaista korpusta BIBREF22 (INLINEFORM0-lauseet koulutusjoukosta) ja ILCI:n englannin ja indin rinnakkaista korpusta (INLINEFORM1-lauseet). ILCI:n (Indian Language Corpora Initiative) monikielinen rinnakkaiskorpus BIBREF23 kattaa useita intialaisia kieliä terveyden ja matkailun aloilta. Käytämme validointiin IITB:n rinnakkaiskorpuksen 520 lauseen dev-joukkoa. Kussakin lapsitehtävässä käytämme testijoukkona ILCI-korpuksen INLINEFORM2-lauseita.

**Tulos**

Millä tietokokonaisuudella (tietokokonaisuuksilla) he tekevät kokeita?

**Esimerkki 5.1256**

Mittaamme, kuinka kontekstisidonnainen sanaesitys on, käyttämällä kolmea eri mittaria: itsesimilaarisuutta, lauseen sisäistä samankaltaisuutta ja suurinta selitettävissä olevaa varianssia.

**Tulos**

Mitä kokeita ehdotetaan sen testaamiseksi, että ylemmät kerrokset tuottavat kontekstisidonnaisia sulautuksia?

**Esimerkki 5.1257**

Perusasetelman ja Oraakkelin lisäksi, joissa otetaan huomioon vain ASR:n 1 paras hypoteesi, suoritamme myös kokeita, joissa käytetään ASR:n $n$ parasta hypoteesia arvioinnissa.

**Tulos**

Mitä ovat yksinkertaisten mallien sarjat?

**Esimerkki 5.1258**

Runon tuottamiseen kuvista käytämme olemassa olevaa toimijakriittistä arkkitehtuuria BIBREF1. Shakespearen nykyaikaisten englanninkielisten tekstien luomiseksi kokeilimme erityyppisiä sekvenssistä sekvenssiin -malleja. Käytämme sekvenssistä sekvenssiin -mallia, joka koostuu yksikerroksisesta yksisuuntaisesta LSTM-koodaajasta ja yksikerroksisesta LSTM-dekoodaajasta sekä valmiiksi koulutetuista jälkikäteen sovitetuista sanojen upotuksista, jotka on jaettu lähde- ja kohdelauseiden välillä. Koska vastaavilla Shakespearen ja nykyenglannin lauseilla on merkittävä sanaston päällekkäisyys, laajennamme edellä mainittua sekvenssistä sekvenssiin -mallia käyttämällä osoitinverkkoja BIBREF11 , jotka tarjoavat sijaintiin perustuvan huomion ja joita on käytetty mahdollistamaan merkkien kopiointi suoraan syötteestä.

**Tulos**

Mitä malleja käytetään maalauksen upottamiseen ja mitä kielityylin siirtämiseen?

**Esimerkki 5.1259**

Mukauttaminen elintarvikealan kuvateksteihin Näin ollen elintarvikealan tietokokonaisuudessa on 3 806 kuvaa harjoittelua varten ja 1 775 kuvaa validointia varten. Muihin kuin elintarvikkeisiin liittyvässä tietokokonaisuudessa on 78 976 kuvaa harjoittelua varten ja 38 749 validointia varten.

**Tulos**

Kuinka monta esimerkkiä kohdealueella on?

**Esimerkki 5.1260**

Näyttää siis mahdolliselta, että käyttämällä samanlaista mallia voimme tuottaa tekstinäytteitä, joiden ehtona on joukko käyttäjän määrittämiä avainsanoja. Esimerkkinä sanotaan, että käyttäjä haluaa, että tuotettu tuloste sisältää avainsanat $\lbrace subway, manhattan\rbrace $ .

**Tulos**

Mitkä ovat käyttäjän määrittämät avainsanat?

**Esimerkki 5.1261**

Kuten aiemminkin, harjoittelu- ja testijoukkoihin sisältyy noin 30 000 ja 5 000 lauseparia.

**Tulos**

Kuinka monta näytettä he tuottivat keinotekoista kieltä varten?

**Esimerkki 5.1262**

Vaikka $\beta $-VAE tarjoaa ELBO:n regularisoinnin lisäkertoimen $\beta \in {\rm I\!R}^+$ avulla, sen tavoitefunktion yksinkertainen laajennus BIBREF16 sisältää ylimääräisen hyperparametrin $C$ KL-termin suuruuden hallitsemiseksi,jossa $C\!\! \ in \!\! {\rm I\!R}^+$ ja $| . |$ tarkoittaa absoluuttista arvoa.

**Tulos**

Miltä näyttää kirjoittajien ehdottama KL-divergenssitermiä koskeva nimenomainen rajoitus?

**Esimerkki 5.1263**

Annotaatiojärjestelmämme tarjoaa koulutusyhteisölle mahdollisuuksia tehdä lisätutkimuksia opiskelijoiden puheen piirteiden, opiskelijoiden oppimisen ja keskustelun laadun välisestä suhteesta. Kun automatisoidut luokittelijat on kehitetty, tällaisia puheiden ja oppimisen välisiä suhteita voidaan tutkia laajamittaisesti. Lisäksi automaattinen merkintä standardikoodausjärjestelmän avulla voi tukea tulosten yleistämistä eri tutkimusten välillä ja mahdollisesti johtaa automaattisten työkalujen kehittämiseen opettajille ja opiskelijoille. Erityisesti tätä ongelmaa varten suunnitellun merkintäjärjestelmän kehittäminen on ensimmäinen askel kohti sellaisten korpusten keräämistä ja merkitsemistä, joita NLP-yhteisö voi käyttää alan edistämiseksi tällä erityisalueella.

**Tulos**

mitä mahdollisuuksia korostetaan?

**Esimerkki 5.1264**

Sanojen upotukset ovat kasvattaneet suosiotaan NLP-sovelluksissa, koska erityisesti big data -ympäristöön suunnitellut mallit ovat menestyneet. Huomattakoon, että "suuria" tietokokonaisuuksia ei aina ole saatavilla, erityisesti yhteiskuntatieteellisissä NLP-sovelluksissa, joissa kiinnostavia tietoja ei useinkaan saada laajamittaisista lähteistä, kuten internetistä ja sosiaalisesta mediasta, vaan lähteistä, kuten lehdistötiedotteista BIBREF11 , akateemisista aikakauslehdistä BIBREF10 , kirjoista BIBREF12 ja nauhoitettujen puheiden BIBREF13 , BIBREF14 , BIBREF15 , on ehdotettu mallipohjaista menetelmää tulkittavissa olevien korpuskohtaisten sanojen upotusten kouluttamiseksi laskennallisessa yhteiskuntatieteessä käyttäen sekamuotoisia jäsenyysrepresentaatioita, Metropolis-Hastings-Walker-näytteenottoa ja NCE:tä. Kokeelliset tulokset ennusteista, valvotusta oppimisesta ja tapaustutkimukset unionin tilaa koskevista puheista ja NIPS-artikkeleista osoittavat, että menetelmällä voidaan saada korkealaatuisia upotuksia ja aiheita. Tulokset korostavat sitä, että suuri data ei aina ole parasta, sillä aluespesifinen data voi olla hyvin arvokasta, vaikka se olisi pientä.

**Tulos**

Miksi big data ei sovellu tähän tehtävään?

**Esimerkki 5.1265**

Klusteroimme upotukset INLINEFORM0 -Means-menetelmällä.

**Tulos**

Miten klusteri poimittiin?

**Esimerkki 5.1266**

Twitterissä näemme tuloksia, jotka ovat yhdenmukaisia RCV-tulosten kanssa vasemmistosta keskelle ulottuvan poliittisen kirjon osalta. Poikkeuksena, joka erottuu selvästi, ovat oikeistoryhmät ENL ja EFDD, jotka näyttävät olevan kaikkein yhtenäisimpiä. Tämä on täysin päinvastaista kuin mitä RCV:n tiedoissa havaittiin. Arvelemme, että tämä ilmiö johtuu siitä, että Euroopan oikeistoryhmät tukeutuvat sekä Euroopan että kansallisella tasolla suurelta osin sosiaaliseen mediaan levittääkseen Euroopan yhdentymistä kritisoivia kertomuksiaan.

**Tulos**

Mainitsivatko kirjoittajat tutkimuksessaan mahdollisia sekaannuksia?

**Esimerkki 5.1267**

Psykolingvististen tutkimusten BIBREF5 , BIBREF4 , BIBREF7 innoittamana ehdotettiin puolimerkkeihin perustuvaa RNN:ää (ScRNN), joka käsittelee sanoja sisältävän lauseen, jossa on kirjoitusvirheitä, ja ennustaa oikeat sanat jokaisessa vaiheessa. Olkoon $s = \lbrace w\_1, w\_2, \dots , w\_n\rbrace $ syöttölause, joka on sarja sanoja $w\_i$ . Kukin tulosana ( $w\_i$ ) esitetään ketjuttamalla (i) ensimmäisen merkin yksi kuuma vektori ( $\mathbf {w\_{i1}}}$ ); (ii) viimeisen merkin yhden kuuman esityksen ( $\mathbf {w\_{il}}}$ , jossa $l$ on sanan $w\_i$ pituus ); ja (iii) sisäisten merkkien merkkipussin esityksen ( $\sum \_{j=2}^{l-1}\mathbf {w\_{ij}})$ . ScRNN käsittelee ensimmäiset ja viimeiset merkit erikseen, ja se ei välitä sisäisten merkkien järjestyksestä. Kukin sana, joka esitetään vastaavasti, syötetään sitten BiLSTM-soluun. Jokaisessa sekvenssiaskeleessa koulutuskohteena on oikea vastaava sana (lähtöulottuvuus on sama kuin sanaston koko), ja malli optimoidaan risti-entropiahäviön avulla.

**Tulos**

Mikä on puolimerkkiarkkitehtuuri?

**Esimerkki 5.1268**

Validoimme lähestymistapamme Gigaword-korpuksella, joka koostuu 3,8 miljoonan artikkelin otsikoista (joita pidetään koko tekstinä) ja otsikoista (tiivistelminä) koostuvasta harjoitusjoukosta sekä 200 000 validointiparista, ja raportoimme testisuorituksen samalla 2 000 artikkelin joukolla, jota käytettiin BIBREF7:ssä.

**Tulos**

Mitä tietokokonaisuutta he käyttävät arvioinnissa?

**Esimerkki 5.1269**

Tässä jaksossa kuvaamme datan lisäämisen menetelmiä, joita käytämme koulutusdatan määrän lisäämiseksi, jotta NMT-järjestelmämme eivät kärsisi niin paljon japanin INLINEFORM0-vietnamin käännöstyön vähäisistä resursseista. TakaisinkääntäminenYksi lähestymistavoista yksikielisen datan hyödyntämiseksi on käyttää konekäännösjärjestelmää kääntämään nämä tiedot synteettisen rinnakkaisdatan luomiseksi. Tavallisesti kohdekielen yksikielinen aineisto käännetään, mistä menetelmän nimi johtuu: Back Translation BIBREF11 . Mix-Source ApproachToinen tiedon lisäämisen menetelmä, jota pidetään hyödyllisenä tässä vähäisten resurssien tilanteessa, on mix-source-menetelmä BIBREF12 .

**Tulos**

mitä menetelmiä käytettiin tietojen harvinaisuuden vaikutusten vähentämiseksi?

**Esimerkki 5.1270**

AEM:n tehokkuuden validoimiseksi tapahtumien poimimisessa sosiaalisesta mediasta (esim. Twitter) ja uutismediasivustoista (esim. Google news) käytetään kolmea tietokokonaisuutta (FSD BIBREF12 , Twitter ja Google).

**Tulos**

Mitä tietokokonaisuuksia käytetään?

**Esimerkki 5.1271**

Logistinen regressio: Yhdistetään asiakirjan Bag-Of-Words-edustus ja kysymyksen Bag-Of-Words-edustus. LSTM: Aloitetaan asiakirjan indeksisarjan ja kysymyksen indeksisarjan ketjuttamisella. Syötämme LSTM-verkkoa tällä vektorilla ja käytämme lopputilaa syötteen esityksenä. Lopuksi sovellamme logistista regressiota tähän esitykseen lopullisen päätöksen tuottamiseksi. End-to-end-muistiverkot: Tämä arkkitehtuuri perustuu kahteen eri muistisoluun (tulo ja lähtö), jotka sisältävät asiakirjan esityksen. Syvä projektiivinen lukija: Tämä on itse suunnittelemamme malli, joka on pitkälti saanut vaikutteita tehokkaasta R-verkon lukijasta BIBREF12 .

**Tulos**

Mitä perustasoja esitetään?

**Esimerkki 5.1272**

Suoritimme kokeita synteettisesti luoduilla tietokokonaisuuksilla, koska näin voimme paremmin hallita tietojen jakaumaa. Vertailimme erityisesti lähestymistapamme avulla saatuja hyötyjä suhteessa jakauman varianssiin. Loimme tietokokonaisuuden seuraavasta generatiivisesta prosessista. [H] Generatiivinen prosessi [1] Generoi dataPoimi k pistettä INLINEFORM0 toimialueen -1 keskiarvoiksi ja vastaava joukko k pistettä INLINEFORM1 toimialueen-2 keskiarvoiksi sekä kovarianssimatriisit INLINEFORM2iter INLINEFORM0 upto num INLINEFORM1 näytteet Näytteenottoluokka INLINEFORM2 Näyte INLINEFORM3 Näyte INLINEFORM4 Lisää q ja a niin näytteenottoluokka listaan q,a-parien joukkoon Generoimme tietokokonaisuuden edellä mainitulla näytteenottoprosessilla, jossa keskiarvot on valittu INLINEFORM5-kokoiselle 2-ulotteiselle ruudukolle, jossa varianssi on asetettu INLINEFORM6:ksi kussakin ulottuvuudessa.10000 otospistettä luotiin. Edellä mainitun algoritmin parametriksi INLINEFORM7 asetettiin 0,5 ja k:ksi 9 (koska pisteet voitiin tuottaa yhdestä yhdeksästä gaussista, joiden keskipisteet olivat INLINEFORM8 -ruudukossa).

**Tulos**

Miten synteettinen tietokokonaisuus luodaan?

**Esimerkki 5.1273**

Tässä jaksossa kuvaamme ensin ilmiötä, joka liittyy poliittisten puolueiden nimien mainitsemiseen käyttäjien profiiliominaisuuksissa. Tämän jälkeen analysoidaan profiileja, joiden profiiliattribuuteissa mainitaan nimenomaisesti poliittisia kahvoja. Molemmat näistä ovat orgaaninen tapa osoittaa tukea puolueelle, eikä niihin liity puolueiden suoraa kampanjointia. Katsomme, että koko kampanja Chowkidarin lisäämisestä profiiliattribuutteihin osoittaa epäorgaanista käyttäytymistä, jossa poliittiset johtajat toimivat katalysaattoreina. Mielestämme profiiliattribuutin muuttaminen pääministerin kampanjan mukaisesti on esimerkki epäorgaanisen käyttäytymisen leviämisestä BIBREF6, BIBREF9. Väitämme kuitenkin, että vaalikampanjaan liittyvien avainsanojen lisääminen profiiliin on eräänlaista epäorgaanista käyttäytymistä.

**Tulos**

Mitkä ovat orgaaniset ja epäorgaaniset tavat osoittaa poliittinen suuntautuminen profiilimuutoksilla?

**Esimerkki 5.1274**

Kun on annettu todistustila $= (\_, \_)$ , jossa $\_$ ja $\_$ merkitsevät korvausjoukkoa ja todistuspistemäärää, yhdistäminen lasketaan seuraavasti: Tuloksena saatu todistuspistemäärä $g$ on:$$ \begin{aligned} \max \_{f \in \mathcal {K}} & \; {unify}\_(g, [f\_{p}, f\_{s}, f\_{o}], (\emptyset , ))) \\ & = \max \_{f \in \mathcal {K}} \; \min \big \lbrace , \operatorname{k}(\_{\scriptsize {grandpaOf}:}, \_{f\_{p}:}),\\\ &\qquad \qquad \qquad \qquad \operatorname{k}(\_{{abe}:}, \_{f\_{s}:}), \operatorname{k}(\_{{bart}:}, \_{f\_{o}:}) \big \rbrace , \end{aligned}$$ (Eq. 3)missä $f \triangleq [f\_{p}, f\_{s}, f\_{o}]$ on fakta $\mathcal {K}$:ssa, joka tarkoittaa $f\_{s}$:n ja $f\_{o}$:n välistä $f\_{p}$-tyyppistä suhdetta $f\_{s}$ ja $f\_{o}$ välillä, $\_{s:}$ on symbolin $s$ sulautusrepresentaatio, $$ tarkoittaa alkuperäistä todistuspistemäärää ja $\operatorname{k}({}\cdot {}, {}\cdot {})$ tarkoittaa RBF-ydintä.

**Tulos**

Miten todistuspisteet lasketaan?

**Esimerkki 5.1275**

Ensimmäinen tässä työssä arvioimamme luokittelijatyyppi on feedforward-neuraaliverkot (DNN), jotka koostuvat kolmesta piilokerroksesta, joissa kussakin on 512 rektifioitua lineaarista yksikköä (ReLU), joissa on softmax-aktivointifunktio. Toisena arvioitavana luokittelijana käytämme konvoluutiohermoverkkoja (convolutional neural networks, CNN), joissa on 2 konvoluutio- ja max pooling-kerrosta, joita seuraa 2 täysin yhdistettyä ReLU-kerrosta, joissa on 512 solmua.

**Tulos**

Minkä mallin avulla he luokittelevat foneettisia segmenttejä?

**Esimerkki 5.1276**

Haluaisimme kiittää WASSA-2017 Shared Task on Emotion Intensity -tapahtuman järjestäjiä tietojen, ohjeiden ja oikea-aikaisen tuen tarjoamisesta.

**Tulos**

mitä tietokokonaisuutta käytettiin?

**Esimerkki 5.1277**

UCL:n ryhmällä oli pääsy mikrokertomuksiin sekä kontekstikohtaisiin metatietoihin, kuten väestötietoihin ja hanketietoihin. Poliittisten päättäjien kansainvälistä vertailua varten tiimi käänsi monikieliset vastaukset englanniksi käyttämällä konekäännöstä, tässä tapauksessa Translate API:ta (Yandex Technologies). Esikäsittelyvaiheessa poistettiin sanat, joilla ei ollut toiminnallista merkitystä (esim. "minä"), harvinaiset sanat, jotka esiintyivät vain yhdessä kertomuksessa, numerot ja välimerkit. Jäljelle jääneistä sanoista poistettiin substantiivien monikkomuodot ja verbien taivutusmuodot.

**Tulos**

Mitä luonnollisen kielen käsittelyn elementtejä ehdotetaan laadullisen tiedon analysoimiseksi?

**Esimerkki 5.1278**

Havaitsimme esimerkiksi, että talvisanat olivat yhdessä ja kaukana kesäsanoista. Viikonpäivät olivat myös ryhmiteltyinä ja kaukana viikonloppupäivistä.

**Tulos**

Mitä geometrisia ominaisuuksia sulautumilla on?

**Esimerkki 5.1279**

LiLi:llä olisi oltava seuraavat ominaisuudet:

**Tulos**

Mitkä ovat yleistiedon oppimismoottorin osat?

**Esimerkki 5.1280**

Kaikissa 26 tietokokonaisuudessa RCRN päihittää tavallisten BiLSTM:ien lisäksi myös 3L-BiLSTM:t, joiden parametrointi on suunnilleen sama.

**Tulos**

Onko heidän mallissaan enemmän parametreja kuin muissa malleissa?

**Esimerkki 5.1281**

Loogisten johtopäätösten osalta monotoninen päättely BIBREF6 , BIBREF7 , joka on eräänlainen sanojen korvaamiseen perustuva päättelytapa, edellyttää kykyä kuvata leksikaalisten ja syntaktisten rakenteiden välistä vuorovaikutusta.

**Tulos**

Mitä on monotoninen päättely?

**Esimerkki 5.1282**

W2V:n kouluttamiseen tarvittava tietokokonaisuus saatiin käyttämällä tietoja, jotka oli poimittu italialaisen Wikipedian dumppausdatasta (päivätty 2019.04.01), italialaisen Google Newsin pääluokista (MAAILMA, KANSAT, LIIKETOIMINTA, TEKNOLOGIA, VIIHDE, URHEILU, TIETO, TERVEYS) ja joistakin anonymisoiduista keskusteluista, jotka käytiin käyttäjien ja asiakaspalvelukeskustelurobotin (Laila) välillä. Tietokanta (joka koostuu 2,6 gigatavusta raakatekstiä) sisältää 421\ 829\ 960$ sanaa, jotka on jaettu 17\ 305\ 401$ lauseisiin. Kaikki sanat muutettiin pieniksi kirjaimiksi (kaksinkertaisen esiintymisen välttämiseksi), jolloin sanavarastoksi saatiin lopulta 618 \ 224 \ 224 \ dollarin sanat.

**Tulos**

Mitä tietokokonaisuutta käytetään Word2Vec-algoritmin syötteenä?

**Esimerkki 5.1283**

WordDec-mallimme (WDec) F1-pisteet ovat 3,9 \%$ ja 4,1 \%$ korkeammat kuin HRL:n tulokset NYT29- ja NYT24-tietokannoissa. Vastaavasti PtrNetDecoding (PNDec) -mallilla saavutetaan F1-pisteet, jotka ovat 3,0 \%$ ja 1,3 \%$ korkeammat kuin HRL:llä NYT29- ja NYT24-tietokannoissa.

**Tulos**

Kuinka korkeat F1-pisteet ovat aiempaan työhön verrattuna?

**Tulos**

Kumpi kahdesta ehdotetusta lähestymistavasta toimi kokeissa paremmin?

**Esimerkki 5.1284**

Vertailemme FlowSeqin suorituskykyä ensin vahvoihin perusmalleihin, kuten NAT w/ Fertility BIBREF6, NAT-IR BIBREF7, NAT-REG BIBREF25, LV NAR BIBREF26, CTC Loss BIBREF27 ja CMLM BIBREF8.

**Tulos**

Mitä ei-autoregressiivisiä NMT-malleja käytetään vertailussa?

**Esimerkki 5.1285**

Vuoropuheluaktit on luokiteltu moniin eri taksonomioihin: puheaktit BIBREF14 viittaavat lausahdukseen, ei ainoastaan tiedon esittämiseen, vaan myös suoritettavaan toimintaan. Myöhemmin puheaktit on muutettu viiteen luokkaan (Assertive, Directive, Commissive, Expressive, Declarative) BIBREF15. Keskustelutietojen merkitsemiseen on olemassa monia tällaisia vakiotaksonomioita ja -järjestelmiä, ja useimmat niistä noudattavat diskurssin kompositionaalisuutta. Nämä järjestelmät ovat osoittautuneet tärkeiksi diskurssi- tai keskustelunanalyysin kannalta BIBREF16. Vuoropuhelujärjestelmien ja diskurssianalyysin lisääntyneen kehityksen aikana viime vuosikymmeninä otettiin käyttöön vakiotaksonomia, jota kutsutaan nimellä Dialogue Act Markup in Several Layers (DAMSL) tag set. DAMSL:n mukaan jokaisella DA:lla on eteenpäin suuntautuva toiminto (kuten lausunto, infokysely, kiitos) ja taaksepäin suuntautuva toiminto (kuten hyväksyminen, hylkääminen, vastaus) BIBREF17.

**Tulos**

Miten vuoropuhelulakimerkinnät määriteltiin?

**Esimerkki 5.1286**

Yhteistyöelimen kokonaissuorituskyky oli huomattavasti parempi kuin muiden mallien, ja sen keskimääräinen tarkkuusero oli pienin, 22,5 pistettä.

**Tulos**

Mikä malli yleisti parhaiten?

**Esimerkki 5.1287**

Kutsumme näitä tekniikoita Recurrence over BERT (RoBERT) ja Transformer over BERT (ToBERT). Havaitsimme, että ToBERT päihittää RoBERT:n valmiiksi koulutetuilla BERT-ominaisuuksilla ja hienosäädetyillä BERT-ominaisuuksilla kaikissa tehtävissämme.

**Tulos**

Toimiiko RNN-kerros BERTin päällä paremmin vai muuntajakerros?

**Esimerkki 5.1288**

Nimettyjen entiteettien tunnistaminen (NER) biolääketieteen alalla sisältää yleensä sellaisten entiteettien tunnistamisen kuin proteiinit, geenit, sairaudet, hoidot, lääkkeet jne.

**Tulos**

Mikä on NER?

**Esimerkki 5.1289**

Järjestelmämme menestyi huomattavasti paremmin kuin perustason järjestelmä yhteisessä tehtävässä, sillä perustason järjestelmä sai huomattavasti huonommat pisteet sukupuolitehtävässä kuin lajitehtävässä.

**Tulos**

Missä tehtävässä malli toimii parhaiten?

**Esimerkki 5.1290**

Havaitseminen: Hypernymian havaitsemisessa tehtävänä on luokitella, ovatko sanaparit hypernymiasuhteessa. Tätä tehtävää varten arvioimme kaikkia malleja viidellä vertailutietoaineistolla: Ensinnäkin käytämme blessin substantiivi-substantiivi -osajoukkoa, joka sisältää hypernymia-merkintöjä 200 konkreettiselle, enimmäkseen yksiselitteiselle substantiiville. Toiseksi arvioimme leds BIBREF13 -mallilla, joka koostuu 2 770 substantiiviparista, jotka on tasapainotettu positiivisten hypernymiaesimerkkien ja satunnaisesti sekoitettujen negatiivisten esimerkkien välillä. Suunta: Suunnan ennustamisessa tehtävänä on tunnistaa, mikä termi on laajempi tietyssä sanaparissa. Tätä tehtävää varten arvioimme kaikkia malleja kolmella BIBREF16 -tietokannalla: On bless, tehtävänä on ennustaa suunta kaikille tietokannan 1337 positiiviselle parille. Parit lasketaan oikeiksi vain, jos hypernyymisuunta saa korkeammat pisteet kuin käänteissuunta, eli INLINEFORM0 . Varataan 10 % datasta validointia varten ja testataan loput 90 %. Wbless-tietokannassa noudatamme aiempaa työtä BIBREF17 , BIBREF18 ja suoritamme 1000 satunnaista iteraatiota, joissa 2 % datasta käytetään validointijoukkona luokittelukynnyksen oppimiseksi, ja testaamme lopuilla datan osilla. Raportoimme kaikkien iteraatioiden keskimääräisen tarkkuuden. Lopuksi arvioimme biblessin BIBREF16 , joka on wblessin muunnos, jossa hypernyymi- ja hyponyymiparit on nimenomaisesti merkitty niiden suunnan mukaan. Graded Entailment: Graded entailment -menetelmässä tehtävänä on kvantifioida se, missä määrin hypernymy-suhde pitää paikkansa. Tässä tehtävässä noudatamme aiempaa työtä BIBREF19 , BIBREF18 ja käytämme hyperlexin BIBREF20 substantiiviosaa, joka koostuu 2 163 substantiiviparista, jotka on kommentoitu INLINEFORM0 on-a INLINEFORM1 asteikolla INLINEFORM2 .

**Tulos**

Mitä vertailutietoaineistoja käytetään?

**Esimerkki 5.1291**

Arvioimme menetelmäämme kahdella laajasti tutkitulla tietokokonaisuudella, jotka ovat Waseem ja Hovey BIBREF5 sekä Davidson et al. BIBREF9.

**Tulos**

Mitä julkisesti saatavilla olevia tietokokonaisuuksia käytetään?

**Esimerkki 5.1292**

Verrattuna luetun ymmärtämiseen perustuvaan laadunvarmistukseen (RCQA), jossa vastaukset kysymykseen löytyvät yleensä annetusta pienestä kappaleesta, OpenBookQA-asetelmassa avoimen kirjan osa on paljon laajempi (kuin pieni kappale), eikä se ole täydellinen, koska siinä saatetaan tarvita lisää yleistä tietoa.

**Tulos**

Miten OpenBookQA eroaa muusta luonnollisen kielen laadunvarmistuksesta?

**Esimerkki 5.1293**

CreateDebate-tietokokonaisuus kerättiin englanninkieliseltä verkkokeskustelufoorumilta, jossa keskusteltiin neljästä aiheesta: abortti (ABO), homojen oikeudet (GAY), Obama (OBA) ja marihuana (MAR).

**Tulos**

Mitkä aiheet sisältyvät keskustelutietoihin?

**Esimerkki 5.1294**

Koulutusaineistona käytämme BIBREFin julkaisemia Daily Mail -uutisartikkeleita9 .

**Tulos**

mitä tietokokonaisuutta käytettiin?

**Esimerkki 5.1295**

Tunnetietoisen chatbotin arviointi jakautuu kahteen eri osaan, laadulliseen ja määrälliseen arviointiin. Kvalitatiivisessa arvioinnissa keskitytään ohjelmiston toimivuuden arviointiin, kun taas kvantitatiivisessa keskitytään enemmän mittaamaan chatbottien suorituskykyä numerolla. Useiden aiempien tutkimusten perusteella havaitsimme, että useimmissa teoksissa käytettiin ISO 9241 -standardia chatbottien laadun arvioinnissa keskittymällä käytettävyysnäkökohtiin. Tämä näkökohta voidaan ryhmitellä kolmeen painopisteeseen, mukaan lukien tehokkuus, vaikuttavuus ja tyytyväisyys, jotka koskevat järjestelmien suorituskykyä määritettyjen tavoitteiden saavuttamisessa. Automaattisessa arvioinnissa joissakin tutkimuksissa keskitytään järjestelmän arviointiin tunnetasolla BIBREF15 , BIBREF28 . Siksi järjestelmän suorituskyvyn mittaamiseen käytetään joitakin yleisiä mittareita, kuten tarkkuutta, palautusta ja tarkkuutta, verrattuna kultaiseen merkkiin. Tämä arviointi on samanlainen kuin tunteiden luokittelutehtävät, kuten aiemmat SemEval 2018 BIBREF32 ja SemEval 2019 . Muissa tutkimuksissa on myös ehdotettu perpleksiteetin käyttöä mallin arvioimiseksi sisältötasolla (sen määrittämiseksi, onko sisältö merkityksellistä ja kieliopillista) BIBREF14 , BIBREF39 , BIBREF28 . Tätä arviointimittaria käytetään laajalti arvioitaessa dialogipohjaisia järjestelmiä, jotka perustuvat probabilistiseen lähestymistapaan BIBREF61 . Toisessa BIBREF14:n teoksessa käytettiin BLEU:ta koneellisen vastauksen arvioimiseksi ja sen vertaamiseksi kultaiseen vastaukseen (varsinaiseen vastaukseen), vaikka BIBREF62 ei suosittele BLEU:n käyttämistä keskustelunmuodostustehtävän mittaamiseen, koska se korreloi heikosti ihmisen arvioinnin kanssa. Tässä arvioinnissa chatbottien suorituskyvyn mittaamiseen käytetään ihmisen arviointia useiden kriteerien perusteella. BIBREF15 käytti kolmea kommentoijaa arvioimaan chatbottien vastauksia kahdella kriteerillä, jotka olivat sisältö (asteikko 0,1,2) ja tunteet (asteikko 0,1). Sisältö keskittyy sen mittaamiseen, onko vastaus luonnollisesti hyväksyttävä ja voisiko ihminen tuottaa sen uskottavasti. Tutkijat ovat jo ottaneet käyttöön ja suosittelevat tätä metriikkaa ja haastavia keskustelutehtäviä, kuten BIBREF-julkaisussa38 ehdotetaan.

**Tulos**

Miten EAC arvioidaan?

**Esimerkki 5.1296**

Toinen turkkilainen tietokokonaisuus on Twitter-korpus, joka koostuu turkkilaisia matkaviestinoperaattoreita koskevista twiiteistä. Nämä twiitit ovat useimmiten paljon äänekkäämpiä ja lyhyempiä kuin elokuvakorpuksen arvostelut. Yhteensä twiittejä on 1 716 kappaletta. Niistä 973 on negatiivisia ja 743 positiivisia. Kaksi ihmistä on kommentoinut nämä twiitit manuaalisesti, ja merkinnät ovat joko positiivisia tai negatiivisia.

**Tulos**

Mitä tietoja Twitter-tietokannasta annetaan?

**Esimerkki 5.1297**

GPT-2 saa korkeimman pistemäärän ja $n$-grammi alhaisimman. Transformer-XL ja LSTM LM toimivat keskellä ja suunnilleen samalla tasolla keskenään. Taulukossa TABREF14 ilmoitetaan kaikkien mallien ja ihmisen suorittaman arvioinnin 12 kategorian tarkkuustulokset.

**Tulos**

Mikä on mallien suorituskyky tehtävissä?

**Esimerkki 5.1298**

Vertailemme seuraamusluokittimen osalta Decomposable Attention BIBREF2 , BIBREF3 sellaisena kuin se on toteutettu virallisessa perusversiossa, ESIM BIBREF4 ja muunnosverkkoa, jossa on valmiiksi koulutetut painot BIBREF5 .

**Tulos**

Mitä valmiiksi koulutettua muuntajaa he käyttävät?

**Esimerkki 5.1299**

Edellä esitetyn analyysin mukaan Edellä esitetyn analyysin mukaan ehdotimme painotettua versiota DIRL:stä, jotta voimme ratkaista ongelman, joka aiheutuu $\rm {P}(\rm {Y})$:n siirtymisestä DIRL:ään. Tämän kehyksen keskeisenä ajatuksena on ensin yhdenmukaistaa $\rm {P}(\rm {Y})$ eri osa-alueiden välillä ennen osa-alue-invariantin oppimisen suorittamista ja ottaa sitten huomioon $\rm {P}(\rm {Y})$:n siirtymä etikettien ennustamismenettelyssä. Tarkemmin sanottuna siinä otetaan käyttöön luokkapaino $\mathbf {w}$ lähdealueen esimerkkien painottamiseksi luokittain. Painotetun lähdealueen perusteella toimialueen siirto-ongelma ratkaistaan kahdessa vaiheessa. Tämän käytännön motivaationa on säätää lähde- tai kohdealueen datajakaumaa, jotta $\rm {P}(\rm {Y})$:n siirtyminen eri alueiden välillä voidaan lieventää ennen DIRL:n soveltamista. Jos meillä on vain lähdealueen tietojen merkinnät, päätämme mukauttaa lähdealueen tietojen jakaumaa. Tätä tarkoitusta varten otamme käyttöön koulutettavan luokkapainon $\mathbf {w}$, jolla lähdealueen esimerkkejä painotetaan uudelleen luokittain DIRL:ää suoritettaessa, kun $\mathbf {w}\_i > 0$. Tarkemmin sanottuna toivomme, että:ja merkitsemme $\mathbf {w}^\*$ $\mathbf {w}$:n arvoa, joka saa tämän yhtälön pitämään paikkansa. Näemme, että kun $\mathbf {w}=\mathbf {w}^\*$, DIRL:n tehtävänä on sovittaa $\rm {P}\_S(G(\rm {X})|\rm {Y})$ yhteen $\rm {P}\_T(G(\rm {X})|\rm {Y})$:n kanssa ilman $\rm {P}(\rm {Y})$:n siirtoa. Analyysimme mukaan tiedämme, että $\rm {P}(\rm {Y})$:n siirtymisen vuoksi valvotun oppimisen $\mathcal {L}\_{sup}$ ja toimialuevariantin oppimisen $\mathcal {L}\_{inv}$ harjoituskohteiden välillä on ristiriita. Ristiriidan aste pienenee, kun $\rm {P}\_S(\rm {Y})$ lähestyy $\rm {P}\_T(\rm {Y})$. Siksi mallin harjoittelun aikana $\mathbf {w}$ optimoidaan kohti $\mathbf {w}^\*$, koska se saa painotetun lähdealueen $\rm {P}(\rm {Y})$ lähelle $\rm {P}\_T(\rm {Y})$:ta, jotta konflikti voidaan ratkaista.

**Tulos**

Miten eri aloja painotetaan WDIRL:ssä?

**Esimerkki 5.1300**

Molemmat päihittivät huomiotta jättämisen perustason huomattavasti. Huomion perustasoa varten käytämme tavallista parametrisoitua huomion BIBREF2 .

**Tulos**

Mitä perustason menetelmiä käytetään?

**Esimerkki 5.1301**

Arvioimme ehdotetun menetelmän suorituskykyä käyttämällä kahta kriteeriä: i) rank-korrelaatio BIBREF25 visuaalisen maadoituksen arvioimiseksi ja ii) tarkkuus kysymyksiin vastaamisen arvioimiseksi. Intuitiivisesti sijoituskorrelaatio mittaa ihmisen ja mallin huomiokarttojen samankaltaisuutta sijoitukseen perustuvalla mittarilla. Korkea sijoituskorrelaatio tarkoittaa, että malli "katsoo" kuva-alueita, jotka vastaavat ihmisen samaan kysymykseen vastaamiseen käyttämää visuaalista informaatiota. Ennustetun vastauksen tarkkuuden kannalta INLINEFORM0 arvioidaan seuraavasti: DISPLAYFORM0

**Tulos**

Miten he mittaavat manuaalisten ja mallin tuottamien pohjakosketusten välistä korrelaatiota?

**Esimerkki 5.1302**

Jäljelle jää vain kysymys siitä, mikä tekee kahdesta ympäristöstä tarpeeksi samanlaisia, jotta voidaan päätellä yhteisen luokan olemassaolo. Tästä kysymyksestä on olemassa laaja kirjallisuus (mukaan lukien edellä mainitut kielen mallintamista, valvomatonta jäsennystä ja yhdenmukaistamista koskevat työt), mutta tässä työssä käytämme hyvin yksinkertaista kriteeriä: fragmentit ovat keskenään vaihdettavissa, jos ne esiintyvät vähintään yhdessä täsmälleen samanlaisessa leksikaalisessa ympäristössä.

**Tulos**

Miten ne määrittelevät samankaltaiset ympäristöt fragmentteja varten tietojen lisäysjärjestelmässä?

**Esimerkki 5.1303**

Koska englanti ei myöskään merkitse kieliopillista sukupuolta, englantia varten kehitettyjä lähestymistapoja ei voida siirtää morfologisesti rikkaisiin kieliin, joissa on sukupuolisopimusta BIBREF8 .

**Tulos**

Miksi englanninkielinen lähestymistapa ei toimi muilla kielillä?

**Esimerkki 5.1304**

GLUE-tehtävistä nämä huipputason järjestelmät saavuttavat suurimmat voitot hyväksyttävyystehtävässä, jossa käytetään kielellisen hyväksyttävyyden korpusta BIBREF0 . CoLA sisältää esimerkkilauseet kielitieteellisistä julkaisuista, jotka asiantuntijat ovat leimanneet kieliopillisen hyväksyttävyyden kannalta ja jotka on kirjoitettu niin, että niissä näkyy hienovaraisia kieliopillisia piirteitä. Koska minimaaliset syntaktiset erot voivat erottaa hyväksyttävät lauseet hyväksyttävistä lauseista, joita ei voida hyväksyä (Mistä Bo kirjoitti kirjan? / \*Mitä kirjaa Bo kirjoitti?), ja hyväksyttävyysluokittelijat ovat luotettavampia, kun ne on koulutettu GPT:llä ja BERT:llä kuin rekursiivisilla malleilla, on loogista, että GPT:llä ja BERT:llä on parempi implisiittinen tietämys hyväksyttävyyden kannalta merkityksellisistä syntaktisista ominaisuuksista.

**Tulos**

Miten CoLA on kieliopillisesti kommentoitu?

**Esimerkki 5.1305**

Oppimismalli hakua varten koulutetaan oraakkelilla, joka on rakennettu etävalvonnan avulla. Koulutusjoukon vastausmerkintöjen avulla voidaan löytää sopivia artikkeleita, jotka sisältävät kysymyksessä pyydetyt tiedot. Ensin kysymys ja artikkeli upotetaan vektorisarjoiksi käyttäen samaa menetelmää kuin ymmärtämismallissa. Emme käytä tässä anonymisointia yksinkertaisuuden säilyttämiseksi. Muussa tapauksessa anonymisointimenettely olisi toistettava useita kertoja mahdollisesti suuren asiakirjakokoelman osalta. Nämä vektorisekvenssit syötetään seuraavaksi Bi-GRU:lle, joka tuottaa tulosteet $v$ (kysymyksen osalta) ja $H\_c$ (asiakirjan osalta) edellisen jakson tapaan.

**Tulos**

Miten neuraalista mallia voidaan käyttää hakuun, jos syötteenä on koko Wikipedia?

**Esimerkki 5.1306**

Vertailemme luokittelu- ja regressiomenetelmiä ja osoitamme, että luokittelu tuottaa parempia tuloksia kuin regressio, mutta tulosten laatu riippuu siitä, miten tietojen merkinnät merkitään.

**Tulos**

Suoriutuivatko luokittelumallit paremmin kuin aiemmat regressiomallit?

**Esimerkki 5.1307**

Esittelemme LOG-Cadin, neuroverkkopohjaisen kuvausgeneraattorin (kuva KUVA 1), jolla tämä tehtävä voidaan ratkaista suoraan. Kun annetaan sana ja sen konteksti, generaattorimme hyödyntää kohdesanan upotusta, joka on esivalmennettu massiivisesta tekstistä (globaalit kontekstit), samalla kun se koodaa myös annetun paikallisen kontekstin, ja yhdistää molemmat luodakseen luonnollisen kielellisen kuvauksen.

**Tulos**

Käytetäänkö niissä esivalmisteltuja sanasulkeumia?

**Esimerkki 5.1308**

Yhdistetty mallimme, joka käyttää sekä latenttia että eksplisiittistä rakennetta, toimii parhaiten, sillä se parantaa ROUGE-L:n tuloksia huomattavasti, 1,08 pistettä perusosoitingeneraattorimalliin verrattuna ja 0,6 pistettä ROUGE-1:een verrattuna.

**Tulos**

Kuinka paljon ne paranevat edelliseen huipputekniikkaan verrattuna?

**Esimerkki 5.1309**

Tämän tietokokonaisuuden vitsikorpus sisältää tuhansia ainutlaatuisia vitsejä eri kategorioista (scifi, urheilu jne.) ja tyypeistä (sanaleikit, limerikit jne.).

**Tulos**

Mistä todelliset tuotantotiedot ovat peräisin?

**Esimerkki 5.1310**

Perustasomme on kolmikerroksinen LSTM-LM, jossa on 1 150 piilotettua yksikköä sisäisillä kerroksilla ja joka on koulutettu tavanomaisella ristiinentropiahäviöllä.

**Tulos**

Mitä neuraalisia kielimalleja tutkitaan?

**Esimerkki 5.1311**

Taulukosta TABREF20 käy ilmi, että WoZ- ja DSTC2-tietoaineistoissa SIM-mallilla on sama määrä parametreja, joka on vain 23 prosenttia ja 19 prosenttia GLAD-mallin parametreista.

**Tulos**

Miten he mittaavat mallin koon?

**Esimerkki 5.1312**

Koska tämä kommentoitujen potilasmuistiinpanojen korpus koostuu alkuperäisistä terveydenhuollon tiedoista, jotka sisältävät vuoden 1996 HIPAA-lain (Health Information Portability and Accountability Act of 1996) BIBREF16 mukaisia suojattuja terveystietoja ja jotka voidaan liittää MIMIC-III-tietokantaan, henkilöiden, jotka haluavat päästä käsiksi tietoihin, on täytettävä kaikki MIMIC-III-tietokantaan sisältyviin tietoihin pääsyä koskevat vaatimukset. Näiden edellytysten täyttämiseksi tietokantaan pääsyä haluavan henkilön on suoritettava "Data or Specimens Management" -kurssi sekä allekirjoitettava MIMIC-III-tietokannan verkkosivulla esitetty käyttäjäsopimus, jossa tämän tietokannan uusin versio sijaitsee nimellä "Annotated Clinical Texts from MIMIC" BIBREF17. Tähän korpukseen pääsee käsiksi myös GitHubissa, kun kaikki edellä mainitut vaatimukset on täytetty.

**Tulos**

Onko tämä tietokokonaisuus julkisesti saatavilla kaupalliseen käyttöön?

**Esimerkki 5.1313**

Lisäksi Bertramin add- ja add-gated-muunnokset toimivat yllättävän hyvin useammin esiintyvien sanojen osalta: WNLaMPro-mediumin pisteet paranevat 50 % verrattuna BERT$\_\text{base}$:iin ja 31 % verrattuna Attentive Mimickingiin.

**Tulos**

Kuinka paljon harvinaisten ja keskitiheiden sanojen edustavuus on parantunut verrattuna erilliseen yhteistyöhön ja aiempaan työhön?

**Esimerkki 5.1314**

Osallistujat, joille näytettiin määritelmä, ehdottivat todennäköisemmin twiitin kieltämistä. Itse asiassa ryhmän yksi osallistujat antoivat hyvin harvoin eri vastauksia kysymyksiin yksi ja kaksi (18 tapausta 500:sta eli 3,6 %). Tämä viittaa siihen, että kyseisen ryhmän osallistujat yhdenmukaistivat oman mielipiteensä määritelmän kanssa.

**Tulos**

Miten kirjoittajat osoittivat, että vihapuheen määritelmän esittäminen sai kommentoijat osittain yhdenmukaistamaan oman mielipiteensä määritelmän kanssa?

**Esimerkki 5.1315**

Vaikka kilpailussa ehdotetaan kahta erilaista skenaariota, itse asiassa molempia ohjaa snomed ct -ontologia - osatehtävässä 1 entiteetit on tunnistettava offsettien avulla ja ne on yhdistettävä ennalta määriteltyihin neljään luokkaan (PROTEIINIT, NORMALIZABLES, NO\_NORMALIZABLES ja UNCLEAR); osatehtävässä 2 on annettava luettelo kaikista tekstissä esiintyvistä entiteettien snomed ct -tunnisteista (sctid), jota yhteisen tehtävän järjestäjät kutsuvat käsiteindeksoinniksi.

**Tulos**

Mitkä ovat PharmaCoNERin kaksi osatehtävää?

**Esimerkki 5.1316**

Pääsyy tähän on se, että tietokokonaisuudet sisältävät vain pienen osan moniulotteisista lauseista, joissa on erilaisia polariteetteja. Huomiohäiriö ei vaikuta kovinkaan paljon sentimentin ennustamiseen yksinäkökulmaisissa lauseissa tai moninäkökulmaisissa lauseissa, joissa on samat polariteetit.

**Tulos**

Arvioidaanko mallia perusmallia vastaan myös yhden aspektin lauseilla?

**Esimerkki 5.1317**

Logoscope hakee päivittäin sanomalehtiartikkeleita useista ranskankielisistä RSS-syötteistä.

**Tulos**

Kuinka usein sanomalehtien verkkosivuja selataan päivittäin?

**Esimerkki 5.1318**

Kun kaikki sanat, joita vähintään kolme erillistä käyttäjää ei ole käyttänyt harjoitusjoukossa, on jätetty pois, muodostetaan AllWords-malli laskemalla kaikkien jäljelle jäävien sanojen frekvenssit ja harjoittelemalla multinomiaalinen Naive Bayes -luokitin.

**Tulos**

Mitä mallia he käyttivät järjestelmässään?

**Esimerkki 5.1319**

Tästä syystä käytämme englanninkielisen aineiston perusalgoritmina BIBREF0-tietokannan tuloksia, ja venäjänkielisen aineiston osalta käytimme BIBREF8-tietokannassa kuvattua todennäköisyysperusteista kielimallia.

**Tulos**

Mitä kieliä paperissa käytetään?

**Tulos**

Arvioidaanko RNN-mallia suhteessa johonkin perustasoon?

**Esimerkki 5.1320**

Ensimmäinen osa kuvataan tehtäväksi A, jonka tarkoituksena on tunnistaa aikatietojen rajat ja luokitella ne johonkin seuraavista luokista: päivämäärä, aika, kesto, joukko. Valitsimme kustakin sanojen upotusten ryhmästä (EE, EP, EC) kolme parasta tulosta taulukosta TABREF19, jossa esitetään kaikkien mallien F1-pisteet. Tämän jälkeen arvioimme näitä tuloksia käyttämällä yksityiskohtaisempia mittareita aikatekstien osalta, jotka on esitetty BIBREF:ssä27 .

**Tulos**

Mitä kokeita esitetään?

**Esimerkki 5.1321**

Kokeilussamme verrataan joitakin laajalti käytettyjä tekstin klusterointimenetelmiä lähestymistapaamme. K-means-, Skip-thought Vectors-, Recursive Neural Network- ja Paragraph Vector -pohjaisten klusterointimenetelmien lisäksi neljä perustason klusterointimenetelmää perustuu suoraan suosittuihin valvomattomiin dimensioiden pienentämismenetelmiin, jotka on kuvattu SECREF11 -jaksossa.

**Tulos**

Mitä suosittuja klusterointimenetelmiä he kokeilivat?

**Esimerkki 5.1322**

BIBREF12 havaitsi kiinalaisessa korpuksessa, että sanamerkintä "End" toimii paremmin kuin "Begin". Tämä motivoi meitä suorittamaan taaksepäin suuntautuvan ahneen haun kunkin lauseen merkintäsekvenssistä sanojen rajojen tunnistamiseksi. Jos kaksi lauseessa segmentoitua sanaa tunnistetaan substantiiveiksi ja toinen sana on välittömästi ennen toista, yhdistämme niiden rajat ja luomme uuden sanaehdokkaan olion tunnistusta varten. Tämän strategian etuna on löytää nimettyjä entiteettejä, joiden sanan pituus on pitkä. Se myös vähentää eri segmentointikriteerien aiheuttamaa vaikutusta.

**Tulos**

Mitä rajakokoonpanomenetelmää käytetään?

**Esimerkki 5.1323**

Käytimme tehtävään kahta tietokokonaisuutta - AMR Bank BIBREF10 ja CNN-Dailymail ( BIBREF11 BIBREF12 ).

**Tulos**

Mitä tietokokonaisuutta tässä asiakirjassa käytetään?

**Esimerkki 5.1324**

Tässä asiakirjassa esitetyt tiedot kerättiin ja validoitiin Mozillan Common Voice -aloitteen avulla. Käyttämällä joko Common Voice -sivustoa tai iPhone-sovellusta osallistujat nauhoittavat äänensä lukemalla näytöllä näkyviä lauseita (ks. kuva (KUVA 5)). Myöhemmin muut osallistujat vahvistavat äänitteet yksinkertaisen äänestysjärjestelmän avulla. Kuvassa (FIGREF6) esitetyssä validointirajapinnassa osallistujat merkitsevät $<$audio,transcript$>$-parit joko oikeiksi (ääni ylös) tai vääriksi (ääni alas).

**Tulos**

Mitä joukkoistamisalustaa käytetään tietojen keräämiseen ja validointiin?

**Esimerkki 5.1325**

Harkitsemme seuraavia mittareita eri esitysten automaattista arviointia varten. Yhteisen tavoitteen tarkkuutta on käytetty ensisijaisena metriikkana esitysten sijoittamisessa paremmuusjärjestykseen.Active Intent Accuracy: Pyydettyjen aikojen F1: Pyydettyjen aikojen makrokeskiarvoinen F1-pistemäärä kaikista hyväksyttävistä aikojen aikomuksista. Käännöksiä, joissa ei ole pyydettyjä aikoja pohjatodellisuudessa ja ennusteissa, ei oteta huomioon.Average Goal Accuracy: Kullekin vuorolle ennustetaan yksi arvo jokaiselle dialogin tilassa olevalle aikavälille. Tämä on keskimääräinen tarkkuus, jolla slotin arvo ennustetaan oikein.Joint Goal Accuracy: Tämä on keskimääräinen tarkkuus, jolla ennustetaan oikein kaikki tietyn palvelun lähtöarvot vuorossa.

**Tulos**

Miten malleja arvioidaan?

**Esimerkki 5.1326**

Käyntien vektorikuvausten välisen euklidisen etäisyyden perusteella sovellettiin ja verrattiin kahta klusterointialgoritmia: k-means ja hierarkkinen klusterointi Wardin menetelmällä klustereiden yhdistämiseksi BIBREF23 .

**Tulos**

Mitä klusterointitekniikkaa he käyttävät osanottajien vierailuteksteihin?

**Esimerkki 5.1327**

Panoksemme hyödyntää valmiiksi koulutettuja sanojen upotuksia (GloVe, koulutettu wikipedia+gigaword-korpuksen pohjalta), DepecheMood-vaikutteista leksikkoa ja konvoluutiohermoverkkoja.

**Tulos**

Mitä sulautumia he käyttävät?

**Tulos**

Mikä on heidän sanastonsa lähde?

**Esimerkki 5.1328**

Työskentelemme neljän yksittäisen tietokokonaisuuden kanssa. Tietoaineistot sisältävät yhteen-, vähennys-, kerto- ja jakotehtäviä.AI2 BIBREF2. AI2 on kokoelma 395 yhteen- ja vähennyslaskuongelmaa, jotka sisältävät numeerisia arvoja, joista osa voi olla kysymyksen kannalta epäolennaisia.CC BIBREF19. Common Core -tietokokonaisuus sisältää 600 2-vaiheista kysymystä. Pennsylvanian yliopiston Cognitive Computation Group kokosi nämä kysymykset.IL BIBREF4. Illinoisin tietokokonaisuus sisältää 562 1-vaiheista algebran sanakysymystä. Cognitive Computation Group kokosi myös nämä kysymykset.MAWPS BIBREF20. MAWPS on suhteellisen laaja kokoelma, joka on peräisin pääasiassa muista MWP-tietokannoista. Käytämme 2 373:aa tämän joukon 3 915:stä MWP:stä.

**Tulos**

Mitä tietokokonaisuuksia ne käyttävät?

**Esimerkki 5.1329**

Tässä jaksossa raportoidaan kokeista, jotka on tehty edellä kuvatun järjestelmän ja tietojen avulla. Ensin esitellään englanninkieliset tulokset kolmesta ABSA-julkaisusta sekä vertailu aiempaan työhön. Sen jälkeen teemme saman viidelle muulle kielelle, jotka sisältyvät ABSA 2016 -painokseen: hollanti, ranska, venäjä, espanja ja turkki. Paikalliset ja klusterointiominaisuudet, jotka on kuvattu kohdassa SECREF11 , ovat samat jokaiselle kielelle ja arviointiasetelmalle. Ainoa muutos on eri kielissä käytettävät klusterointisanakirjat. Kuten kohdassa SECREF11 todetaan, paras klusteriyhdistelmä valitaan harjoitusaineistosta viisinkertaisella ristiinvalidoinnilla (CV). Ensin kokeillaan jokaista permutaatiota Clark- ja Word2vec-klustereilla. Kun paras yhdistelmä on saatu, kokeillaan Brownin klustereilla, jolloin saadaan lopullinen malli kullekin kielelle ja tietokokonaisuudelle.

**Tulos**

Mitä kuutta kieltä kokeillaan?

**Esimerkki 5.1330**

Lopuksi muutamme luokiteltavan tekstin skalaareiksi, jotka kuvaavat niiden etäisyyttä rakennetusta vihavektorista, ja käytämme näitä satunnaismetsäluokittimen syötteenä.

**Tulos**

Mitä luokittelulaitetta he käyttivät?

**Esimerkki 5.1331**

Vertaillaksemme neuromalliamme perinteisiin lähestymistapoihin, kokeilimme useita olemassa olevia malleja, kuten: Logistinen regressio (LR), erotteleva todennäköisyysmalli, ja Random Forest (RF), päätöspuiden yhdistelmämalli.

**Tulos**

Mikä oli heidän lähtötasonsa?

**Esimerkki 5.1332**

Useimmissa aiemmissa töissä keskitytään loukkaavan kielen eri osa-alueisiin, kuten loukkaavaan kieleen BIBREF0 , BIBREF1 , (verkko)aggressioon BIBREF2 , (verkko)kiusaamiseen BIBREF3 , BIBREF4 , myrkyllisiin kommentteihin INLINEFORM0 , vihapuheeseen BIBREF5 , BIBREF6 , BIBREF7 , BIBREF7 , BIBREF8 , BIBREF8 , BIBREF8 , BIBREF8 , BIBREF8 , BIBREF9 , BIPREF9 , BIBREF10 , ja hyökkäävään kielenkäyttöön BIPREF11 . Aiemmissa töissä on keskitytty näihin loukkaavan kielen näkökohtiin Twitterissä BIBREF3 , BIBREF7 , BIBREF8 , BIBREF11 , Wikipedian kommenteissa ja Facebook-postauksissa BIBREF2 .

**Tulos**

Mikä on loukkaavan kielenkäytön määritelmä?

**Esimerkki 5.1333**

Huomattavaa on, että vaikka arkkitehtuurimme perustuu ja on integroitu BERT$\_\text{base}$-malliin, se jopa päihittää erillisen BERT$\_\text{large}$-mallin huomattavasti.

**Tulos**

Mihin muihin malleihin kuin erilliseen BERT-malliin uutta mallia verrataan?

**Esimerkki 5.1334**

Annotaatioprojektiomenetelmä, jota noudatamme tässä työssä, on yksi tapa ratkaista tämä ongelma. Se otettiin käyttöön POS-merkintää, substantiivilauseiden perussanojen sulkeistamista, NER-merkintää ja taivutusmorfologista analyysia varten BIBREF29 , mutta sitä on käytetty myös riippuvuusjäsennykseen BIBREF30 , roolimerkintään BIBREF31 , BIBREF32 ja semanttiseen jäsennykseen BIBREF26 .

**Tulos**

Testaavatko kirjoittajat annotaatioprojisointitekniikoitaan muissa tehtävissä kuin AMR:ssä?

**Esimerkki 5.1335**

Esimerkiksi BIBREF16 ja BIBREF17 osoittivat, että vain hypoteeseja sisältävä perustaso suoriutuu satunnaista paremmin leksikaaliseen valintaan ja lauseen pituuteen liittyvien vihjeiden ansiosta. Vastaavasti BIBREF18 osoitti, että NLI-malleilla on taipumus ennustaa entailment lausepareille, joissa on paljon leksikaalista päällekkäisyyttä.

**Tulos**

Mitkä ovat esimerkkejä näistä esineistä?

**Esimerkki 5.1336**

CNN-C tarkoittaa CNN:ää, jossa on (kiinalaisten) merkkien upotus.CNN-W tarkoittaa CNN:ää, jossa on (kiinalaisten) sanojen upotus.CNN-Lex-C tarkoittaa algoritmia, jossa CNN:ään integroidaan myös polaariset sanat, ja jota ovat ehdottaneet Shin et al. BIBREF24 . CNN-Lex-W tarkoittaa algoritmia, joka integroi myös polaariset sanat CNN:ään, jota Shin et al. ovat ehdottaneet BIBREF24 . Bi-LSTM-C tarkoittaa BI-LSTM:ää, jossa on (kiinalaisten) merkkien upotus.Bi-LSTM-W tarkoittaa Bi-LSTM:ää, jossa on (kiinalaisten) sanojen upotus.Lex-rule tarkoittaa kuvassa 1 esitettyä sääntöpohjaista lähestymistapaa. BOW tarkoittaa perinteistä algoritmia, joka perustuu sanasäkkiominaisuuksiin.

**Tulos**

Mihin muihin malleihin niitä verrataan?

**Esimerkki 5.1337**

Muunnetun semanttisen avaruuden luotettavuuden todentamiseksi ehdotamme sanojen samankaltaisuustietokantoihin perustuvaa arviointivertailua. Kun otetaan huomioon rikastettu avaruus INLINEFORM0 ja samankaltaisuustietokanta INLINEFORM1 , lasketaan kunkin sanaparin INLINEFORM2 samankaltaisuus näiden kahden avaruuden vastaavien muunnettujen vektoreiden INLINEFORM3 ja INLINEFORM4 kosinuskohtaisena samankaltaisuutena, jossa INLINEFORM5 ja INLINEFORM6 LS:lle ja INLINEFORM7 ja INLINEFORM8 CCA:lle. Korkea suorituskyky tässä vertailuarvossa osoittaa, että kartoitus on onnistunut sijoittamaan semanttisesti samankaltaiset termit lähelle toisiaan, kun taas toisistaan poikkeavat termit ovat suhteellisen kaukana toisistaan avaruudessa. Toistamme laskennan kullekin parille päinvastaiseen suuntaan.

**Tulos**

Miten suorituskykyä mitataan?

**Esimerkki 5.1338**

Sekä Yhdistyneen kuningaskunnan että Ranskan sulauttamismatriisien ensimmäiset analyysit osoittivat, että sanat ryhmiteltiin yleensä asiayhteyden tai sähkönkulutukseen kohdistuvan vaikutuksen mukaan. Havaitsimme esimerkiksi, että talvisanat olivat yhdessä ja kaukana kesäsanoista.

**Tulos**

Onko olemassa esimerkkejä, joissa geometrinen ominaisuus näkyy sanojen samankaltaisuuden kontekstissa?

**Esimerkki 5.1339**

Konekäännöstä käytetään kemian tietotekniikassa "kääntämisessä" yhdeltä kieleltä (esim. reaktiot) toiselle (esim. tuotteet). Variational Auto-encoder (VAE) on toinen laajalti käytetty tekstinmuodostusarkkitehtuuri BIBREF101. Generative Adversarial Network (GAN) -mallit tuottavat uusia molekyylejä käyttämällä kahta komponenttia: generaattoriverkko tuottaa uusia molekyylejä, ja diskriminaattoriverkon tarkoituksena on erottaa generoidut molekyylit todellisista molekyyleistä BIBREF107.

**Tulos**

Ovatko nämä mallit yleensä puolivalvottuja vai valvomattomia?

**Esimerkki 5.1340**

Otimme japanilaisia sanapareja Evaluation Dataset of Japanese Lexical Simplification kodaira -aineistosta.

**Tulos**

mistä tiedot ovat peräisin?

**Esimerkki 5.1341**

Kaikki tiedot ladattiin Twitteristä kahdessa erillisessä erässä käyttämällä "twint"-kaapimistyökalua BIBREF5 otokseen historiallisia twiittejä useilla eri hakusanoilla; kyselyihin sisältyi aina joko "ilmastonmuutos" tai "ilmaston lämpeneminen", ja lisäksi niihin sisältyi katastrofikohtaisia hakusanoja (esim. "pommisykloni", "lumimyrsky", "lumimyrsky" jne.).

**Tulos**

Raportoidaanko tulokset vain englanninkielisistä tiedoista?

**Esimerkki 5.1342**

Voimme nähdä, että 1.10älykkyyden käyttö vaihtelee eri aloilla: tietotekniikan alalla samankaltaisimpia sanoja ovat 1.10artificial ja 1.10ai; rahoituksen alalla samankaltaisia sanoja ovat 1.10abilities ja 1.10consciousness.

**Tulos**

Mitä sanoja käytetään eri tavoin ArXivissa?

**Esimerkki 5.1343**

Osoitimme Katecheon käyttökelpoisuuden ottamalla järjestelmän käyttöön kysymyksiin vastaamiseen kahdessa aiheessa, lääketieteessä ja kristinuskossa.

**Tulos**

kuinka monta verkkotunnusta he kokeilivat?

**Esimerkki 5.1344**

Tässä tutkimuksessa keskityimme suosittelujärjestelmien LDA-lähestymistapoihin, ja koska tutkimus on tärkeää, olemme tutkineet viimeaikaisia vaikuttavia artikkeleita tästä aiheesta ja esittäneet viimeaikaisen tutkimuksen LDA:han perustuvien suosittelujärjestelmien taksonomian. arvioimme ISWC- ja WWW-konferenssien artikkeleita DBLP-sivustolta ja käytimme Gibbs-näytteenottoalgoritmia arviointiparametrina. Onnistuimme löytämään LDA-aiheiden ja artikkelien ominaisuuksien välisen suhteen ja saimme myös tutkijoiden kiinnostuksen tutkimusalaa kohtaan. Likimääräisen päättelyn ja LDA:n oppimisen suorittamiseksi on olemassa monia päättelymenetelmiä LDA:n aihepiirimallia varten, kuten Gibbs-näytteenotto, romahtanut Variational Bayes, Expectation Maximization. Gibbs-näytteenotto on suosittu tekniikka sen yksinkertaisuuden ja alhaisen latenssin vuoksi. Suurten aiheiden lukumäärän tapauksessa Gibbs sampling voi kuitenkin muuttua hankalaksi. Tässä artikkelissa käytämme Gibbs Sampling -menetelmää kokeessamme luvussa 5.

**Tulos**

Miten he käyttävät LDA:ta ja Gibbs-näytteenottoa ISWC- ja WWW-julkaisujen arviointiin?

**Esimerkki 5.1345**

Näin ollen olemme määritelleet kohteiksemme Galatasarayn (eli Kohde-1) ja Fenerbahçen (eli Kohde-2), jotka ovat kaksi Turkin suosituinta jalkapalloseuraa.

**Tulos**

Mitkä urheiluseurat ovat kohteita?

**Esimerkki 5.1346**

Odotamme, että selityksen suorituskyvyn pitäisi korreloida ennustuskyvyn kanssa. Koska mahdollista vastausta koskevaa tietoa tarvitaan ensisijaisesti päättämään, onko verkolla riittävästi tietoa vastata haastekysymykseen arvaamatta, ja relevanttia muuttujaa koskevaa tietoa tarvitaan, jotta verkko tietäisi, mitä kysyä, analysoimme verkon suorituskykyä kyselyssä ja vastaamisessa erikseen. Muistiverkolla on erityisiä vaikeuksia oppia kyselemään relevantteja muuttujia, sillä se saavuttaa vain noin 0,5 prosentin tarkkuuden kysyttäessä. Samaan aikaan se oppii vastaamaan erittäin hyvin, sillä se saavuttaa yli 0,9 prosentin tarkkuuden. Koska pyydämme sitä selittämään näitä vuorovaikutuksen kahta osaa kahdessa tilassa, havaitsemme, että selitysten laatu korreloi vahvasti verkon suorittaman algoritmin laadun kanssa.

**Tulos**

Miten he mittaavat korrelaatiota ennusteen ja selityksen laadun välillä?

**Esimerkki 5.1347**

Tässä artikkelissa tarkastellaan mahdollisuuksia ennustaa käyttäjän toimiala - tietyn alan yritysten kokonaisuus - tunnistamalla sosiaalisessa mediassa esiintyvää toimialakohtaista tekstiä.

**Tulos**

Mitä he tarkoittavat henkilön toimialalla?

**Esimerkki 5.1348**

1. BERT-pohjainen hienosäätö: Ensimmäisessä lähestymistavassa, joka on esitetty kuvassa KUVA 8, BERT-pohjaan tehdään hyvin vähän muutoksia. Tässä arkkitehtuurissa käytetään ainoastaan yhteistyöelimen antamaa [CLS]-merkin ulostuloa. [CLS]-ulostulo, joka vastaa 12. muuntajakooderin [CLS]-merkkilähdettä, joka on 768-kokoinen vektori, annetaan syötteenä täysin kytketylle verkolle, jossa ei ole piilokerrosta. Piilokerrokseen sovelletaan luokittelua varten softmax-aktivointifunktiota. 2. Lisätään epälineaarisia kerroksia: Tässä päivitetään ensimmäistä arkkitehtuuria ja luodaan arkkitehtuuri, jossa on kestävämpi luokittelija ja jossa käytetään täysin kytkettyä verkkoa, jossa ei ole piilokerrosta, vaan täysin kytkettyä verkkoa, jossa on kaksi piilokerrosta, joiden koko on 768. Tässä käytetään täysin kytkettyä verkkoa, jossa on kaksi piilokerrosta, joiden koko on 768. Kahdessa ensimmäisessä kerroksessa käytetään Leaky Relu -aktivointifunktiota, jonka negatiivinen kaltevuus = 0,01, mutta viimeisessä kerroksessa käytetään ensimmäisen arkkitehtuurin tapaan softmax-aktivointifunktiota, kuten kuvassa KUVA 8.3 esitetään. Lisätään Bi-LSTM-kerros: Toisin kuin aiemmissa arkkitehtuureissa, joissa käytetään vain [CLS] luokittimen syötteenä, tässä arkkitehtuurissa viimeisimmän muuntajakooderin kaikkia ulostuloja käytetään siten, että ne annetaan syötteenä kaksisuuntaiselle rekursiiviselle neuroverkolle (Bi-LSTM), kuten kuvassa FIGREF8 on esitetty. Syötteen käsittelyn jälkeen verkko lähettää lopullisen piilotetun tilan täysin kytketylle verkolle, joka suorittaa luokittelun käyttäen softmax-aktivointifunktiota. 4. Lisää CNN-kerros: Tässä kuvassa FIGREF8 esitetyssä arkkitehtuurissa käytetään kaikkien muuntajakoodereiden ulostuloja sen sijaan, että käytettäisiin viimeisimmän muuntajakooderin ulostuloa. Kunkin muuntajakoodaajan ulostulovektorit yhdistetään ja tuotetaan matriisi. Konvoluutio-operaatio suoritetaan ikkunalla, jonka koko on (3, BERT:n piilokoko, joka on 768 BERTbase-mallissa), ja maksimiarvo tuotetaan kullekin muuntajakooderille soveltamalla max poolingia konvoluutiotulokseen. Yhdistämällä nämä arvot luodaan vektori, joka annetaan syötteenä täysin kytketylle verkolle. Soveltamalla softmaxia syötteeseen suoritetaan luokitteluoperaatio.

**Tulos**

Mitä uusia hienosäätömenetelmiä esitellään?

**Esimerkki 5.1349**

ConvE:ssä vain $v\_h$ ja $v\_r$ muotoillaan uudelleen ja yhdistetään sitten tulomatriisiksi, joka syötetään konvoluutiokerrokseen.

**Tulos**

Kokeilivatko kirjoittajat useiden konvoluutiokerrosten pinoamista?

**Esimerkki 5.1350**

Kokeissamme käytämme APDA:n yhteisten tehtävien järjestäjien julkaisemia tietoja. Tietokokonaisuus on jaettu järjestäjien mukaan train- ja testitietokokonaisuuksiin. Koulutusjoukkoon jaetaan merkinnät kolmea tehtävää varten: ikä, murre ja sukupuoli.

**Tulos**

Mitä kolmea tietokokonaisuutta käytetään tutkimuksessa?

**Esimerkki 5.1351**

Chatino on Meksikon Oaxacassa puhuttu kieliryhmä. Chatino-kielet muodostavat yhdessä zapotekkien kieliryhmän kanssa otomanguean kieliperheen zapotekkihaaran.

**Tulos**

Mihin kieliperheeseen chatino kuuluu?

**Esimerkki 5.1352**

Toiminnallinen eroavuuspistemäärä laskettiin käyttämällä CoNLL 2017 Universal Dependencies -tehtävän BIBREF20 testijoukon lauseita kyseisillä kielillä, joiden UPOS-sekvenssit on annettu. Laskimme lähimpien naapureiden kokeen kaikille kielille edellä mainittujen mallien harjoitusaineistossa.

**Tulos**

Mitä arviointimittareita he käyttävät kielen mallintamisessa?

**Esimerkki 5.1353**

Toisessa algoritmissa käytetään viimeaikaisia ajatuksia tietojen valinnasta konekääntämisessä BIBREF7 . Kehitämme näiden ongelmien ratkaisemiseksi uudenlaisen uudelleenpainotusmenetelmän, jossa hyödynnetään ajatuksia, jotka ovat saaneet inspiraationsa konekääntämisen tiedonvalinnasta BIBREF26 , BIBREF7 .

**Tulos**

Mikä on tietojen valinta paperi konekääntäminen

**Esimerkki 5.1354**

Tietojemme mukaan ainoa tieteellisen tiivistämisen vertailuarvo on TAC 2014:n tiivistämisraiteelta. Rouge-varianttien ja metriikkamme (Sera) tehokkuuden arvioimiseksi käytämme tätä vertailukohtaa, joka koostuu 20 aihealueesta, joista jokaisessa on biolääketieteellisen lehden artikkeli ja 4 ihmisen kirjoittamaa kultaista tiivistelmää. Tarkastellaan seuraavaa esimerkkiä: Endogeenisiä pieniä RNA:ita (miRNA) seulottiin geneettisesti ja tutkittiin niiden miRNA:iden löytämiseksi, jotka liittyvät kasvainten syntyyn.

**Tulos**

Raportoivatko kirjoittajat tulokset vain englanninkielisistä tiedoista?

**Esimerkki 5.1355**

Vaikka on selvää, että sulautuksiamme voidaan käyttää uusien ennustemallien ominaisuuksina, on myös hyvin helppoa sisällyttää oppimamme Dolores-sulautumat olemassa oleviin tietämysgraafeihin perustuviin ennustemalleihin. Ainoa vaatimus on, että malli hyväksyy syötteenä upotuskerroksen (olioille ja suhteille). Jos malli täyttää tämän vaatimuksen (ja monet tietämysgraafeihin perustuvat neuraaliset mallit täyttävät tämän vaatimuksen), voimme käyttää Dolores-istutuksia korvaavana vaihtoehtona. Aloitamme vain vastaavan upotuskerroksen Doloresin upotuksilla. Jäljempänä esitettävässä arvioinnissa osoitamme, miten useita uusimpia malleja voidaan parantaa eri tehtävissä yksinkertaisesti ottamalla Dolores käyttöön alkuperäisen upotuskerroksen tilalle.

**Tulos**

Miten kuvaajan mielekkäät ketjut valitaan?

**Esimerkki 5.1356**

Luomme kuvaajan käyttäen kaikkia tietoja, ja harjoitussarjan twiiteillä on alkuperäinen kielitunnistejakauma. Naiivi lähestymistapa twiitti-twiitti-aligraafin rakentamiseen vaatii O( INLINEFORM0 ) vertailua, jossa mitataan kunkin twiitin samankaltaisuutta kaikkien muiden kanssa. Sen sijaan suoritimme INLINEFORM1 -lähimmän naapurin luokittelun kaikille twiiteille, jotka esitettiin unigrammien säkkinä, ja vertasimme kutakin twiittiä ja INLINEFORM2 -naapureita. Käytimme Junto (mad) BIBREF5 -järjestelmää etsiessämme merkintöjä merkityistä solmuista merkitsemättömiin solmuihin. Konvergenssin jälkeen normalisoimme uudelleen alun perin merkitsemättömien solmujen merkintäpisteet INLINEFORM4 -arvon löytämiseksi.

**Tulos**

Miten merkinnät levitetään tätä lähestymistapaa käytettäessä?

**Esimerkki 5.1357**

Rakennamme kolme tietokokonaisuutta, jotka perustuvat IMDB- ja Yelp-arvosteluihin. IMDB-tietokanta binarisoidaan ja jaetaan harjoitus- ja testijoukkoon, joissa kummassakin on 25 000 arvostelua (2 000 arvostelua harjoitusjoukosta on varattu kehittämistä varten). Yelpin osalta binarisoimme arvostelut ja luomme kaksi tietokokonaisuutta, joissa säilytämme vain arvostelut, joissa on $\le $ 50 -merkkejä (yelp50) ja $\le $ 200 -merkkejä (yelp200).

**Tulos**

Mitä tietokokonaisuuksia ne käyttävät?

**Esimerkki 5.1358**

Esittelemämme QA-CTS-tehtävä yhtenäistää perinteisen CTS-tehtävän tulostusmuodon ja tekee harjoitustiedoista yhteiskäyttöisiä, mikä rikastuttaa harjoitustietoja. Kaikki kysymys-vastausparit ovat neljän kliinikon kommentoimia ja tarkistamia kolmenlaisilla kysymystyypeillä, nimittäin kasvaimen koolla, proksimaalisella resektiomarginaalilla ja distaalisella resektiomarginaalilla. Kokeelliset tulokset reaalimaailman aineistolla osoittavat, että ehdotettu malli kilpailee vahvojen perusmallien kanssa kaikissa kolmessa erityistehtävässä.

**Tulos**

Mitkä ovat yhdistettävät erityistehtävät?

**Esimerkki 5.1359**

On myös huomattava, että signaalit voidaan tunnistaa, vaikka malli luokittelisi suhteet väärin, koska ${\Delta }\_s$ ei perustu oikeaan luokitteluun: se ainoastaan mittaa sanan osuutta kontekstissa oikean merkinnän pistemäärään. Jos tarkastelemme kunkin sanan vaikutusta oikean relaation pistemäärään, tämän vaikutuksen pitäisi edelleen korreloida ja korreloi ihmisen arvioiden kanssa, jotka perustuvat siihen, mitä järjestelmä saattaa merkitä toiseksi tai kolmanneksi parhaaksi luokaksi.

**Tulos**

Missä ehdotettu metriikka on päällekkäinen juman-tuomion kanssa?

**Esimerkki 5.1360**

Koska INLINEFORM1:n merkinnät ovat tuntemattomia, EGL laskee gradienttinormin odotuksen kaikkien mahdollisten merkintöjen osalta. BIBREF3 tulkitsee EGL:n "odotetuksi mallin muutokseksi". Seuraavassa luvussa virallistetaan EGL:n intuitiota ja osoitetaan, että se seuraa luonnollisesti estimaattorin varianssin pienentämisestä. Yhtälö ( EQREF7 ) osoittaa, että INLINEFORM0:n pienentämiseksi testiaineistossa meidän on minimoitava odotettu varianssi INLINEFORM1 testijoukossa. Käytännön ongelmana on se, että emme tiedä INLINEFORM0:ta etukäteen. Voisimme sen sijaan korvata INLINEFORM1-estimaatin esivalmistellusta mallista, jossa on järkevää olettaa INLINEFORM2 olevan lähellä todellista INLINEFORM3 . Erävalinta toimii tällöin siten, että otetaan näytteet, joilla on suurimmat gradienttinormit, DISPLAYFORM0RNN:lle voidaan saada kunkin mahdollisen etiketin gradientit takaisinlevittämällä. Toinen käytännön ongelma on se, että EGL marginalisoi kaikki mahdolliset merkinnät, mutta puheentunnistuksessa merkintöjen määrä skaalautuu eksponentiaalisesti aika-askeleiden määrään. Siksi marginalisoimme vain INLINEFORM0 todennäköisimmät merkinnät. Ne saadaan palkkihakupurkujen dekoodauksella, kuten BIBREFissä7 . BIBREF3:n EGL-menetelmä on lähes sama kuin yhtälö ( EQREF8 ), paitsi että gradientin normi ei ole BIBREF3:ssa neliöllinen.

**Tulos**

Miten he laskevat varianssin mallin tuloksista?

**Esimerkki 5.1361**

Tämän tutkimuksen tavoitteena on tarjota empiiristä näyttöä siitä, että saksankielisten tekstien avoimia tietolähteitä hyödyntämällä on mahdollista ennustaa poliittisia ennakkoluuloja automaattisesti ja todennäköistä tarkemmin.

**Tulos**

Mistä maista ja kielistä poliittiset puheet ja manifestit ovat peräisin?

**Esimerkki 5.1362**

koneen ymmärtäminen Nelufar

**Tulos**

Mistä MC on lyhenne?

**Esimerkki 5.1363**

Twitter API BIBREF15 -rajapinnasta poimitaan yhteensä 2 50 000 Microsoftia koskevaa twiittiä 31. elokuuta 2015 - 25. elokuuta 2016 väliseltä ajalta. Mukaan otettiin myös Microsoftia koskevat Twitter-uutiset ja tuotejulkaisuja koskevat twiitit. Microsoftin osakkeen avaus- ja päätöskurssit 31. elokuuta 2015 ja 25. elokuuta 2016 väliseltä ajalta on saatu Yahoo! Finance BIBREF16 .

**Tulos**

Mitä tietokokonaisuutta käytetään mallin kouluttamiseen?

**Esimerkki 5.1364**

Tarkastelemme kolmea vertailutietoaineistoa: ( INLINEFORM0 ) Cora, BIBREF44 -ohjelmalla luotu tekstitietoja sisältävä viittausverkko. Hepth, Arxivista peräisin oleva korkean energian fysiikan teoriaa käsittelevä paperiviittausverkko, jossa paperien tiivistelmät ovat tekstitietoa. ( INLINEFORM2 ) Zhihu, BIBREF9:n rakentama Q&A-verkostotietokanta, jossa on 10 000 aktiivista käyttäjää tekstikuvauksineen ja heidän yhteistyölinkkeineen.

**Tulos**

Mitä tekstiaineistoa he käyttävät?

**Esimerkki 5.1365**

Käytämme yhtä julkista tietokokonaisuutta Social Honeypot -tietokokonaisuutta ja yhtä itse kerättyä tietokokonaisuutta Weibo-tietokokonaisuutta validoidaksemme ehdotettujen ominaisuuksien tehokkuuden. Ennen kuin suoritamme kokeet suoraan käytetyillä tietokokonaisuuksilla, poistamme ensin joitakin tilejä, joilla on vain vähän viestejä, koska twiittien määrä viittaa hyvin paljon roskapostittajiin. Englanninkielisestä Honeypot-tietokannasta poistetaan stopwords, välimerkit, muut kuin ASCII-sanat ja tehdään stemming. Kiinalaisen Weibo-tietokannan segmentointi suoritetaan kiinalaisen tekstin segmentointityökalun Jieban avulla. Esikäsittelyvaiheiden jälkeen Weibo-tietokanta sisältää 2197 laillista käyttäjää ja 802 roskapostittajaa, ja honeypot-tietokanta sisältää 2218 laillista käyttäjää ja 2947 roskapostittajaa.

**Tulos**

Mikä on vertailuaineisto ja onko sen laatu korkea?

**Esimerkki 5.1366**

Edellä esitetyn mukaisesti tehtävässä on havaittava semanttinen muutos kahden korpuksen välillä. Jaetussa tehtävässä käytetyt kaksi korpusta vastaavat BIBREF0:n diakronista korpusparia: DTA18 ja DTA19. Ne koostuvat DTA-korpuksen BIBREF11 osista, joka on vapaasti saatavilla oleva lemmatoitu, POS-tagitettu ja oikeinkirjoitusnormalisoitu saksan kielen diakroninen korpus, joka sisältää tekstejä 1500-luvulta 1900-luvulle. DTA18 sisältää 26 miljoonaa lausetta, jotka on julkaistu vuosina 1750-1799, ja DTA19 40 miljoonaa lausetta, jotka on julkaistu vuosina 1850-1899.

**Tulos**

Mitä korpusta käytetään tehtävässä?

**Esimerkki 5.1367**

Kaikista yksittäisistä malleista LFT toimii parhaiten, ja sen jälkeen tulee MinAvgOut. RL on myös vertailukelpoinen aiempien huippumallien VHRED (attn) ja Reranking-RL kanssa. Uskomme, että tämä johtuu siitä, että LFT ei käytä voimaa vetääkseen mallin ennusteita pois perustotuuden merkkien luota, vaan se vain tekee itsensä tietoiseksi siitä, kuinka tylsiä kukin vastaus on. Näin ollen sen vastaukset vaikuttavat merkityksellisemmiltä kuin kahden muun lähestymistavan vastaukset. Lisäksi hybridimalli (viimeinen rivi) päihittää kaikki muut mallit selvästi.

**Tulos**

Mikä neljästä ehdotetusta mallista toimi parhaiten?

**Esimerkki 5.1368**

Käytämme ensin ennusteita, jotka perustuvat kirjallisuudessa aiemmin ehdotettuihin sääntöihin: sanan pituus, foneemien määrä, tavujen määrä, aakkosjärjestys ja taajuus.

**Tulos**

Mitä aiemmin ehdotettuja sääntöjä käytetään kaksisuuntaisen järjestyksen ennustamiseen?

**Esimerkki 5.1369**

Verrattuna BERT-BASE-malliin BERT-SPC parantaa huomattavasti aspektin polariteetin luokittelun tarkkuutta ja F1-pistemäärää.

**Tulos**

Kuinka paljon parempi on ehdotetun mallin suorituskyky verrattuna nykyiseen tekniikkaan näissä eri kokeissa?

**Esimerkki 5.1370**

Tämän seurauksena siirtokoulutus, jossa on vain 1000 tunnin tiedot, voi vastata 7300 tunnin tiedot sisältävän täyden koulutuksen suorituskykyä.

**Tulos**

kuinka pienellä aineistolla he harjoittelivat?

**Esimerkki 5.1371**

Taulukossa TABREF16 esitetään kasvojen ja äänen tunnetunnistuksen sekoitusmatriisit koko AMMER-aineistolla, ja taulukossa TABREF17 esitetään kunkin menetelmän luokkakohtaiset tulokset, mukaan lukien kasvo- ja äänitiedot sekä mikro- ja makrokeskiarvot. Kasvojen ilmeiden luokittelu tuottaa makrokeskiarvona $\text{F}\_1$ 33 %:n tuloksen kolmelle tunteelle: ilo, epävarmuus ja ärsytys (P=0,31, R=0,35).

**Tulos**

Miten kasvojen ja äänitiedon analysointia arvioidaan?

**Esimerkki 5.1372**

Aloitimme indeksoinnin 100 kysymyksellä, jotka valittiin satunnaisesti eri aihealueilta, jotta eri kysymystyypit saataisiin katettua. Kysymysten indeksoinnissa noudatetaan BFS-mallia aiheeseen liittyvien kysymysten linkkien kautta. Saimme 822 040 yksilöllistä kysymystä 80 253 eri aihealueelta ja yhteensä 1 833 125 vastausta näihin kysymyksiin.

**Tulos**

Keskittyvätkö kokeet tiettyyn alaan?

**Esimerkki 5.1373**

Ensimmäinen vaihe on mallin esivalmennus yhdellä MCQA-aineistolla, jota kutsutaan lähdetehtäväksi ja joka sisältää yleensä runsaasti harjoitusdataa. Toisessa vaiheessa samaa mallia hienoviritetään toisella MCQA-aineistolla, jota kutsutaan kohdetehtäväksi ja josta me itse asiassa välitämme, mutta joka sisältää yleensä paljon vähemmän harjoitusdataa.

**Tulos**

Kuinka suuri on lähteen ja kohteen tietokokonaisuuden koko?

**Esimerkki 5.1374**

NMT:hen käytimme KyotoNMT-järjestelmää BIBREF16 . NMT-koulutusasetukset ovat samat kuin WAT 2016 -tapahtumaan osallistuneiden parhaiden järjestelmien asetukset. Lähde- ja kohdesanastojen, lähde- ja kohdesivun upotusten, piilotettujen tilojen, huomiomekanismin piilotettujen tilojen ja syvän softmax-ulostulon, jossa on 2-maxout-kerros, koot olivat vastaavasti 32 000, 620, 1000, 1000 ja 500. Käytimme 2-kerroksisia LSTM:iä sekä lähde- että kohdepuolella. Oppimisalgoritmina käytettiin ADAMia, jossa kerrosten välisen pudotuksen pudotusprosentti oli 20 % ja L2-regularisointi, jossa painon hajoamiskerroin oli 1e-6. Minierän koko oli 64, ja yli 80 merkkiä pidemmät lauseet hylättiin.

**Tulos**

Millaisia neuroverkkoja he käyttivät tässä tutkimuksessa?

**Esimerkki 5.1375**

Kooderi on kaksisuuntainen Long-Short Term Memory (LSTM) solu BIBREF14 ja dekooderi yksittäinen LSTM-solu, jossa on huomiomekanismi.

**Tulos**

Kuinka monta huomiokerrosta heidän mallissaan on?

**Esimerkki 5.1376**

Koska saman luokan tekstien sanat kuuluvat samaan kontekstiin, on mahdollista mallintaa kunkin luokan sanavektorit sanojen aliavaruuksina ja vertailla niitä tehokkaasti samankaltaisuuden suhteen käyttämällä sanojen aliavaruuksien välisiä kanonisia kulmia.

**Tulos**

Mitä sana aliavaruus voi edustaa?

**Esimerkki 5.1377**

Esittelemme tässä L-PCFG:n, jossa on kaksi latenttien tilojen kerrosta: toinen kerros on tarkoitettu tavanomaisen syntaktisen tiedon tallentamiseen, ja toisella kerroksella pyritään tallentamaan semanttista ja topikaalista tietoa käyttämällä suurta joukkoa tiloja, joilla on erityiset ominaisuusfunktiot.

**Tulos**

Mitä latentteja muuttujia PCFG:ssä mallinnetaan?

**Esimerkki 5.1378**

WMT14 En-Fr- ja En-De-tietokannat WMT 2014 -mallin testaamiseen käytetään suurta englannin ja ranskan käännöstietokantaa, jossa on 36 miljoonaa dollaria lausepareja. Keskikokoisen tietokokonaisuuden osalta lainaamme BIBREF0:n asetusta ja otamme käyttöön WMT 2014 English-German translation dataset, joka koostuu 4,5 miljoonasta lauseparista, ja BPE-sanaston kooksi asetetaan 32 000 dollaria. IWSLT De-En- ja En-Vi-tietokannat Lisäksi teemme kokeita kahdella pienellä IWSLT-tietokannalla testataksemme MUSEn pientä versiota muiden vastaavien mallien kanssa. IWSLT 2014 German-English translation dataset koostuu 160 000 $$ lausepareista. Otamme käyttöön myös yhteisen lähde- ja kohdekielisen BPE-faktoroinnin, jonka sanaston koko on $32K$. IWSLT 2015 -englanti-vietnam -käännöstietokanta koostuu $133K$:n harjoituslausepareista.

**Tulos**

Mitä tietokokonaisuuksia käytetään?

**Esimerkki 5.1379**

Tehtävien käsittely: tietojen etsintätehtävien muuntaminen algebrallisiksi operaatioiksi sulautusavaruudessa tehtäväkohtaisia muunnosmalleja noudattamalla. Joitakin tärkeitä tehtäviä ja niiden muunnosmalleja käsitellään jaksossa SECREF5. Kyselyjen käsittely: semanttisten kyselyjen suorittaminen sulautusavaruudessa ja tulosten palauttaminen. Huomaa, että sulautusvektoreiden algebralliset operaatiot ovat lineaarisia ja ne voidaan suorittaa rinnakkain. Siksi semanttinen kysely on tehokas.

**Tulos**

Millaista tietojen etsintää näiden semanttisten rakenteiden analyysi tukee?

**Esimerkki 5.1380**

Kokeet ::: doc2vec BIBREF37. Doc2vec (tai paragraph vector) on word2vecin laajennus, joka oppii vektoreita asiakirjoille täysin ilman valvontaa. Asiakirjojen upotukset syötetään sitten logistiselle regressioluokittelijalle.CNN BIBREF38. Tietokonenäköalalla tunnettua konvoluutiohermoverkkoarkkitehtuuria sovelletaan tekstiin. Paikkatietoulottuvuuksia on yksi, ja sanojen upotuksia käytetään kanavina (syvyysulottuvuuksina).DAN BIBREF39. Deep Averaging Network (Syvä keskiarvoverkko) siirtää syötesanojen upotusten painottamattoman keskiarvon useiden tiheiden kerrosten ja lopullisen softmaxin läpi.Tree-LSTM BIBREF40 on tavanomaisen LSTM-arkkitehtuurin yleistäminen konstituutio- ja riippuvuuslukupuille.DRNN BIBREF41. Rekursiiviset neuroverkot pinotaan ja niitä sovelletaan jäsennyspuihin.LSTMN BIBREF42 on LSTM-mallin laajennus, jossa muistisolu korvataan muistiverkolla, joka tallentaa sanojen representaatiot.C-LSTM BIBREF43 yhdistää konvoluutio- ja rekursiiviset neuroverkot. CNN:n tuottamat alueiden upotukset syötetään LSTM:ään.SPGK BIBREF44 mallintaa myös dokumentit sanojen yhteisesiintymisverkkoina. Se laskee graafin ytimen, joka vertaa lyhimpiä polkuja, jotka on poimittu sanojen yhteisesiintymisverkoista, ja käyttää sen jälkeen SVM:ää asiakirjojen luokitteluun.WMD BIBREF45 on sovellus tunnetusta Earth Mover's Distance -menetelmästä tekstiin. S-WMD BIBREF46 on valvottu laajennus Word Mover's Distance -menetelmään. Semantic-CNN BIBREF47. LSTM-GRNN BIBREF26 on hierarkkinen malli, jossa lauseiden upotukset saadaan CNN:llä ja GRU-RNN:lle syötetään lauseiden representaatiot asiakirjavektorin saamiseksi.HN-ATT BIBREF27 on toinen hierarkkinen malli, jossa samaa koodausarkkitehtuuria (kaksisuuntainen GRU-RNN) käytetään sekä lauseiden että asiakirjojen koodaamiseen eri parametreilla. RNN-merkintöihin kullakin tasolla sovelletaan itsehuomautusmekanismia.

**Tulos**

Mikä on nykyaikainen järjestelmä?

**Esimerkki 5.1381**

Arvioimme lähestymistapaamme kuudella kohdekielellä: ranska (fr), venäjä (ru), arabia (ar), kiina (zh), hindi (hi) ja vietnam (vi).

**Tulos**

Mille kielille malli siirretään?

**Esimerkki 5.1382**

Kutsumme mallia modifioiduksi Toulminin malliksi. Se sisältää viisi argumentin osatekijää, nimittäin väitteen, lähtökohdan, perustelun, vastaväitteen ja kumoamisen. Kun asiakirjaa merkitään, mikä tahansa merkkiväli voidaan merkitä argumenttikomponentilla; komponentit eivät ole päällekkäisiä.

**Tulos**

Mitä argumenttikomponentteja ML-menetelmillä pyritään tunnistamaan?

**Esimerkki 5.1383**

Vertaamme lähestymistapaamme kahteen muuhun, joista ensimmäinen käyttää lähes samaa twiittidataa kuin me käytämme harjoitteluun, ja toinen on CrowdFlower-tietokanta, johon on merkitty tunteita.Ensimmäisessä Wang et al. BIBREF21 latasi yli 5 miljoonaa twiittiä, jotka sisälsivät yhden 131 emotionaalisesta hashtagista, jotka perustuvat Parrottin kolmiportaiseen tunteiden luokitteluun seitsemään kategoriaan: ilo, surullisuus, suuttumus, viha, rakkaus, pelko, kiitollisuus, yllätys. Toisessa raportoidut tulokset ovat BIBREF33:n julkaisusta, jossa he käyttivät maksimieentropialuokittelijaa ja bag of words -mallia luokittelemaan erilaisia tunnetietoaineistoja. Tässä raportoidaan vain osa heidän CrowdFlower-tietokokonaisuutta koskevista tuloksistaan, jotka voidaan yhdistää johonkin seitsemästä tunnemerkinnästämme.

**Tulos**

Mitä perustasoa käytetään?

**Esimerkki 5.1384**

Tässä jaksossa vertaamme malliamme uusimpiin järjestelmiin, mukaan lukien järjestelmät, joissa on eriasteista valvontaa. Perusversioihin kuuluvat mm: (1) Procrustes BIBREF11, joka oppii lineaarisen kartoituksen Procrustes Analysis BIBREF36:n avulla. (2) GPA BIBREF37, joka on Procrustes-analyysin laajennus. (3) GeoMM BIBREF38, geometrinen lähestymistapa, jossa opitaan Mahalanobis-metriikka tarkentamaan samankaltaisuuden käsitettä. (4) GeoMM$\_{semi}$, iteratiivinen GeoMM heikolla valvonnalla. (5) Adv-C-Procrustes BIBREF11, joka tarkentaa Adv-C:llä opittua kartoitusta iteratiivisella Procrustesilla, joka oppii uuden kartoitusmatriisin rakentamalla kaksikielisen sanaston iteratiivisesti. (6) Unsup-SL BIBREF13, jossa yhdistetään heikko valvomaton kartoitus ja kestävä itseoppiminen. (7) Sinkhorn-BT BIBREF28, jossa yhdistetään sinkhorn-etäisyys BIBREF29 ja takaisinkääntäminen.

**Tulos**

Mitkä ovat nykyiset uusimmat menetelmät, joissa näitä kahta tehtävää tarkastellaan toisistaan riippumatta?

**Esimerkki 5.1385**

Yhteisviittauksen päätöslauselma: Joskus on tarpeen käyttää rinnakkaisviittausten ratkaisutekniikoita tehokkaan haun varmistamiseksi, jotta voidaan tukea monikierroksista vuorovaikutusta. Macaw-ohjelmassa tunnistetaan kaikki rinnakkaisviittaukset käyttäjän viimeisimmästä pyynnöstä keskusteluhistoriaan. Samoja rinnakkaisviittausten ratkaisun tuloksia voidaan käyttää eri kyselyjen muodostuskomponenteissa. Tämä voi olla yleinen tai toimintokohtainen komponentti: Tämä komponentti luo kyselyn käyttäjän ja järjestelmän aiempien vuorovaikutustilanteiden perusteella. Kyselyn luomisessa voidaan hyödyntää rinnakkaisviittausten ratkaisua kyselyn laajentamiseen tai uudelleenkirjoittamiseen: Tämä on keskeinen luokittelukomponentti, joka hakee asiakirjoja tai kohtia suuresta kokoelmasta. Macaw voi hakea asiakirjoja mielivaltaisesta asiakirjakokoelmasta käyttämällä Indri python-rajapintaa BIBREF9, BIBREF10. Tarjoamme myös tuen verkkohaulle Bing Web Search API:n avulla. Macaw mahdollistaa myös monivaiheisen asiakirjojen uudelleenjärjestyksen: Haetut asiakirjat voivat olla liian pitkiä esitettäväksi joidenkin käyttöliittymien avulla. Tulosten tuottaminen on periaatteessa jälkikäsittelyvaihe, joka suoritetaan haetulla tulosluettelolla. Kysymyksiin vastaamisessa voidaan käyttää vastausten valinta- tai tuottamistekniikoita, kuten koneellisia luetun ymmärtämisen malleja. Esimerkiksi Macaw käyttää DrQA-mallia BIBREF11 kysymyksiin vastaamiseen.

**Tulos**

Mitä toimintoja Macaw tarjoaa?

**Esimerkki 5.1386**

Korpus on vapaasti saatavilla seuraavasta linkistä osoitteessa

**Tulos**

Onko tämä tietokokonaisuus julkisesti saatavilla?

**Esimerkki 5.1387**

Vertailua varten valitsimme parhaan tekstimallin kullekin esitykselle. Odotetusti suurin parannus ($22-26\%$ E.R) saadaan, kun tekstipohjaiset valvomattomat mallit yhdistetään kuvamalleihin.

**Tulos**

Kuinka paljon parempi on päättely, johon on lisätty kuvaesitys, verrattuna pelkkään tekstiesitykseen?

**Esimerkki 5.1388**

MP-kehys perustuu rekursiivisen naapuruusaggregoinnin ydinajatukseen. Toisin sanoen jokaisen vertexin esitys päivitetään jokaisessa iteraatiossa sen naapureilta saatujen viestien perusteella. Kaikkia spektraalisia GNN:iä voidaan kuvata MP-kehyksen avulla.GNN:iä on sovellettu menestyksekkäästi bioinformatiikan ja sosiaalisten verkostojen dataan solmujen luokitteluun, linkkien ennustamiseen ja graafien luokitteluun. Kuitenkin vain muutamissa tutkimuksissa on keskitytty MP-kehyksen soveltamiseen tekstin esittämisen oppimiseen. Tässä artikkelissa ehdotetaan yhtä tällaista sovellusta. Graafien yli tapahtuvan viestien välittämisen käsite on ollut käytössä jo useita vuosia BIBREF0, BIBREF1, samoin kuin graafeihin perustuvien neuroverkkojen käsite.

**Tulos**

Mikä on viestinvälityskehys?

**Esimerkki 5.1389**

BIBREF21:n mukaisesti otamme REINFORCEa soveltaessamme jokaisesta kuvasta $K$:n otoksen kuvateksteistä: ${\hat{c}}\_1 \ldots {\hat{c}}\_K$, ${\hat{c}}}\_k \sim p\_{\theta }(c|I)$,Kunkin otoksen kuvatekstin perustasona käytetään loput otokset sisältävien otosten keskimääräistä palkkiota.

**Tulos**

Mitä perustason funktiota käytetään REINFORCE-algoritmissa?

**Esimerkki 5.1390**

Arvioidaksemme kaikkien vertailtujen menetelmien tuottamia vastauksia, laskemme seuraavat automaattiset mittarit testijoukolle: BLEU: BLEU-n: BLEU-n mittaa keskimääräistä n-grammin tarkkuutta vertailuvastausten joukossa. Raportoimme BLEU-n:n, kun n=1,2,3,4.Distinct-1 & distinct-2 BIBREF5: Laskemme tuotetuissa vastauksissa olevien erillisten uni- ja bi-grammien lukumäärät ja jaamme lukumäärät testijoukon tuotettujen uni- ja bi-grammien kokonaismäärällä. Näitä mittareita voidaan pitää automaattisena mittarina vastausten monimuotoisuuden arvioimiseksi.

**Tulos**

Mitä automaattisia mittareita käytetään?

**Esimerkki 5.1391**

Luokittelijoiden harjoitteluun käytettiin osaa ILCI English-Hindi Tourism -rinnakkaiskorpuksesta (1500 lausetta).

**Tulos**

Minkä kokoista rinnakkaista korpusta käytetään mallirajoitusten kouluttamiseen?

**Esimerkki 5.1392**

Doukhan2018openin työn jälkeen halusimme tutkia korporaatioita tarkastelemalla kunkin sukupuolikategorian puhujien määrää ja heidän puheensa kestoa, sillä molemmat muuttujat ovat hyviä ominaisuuksia, joiden avulla voidaan ottaa huomioon sukupuolen edustus.

**Tulos**

Millaisia esityksiä tässä asiakirjassa esitetään?

**Esimerkki 5.1393**

Tämän asiakirjan pääkysymykset, joihin pyritään vastaamaan, ovat seuraavat: Haittaako yhteistyöelimen pakkaaminen sen kykyä siirtyä uusiin tehtäviin? Ja tekeekö hienosäätö yhteistyöelimestä enemmän vai vähemmän pakattavaa?Näiden kysymysten tutkimiseksi puristimme englanninkielisen yhteistyöelimen käyttäen suuruusluokan painojen karsintaa BIBREF8 ja havainnoimme tuloksia siirto-oppimisesta General Language Understanding Evaluation (GLUE) -vertailumittariin BIBREF9 , joka on monipuolinen joukko luonnollisen kielen ymmärtämistehtäviä, kuten tunneanalyysi, NLI ja tekstuaalisen samankaltaisuuden arviointi. Valitsimme suuruusluokan painojen karsinnan, joka pakkaa malleja poistamalla lähellä 0:a olevia painoja, koska se on yksi hienojakoisimmista ja tehokkaimmista pakkausmenetelmistä ja koska karsinnan tarkasteluun on monia mielenkiintoisia tapoja, joita tarkastelemme seuraavassa jaksossa. Havaintomme ovat seuraavat: Alhainen karsinta (30-40 %) ei lisää esiharjoittelun häviötä eikä vaikuta lainkaan siirtoon myöhempiin tehtäviin. Keskisuuret karsinnat lisäävät harjoittelua edeltävää häviötä ja estävät hyödyllisen harjoittelua edeltävän tiedon siirtymisen myöhempiin tehtäviin.

**Tulos**

Miten he havaitsivat, että yhteistyöjärjestelmän hienosäätö tiettyä tehtävää varten ei paranna sen karsittavuutta?

**Esimerkki 5.1394**

Tämä kerros koostuu kahdesta alamoduulista: Multi-Head Attention (MHA) ja Point-wise Convolution Transformation (PCT).

**Tulos**

Käytetäänkö niissä monihuomiopäitä?

**Esimerkki 5.1395**

Koulutusaineistona käytettiin 3 miljoonaa verkkosivua tästä korpuksesta, jotka käsiteltiin CCG-jäsennysohjelmalla loogisten muotojen BIBREF10 tuottamiseksi. Käytimme myös Krishnamurthyn ja Mitchellin luomaa testijoukkoa, joka sisältää 220 kyselyä, jotka on tuotettu samalla tavalla kuin harjoitusaineisto ClueWebin erillisestä osiosta. Koska he eivät kuitenkaan julkaisseet kehitysjoukkoa datansa kanssa, käytimme tätä joukkoa kehitysjoukkona. Tämä lopullinen testijoukko sisältää 307 kyselyä.

**Tulos**

Kuinka suuri on heidän tietokokonaisuutensa?

**Esimerkki 5.1396**

Negatiivisten arvojen käsittelemiseksi ehdotamme leikattua $\mathit {PMI}$ -mallia, joka vastaa $\mathit {PPMI}$ -mallia, kun $z = 0$. Käytännössä havaitsemme, että tämä toimii huonosti, jos se tehdään symmetrisesti, joten otamme käyttöön muunnoksen nimeltä $\mathit {NNEGPMI}$, joka ainoastaan normalisoi $\mathit {\texttt {-}PMI}$:

**Tulos**

Mitä uusia PMI-muunnoksia otetaan käyttöön?

**Esimerkki 5.1397**

Huomaa, että GANE-AP tuottaa parempia tuloksia kuin GANE-OT, mikä viittaa siihen, että huomion jäsentelymekanismi voi parantaa matalan tason keskinäisen huomion matriisia entisestään.

**Tulos**

Kumpi heidän ehdottamistaan huomiomenetelmistä toimii kokonaisuudessaan paremmin?

**Esimerkki 5.1398**

Määrittelemme metriikan nimeltä Semantic Text Exchange Score (STES), jolla arvioidaan mallin yleistä kykyä suorittaa STE, sekä säädettävän parametrin, jolla voidaan hallita semanttisen muutoksen määrää.

**Tulos**

Onko STES:ää käytetty aiemmin kirjallisuudessa samankaltaisten tehtävien arviointiin?

**Esimerkki 5.1399**

Löytääksemme suuren määrän pakolaiskriisiä koskevaa vihapuhetta käytimme 10 hashtagia, joita voidaan käyttää loukkaavalla tai loukkaavalla tavalla.

**Tulos**

Miten mahdolliset vihamieliset viestit tunnistettiin?

**Esimerkki 5.1400**

Käytämme arviointimittarina mikrokeskiarvoista F1-pistemäärää kaikilla merkinnöillä.

**Tulos**

Mitä mittareita käytetään arvioinnissa?

**Esimerkki 5.1401**

Tässä käytetään vain sanakirjamääritelmiin liittyviä tietoja ja jätetään huomiotta polariteettipisteet. Kun kuitenkin hyödynnämme valvottua pistemäärää (+1 tai -1), vastakkaisia polariteetteja sisältävät sanat (esim. "onnellinen" ja "onneton") etääntyvät toisistaan, kun ne käännetään koordinaattialueiden välillä.

**Tulos**

Miten sanojen valvotut pisteet lasketaan?

**Esimerkki 5.1402**

Osatehtävässä C tavoitteena on luokitella loukkaavan kielenkäytön kohde. Tässä BIBREF17 -tehtävässä otetaan huomioon vain osatehtävässä B kohdennetuiksi loukkauksiksi (TIN) leimatut viestit. Näytteet merkitään yhdellä seuraavista: Yksilö (IND): Viestit, jotka kohdistuvat nimettyyn tai nimeämättömään henkilöön, joka on osa keskustelua. Englanniksi tämä voi olla esimerkiksi viesti: @USER Is a FRAUD Female @USER group paid for and organized by @USER. Tanskan kielessä tämä voisi olla esimerkiksi USER du er sku da syg i hoved. Nämä esimerkit osoittavat edelleen, että tämä luokka kattaa verkkokiusaamisen piirteet, sellaisina kuin ne on määritelty kohdassa "Tausta". ryhmä (GRP): Viestit, jotka kohdistuvat ihmisryhmään etnisen alkuperän, sukupuolen tai seksuaalisen suuntautumisen, poliittisen suuntautumisen, uskonnollisen vakaumuksen tai muiden ominaisuuksien perusteella. Englanniksi tämä voisi olla esimerkiksi viesti #Antifa are mentally unstable cowards, pretending to be relevant. Tanskan kielessä tämä voisi olla esimerkiksi Åh nej! Svensk lorteret!Muu (OTH): Loukkaavan kielenkäytön kohde ei täytä kummankaan edellisen kahden kategorian kriteerejä. BIBREF17 . Englanniksi tämä voisi olla esimerkiksi viesti And these entertainment agencies just gonna be an ass about it... Tanskan kielellä tämä voisi olla esimerkiksi viesti Netto er jo et tempel over lort.

**Tulos**

Kuinka monta loukkaavan kielenkäytön luokkaa oli?

**Esimerkki 5.1403**

Käytetyt koneoppimistekniikat vaihtelivat maksimientropialuokittelijoista (BIBREF4) tukivektorikoneisiin (BIBREF5,BIBREF6,BIBREF7,BIBREF8), kun taas syväoppimismenetelmiin kuuluivat rekursiiviset neuroverkot (BIBREF9,BIBREF10), konvolutiiviset neuroverkot (BIBREF11) ja viimeisimpänä siirto-oppimiseen perustuvat arkkitehtuurit, kuten Bidirectional Encoder Representation from Transformers (BERT) (BIBREF12). Kuvissa FIGREF1 ja FIGREF1 on yhteenveto spekulaation havaitsemista ja laajuuden selvittämistä käsittelevistä artikkeleista (BIBREF13, BIBREF5, BIBREF9, BIBREF3, BIBREF14, BIBREF15, BIBREF16, BIBREF17, BIBREF16, BIBREF6, BIBREF11, BIBREF18, BIBREF18, BIBREF10, BIBREF19, BIBREF7, BIBREF4 ja BIBREF8).

**Tulos**

Mitkä olivat lähtötasot?

**Esimerkki 5.1404**

BD-4SK-ASR-tietokanta ::: Loimme kielen transkriptioista. Malli luotiin käyttämällä CMUSphinxiä, jossa (kiinteä) diskonttomassa on 0,5, ja backoffit lasketaan suhdelukumenetelmällä. Malli sisältää 283 unigrammia, 5337 bigrammia ja 6935 trigrammia.

**Tulos**

Mitkä ovat kokeen tulokset?

**Esimerkki 5.1405**

Kuten edellä esiteltiin, "One Million Post" -korpus tarjoaa merkintöjä yli 11 000 käyttäjäkommentille. Vaikka jaetussa tehtävässä määriteltyä "loukkaavaa kieltä" kuvaavaa suoraan vertailukelpoista luokkaa ei olekaan, on olemassa kaksi läheisesti toisiinsa liittyvää luokkaa. Poistamme aineistosta kaikki ne kommentit, joiden kommentoijien enemmistö on sitä mieltä, että ne sisältävät joko "sopimatonta" tai "syrjivää" sisältöä tai ei mitään edellä mainituista. Käsittelemme kahta ensimmäistä tapausta esimerkkeinä `rikkomuksesta' ja jälkimmäistä tapausta esimerkkeinä `muut'.

**Tulos**

Mitkä ovat lähes loukkaavat kielikategoriat?

**Esimerkki 5.1406**

Kaikki perusmallit ovat Transformer-malleja niiden peruskonfiguraatiossa BIBREF11, jossa käytetään 6 kooderi- ja dekooderikerrosta, mallin ja piilotetun mallin dimensiot ovat 512 ja 2048 ja kaikkien huomiokerrosten 8 päätä. Alkuvaiheen monitehtäväkokeissa kaikki malliparametrit jaettiin BIBREF12:lla, mutta suorituskyky laski useita BLEU-pisteitä perusmalleihin verrattuna.

**Tulos**

Mitä mukautumattomia peruslinjoja käytetään?

**Esimerkki 5.1407**

Kirjallisuudessa kuvatuissa oikeinkirjoituksen korjaustesteissä on yleensä keskitytty yhteen lähestymistapaan, jota on sovellettu tiettyyn korpukseen. Esimerkkeinä voidaan mainita mammografiaraporttien ja twiittien oikeinkirjoituksen tarkistusta käsittelevät teokset BIBREF7 , BIBREF4 . Näissä töissä korostettiin, että on tärkeää räätälöidä korjausjärjestelmät niiden korpusten erityisongelmiin, joihin niitä sovelletaan. Esimerkiksi mammografiaraportit kärsivät huonosta kirjoitusasusta, joka tässä tapauksessa on suhteellisen kiireessä tehtyä toistotyötä. Twiitit taas sisältävät usein hymiöitä ja neologismeja, jotka voivat olla hankalia sääntöihin ja sanakirjoihin perustuville ratkaisuille, kuten LanguageToolille. Jälkimmäinen soveltuu sinänsä melko hyvin puolankielisiin teksteihin, koska monet laajennukset tämän sovelluksen rakenteeseen ovat saaneet innoituksensa puolan kielen morfologian ongelmista BIBREF3 .

**Tulos**

Mitä erityisiä virheenkorjausratkaisuja on aiemmin ehdotettu erikoistuneita korporaatioita varten?

**Esimerkki 5.1408**

Tässä artikkelissa olemme formalisoineet automaattisen tyhjiin täytettävien tietokilpailujen tuottamisen ongelman käyttämällä kahta hyvin määriteltyä oppimismenetelmää: sekvenssien luokittelua ja sekvenssien merkitsemistä.

**Tulos**

Mitä kahta järjestelmää käytetään?

**Esimerkki 5.1409**

Arvioinnissa annotaattoreina toimi 50 natiivia henkilöä, jotka olivat hyvin perehtyneitä sekä englannin että tamilin kieliin. Vertailua varten otettiin noin 100 lauseen näytekokoelma testijoukon tuloksista. Tämä joukko sisälsi satunnaistetun valikoiman käännöstuloksia arvioinnin objektiivisuuden varmistamiseksi. RNNMorph-tulosten sujuvuus- ja riittävyystulokset on taulukoitu. Riittävyysluokitus laskettiin viisiportaisella asteikolla, jolla mitattiin, kuinka paljon merkitystä käännös välittää (kaikki, suurin osa, paljon, vähän, ei lainkaan). Sujuvuusluokitus laskettiin kieliopin oikeellisuuden perusteella viisiportaisella asteikolla (virheetön, hyvä, ei-äidinkielenomainen, epäselvä, käsittämätön). Vertailua varten RNNMorph- ja RNNSearch + Word2Vec -mallien lausetason käännökset asetettiin erikseen paremmuusjärjestykseen toistensa välillä, mikä mahdollisti kahden käännöksen tasapisteiden syntymisen paremmuusjärjestyksessä.

**Tulos**

Miten inhimilliset tuomiot koottiin?

**Esimerkki 5.1410**

Parantaaksemme suorituskykyä entisestään otamme käyttöön järjestelmän yhdistelmän dekoodausristikoiden tasolla. Yhdistämällä järjestelmiä voimme hyödyntää kunkin eri alueille optimoidun mallin vahvuuksia. Paras tulos vlsp2018:ssa (4,85 % WER) saadaan yhdistelmällä painot 0,6:0,4, jossa 0,6 annetaan yleiskielimallille ja 0,4 keskustelumallille. Vlsp2019-joukossa suhdetta muutetaan hieman 0,7:0,3, jolloin paras tulos on 15,09 %.

**Tulos**

Mitä kielimallien yhdistämistekniikkaa käytetään tutkimuksessa?

**Esimerkki 5.1411**

Olemme myös erityisen kiinnostuneita näkemään tämän mallin soveltamisen eri kieliin.

**Tulos**

Ovatko kirjoittajat kokeilleet tätä lähestymistapaa muilla kielillä?

**Esimerkki 5.1412**

ASR- ja ST-mallimme noudattavat berard2018endin arkkitehtuuria, mutta niissä on 3 dekooderikerrosta kuten pino2019harnessingissa.

**Tulos**

Mikä on heidän mallinsa arkkitehtuuri?

**Esimerkki 5.1413**

Mallinnamme ASP-sijoitustehtävän AEP-tehtävän seuraajaksi. Tehtävänä on määrittää kaikille "merkityksellisille" uutisyksikköpareille oikea yksikköosio. Wikipedian oliosivun jokainen osio edustaa eri aihetta. Esimerkiksi Barack Obamalla on osiot `Early Life', `Presidency', `Family and Personal Life' jne. Artikkeli-osio Perustotuus. Aineisto koostuu kolmikosta INLINEFORM0 , jossa INLINEFORM1 , jossa oletetaan, että INLINEFORM2 on jo määritetty relevantiksi. Meillä on siis moniluokkainen luokitusongelma, jossa meidän on määritettävä INLINEFORM3:n jakso, jossa INLINEFORM4:ää siteerataan.

**Tulos**

Miten he määrittävät tarkan osion, jossa käytetään syöttöartikkelia?

**Esimerkki 5.1414**

Käytämme ranskankielisen OSCAR-korpuksen käsittelemätöntä versiota, jossa on 138 gigatavua pakkaamatonta tekstiä ja 32,7 miljardia SentencePiece-merkkiä.

**Tulos**

Mihin CamemBERT on koulutettu?

**Tulos**

Mitä tietoja käytetään CamemBERT-koulutuksessa?

**Esimerkki 5.1415**

Näin ollen ehdotamme kahdenlaista suhdekohtaista metatietoa: suhdemetatietoa ja gradienttimetatietoa, jotka vastaavat edellä mainittuja kahta näkökulmaa. Ehdotetussa MetaR-kehyksessä relaation metatieto on korkeamman asteen esitys suhteesta, joka yhdistää pää- ja loppuoliot. Gradientti-meta on relaatio-metan häviögradientti, jota käytetään nopeaan päivitykseen ennen kuin relaatio-meta siirretään epätäydellisiin kolmioihin ennusteen aikana.

**Tulos**

Mitä metatietoja siirretään?

**Esimerkki 5.1416**

Kuten aiemmin mainittiin, tarpeisiimme ei ole saatavilla saksalaista vihapuhekorpusta, varsinkaan Euroopan pakolaiskriisiä koskevasta hyvin tuoreesta aiheesta. Siksi meidän oli koottava oma korpuksemme.

**Tulos**

Mitkä kielet sisältyvät vihamielisen sisällön tietokantaan?

**Esimerkki 5.1417**

Kaksi suosittua segmentointimenetelmää ovat morfeemien segmentointi BIBREF4 ja tavuparikoodaus (BPE) BIBREF5. Sanojen segmentoinnin jälkeen lisäämme lisäksi jokaisen erotetun sanayksikön taakse erityisen symbolin, jonka tarkoituksena on auttaa NMT-mallia tunnistamaan morfeemien rajat ja vangitsemaan semanttinen informaatio tehokkaasti. Hyödynnämme Zemberekiä morfologisen disambiguointityökalun kanssa turkkilaisten sanojen segmentoimiseksi morfeemiyksiköihin ja morfologian analysointityökalua BIBREF12 uiguurin sanojen segmentoimiseksi morfeemiyksiköihin.

**Tulos**

Miten sanojen segmentointimenetelmä toimii?

**Esimerkki 5.1418**

Saimme noin 300 000 lauseen korpuksen, josta saatiin noin 1,5 miljoonaa yksittäistä kysymystä sisältävää harjoitusesimerkkiä.

**Tulos**

Mikä on tietokokonaisuuden koko?

**Esimerkki 5.1419**

Niille, jotka haluavat Seshatin tiukemmin valvotun asennuksen järjestelmäänsä, annamme myös täydelliset tiedot manuaalisista asennusvaiheista verkkodokumentaatiossamme).

**Tulos**

Onko tämä ohjelmisto yleisön saatavilla?

**Esimerkki 5.1420**

Katecheon nykyisessä versiossa käytetään Bi-Directional Attention Flow (BiDAF) -mallia luetun ymmärtämiseen BIBREF6 . Katecheon tuleviin versioihin sisältyy mahdollisuus vaihtaa luetun ymmärtämisen malli uudempiin arkkitehtuureihin, jotka perustuvat esimerkiksi BERT BIBREF8 tai XLNet BIBREF9 -malleihin tai mukautettuihin koulutettuihin malleihin.Arkkitehtuuri ja konfigurointi

**Tulos**

mitä esivalmennettuja malleja käytettiin?

**Esimerkki 5.1421**

Kokeellinen protokolla ::: Aineistot ::: TTS System dataset: Koulutimme TTS-järjestelmämme sekoituksella neutraalia ja uutisankkurityylistä puhetta. Yhteensä 24 tuntia harjoitusaineistoa, joka jakautui 20 tuntiin neutraalia puhetta (22000 lausumaa) ja 4 tuntiin uutisankkurin tyylistä puhetta (3000 lausumaa).(ii) Sulauttamisen valinta-aineisto: Koska arviointi suoritettiin vain uutisankkurin puhetyylillä, rajoitamme kielellisen hakuavaruuden uutisankkurin tyyliin liittyviin lausumiin: Kokeellinen protokolla :::: Aineistot ::: Järjestelmät arvioitiin kahdella tietokokonaisuudella:(i) Common Prosody Errors (CPE): Tietokokonaisuus, jossa Prostron-perusmalli ei pysty tuottamaan sopivaa prosodiaa. Tämä tietokokonaisuus koostuu monimutkaisista lausumista, kuten yhdyssanoista (22 %), "tai"-kysymyksistä (9 %) ja "wh"-kysymyksistä (18 %). Joukkoa on täydennetty hankkimalla monimutkaisia lauseita (51 %) BIBREF-tietokannasta24. ii) LFR: Kuten BIBREF-tietokannassa25 osoitettiin, lauseiden arviointi yksinään ei riitä, jos halutaan arvioida pitkän puheen laatua. Näin ollen LFR-arviointeja varten kuratoimme uutisnäytteiden tietokokonaisuuden. Uutistyyliset lauseet ketjutettiin kokonaisiksi uutisjutuiksi, jotta ne kuvaisivat suunnitellun käyttötapauksen kokonaiskokemusta.

**Tulos**

Mitä tietokokonaisuutta käytetään tämän menetelmän harjoittelussa/testauksessa?

**Esimerkki 5.1422**

Software Ontology (SWO) BIBREF5 on otettu mukaan, koska sen CQ-joukko on huomattavan suuri ja se oli osa Ren et al. analysoitujen CQ-joukkoa. Dem@Care BIBREF8:n ja OntoDT BIBREF9:n CQ-joukot otettiin mukaan, koska ne olivat saatavilla. Stuff BIBREF6- ja African Wildlife (AWO) BIBREF7 -ontologioiden CQ:t lisättiin joukkoon, koska yksi kirjoittajista on kehittänyt nämä ontologiat (mikä helpottaa tarvittaessa syvällistä analyysia), ne kattavat muita aiheita ja ovat eri tyyppisiä (opasontologia (AWO) ja ydinontologia (Stuff)), mikä osaltaan edistää lähteiden valinnan monimuotoisuutta.

**Tulos**

Kuinka monesta ontologian alueesta ne keräävät tietoja?

**Esimerkki 5.1423**

Menetelmää arvioidaan BERTbase-mallilla, jossa on 12 kerrosta, 12 itsetarkkailupäätä ja piilokoko 768. Koulutusnopeuden nopeuttamiseksi käytetään kaksivaiheista koulutusta BIBREF1. Ensimmäisessä vaiheessa käytetään 128 lauseen maksimipituutta ja toisessa vaiheessa 512 lauseen pituutta. Kahden vaiheen koulutusvaiheiden määrät ovat 50 000 ja 40 000 BERTBase-mallissa. Käytimme AdamW BIBREF13 -optimointilaitetta, jonka oppimisnopeus on 1e-4, $\beta \_1$ 0,9, $\beta \_2$ 0,999 ja L2-painon hajoamisnopeus $0,01$. Ensimmäiset 10 % kokonaisvaiheista käytetään oppimisnopeuden lämpenemiseen, minkä jälkeen käytetään lineaarista hajoamiskaaviota. Käytimme kaikilla kerroksilla 0,1:n pudotustodennäköisyyttä. Esiharjoittelussa käytetty data on sama kuin BERT:ssä, eli englanninkielinen Wikipedia (2500 miljoonaa sanaa) ja BookCorpus (800 miljoonaa sanaa) BIBREF14.

**Tulos**

Harjoittelevatko he mallia tarkistuspisteestä alkaen?

**Esimerkki 5.1424**

Koska BERT-kielimallin esivalmennus on kallista, otamme suoraan käyttöön Googlen esivalmennetut parametrit kiinalaisesta yleiskorpuksesta. Nimettyjen entiteettien tunnistusta sovelletaan sekä patologisten raporttien teksteihin että kyselyteksteihin.

**Tulos**

Millä aineistolla kielimalli on esivalmennettu?

**Esimerkki 5.1425**

Intuitiivisin tapa esittää syntaktista tietoa on käyttää suoraan yksittäisiä riippuvuussuhteita, kuten riippuvuuspäätettä ja riippuvuussuhteen merkintää, joita merkitään lyhyesti Dep ja Rel. Jotta riippuvuuspuiden rakenteellinen informaatio säilyisi mahdollisimman hyvin, otamme riippuvuuspuiden argumenttiehdokkaiden ja predikaattien välisen syntaktisen polun kielellisenä tietona. Viitaten BIBREF9:ään käytämme riippuvuuspolkuna (DepPath) puupohjaista sijaintiominaisuutta (TPF) ja lyhintä riippuvuuspolkua (SDP) relaatiopolkuna (RelPath).

**Tulos**

Mitä erilaisia lähestymistapoja syntaktisen tiedon koodaamiseen kirjoittajat esittävät?

**Esimerkki 5.1426**

Syöttötietojen simuloimiseksi kehitettiin rutiinit, jotka perustuvat kirjoittajien kokemukseen todellisista terveydenhuollon tiedoista. Tähän valintaan oli kaksi syytä: Ensinnäkin, terveydenhuollon tiedot voivat olla erittäin monimutkaisia ja vaatia kertaluonteista koodia epätavallisen syötteen käsittelemiseksi, mutta eivät välttämättä niin, että ne muuttaisivat merkittävästi tämän kaltaisen semanttisen rikastusmoottorin perustavanlaatuisia teknisiä valintoja. Toiseksi, terveydenhuollon tietoja säännellään tiukasti, ja terveydenhuollon tietojen saaminen tutkimuskäyttöön voi olla hankalaa ja aikaa vievää.Simuloinnissa käytettiin yksinkertaistettua syöttötietoaineistoa, joka oli eri muodoissa, joita esiintyy usein terveydenhuollossa.

**Tulos**

Minkälaisia reaaliaikaisten syötteiden simulaatioita käytetään validoinnissa?

**Esimerkki 5.1427**

Osoittaaksemme lähestymistapamme tehokkuuden esitämme tuloksia SICK-tietokannassa BIBREF1, joka on yleinen vertailukohde logiikkapohjaiselle NLI:lle, ja havaitsemme, että MonaLog on kilpailukykyinen monimutkaisempien logiikkapohjaisten lähestymistapojen kanssa (monet niistä vaativat täydellistä semanttista jäsennystä ja monimutkaisempia loogisia koneistoja).

**Tulos**

Päihittävätkö ne SICKin nykyisen tekniikan tason?

**Esimerkki 5.1428**

Kahden puhujan päällekkäinen puhe tuotetaan keinotekoisesti sekoittamalla näitä aaltomuotoisia segmenttejä. Maksimoidaksemme puheen päällekkäisyyden kehitimme menettelyn, jolla samankokoisia segmenttejä sekoitetaan noin 0 dB:n tasolla. Ensin lajittelemme puhesegmentit pituuden mukaan. Sitten otamme segmentit pareittain ja nollaamme lyhyemmän segmentin niin, että molemmilla on sama pituus. Sitten nämä parit sekoitetaan yhteen päällekkäisen puheaineiston luomiseksi.

**Tulos**

Miten nämä kaksi tietokokonaisuutta ovat keinotekoisesti päällekkäisiä?

**Esimerkki 5.1429**

Käytämme graafia sosiaalisen median kontekstin mallintamiseen, jossa twiitit liittyvät toisiinsa, kirjoittajat twiitteihin ja muihin kirjoittajiin. Kuvassa FIGREF7 esitetään graafi, joka koostuu kolmenlaisista solmuista: twiitit (T), käyttäjät (U) ja "maailma" (W). Solmujen välille luodaan särmiä, jotka painotetaan seuraavasti: T-T twiittien välinen unigram-kosiinimaisuus, T-U painotettu 100 twiitin ja sen kirjoittajan välillä, U-U painotettu 1 kahden käyttäjän välillä, joilla on "seuraa"-suhde, ja U-W painotettu 0,001, jotta varmistetaan, että graafi on yhdistetty hullua algoritmia varten.

**Tulos**

Mitä tietoa twiitin kirjoittajien sosiaalinen graafi sisältää?

**Esimerkki 5.1430**

Ranskaa, venäjää ja arabiaa voidaan pitää kielinä, joilla on paljon resursseja, kun taas hindistä on paljon vähemmän tietoja, ja sitä voidaan pitää kielinä, joilla on vähän resursseja.

**Tulos**

Onko järjestelmää testattu vähäisten resurssien kielillä?

**Esimerkki 5.1431**

Aloitimme näiden sanojen upotukset 300-ulotteisilla Glove- upotuksilla BIBREF31 .

**Tulos**

Käytetäänkö niissä esivalmisteltuja upotuksia?

**Esimerkki 5.1432**

Perustason määrittämiseksi tarkastelemme satunnaista menetelmää, joka ennustaa positiivisen merkinnän 0,15 todennäköisyydellä (positiivisten tapausten perusprosentti). Yleinen suorituskyky (kuva FIGREF28). Vaikka sanatason tehtävämme on erittäin epätasapainoinen, kaikki mallimme ovat selvästi parempia kuin satunnainen perustaso. Odotetusti sisällön sanoja on paljon vaikeampi ennustaa kuin stop-sanoja, mutta paras F1-tulos sisällön sanojen osalta on yli kaksinkertainen satunnaiseen perusmalliin verrattuna (0,286 vs. 0,116). Huomattavaa on, että vaikka parannamme satunnaista perustasoa huomattavasti, parhaatkin F1-pisteet ovat suhteellisen alhaisia, ja tämä pätee käytetystä mallista riippumatta. Huolimatta siitä, että tehtäviin sisältyy enemmän merkkejä kuin tavanomaisiin merkintätehtäviin (esim. BIBREF41 ja BIBREF42), sen ennustaminen, tuleeko sana kaikuisaksi selityksissä, on edelleen haastava ongelma.

**Tulos**

Mitkä ovat tämän uuden tehtävän yleiset perustulokset?

**Esimerkki 5.1433**

Esittelemme luettelon kahdeksasta eri osaamisesta, jotka lukemisjärjestelmän tulisi hallita, jotta se voisi käsitellä arvosteluja ja tekstiasiakirjoja yleensä. Nämä kahdeksan tehtävää edellyttävät erilaista osaamista ja eritasoista asiakirjan ymmärtämistä, jotta niihin voidaan vastata hyvin. Esimerkiksi sen havaitseminen, mainitaanko arvostelussa jokin näkökohta, edellyttää vähemmän arvostelun ymmärtämistä kuin kyseisen näkökohdan arvosanan ennustaminen. Taulukossa TABREF10 esitellään kahdeksan tehtävää, jotka olemme ottaneet käyttöön tässä tietokokonaisuudessa, sekä esimerkki kutakin tehtävää vastaavasta kysymyksestä.

**Tulos**

Millaisia kysymyksiä aineistossa on?

**Esimerkki 5.1434**

Regressorina käytettiin tukivektoriregressiota (SVR) ja luokittelijana tukivektoriluokitusta (SVC). Taulukkoon sisältyy neuraalinen regressori (NNR) ja neuraalinen luokittelija (NNC). Neuraalinen luokittelija on koulutettu kahdessa asetelmassa: "NNC top 5" käyttää SECREF3-jaksossa kuvattuja luokittelutunnisteita ja "NNC SU4 F1" käyttää regressiotunnisteita eli kunkin lauseen ROUGE-SU4 F1 -pisteitä.

**Tulos**

Millaisia luokittelumenetelmiä kokeiltiin tätä tehtävää varten?

**Esimerkki 5.1435**

Asiantuntijoiden määrä vaihteli mallien välillä käyttämällä tavallisia MoE-kerroksia, joissa oli 4, 32 ja 256 asiantuntijaa, ja hierarkkisia MoE-kerroksia, joissa oli 256, 1024 ja 4096 asiantuntijaa.

**Tulos**

Miten päätetään käytettävien asiantuntijoiden oikea määrä?

**Esimerkki 5.1436**

Keräämme Food.com-sivustolta uudenlaisen tietokokonaisuuden, joka sisältää yli 230 000 reseptitekstiä ja yli 1 miljoonaa käyttäjän vuorovaikutusta (arvostelua) 18 vuoden ajalta (2000-2018).

**Tulos**

Mistä he saavat reseptit?

**Tulos**

Miten he saivat vuorovaikutussuhteet?

**Esimerkki 5.1437**

Teimme kokeen selvittääksemme, voimmeko säilyttää tai parantaa luokittelijan suorituskykyä soveltamalla seuraavaa kolmiportaista ominaisuuksien eliminointimenetelmää: Vähentäminen Vähensimme kutakin luokkaa varten koodattua tietokokonaisuutta poistamalla ominaisuudet, jotka esiintyvät koko tietokokonaisuudessa vähemmän kuin kaksi kertaa.Valinta Sovelsimme pienennettyyn tietokokonaisuuteen iteratiivisesti khiin neliö -ominaisuuden valintaa valitsemalla korkeimmalle sijoittuneiden ominaisuuksien ylimmän prosenttiosuuden 5 prosentin askelin tukivektorimallin harjoittelua ja testausta varten käyttäen lineaarista ydintä ja 5-kertaista ositettua ristiinvalidointia.Sijoitus Kuvasimme kumulatiivisesti kunkin korkeimmalle sijoittuneiden ominaisuuksien prosenttiosuuden keskimääräisen F1-pistemäärän suoritukset. Ilmoitamme prosenttiosuuden ja niiden ominaisuuksien lukumäärän, jotka johtavat korkeimman keskimääräisen F1-pistemäärän ensimmäiseen esiintymiseen kussakin luokassa.

**Tulos**

Mitkä ovat kolme vaihetta ominaisuuksien poistamiseksi?

**Esimerkki 5.1438**

Vertailukohtamme, Honk( UID9 ), DeepSpeech-finetune( UID10 ), olivat verrattain alhaisemmat sekä recall että precision.

**Tulos**

Mitkä ovat perustasot?

**Esimerkki 5.1439**

Järjestelmän tulosta, jossa kohdesanat ovat ennustetussa järjestyksessä, verrataan DURel-tietoaineiston kultaiseen sijoitukseen. Mittarina käytettiin Spearmanin $\rho $:n mittaria sen arvioimiseksi, kuinka hyvin mallin tuotos vastaa kultaista järjestystä. Mitä korkeampi Spearmanin järjestyskorrelaatio on, sitä parempi on järjestelmän suorituskyky.

**Tulos**

Miten arviointi suoritetaan?

**Esimerkki 5.1440**

Tuloksena saatu tietokokonaisuus koostuu 22 880 käyttäjästä, 41 094 blogista ja 561 003 viestistä. Taulukossa TABREF2 esitetään aineistomme muita tilastoja.

**Tulos**

Kuinka monta käyttäjää he katsovat?

**Esimerkki 5.1441**

Tästä huolimatta havaitsimme, että upotusten harjoittelu ei ollut merkittävästi hitaampaa kuin SGNS:n kaltaisten 2. järjestyksen vastaavien järjestelmien harjoittelu. Selkeästi sanottuna GPU koulutti CBOW-vektorit (käyttäen jäljempänä esitettyjä kokeellisia asetuksia) 3568 sekunnissa, kun taas CP-S:n ja JCP-S:n koulutukseen kului 6786 ja 8686 sekuntia.

**Tulos**

Mittaavatko ne faktorisointiensa laskenta-aikaa verrattuna muihin sanojen upotuksiin?

**Esimerkki 5.1442**

Vertailimme useita tiivistämismenetelmiä, jotka voidaan jakaa kolmeen ryhmään: valvomattomat, ei-neuraaliset valvotut ja neuraaliset valvotut menetelmät. Perustasona käytimme Lead-N-menetelmää, joka valitsee yhteenvedoksi INLINEFORM0-johtavat lauseet. Kaikkien menetelmien osalta poimimme tiivistelmäksi 3 lausetta, koska se on tutkimuksellisessa analyysissämme löytämiemme kultaisten tiivistelmien lauseiden mediaanimäärä. Lead-3-peruslinja suoriutuu todella hyvin ja päihittää lähes kaikki muut mallit, mikä ei ole yllättävää ja jopa yhdenmukaista muiden töiden kanssa, joiden mukaan uutisten tiivistämisessä Lead-N-peruslinja on yllättävän vaikeasti lyötävissä.

**Tulos**

Mikä oli parhaiten toimiva perustaso?

**Esimerkki 5.1443**

Lopulliset kyselytutkimukseen perustuvat luokittelutulokset osoittivat 96 prosentin tarkkuuden PTSD:n havaitsemisessa ja 1,2 prosentin keskimääräisen neliövirheen sen voimakkuuden arvioinnissa, kun otetaan huomioon, että meillä on neljä voimakkuutta: ei PTSD:tä, matalan riskin PTSD:tä, kohtalaisen riskin PTSD:tä ja korkean riskin PTSD:tä pistemäärillä 0, 1, 2 ja 3.

**Tulos**

Miten PTSD:n voimakkuus määritetään?

**Esimerkki 5.1444**

Kahdessa yksittäisen lauseen tehtävässä - syntaksiin suuntautuneessa CoLA-tehtävässä ja SST-tunnetehtävässä - suorituskyky heikkeni jonkin verran.

**Tulos**

Haittaako valvottuja tehtäviä koskeva lisäkoulutus suorituskykyä joissakin tehtävissä?

**Esimerkki 5.1445**

Nämä tietokokonaisuudet ovat osa SemEval-2016 Challenge Task 5 BIBREF27 , BIBREF28 . Taulukossa TABREF7 esitetään kunkin testikorpuksen havaintojen määrä.

**Tulos**

mitä tietokokonaisuuksia arvioinnissa käytettiin?

**Esimerkki 5.1446**

Työssämme käyttämämme Twitterin tietokokonaisuus koostuu koulutussarjasta, validointisarjasta ja testisarjasta. Se julkaistiin "First workshop on categorizing different types of online harassment languages in social media" -tapahtumaa varten. Koko tietokokonaisuus on jaettu kahteen luokkaan, jotka ovat häirintäviestit ja ei-häirintäviestit. Lisäksi häirinnän tyyppi huomioon ottaen twiitit on jaettu kolmeen alaluokkaan, jotka ovat epäsuora häirintä, seksuaalinen ja fyysinen häirintä.

**Tulos**

Millaista verkkohäirintää tutkitaan?

**Esimerkki 5.1447**

Tietokokonaisuutemme on kommentoitu Ruijinin sairaalan ruoansulatuskanavan kirurgian osaston toimittamien kiinalaisten patologisten raporttien perusteella. Se sisältää 17 833 lausetta, 826 987 merkkiä ja 2 714 kysymys-vastausparia. Neljä kliinikkoa on kommentoinut ja tarkistanut kaikki kysymys-vastausparit kolmen kysymystyypin avulla: kasvaimen koko, proksimaalinen resektiomarginaali ja distaalinen resektiomarginaali.

**Tulos**

Onko kaikki tämän aineiston teksti kysymystä, vai onko kysymysten välissä irrallisia lauseita?

**Esimerkki 5.1448**

Käytämme BIBREF73 -painotettua termien sanastoa, jossa käytetään 75 394 Facebook-käyttäjän tietokokonaisuutta, jossa käyttäjät ovat ilmoittaneet statuksensa, ikänsä ja sukupuolensa. Näiden kahden painotetun sanaston avulla ennustamme INLINEFORM0:n demografiset tiedot (ikä tai sukupuoli) (merkitään INLINEFORM1 ) seuraavan yhtälön avulla: INLINEFORM2 jossa INLINEFORM0 on termin sanakirjapaino ja INLINEFORM1 edustaa termin esiintymistiheyttä käyttäjän tuottamassa INLINEFORM2 -lomakkeessa ja INLINEFORM3 mittaa sanojen kokonaismäärää INLINEFORM4 -lomakkeessa.

**Tulos**

Miten nämä puitteet helpottavat demografisia päätelmiä sosiaalisesta mediasta?

**Esimerkki 5.1449**

Vertailemme tavallisen Transformer Base -mallin ja puoliparametrisen NMT-lähestymistapamme suorituskykyä englannin ja ranskan välisessä käännöstehtävässä.

**Tulos**

Mihin järjestelmiin he vertaavat tuloksiaan?

**Esimerkki 5.1450**

Ehdotamme myös tehokasta mallia, jolla kliiniset nimetyt entiteettitiedot voidaan integroida valmiiksi koulutettuun kielimalliin. Tässä jaksossa esitellään tehokas malli kysymyksiin vastaamiseen perustuvaa kliinisen tekstin jäsentämistä varten (QA-CTS). Kuten kuvassa FIGREF8 on esitetty, kappaleen teksti $X$ siirretään ensin kliinisen nimettyjen entiteettien tunnistuksen (CNER) malliin BIBREF12 nimettyjen entiteettien tietojen keräämiseksi ja yhden kuuman CNER-tulosteen merkintäsekvenssin saamiseksi kyselytekstille $I\_{nq}$ ja kappaleen tekstille $I\_{nt}$ BIEOS (Begin, Inside, End, Outside, Single) -merkintäjärjestelmällä. $I\_{nq}$ ja $I\_{nt}$ integroidaan sen jälkeen yhteen $I\_n$:ksi. Samaan aikaan kappaleen teksti $X$ ja kyselyteksti $Q$ järjestetään ja siirretään kontekstualisoidulle esitysmallille, joka on tässä esivalmennettu kielimalli BERT BIBREF26, jotta saadaan kontekstualisoitu esitysvektori $V\_s$ sekä tekstille että kyselylle. Tämän jälkeen $V\_s$ ja $I\_n$ integroidaan yhteen ja syötetään eteenpäin syöttävään verkkoon vastaukseen liittyvän tekstin alku- ja loppuindeksin laskemiseksi.

**Tulos**

Miten he tuovat toimialakohtaisia piirteitä valmiiksi koulutettuun kielimalliin?

**Esimerkki 5.1451**

Myöhemmin BIBREF8 esitteli RNN:n, jossa on ulkoinen pino-muisti, jolla opitaan yksinkertaisia kontekstittomia kieliä, kuten $a^n b^m$ , $a^nb^ncb^ma^m$ ja $a^{n+m} b^n c^m$ . Samankaltaisissa tutkimuksissa BIBREF15 , BIBREF16 , BIBREF17 , BIBREF10 , BIBREF11 on tutkittu, onko yksinkertaisissa RNN:issä olemassa stabiileja laskentamekanismeja, jotka mahdollistaisivat erilaisten kontekstivapaiden ja kontekstiherkkien kielten oppimisen BIBREF9 , puolestaan ehdotti LSTM-verkkojen (Long Short-Term Memory) muunnelmaa kahden kontekstittoman kielen, $a^n b^n$ , $a^n b^m B^m A^n$ , ja yhden tiukasti kontekstisidonnaisen kielen, $a^n b^n c^n$ , oppimiseen.

**Tulos**

Miten he saavat viralliset kielet?

**Esimerkki 5.1452**

Plackett-Luce Model for SMT RerankingKaksoistamisen jälkeen N-parhaan listan keskimääräinen koko on noin 300 ja siinä on 7491 ominaisuutta. Tämä kokeilu osoittaa, että suurten piirteiden osalta Plackett-Luce-malli korreloi hyvin BLEU-pisteiden kanssa ja lieventää jossain määrin ylisovittamista.

**Tulos**

Mitä kokeita tehdään laajamittaisilla ominaisuuksilla?

**Esimerkki 5.1453**

Arvioimme LinkNBediä ja perusratkaisuja kahdella todellisella tietograafilla: D-IMDB (johdettu suuren mittakaavan IMDB-tietotilannekuvasta) ja D-FB (johdettu suuren mittakaavan Freebase-tietotilannekuvasta).

**Tulos**

Millä tiedoilla mallia arvioidaan?

**Esimerkki 5.1454**

Englannin kielen WSD:tä varten koulutamme WSD-mallimme SemCor BIBREF24:llä ja testaamme sitä Senseval-2:lla (SE2), Senseval-3:lla (SE3), SemEval 2013 task 12:lla (SE13) ja SemEval 2015 task 13:lla (SE15). Käytämme OntoNotes Release 5.0 -versiota, joka sisältää useita annotaatioita, mukaan lukien kiinan kielen sanojen aistimukset.

**Tulos**

Mitä tietokokonaisuuksia käytetään testauksessa?

**Esimerkki 5.1455**

Menetelmät ::: Kontekstitieto on informatiivista siinä mielessä, että yleensä samankaltaisilla sanoilla on taipumus esiintyä samoissa yhteyksissä. Korpuspohjaisessa lähestymistavassa otetaan huomioon nämä kaksi ominaisuutta ja luodaan alalle ominaisia sanojen upotuksia. Menetelmä ::: Sanakirjapohjainen lähestymistapaTurkin kielessä ei ole vakiintuneita tunnesanastoja kuten englannissa. Tässä lähestymistavassa käytimme TDK-sanakirjan (Türk Dil Kurumu - "Turkish Language Institution") sanapolariteetteja.

**Tulos**

Mitä sanapohjaisia ja sanakirjapohjaisia ominaisuuksia käytetään?

**Esimerkki 5.1456**

BIBREF8-julkaisussa esitellään tarkennettu kokoelma Twitteristä kerättyjä twiittejä. Heidän tietokokonaisuutensa, joka on merkitty nimettyjen entiteettien tunnistustehtävää varten, sisältää 8 257 twiittiä. Tässä tietokokonaisuudessa on yhteensä 12 784 entiteettiä. Taulukossa TABREF19 esitetään kuhunkin nimettyyn entiteettiin liittyviä tilastoja koulutus-, kehitys- ja testijoukoissa.

**Tulos**

Mitä tietokokonaisuuksia he käyttivät?

**Tulos**

Mitä sosiaalisen median alustaa tutkitaan?

**Esimerkki 5.1457**

Vertaamme mallimme tuloksia yang2016multi-arkkitehtuurin innoittamaan perustason monitehtäväarkkitehtuuriin. Perusmallissamme ei ole eksplisiittisiä yhteyksiä tehtävien välillä - ainoat jaetut parametrit ovat piilokerroksessa.

**Tulos**

Mikä on lähtötaso?

**Esimerkki 5.1458**

Tämän mahdollistamiseksi käytämme bifokaalista huomiomekanismia, joka laskee huomion makrotason kenttiin ja mikrotason arvoihin. Tämän jälkeen yhdistämme nämä huomiopainot siten, että kentän huomiopaino vaikuttaa myös sen sisällä olevien arvojen huomioon. Fuusioitu bifokaalinen tarkkaavaisuusmekanismiIntuitiivisesti, kun ihminen kirjoittaa kuvauksen taulukosta, hän pitää kirjaa tiedosta kahdella tasolla. Makrotasolla on tärkeää päättää, mikä on sopiva kenttä, johon kiinnittää seuraavaksi huomiota, ja mikrotasolla (eli kentän sisällä) on tärkeää tietää, mihin arvoihin kiinnittää seuraavaksi huomiota. Tämän käyttäytymisen kuvaamiseksi käytämme bifokaalista huomiomekanismia, joka on kuvattu jäljempänä. Yhdistetty huomio: Intuitiivisesti kentälle annettujen huomiopainojen pitäisi vaikuttaa kaikkiin kyseiseen kenttään kuuluviin arvoihin. Tämän varmistamiseksi punnitsemme mikrotason huomiopainot uudelleen vastaavien makrotason huomiopainojen perusteella. Toisin sanoen yhdistämme kahden tason huomiopainot seuraavasti: DISPLAYFORM0jossa INLINEFORM0 on kenttä, joka vastaa INLINEFORM1 -osaa, INLINEFORM2 on makrotason kontekstivektori.

**Tulos**

Mikä on bifokaalinen huomiomekanismi?

**Esimerkki 5.1459**

Koulutamme molemmat järjestelmät WMT15:n saksan ja englannin välisellä harjoitusaineistolla, ks. taulukko TABREF18, jossa on joitakin tilastoja. Tätä tarkoitusta varten käytämme RWTH:n saksan ja englannin datasetin manuaalisia linjauksia kovina linjauksina.

**Tulos**

Mitä tietokokonaisuuksia käytetään?

**Esimerkki 5.1460**

Arvioimme lausekekompositionaalisuusmalleja Mitchell2010:n laatimilla adjektiivi-substantiivi ja substantiivi-substantiivi -lauseiden samankaltaisuustehtävillä käyttäen samaa arviointijärjestelmää kuin alkuperäisessä työssä. Arviointimittarina käytettiin Spearmanin INLINEFORM0 -lukua koostumustehtävistämme johdettujen fraasien samankaltaisuuksien ja ihmisannotoijien välillä (laskettuna erikseen annotoijittain ja sitten keskiarvona kaikkien annotoijien kesken).

**Tulos**

Miten he pisteyttävät fraseaalisen kompositionaalisuuden?

**Esimerkki 5.1461**

Tarkemmin sanottuna tutkimme kolmea regularisointitermiä ongelman ratkaisemiseksi: (1) neutraaleihin piirteisiin liittyvä regularisointitermi, (2) luokkajakauman maksimieentropian regularisointitermi ja (3) KL-divergenssi referenssi- ja ennustetun luokkajakauman välillä.

**Tulos**

Mitkä ovat kolme regularisointitermiä?

**Esimerkki 5.1462**

Käytämme kahta perusominaisuutta: POS-tunnisteet (Parts of Speech): Käytämme NLTK:n POS-taggeria twiittitekstien merkitsemiseen BIBREF0 . Käytämme POS-ominaisuuksina substantiivi-, adjektiivi-, adverbi- ja verbisanojen lukumäärää twiitissä: Käytämme polariteettisanakirjaa BIBREF3 saadaksemme sanojen aikaisemman polariteetin. Sanakirja sisältää positiiviset, negatiiviset ja neutraalit sanat sekä niiden polariteetin voimakkuuden (heikko tai vahva). Sanan polariteetti riippuu sen POS-tunnisteesta. Esimerkiksi sana `excuse' on negatiivinen, kun sitä käytetään `substantiivina' tai `adjektiivina', mutta sillä on positiivinen merkitys, kun sitä käytetään `verbinä'. Käytämme NLTK postaggerin tuottamia tunnisteita, kun valitsemme sanan etukäteispolariteetin sanakirjasta. Käytämme myös stemming-toimintoa (NLTK:n Porter Stemmer -toteutusta), kun teemme sanakirjahakua, jotta voimme lisätä osumien määrää. Käytämme ominaisuuksina heikosti positiivisten sanojen, heikosti negatiivisten sanojen, vahvasti positiivisten sanojen ja vahvasti negatiivisten sanojen lukumäärää twiitissä.Olemme myös tutkineet joitakin kehittyneitä ominaisuuksia, jotka auttavat parantamaan twiittien tunnetilan havaitsemista. hymiöt: Käytämme BIBREF:n2 emoticon-sanakirjaa ja laskemme positiiviset ja negatiiviset emoticonit jokaisesta twiitistä.[1] Urlin sentimentti: Koska lähes kaikki artikkelit on kirjoitettu hyvin muotoillulla englannilla, analysoimme artikkelin ensimmäisen kappaleen sentimenttiä Standfordin sentimenttianalyysityökalulla BIBREF4 . Se ennustaa sentimentin jokaiselle artikkelin sisältämälle lauseelle. Laskemme negatiivisten, positiivisten ja neutraalien lauseiden osuuden ja käytämme näitä kolmea arvoa ominaisuuksina. hashtag: Laskemme hashtagien määrän kussakin twiitissä.Suuraakkoset: Oletamme, että twiittien isolla alkukirjaimella on jonkinlainen yhteys sentimenttiasteeseen. Laskemme twiiteissä olevien isolla alkukirjaimella kirjoitettujen sanojen määrän.Retweet: Tämä on boolean-ominaisuus, joka ilmaisee, onko twiitti uudelleentwiittaus vai ei.User Mention: Boolean-ominaisuus, joka osoittaa, sisältääkö twiitti käyttäjän maininnan.Negation: Sanoja, kuten "ei", "ei", "ei", "ei" kutsutaan negaatiosanoiksi, koska ne kumoavat sitä seuraavan sanan merkityksen. Esimerkiksi `hyvä' muuttuu `ei hyvä'. Tunnistamme kaikki negaatiosanat twiiteistä. Jos negaatiosanaa seuraa polariteettisana, kumoamme kyseisen sanan polariteetin. Jos esimerkiksi sanaa "hyvä" edeltää sana "ei", muutamme polariteetin sanasta "heikko positiivinen" sanaksi "heikko negatiivinen": Käytämme tf-idf-pohjaisia tekstiominaisuuksia ennustamaan twiitin sentimenttiä. Teemme tf-idf-pohjaisen pisteytyksen twiitissä oleville sanoille ja twiitissä esiintyville hashtageille. Käytämme tf-idf-vektoreita luokittelijan kouluttamiseen ja sentimentin ennustamiseen. Tätä käytetään sitten pinottuna ennusteominaisuutena lopullisessa luokittimessa.

**Tulos**

Mitä kielellisiä piirteitä käytetään?

**Esimerkki 5.1463**

Yksinkertainen valinta ei-harvinaiselle porttifunktiolle BIBREF17 on kertoa tulo koulutettavalla painomatriisilla INLINEFORM0 ja soveltaa sitten INLINEFORM1-funktiota. DISPLAYFORM0Lisäämme Softmax-verkkoon kaksi komponenttia: harvuus ja kohina. Ennen Softmax-funktion ottamista lisäämme viritettävää Gaussin kohinaa, sitten säilytämme vain ylimmät k arvot ja asetamme loput INLINEFORM0:een (mikä saa vastaavat porttiarvot olemaan 0). Harvinaisuuden tarkoituksena on säästää laskentaa, kuten edellä on kuvattu. Vaikka tämä harvinaisuuden muoto luo teoreettisesti pelottavia epäjatkuvuuskohtia porttifunktion ulostuloon, emme ole vielä havainneet tämän olevan ongelma käytännössä. Kohinatermi auttaa kuorman tasapainottamisessa, kuten lisäyksessä SECREF51 käsitellään. Kohinan määrää komponenttia kohden ohjataan toisella koulutettavalla painomatriisilla INLINEFORM1 . DISPLAYFORM0 DISPLAYFORM1 Yksinkertainen valinta ei-harvaksi porttifunktioksi BIBREF17 on kertoa tulo koulutettavalla painomatriisilla INLINEFORM0 ja soveltaa sitten INLINEFORM1-funktiota. DISPLAYFORM0Lisäämme Softmax-verkkoon kaksi komponenttia: harva ja kohina. Ennen Softmax-funktion ottamista lisäämme viritettävää Gaussin kohinaa, sitten säilytämme vain ylimmät k arvot ja asetamme loput INLINEFORM0:een (mikä saa vastaavat porttiarvot olemaan 0). Harvinaisuuden tarkoituksena on säästää laskentaa, kuten edellä on kuvattu. Vaikka tämä harvinaisuuden muoto luo teoreettisesti pelottavia epäjatkuvuuskohtia porttifunktion ulostuloon, emme ole vielä havainneet tämän olevan ongelma käytännössä. Kohinatermi auttaa kuorman tasapainottamisessa, kuten lisäyksessä SECREF51 käsitellään. Kohinan määrää komponenttia kohden ohjataan toisella koulutettavalla painomatriisilla INLINEFORM1 . DISPLAYFORM0 DISPLAYFORM1

**Tulos**

Mitä yhtälöitä käytetään koulutettavassa porttiverkossa?

**Esimerkki 5.1464**

Vertaamme personoituja mallejamme kahteen perusmalliin. Ensimmäinen on nimipohjainen lähin naapuri -malli (Nearest-Neighbor, NN). Sovelsimme aluksi BIBREF0:n neuraalista tarkistuslistamallia perustasoksi; käytämme kuitenkin lopulta yksinkertaista Encoder-Decoder-perustasoa, jossa on ainesosahuomio (Enc-Dec), joka tarjoaa vertailukelpoisen suorituskyvyn ja pienemmän monimutkaisuuden.

**Tulos**

Mitkä ovat perusmallit?

**Esimerkki 5.1465**

Ensinnäkin ehdotamme NLP-tehtäviä varten luokkaa rekursiivisia neuroverkkoja, jotka täyttävät differentiaaliyhtälön DISPLAYFORM0 jossa DISPLAYFORM0ja jossa INLINEFORM0 ja INLINEFORM1 ovat opittuja funktioita. INLINEFORM2 vastaa perinteisiä RNN-verkkoja, joiden INLINEFORM3 . INLINEFORM4:n osalta tämä tapahtuu RNN-solujen muodossa, joilla on joko sisäkkäisiä sisäisiä muistoja tai riippuvuuksia, jotka ulottuvat ajallisesti välittömästi edellisen piilotetun tilan ulkopuolelle.

**Tulos**

Mitä uutta luokkaa rekursiivisten verkkojen kaltaisia verkkoja ehdotetaan?

**Esimerkki 5.1466**

Käytämme WMT 2017 English-Chinese test set BIBREF18 -testisarjan kiinalaisten lähdetekstien englanninkielisiä käännöksiä kaikissa tässä artikkelissa esitetyissä kokeissa:

**Tulos**

Mitä kieliä ne tutkivat konekääntämistä varten?

**Esimerkki 5.1467**

Tässä luvussa kuvaamme useita kokeita, joiden tarkoituksena on vertailla suosittujen nimettyjen entiteettien tunnistusalgoritmien suorituskykyä aineistossamme. Koulutimme ja arvioimme Stanfordin NER:ää, spaCy 2.0:aa ja rekursiivista mallia, joka on samanlainen kuin BIBREF13 , BIBREF14 , joka käyttää kaksisuuntaisia LSTM-soluja merkkipohjaiseen ominaisuuksien poimintaan ja CRF:ää, joka on kuvattu Guillaume Genthialin blogikirjoituksessa Sequence Tagging with Tensorflow BIBREF15 . Stanford NER on CRF-luokittelija (Conditional Random Fields), joka perustuu leksikaalisiin ja kontekstuaalisiin piirteisiin, kuten nykyiseen sanaan, merkkitason n-grammeihin, joiden pituus on enintään 6 sen alussa ja lopussa, edellisiin ja seuraaviin sanoihin, sanan muotoon ja sekvenssipiirteisiin BIBREF16 . spaCy 2.0 käyttää CNN-pohjaista siirtymävaihetta nimettyjen entiteettien tunnistamiseen. Pääasiallinen malli, johon keskityimme, oli rekurrenssimalli, jossa on CRF-yläkerros, ja edellä mainitut menetelmät toimivat enimmäkseen vertailukohtina.

**Tulos**

mitä ner-malleja arvioitiin?

**Esimerkki 5.1468**

Tavoitteenamme on verrata ehdotetun lähestymistavan tehokkuutta BIBREF-julkaisussa13 kuvattuun perusjärjestelmään. Arvioimme väärän hylkäämisen (FR) ja väärän hyväksymisen (FA) kompromissia useissa eri kokoisissa ja laskennallisesti monimutkaisissa päästä päähän -malleissa. Kuten kuvassa FIGREF14 esitetyistä Receiver Operating Characteristic (ROC) -käyristä voidaan nähdä, kaksi suurinta päästä päähän -mallia, joissa on kaksivaiheinen koulutus, ovat huomattavasti parempia kuin paljon suuremman ja monimutkaisemman Baseline\_1850K-järjestelmän tunnistuksen laatu. Tarkemmin sanottuna E2E\_318K\_2stage ja E2E\_700K\_2stage osoittavat jopa 60 prosentin suhteellista FR-asteen alenemista Baseline\_1850K:hon verrattuna useimmissa testiolosuhteissa.

**Tulos**

Miten he mittaavat havaitsemisen laatua?

**Esimerkki 5.1469**

Käyttämällä luokan yksittäisten sanojen jakaumaa voimme laatia koko luokan jakaumat ja siten luoda karttoja näille sanaluokille.

**Tulos**

Rakentavatko he mallin, jolla voidaan automaattisesti havaita ihmisten demografiset, lingvistiset tai psykologiset ulottuvuudet?

**Esimerkki 5.1470**

Vertailun vuoksi raportoimme myös Rei2016:n virheentunnistusjärjestelmän suorituskyvyn, joka on koulutettu käyttäen samaa FCE-tietokokonaisuutta.

**Tulos**

Mitä perustasoa käytettiin?

**Esimerkki 5.1471**

Ongelma-avaruuden tutkimiseksi keskitymme erityisesti viiteen binääriseen luokitusongelmaan, joita on käsitelty BIBREF-tietokannoissa17 , BIBREF-tietokannassa18 ja jotka ovat konsonanttien läsnäolo/poissaolo, foneemiset nasaalit, bilabiaalit, korkeat etuvokaalit ja korkeat takavokaalit.

**Tulos**

Mitkä ovat viisi erilaista binääriluokittelutehtävää?

**Esimerkki 5.1472**

Tällä tavoin huomiovektorisekvenssi segmentoidaan useisiin osajaksoihin, ja kukin osajakso edustaa yhden sanan huomiota. Tämän jälkeen kehitämme sopivan aggregointimoduulin, joka yhdistää sanan sisäisen merkkihuomion, koska psykolingvistinen havainto, jonka mukaan lukijat kiinnittävät todennäköisesti likimääräistä huomiota kuhunkin merkkiin yhdessä kiinalaisessa sanassa, motivoi meitä.

**Tulos**

Miten fuusiomenetelmä toimii?

**Esimerkki 5.1473**

Arvioidaksemme ehdotetun kehyksen tehokkuutta ATE-tehtävässä suoritamme kokeita neljällä SemEval ABSA -haasteeseen kuuluvalla vertailutietoaineistolla BIBREF1 , BIBREF18 , BIBREF12 . Taulukossa TABREF24 esitetään niiden tilastot. INLINEFORM0 (SemEval 2014) sisältää arvosteluja kannettavien tietokoneiden alalta ja INLINEFORM1 (SemEval 2014), INLINEFORM2 (SemEval 2015) ja INLINEFORM3 (SemEval 2016) ravintola-alalta. Näissä tietokokonaisuuksissa aspektitermit on leimannut tehtävän järjestäjä.

**Tulos**

Mitä tietokokonaisuutta (tietokokonaisuuksia) he käyttävät mallin kouluttamiseen?

**Esimerkki 5.1474**

Erityisesti ei ole selvää, onko yhteistyöelimen avulla voitu ratkaista sellaisia tehtäviä kuin Pronoun Disambiguation Problem (PDP) ja Winograd Schema Challenge (WSC). Pronominien erittelyä koskevassa tietokokonaisuudessa PDP-60 menetelmämme saavuttaa 68,3 prosentin tarkkuuden, mikä on parempi kuin uusimman tekniikan tason 66,7 prosentin tarkkuus. WSC-tietokannassa, WSC-273, menetelmämme saavuttaa 60,3 %:n tuloksen.

**Tulos**

Millaisia tietokokonaisuuksia ne arvioivat?

**Esimerkki 5.1475**

Ensemble, joka valitsee lauseen käyttäen 100 agenttia, jotka on koulutettu klusteroitujen dialogien perusteella kohdassa SECREF4 kuvatulla tavalla - keskittyvä agentti valitaan käyttämällä regressoria dialogipalkkion ennustajana INLINEFORM0 käyttäen samanlaista neuroverkkoa kuin ChatDQN-agentit, paitsi että viimeisessä kerroksessa on yksi solmu ja piilokerrosten välissä käytetään BIBREF44 -eränormalisointia piilokerrosten välillä, kuten BIBREF36:ssa;

**Tulos**

Kuinka monta agenttia he kokoavat yhteen?

**Esimerkki 5.1476**

Käytämme monialueen polariteettiluokituskokeissa monialueen aistimustietoaineiston (Multi-Domain Sentiment Dataset BIBREF0) toista versiota. Aineisto sisältää Amazonin tuotearvosteluja neljältä eri toimialueelta: Kirjat (B), DVD:t (D), elektroniikka (E) ja keittiökoneet (K).

**Tulos**

Mitä osa-alueita polariteettiluokittelua koskeva tietokokonaisuus sisältää?

**Esimerkki 5.1477**

Koska tavoitteenamme on vertailla merkkipohjaisia ja sanapohjaisia lähestymistapoja, olemme myös toteuttaneet yksinkertaisen sanatason koodaajan twiittejä varten. Syötetty twiitti jaetaan ensin merkkeihin valkoisia välejä pitkin. Voidaan käyttää kehittyneempää tokenizeria, mutta tasapuolisen vertailun vuoksi halusimme pitää kielikohtaisen esikäsittelyn mahdollisimman vähäisenä. Kooderi on periaatteessa sama kuin tweet2vec, mutta syötteenä on merkkien sijasta sanoja. Hakutaulukkoon tallennetaan sanavektorit $V$ (tässä 20K) yleisimmille sanoille, ja loput ryhmitellään yhteen `UNK'-merkin alle.

**Tulos**

Mikä on sanatason lähtötaso, johon niitä verrataan?

**Tulos**

Mikä on sanatason perustaso?

**Esimerkki 5.1478**

Työskentelemme tietokokonaisuuden kanssa, joka koostuu 3 206 uutisartikkelista, joista jokainen on merkitty joko oikeaksi tai väärennetyksi, ja 1 603 oikeaa ja väärennettyä artikkelia on jaettu täydellisesti 50/50. Väärennetyt artikkelit hankittiin verkkosivuilta, jotka voittoa tavoittelematon riippumaton tiedotusvälineiden faktantarkastusjärjestö Verafiles ja Filippiinien kansallinen journalistiliitto NUJP (National Union of Journalists in the Philippines) olivat merkinneet väärennetyiksi uutissivustoiksi. Oikeat artikkelit saatiin Filippiinien valtavirta-uutissivustoilta, kuten Pilipino Star Ngayon, Abante ja Bandera.

**Tulos**

Mikä on tietokokonaisuuden lähde?

**Esimerkki 5.1479**

MADAMIRA BIBREF6 -järjestelmää on äskettäin arvioitu sokealla testisarjalla (25 000 sanaa modernia arabian standardia (MSA) varten, jotka on valittu Penn Arabic Tree Bankista (PATB)), ja raportoitu tarkkuus oli 96,2 % niiden sanojen prosenttiosuutena, joissa valittu analyysi (SAMA-morfologisen analysaattorin BIBREF7 tarjoama) sisältää oikean lemman. Koska MSA kirjoitetaan yleensä ilman diakriittisiä merkkejä ja IR-järjestelmät poistavat yleensä kaikki diakriittiset merkit hakukyselyistä ja indeksoidusta datasta esikäsittelyn perusvaiheena, lisätään toinen sarake diakriittisille lemoille, ja sitä käytetään lemmatisaattorimme arvioimiseen ja vertailuun uusimman lemmatisointijärjestelmän MADAMIRA:n kanssa.

**Tulos**

Mikä on tekniikan nykytila?

**Esimerkki 5.1480**

Niinpä tässä työssä hyödynnämme hybridi-koodiverkkoja (HCN) BIBREF4 , koska HCN:n avulla saavutetaan huippuluokan suorituskyky datatehokkaalla tavalla tehtäväkeskeisissä vuoropuheluissa, ja ehdotamme AE-HCN:iä, jotka laajentavat HCN:ää automaattisella koodaajalla (kuva FIGREF8 ). AE-HCN(-CNN) päihittää HCN:n Test-OOD-testissä suurella marginaalilla, keskimäärin noin 17(20) pisteellä, säilyttäen samalla suorituskyvyn mahdollisimman pienen kompromissin Testiin verrattuna.

**Tulos**

Kuinka paljon heidän menetelmänsä on parempi kuin uusin OOD-tunnistusmenetelmä?

**Esimerkki 5.1481**

Tutkimme manuaalisesti virheitä, joita keinotekoiset neuroverkot tekevät kohdekielen morfologisen taivutuksen osalta sen jälkeen, kun ne on esiharjoitettu eri lähdekielillä. Laadullisessa analyysissämme käytämme validointijoukkoa. Siksi näytämme validointisarjan tarkkuudet taulukossa TABREF19 vertailun vuoksi. Annotoimme manuaalisesti kunkin lähde-kohdekieliyhdistelmän 75 ensimmäisen kehitysesimerkin tuotokset. Kaikki löydetyt virheet luokitellaan johonkin seuraavista luokista.

**Tulos**

Miten tehtävässä suoriutumista arvioidaan?

**Esimerkki 5.1482**

Wav2vec BIBREF22 ehdotti monikerroksista konvoluutiohermoverkkoa, joka optimoitiin melukontrastisen binääriluokittelun avulla, ja sitä sovellettiin WSJ ASR-tehtäviin. Kokeemme koostuivat kolmesta eri asetelmasta: 1) täysin valvottu järjestelmä, jossa käytettiin kaikkea merkittyä dataa; 2) SSL-järjestelmä, jossa käytettiin wav2vec-ominaisuuksia; 3) SSL-järjestelmä, jossa käytettiin ehdottamiamme DeCoAR-ominaisuuksia. Kaikki käytetyt mallit perustuivat syviin BLSTM-malleihin, joissa käytettiin CTC-häviökriteeriä.

**Tulos**

Mitkä ovat WSJ eval92:n ja LibriSpeech test-cleanin perusmallit?

**Esimerkki 5.1483**

Vertaamme malliamme seuraaviin uusimpiin yhteisiin entiteettien ja relaatioiden louhintamalleihin:(1) SPTree BIBREF4: Tämä on kokonaisvaltainen neuraalinen entiteettien ja relaatioiden louhintamalli, jossa käytetään sekvenssi-LSTM:ää ja puu-LSTM:ää. (2) Tagging BIBREF5: Tämä on neuraalinen sekvenssimerkintämalli, joka poimii entiteetit ja relaatiot yhdessä käyttäen LSTM-kooderia ja LSTM-dekooderia. (3) CopyR BIBREF6: Tässä mallissa käytetään enkooderi-dekooderi-lähestymistapaa entiteettien ja relaatioiden yhteiseen uuttamiseen. (4) HRL BIBREF11: Tässä mallissa käytetään vahvistusoppimisalgoritmia (RL), jossa on kaksi hierarkiatasoa tuplien louhintaan. (5) GraphR BIBREF14: Tämä malli pitää jokaista lauseen merkkiä graafin solmuna ja solmuja yhdistäviä reunoja niiden välisinä suhteina. (6) N-gram Attention BIBREF9: Tässä mallissa käytetään koodaaja-dekooderi-lähestymistapaa, jossa on N-gram attention -mekanismi tietämyspohjan täydentämiseen käyttäen etäisesti valvottuja tietoja.

**Tulos**

Mihin aiempaan työhön kirjoittajat viittaavat?

**Esimerkki 5.1484**

Tutkimme erityisesti huomion kiinnittämistä substantiiveihin, verbeihin, pronomineihin, subjekteihin, objekteihin ja negaatiosanoihin sekä erityisiin BERT-merkkeihin eri tehtävissä.

**Tulos**

Mitä käsityöläisominaisuuksia käytetään?

**Esimerkki 5.1485**

Vaikka valinnainen lemmatointivaihe käytettäisiinkin, voidaan silti pyrkiä vähentämään graafin monimutkaisuutta edelleen yhdistämällä samankaltaisia kärkipisteitä. Tätä vaihetta kutsutaan metapisteiden rakentamiseksi. Metahuipun rakentamisvaihe toimii seuraavasti. Olkoon INLINEFORM0 edellä määritetty kärkipisteiden joukko. Metapiste INLINEFORM1 koostuu joukosta huippuja, jotka ovat INLINEFORM2:n eli INLINEFORM3:n elementtejä. Olkoon INLINEFORM4 INLINEFORM5 -niminen metapiste. Rakennetaan tietty INLINEFORM6 siten, että jokaisen INLINEFORM7 :n osalta INLINEFORM8 :n alkuperäiset särmät (ennen sen yhdistämistä metapisteeksi) kytketään uudelleen juuri lisättyyn INLINEFORM9 :ään. Huomaa, että tällaiset särmät yhdistyvät huippuihin, jotka eivät ole osa INLINEFORM10 -järjestelmää. Näin ollen sekä kärkipisteiden että reunojen määrä vähenee huomattavasti. Tämä ominaisuus toteutetaan seuraavalla menettelyllä: Meta-vertex-ehdokkaiden tunnistaminen. Muokkausetäisyyttä ja sanojen pituusetäisyyttä käytetään sen määrittämiseksi, pitäisikö kaksi sanaa yhdistää metapisteeksi (vain jos pituusetäisyyden kynnysarvo täyttyy, lasketaan kalliimpi muokkausetäisyys). Metatekstin luominen. Yhteisinä tunnisteina käytetään alkuperäisten verteksien kantakielisiä versioita, ja jos kantakielisiä versioita on useampi kuin yksi, valitaan tunnistetuista ehdokkaista se verteksi, jolla on korkein centraliteettiarvo graafissa, ja sen kantakielinen versio otetaan käyttöön uutena verteksinä (meta verteksinä).

**Tulos**

Miten metapisteet lasketaan?

**Esimerkki 5.1486**

Toinen tietoerä koostuu tapahtumiin liittyvistä twiiteistä, jotka koskevat viittä Yhdysvalloissa vuonna 2018 tapahtunutta luonnonkatastrofia. Nämä ovat: itärannikon pommisykloni (2.-6.1.), Mendocinon, Kalifornian maastopalot (27.7.-18.9.), Florence-hurrikaani (31.8.-19.9.), Michael-hurrikaani (7.10.-16.10.) ja Kalifornian leiripalot (8.11.-25.11.).

**Tulos**

Mitä viittä luonnonkatastrofia tutkittiin?

**Esimerkki 5.1487**

Vertaileva arvioija koulutetaan maksimiluotettavuusarvioinnin (MLE) tavoitteella, kuten on kuvattu eq DISPLAY\_FORM6:ssa, jossa $\mathcal {X}$ on edellä kuvatulla tavalla muodostettu pareittaisten harjoitusesimerkkien joukko, $Q(x\_1, x\_2) \ in \lbrace >,<,\approx \rbrace $ on parin ($x\_1$, $x\_2$) todellinen etiketti, $D\_\phi ^q(x\_1, x\_2)$ on todennäköisyys sille, että vertailevan erottelijan ennuste on $q$ ($q \in \lbrace >,<,\approx \rbrace $) parille ($x\_1$, $x\_2$).

**Tulos**

Miten he lisäävät ihmisen ennakko-odotuksia koskevan merkinnän hienosäätöprosessiin?

**Esimerkki 5.1488**

Maksimientropiateoriaa sovelletaan vietnaminkielisten sanojen segmentoinnin ratkaisemiseen BIBREF15 , BIBREF22 , BIBREF23 . Vietnamin kielen sanojen segmentointitehtävästä on tehty useita tutkimuksia viime vuosikymmenen aikana. Dinh et al. aloittivat tämän tehtävän WFST-lähestymistavan (Weighted Finite State Transducer) ja neuroverkkomenetelmän avulla BIBREF9 . Lisäksi koneoppimisen lähestymistapoja on tutkittu ja sovellettu laajalti luonnollisen kielen käsittelyyn ja sanojen segmentointiin. Itse asiassa useissa tutkimuksissa on käytetty tukivektorikoneita (SVM) ja ehdollisia satunnaiskenttiä (CRF) sanojen segmentointitehtävään BIBREF7 , BIBREF8 .

**Tulos**

Mitä lähestymistapoja on sovellettu sanojen segmentoinnin ratkaisemiseen vietnamiksi?

**Esimerkki 5.1489**

Tulevaisuudessa pyrimme käyttämään korpusta annotaatiohankkeissa, kuten puheen osien merkitsemisessä ja nimettyjen entiteettien tunnistamisessa. Lisäksi tuotettuja sanasulkeumia hyödynnetään Sindhi WordNetin automaattisessa rakentamisessa.

**Tulos**

Käytetäänkö koulutettuja sanasulkeumia muihin NLP-tehtäviin?

**Esimerkki 5.1490**

Valitsimme vertailuarvoja varten viisi järjestelmää. Valitsimme ensin langid.py-kirjaston, jota käytetään kirjallisuudessa usein järjestelmien vertailuun. Koska työmme koskee neuroverkko-LID-järjestelmää, valitsimme kirjallisuudesta kaksi neuroverkkojärjestelmää, nimittäin BIBREF6:n koodaaja-dekooderi EquiLID-järjestelmän ja BIBREF10:n GRU-neuroverkko LanideNN-järjestelmän. Lopuksi otimme mukaan CLD2:n ja CLD3:n, kaksi toteutusta Naïve Bayes LID -ohjelmistosta, jota Google käyttää Chrome-selaimessaan BIBREF4, BIBREF0, BIBREF8 ja jota käytetään joskus vertailukohtana LID-kirjallisuudessa BIBREF7, BIBREF6, BIBREF8, BIBREF2, BIBREF10.

**Tulos**

Mitä nykyisiä kielten tunnistusjärjestelmiä testataan?

**Esimerkki 5.1491**

Koska konservatiivien ja liberaalien on raportoitu käyttäytyvän eri tavoin sosiaalisilla verkkoalustoilla BIBREF19BIBREF20BIBREF21 , annoimme lisäksi poliittisen puolueellisuusmerkinnän eri yhdysvaltalaisille tiedotusvälineille (ja näin ollen myös uutisartikkeleille) BIBREF2:ssa kuvatun menettelyn mukaisesti. Menetelmämme kestävyyden arvioimiseksi teimme luokittelukokeita harjoittelemalla vain vasemmistolaisuuteen (tai oikeistolaisuuteen) suuntautuneita sekä disinformaatio- että valtavirta-alan tiedotusvälineitä ja testaamalla koko lähdejoukkoa sekä jättämällä pois tietyt lähteet, jotka olivat näytteiden määrässä muita suurempia, jotta vältyttäisiin liialliselta sovittamiselta.

**Tulos**

Miten eri lähteiden poliittinen puolueellisuus otetaan huomioon mallissa?

**Esimerkki 5.1492**

Keräämme HLA-dataa TV Tropes BIBREF3:sta, joka on pop-kulttuurille omistettu tietopohjainen verkkosivusto, joka sisältää tietoa lukuisista hahmoista eri lähteistä. Wikipedian tapaan sen sisältöä tarjoaa ja muokkaa yhteistyössä valtava käyttäjäkunta. Ominaisuudet määräytyvät ihmiskatsojien ja heidän hahmoista saamiensa vaikutelmien perusteella, ja ne korreloivat ihmisen kaltaisten ominaisuuksien kanssa. Uskomme, että TV Tropes soveltuu tarkoitukseemme fiktiivisten hahmojen mallintamiseen paremmin kuin tietolähteet, joita käytetään esimerkiksi BIBREF25 shuster2019engaging -teoksissa, koska TV Tropesin sisällöntuottajat palkitaan sisällön oikeanlaisesta tarjoamisesta yhteisön tunnustuksen kautta.

**Tulos**

Miten dataset mallintaa hahmojen profiileja?

**Esimerkki 5.1493**

Suoritimme inhimillisen arvioinnin Amazon Mechanical Turk -palvelun avulla subjektiivisen laadun arvioimiseksi. Rekrytoimme master-tason työntekijöitä (joilla on hyvät ennakkohyväksyntäprosentit) suorittamaan inhimillisen vertailun kahden järjestelmän tuottamien vastausten välillä (jotka on poimittu satunnaisesti vertailujärjestelmistä). Työntekijöiden on arvioitava jokainen lausahdus asteikolla 1 (huono) - 3 (hyvä) informatiivisuuden ja luonnollisuuden suhteen. Informatiivisuus ilmaisee, missä määrin tuotettu lausuma sisältää kaikki vuoropuhelussa määritellyt tiedot. Luonnollisuus kertoo, onko lausuma yhtä luonnollinen kuin ihmisen lausuma. Arviointiharhojen vähentämiseksi jaamme kunkin kysymyksen kolmelle eri työntekijälle. Lopulta keräsimme yhteensä 5800 tuomaria.

**Tulos**

Mitkä olivat ihmisten arviointiperusteet?

**Esimerkki 5.1494**

Lopuksi, kun tutkitaan eurooppalaisten ja ei-eurooppalaisten kielten välistä sukulaisuutta (vrt. (En/Fr)$\rightarrow $Ar), saadaan samankaltaisia tuloksia kuin yksikielisessä kokeessa (makro F-tulos 62,4 vs. 68,0), ja parhaat tulokset saavutetaan Ar $\rightarrow $(En/Fr) -kielellä. Tämä osoittaa, että molemmilla puolilla on yhteisiä pragmaattisia välineitä ja samalla tavalla ironisten twiittien kerrontatavassa samankaltaisia tekstipohjaisia kuvioita.

**Tulos**

Tunnistavatko kirjoittajat kulttuurieroja ironian käytössä?

**Esimerkki 5.1495**

Vastakohtaisten muutosten jälkeen alkuperäisten kohdemallien (ilman "-adv"-loppusanaa) suorituskyky putoaa dramaattisesti (esimerkiksi BERT:n kokonaistarkkuus Quorassa putoaa 94,6 prosentista 24,1 prosenttiin), mikä paljastaa, että kohdemallit ovat haavoittuvia vastakohtaisille esimerkeillemme.

**Tulos**

Kuinka paljon dramaattisesti tulokset laskevat malleissa, jotka käyttävät luotuja vastakkaisia esimerkkejä?

**Esimerkki 5.1496**

Koska loukkaavan kielenkäytön havaitsemista koskeva tutkimus on suurelta osin keskittynyt englannin kieleen, lähdimme tutkimaan sellaisten mallien suunnittelua, joita voidaan menestyksekkäästi käyttää sekä englannin että tanskan kielessä.

**Tulos**

Mikä on haaste muille kielille kuin englannille

**Esimerkki 5.1497**

Aloitimme arvioimalla tavanomaisia MT-paradigmoja eli PBSMT BIBREF3 ja NMT BIBREF1 . PBSMT:n osalta tarkastelimme myös kahta kehittynyttä menetelmää: apukieleen perustuvaa pivot-pohjaista käännöstä BIBREF10 ja fraasitaulukoiden induktiota yksikielisestä datasta BIBREF14 . NMT:n osalta vertasimme kahta erilaista koodaaja-dekooderi-arkkitehtuuria: tarkkaavaisuuteen perustuvaa RNN-pohjaista mallia (RNMT) BIBREF2 ja Transformer-mallia BIBREF18 . Normaalin yksisuuntaisen mallintamisen lisäksi tutkimme vähäisten resurssien ongelman ratkaisemiseksi kahta monisuuntaista mallia: kaksisuuntaista mallia BIBREF11 ja monista moniin (M2M) -mallia BIBREF8 . Parhaan mallin tunnistamisen jälkeen tutkimme myös takaisinkääntämiseen perustuvan tiedonlisäysmenetelmän hyödyllisyyttä BIBREF17 .

**Tulos**

Mikä oli lähtötaso?

**Esimerkki 5.1498**

Toisella kierroksella keräsimme 293 annotaatiota 12 annotoijalta. Korektorin jälkeen COSTRA 1.0 -tietokannan muodostavat 4262 yksilöllistä lausetta (mukaan lukien 150 siemenlausetta).

**Tulos**

Miten mahdolliset lauseiden muunnokset esitetään tietokannassa uusina lauseina?

**Esimerkki 5.1499**

Suoritamme kokeilumme käyttäen SemEval 2010 Task 14: Word Sense Induction & Disambiguation BIBREF5 -arviointimenetelmää.

**Tulos**

Millainen arviointi suoritetaan?

**Esimerkki 5.1500**

Vuoden 2019 Workshop on Neural Generation of Text (WNGT) Efficiency -tapahtuman jaettua tehtävää BIBREF0 varten Notre Dame Natural Language Processing (NDNLP) -ryhmä tutki menetelmää, jolla parametreihin saadaan aikaan harvennusta, jota kutsutaan automaattiseksi koonmääritykseksi, jotta voidaan vähentää parametrien lukumäärää muuntajassa suhteellisen vähäisellä suorituskyvyn laskulla.

**Tulos**

Mikä on WNGT 2019:n yhteinen tehtävä?

**Esimerkki 5.1501**

Tietokokonaisuutemme on kommentoitu Ruijinin sairaalan ruoansulatuskanavan kirurgian osaston toimittamien kiinalaisten patologisten raporttien perusteella. Se sisältää 17 833 lausetta, 826 987 merkkiä ja 2 714 kysymys-vastausparia.

**Tulos**

Kuinka suuri QA-CTS-tehtävätietokanta on?

**Esimerkki 5.1502**

Tässä jaksossa kuvaamme näitä kolmea osatekijää koskevaa tiedonkeruuta. Aluksi kuvaamme pilottitutkimuksia, jotka teimme kerätessämme tervejärkisiä päättelykysymyksiä (SECREF4 ). Jaksossa SECREF5 käsitellään kysymysten, tekstien ja vastausten keräämistä Amazon Mechanical Turk -palvelun (jäljempänä MTurk) avulla. Jaksossa SECREF17 annetaan tietoa eräistä tarvittavista jälkikäsittelyvaiheista ja tietokokonaisuuden validoinnista.

**Tulos**

miten tiedot kerättiin?

**Esimerkki 5.1503**

Määrittelemme vihapuheen kieleksi, joka nimenomaisesti tai epäsuorasti uhkaa tai halventaa henkilöä tai ryhmää heidän identiteettinsä, kuten sukupuolen, etnisen alkuperän tai seksuaalisen suuntautumisen, perusteella.

**Tulos**

Mikä on heidän vihapuheen määritelmänsä?

**Esimerkki 5.1504**

Keräsimme tietokokonaisuuden, jossa oli 1900 vuoropuhelua ja 8533 kierrosta.

**Tulos**

Mikä on dialogin keskimääräinen pituus?

**Esimerkki 5.1505**

Joukkotyöntekijöitä pyydettiin merkitsemään, oliko kolmikko oikein, eli kuvastiko kolmikko lauseen seurausta.

**Tulos**

Mikä on joukkoistamisen rooli?

**Esimerkki 5.1506**

Taulukossa TABREF18 esitetään GM$\\_$KL-mallin Spearmanin korrelaatioarvot, jotka on arvioitu sanojen samankaltaisuuden vertailuaineistoissa: SL BIBREF20, WS, WS-R, WS-S BIBREF21, MEN BIBREF22, MC BIBREF23, RG BIBREF24, YP BIBREF25, MTurk-287 ja MTurk-771 BIBREF26, BIBREF27 ja RW BIBREF28. Taulukossa TABREF19 esitetään GM$\\_$KL-mallin arviointitulokset merkitystietoaineistoissa, kuten WordNetistä luotu merkitysparitietoaineisto BIBREF29, jossa on sekä positiivisia että negatiivisia merkintöjä, joukkoistettu tietokokonaisuus BIBREF30, joka sisältää 79 semanttista suhdetta, jotka on merkitty merkityksi merkityksi tai ei merkityksi, ja annotoitu distributiivisesti samankaltaiset substantiivit -tietokokonaisuus BIBREF31.

**Tulos**

Mitä laadullisia kokeita on tehty vertailutietokannoilla?

**Esimerkki 5.1507**

Gensimin julkaiseman D2V-algoritmin virittämiseen käytettävissä olevista parametreista valittiin kuusi parametria BIBREF14-optimointia varten. Parametri window\_size vaikuttaa tekstien analysoinnissa käytettävän liukuikkunan kokoon. Parametri alpha edustaa verkon oppimisnopeutta. Sample-asetuksella malli voi vähentää suurtaajuisille sanoille annettua painoarvoa. dm-parametri määrittää koulutuksessa käytettävän arkkitehtuurin (PV-DM tai PV-DBOW). Parametri hs määrittelee, käytetäänkö koulutuksessa hierarkkista softmaxia vai negatiivista näytteenottoa. Lopuksi parametri vector\_size vaikuttaa tuloksena saatavan vektorin muodostavien ulottuvuuksien määrään.

**Tulos**

Mitkä kuusi parametria optimoitiin ruudukkohakua käyttäen?

**Esimerkki 5.1508**

Perusasetukset Vertaamme perusteellisesti lähestymistapojamme seuraaviin perusasetuksiin: Suora lähde$\suora $kohde: Monikielinen: Yksi yhteinen NMT-malli useille käännössuunnille BIBREF6.Moni-moneen: Koulutettu kaikkia mahdollisia suuntia varten lähde-, kohde- ja pivot-kielten välillä: Koulutettu vain kohdekielen suuntiin, eli lähde$\rightarrow $target ja pivot$\rightarrow $target, mikä yleensä toimii paremmin kuin many-to-many-järjestelmät BIBREF27.

**Tulos**

Mitkä ovat monikieliset mallit, jotka suoriutuivat kokeessa paremmin?

**Esimerkki 5.1509**

Wikipedian versiotallenteet, jotka Leskovec et al. leskovec2010governance esitteli aiemmin, sisältävät kahdeksan gigatavua (pakattu koko) versiomuokkauksia metatietoineen.

**Tulos**

Kuinka suuri on Wikipedian revision dump -tietokanta?

**Esimerkki 5.1510**

Tietokannan lopulliset merkintäluokat ovat: Ilo, suru, viha, pelko, odotus, yllätys, rakkaus, inho, neutraali.

**Tulos**

Kuinka monta tunnetta he katsovat?

**Esimerkki 5.1511**

Ehdolliset satunnaiskentätEhdolliset satunnaiskentät (CRF) BIBREF10 ovat vakiintunut lähestymistapa, kun käsitellään peräkkäisiä tietoja sekvenssien merkitsemisen yhteydessä. BiLSTM-CRF Ennen syvien neuraalisten kielimallien syntymistä BiLSTM-CRF-mallit BIBREF11 olivat saavuttaneet huippuluokan tuloksia sekvenssien merkitsemistehtävässä. MonitehtäväoppiminenMulti-Task Learning (MTL) BIBREF15 on tullut suosituksi syväoppimisen edistymisen myötä. BioBERTDeep neuraaliset kielimallit ovat viime aikoina kehittyneet menestyksekkääksi menetelmäksi tekstin esittämiseen. Erityisesti Bidirectional Encoder Representations from Transformers (BERT) päihitti aiemmat huipputason menetelmät huomattavasti eri NLP-tehtävissä BIBREF17.

**Tulos**

Mitä perustasoja ne ottavat käyttöön?

**Esimerkki 5.1512**

Sen tuottamiseksi tutkimme kahta implisiittistä käyttäjäpalautteen merkintästrategiaa: viiden minuutin uudelleenkäyttöä ja yhden päivän palautusta.

**Tulos**

Mitä palautelappuja käytetään?

**Esimerkki 5.1513**

Suunnittelemme kolme tehtävää seq2seq-mallin esiharjoittelua varten: lauseiden uudelleenjärjestäminen (SR), seuraavan lauseen tuottaminen (NSG) ja peitettyjen asiakirjojen tuottaminen (MDG).

**Tulos**

Mitä on peitettyjen asiakirjojen tuottaminen?

**Esimerkki 5.1514**

Korpus koostuu 53 asiakirjasta, joissa on keskimäärin 156,1 lausetta asiakirjaa kohti ja joissa jokaisessa on keskimäärin 19,55 merkkiä. Korpus sisältää yhteensä 8 275 lausetta ja 167 739 sanaa.

**Tulos**

Kuinka suuri korpus on?

**Esimerkki 5.1515**

Raportteja julkaisevat Météo France ja Met Office, sen brittiläinen vastine. Ne ovat julkisesti saatavilla kyseisten organisaatioiden verkkosivuilla. Molemmat korporaatiot kattavat saman ajanjakson kuin vastaavat aikasarjat, ja koska ne ovat luonteeltaan päivittäisiä, saadaan yhteensä 4 261 ja 4 748 asiakirjaa.

**Tulos**

Kuinka suuri on koulutukseen/testaukseen käytetty tietokokonaisuus?

**Esimerkki 5.1516**

Kuten edellä on kuvattu, mukautimme rupnik2016news-sivuston klusterilinkitystietokannan arvioidaksemme kieltenvälistä online-klusterointimenetelmäämme.

**Tulos**

Mitkä ovat tietokokonaisuuksien lähteet?

**Esimerkki 5.1517**

Tässä luvussa sovellamme syväoppimisen ja kielitieteen yhdistelmää sekä aspektin erottamisen että subjektiivisuuden havaitsemisen ongelman ratkaisemiseen.

**Tulos**

Miten aspektit tunnistetaan aspektien louhinnassa?

**Esimerkki 5.1518**

Vaikka huumori on yleismaailmallinen konstruktio, on olemassa suuria eroja sen välillä, mitä kukin yksilö voi pitää humoristisena. Pyrimme keskittymään väestön osajoukkoon, jossa voimme mitata reaktioita kvantitatiivisesti: suosittuun Redditin r/Jokes-ketjuun. Tämä foorumi on erittäin suosittu - siellä julkaistaan kuukausittain kymmeniä tuhansia vitsejä ja siellä on yli 16 miljoonaa jäsentä. Vaikka on olemassa suurempia vitsitietokantoja, r/Jokes-ketju on vertaansa vailla sen sisältämien luokiteltujen vitsien määrässä. Tietojemme mukaan vastaavaa luokiteltujen vitsien lähdettä ei ole millään muulla kielellä. Nämä Reddit-postaukset koostuvat vitsin rungosta, iskulauseesta ja reaktioiden eli upvotes-äänten määrästä. Vaikka tämäntyyppinen huumori saattaa olla miellyttävintä vain osalle väestöstä, se on tehokas tapa mitata reaktioita vitseihin suuressa ryhmässä.

**Tulos**

Millaista huumoria he ovat arvioineet?

**Esimerkki 5.1519**

Useat artikkelit ovat osoittaneet, että "matalat" naive Bayes-luokittelijat BIBREF12, BIBREF8, BIBREF13, BIBREF14, SVM:t BIBREF15 ja vastaavat mallit toimivat erittäin hyvin LID:n tekemiseen. Esimerkiksi DSL 2017 -julkaisussa BIBREF1 annetaan yleiskatsaus kaikkien jaetusta tehtävästä kilpailleiden tiimien ratkaisuihin, ja voittajaksi selviytynyt lähestymistapa BIBREF16 käytti SVM:ää, jossa oli hahmojen n-grammi- ja puheosien tunniste-ominaisuuksia sekä joitakin muita kehitettyjä ominaisuuksia. DSL 2015:n voittajalähestymistapa käytti ensemble-naive Bayes -luokittelijaa. Fasttext-luokittelija BIBREF17 on ehkä yksi tunnetuimmista tehokkaista "pinnallisista" tekstiluokittelijoista, joita on käytetty LID:ssä. Useissa artikkeleissa on ehdotettu hierarkkisia pinottuja luokittelijoita (myös leksikoita), jotka esimerkiksi luokittelevat tekstin ensin kieliryhmän ja sitten tarkan kielen mukaan BIBREF18, BIBREF19, BIBREF8, BIBREF0. Tutkijat ovat tutkineet syvempiä LID-malleja li

**Tulos**

Mikä on aiempien töiden lähestymistapa?

**Esimerkki 5.1520**

Seuraavaksi osoitamme, että MASC-arkkitehtuurin avulla voidaan saavuttaa erittäin korkea paikannustarkkuus, ja se on huomattavasti parempi kuin perusratkaisu, joka ei sisällä tätä mekanismia.

**Tulos**

Mitä arviointimittareita tekijät tarkastelivat?

**Esimerkki 5.1521**

Perusarvot. Koska yksikään nykyisistä KBC-menetelmistä ei pysty ratkaisemaan OKBC-ongelmaa, valitsemme perustasoksi LiLin eri versiot: LiLi: LiLi-versio, jossa koulutamme yhden ennustemallin INLINEFORM0 kaikille testisuhteille: Emme siirrä (aiemmin opittuja) painoja INLINEFORM0:n alustamista varten, eli poistamme LL.F-th) käytöstä: Tässä käytämme kiinteää ennustuskynnystä 0,5 relaatiokohtaisen kynnyksen INLINEFORM0 sijasta.BG: Puuttuvat tai yhdistävät linkit (kun käyttäjä ei vastaa) täytetään "@-RelatedTo-@" -merkinnällä sokeasti, ei arvausmekanismia.w/o PTS: LiLi ei pyydä lisävihjeitä aiempien tehtävien valinnan kautta taitojen parantamista varten.

**Tulos**

Mitä lähtötasoa kokeissa käytetään?

**Esimerkki 5.1522**

Lämmittelymenettelynä kaksi kommentoijaa (kirjoittaja ja valvoja) kommentoivat ensimmäiset 100 viestiä ja vertailivat tuloksia keskenään. Loppuosan annotointitehtävästä suoritti kirjoittaja, ja tuloksena oli 3600 annotoitua näytettä.

**Tulos**

Keitä olivat kommentoijat?

**Esimerkki 5.1523**

Varmistaaksemme tekniikkamme luotettavuuden harvinaisten sanojen kattavuuden laajentamisessa teimme kokeita harvinaisten sanojen samankaltaisuustietokannalla BIBREF6 . Tietokanta sisältää 2034 paria harvinaisia sanoja, kuten ulcerate-change ja nurturance-care, joita 10 arvioijaa arvioi asteikolla [0,10]. Taulukossa TABREF15 esitetään tulokset aineistosta kolmelle esivalmistetulle sanasulkeumalle (ks. SECREF2 ) niiden alkuperäisessä muodossa sekä WordNetistä saaduilla lisäsanoilla täydennettynä.

**Tulos**

Miten harvinaiset sanat määritellään?

**Esimerkki 5.1524**

Taajuussuodattimen suhde INLINEFORM0 asetetaan matalataajuisten sanojen (harvinaisten sanojen) suodattamiseksi pois hakutaulukosta.

**Tulos**

miten harvinaiset sanat määritellään?

**Esimerkki 5.1525**

Parafraaseja saadaan kääntämällä englanninkielinen merkkijono vieraalle kielelle ja kääntämällä se sitten takaisin englanniksi.

**Tulos**

Vaikuttaa siltä, että kysymysten parafraasien oppimista, neuraalista pisteytysmallia ja vastausten valintamallia ei voida kouluttaa päästä päähän. Miten ne koulutetaan?

**Esimerkki 5.1526**

Se sisältää yhteensä 708 tuntia ranskalaisia (Fr), saksalaisia (De), hollantilaisia (Nl), venäjänkielisiä (Ru), espanjalaisia (Es), italialaisia (It), turkkilaisia (Tr), persialaisia (Fa), ruotsalaisia (Sv), mongolialaisia (Mn) ja kiinalaisia (Zh) puheita, joista ranskalaiset ja saksalaiset puheet ovat olemassa olevista julkisista korpuksista pisimpään kestäneet.

**Tulos**

Mitkä kielet kuuluvat korpukseen?

**Esimerkki 5.1527**

Sekä BasicDCGAN:n että MultitaskDCGAN:n havaittu suorituskyky, kun käytetään 3-luokkia, on verrattavissa uusimpaan tekniikkaan: 49,80 % verrattuna 49,99 %:iin, joka on raportoitu BIBREF16:ssa.

**Tulos**

Millä mallilla saavutetaan tässä tehtävässä uusin mahdollinen suorituskyky?

**Esimerkki 5.1528**

Tarkastelemme kahdenlaisia arkkitehtuureja: kaksisuuntainen kielimalli, jolla täydennetään sekvenssistä sekvenssiin -kooderia, ja yksisuuntainen malli, jolla täydennetään dekooderia. Molemmissa käytetään itsestään huomiota herättävää BIBREF16 -mallia, ja yksisuuntainen malli sisältää INLINEFORM0-muuntajalohkoja, joiden jälkeen sanaluokittelija ennustaa oikealla olevan seuraavan sanan. kaksisuuntainen kielimalli, joka täydentää sekvenssistä sekvenssiin -kooderia.

**Tulos**

Mitä kielimalliarkkitehtuureja käytetään?

**Esimerkki 5.1529**

Tarjoamme kolme paria lyhyitä ja pitkiä tietokokonaisuuksia eri aloilta (elokuvat ja ravintolat) ja eri kieliltä (englanti ja korea), jotka soveltuvat tehtävään: Mov\_en, Res\_en ja Mov\_ko. Suurin osa aineistoista on peräisin aiemmasta kirjallisuudesta, ja ne on kerätty eri tavoin Mov\_en-aineistot on kerätty eri verkkosivustoilta; lyhyt aineisto koostuu BIBREF19:n Rotten Tomatoes -verkkosivuston dokumenttitason arvosteluista käsin poimimista lauseista, kun taas pitkä aineisto koostuu BIBREF20:n saamista arvosteluista IMDB-sivustolta.

**Tulos**

Mitä eri aloja ja kieliä uusissa tietokokonaisuuksissa on?

**Esimerkki 5.1530**

Vertaamme lauseiden luokittelua LSTM:n perustasoon ja arvioimme suurinta tällä hetkellä saatavilla olevaa PICO-lauseaineistoa BIBREF13. SCIBERT valittiin ylimääräiseksi perusmalliksi hienosäätöä varten, koska se edusti parhaiten upotettuja PICO-lauseiden tietoja.

**Tulos**

Mitä lähtökohtia he ottivat huomioon?

**Esimerkki 5.1531**

Lasketaan ensin yksinkertaisesti sekä INLINEFORM0:n että INLINEFORM1:n tuottamien ennusteiden keskiarvo SECREF-osiossa36 esitetyissä kokeissa.

**Tulos**

Mitä kahta mittaria ehdotetaan?

**Esimerkki 5.1532**

Toisessa vaiheessa koulutamme neuraalisen yksikielisen käännösjärjestelmän, joka kääntää PBMT-järjestelmän tuloksesta INLINEFORM0 parempaan kohdelauseeseen INLINEFORM1 .

**Tulos**

Koulutetaanko NMT-malli PBMT:n tuotoksilla?

**Esimerkki 5.1533**

Uusien avainsanojen tunnistamiseksi valituista mikroposteista hyödynnämme jälleen joukkoistamista, sillä ihmiset ovat yleensä koneita parempia antamaan erityisiä selityksiä BIBREF18, BIBREF19. Joukkoistamistehtävässä työntekijöitä pyydetään ensin etsimään ne mikropostit, joissa mallin ennusteet katsotaan oikeiksi. Sitten työntekijöitä pyydetään etsimään näistä mikroposteista avainsana, joka parhaiten osoittaa mallin ennustaman mikropostin luokan.

**Tulos**

Miten avainsanakohtainen odotus herätetään joukosta?

**Esimerkki 5.1534**

Morfosyntaktinen annotaatio tehtiin automaattisesti, mikä voi aiheuttaa virheitä erityisesti silloin, kun kyseessä on meluisa verkkoteksti.

**Tulos**

Käyttivätkö he kommenttien tekemiseen joukkoistamisalustaa?

**Esimerkki 5.1535**

Yhteistyöelimen hienosäätämiseksi käytimme omaa korpusta, joka koostuu sadoista tuhansista oikeudellisista sopimuksista.

**Tulos**

Kuinka suuri on yhteistyöelimen hienosäätöön käytettävä tietokokonaisuus?

**Esimerkki 5.1536**

Tässä jaksossa esitellään yksityiskohtaisesti analyysi, joka on tehty Twitter-viestejä koskevista tiedoista tammikuusta 2020 tähän päivään asti, eli siitä lähtien, kun uutinen Kiinassa puhjenneesta Coronavirus-tartunnasta levisi ympäri kansakuntia. Lisäksi aineisto kattaa erityisesti englannin-, espanjan- ja ranskankieliset twiitit.

**Tulos**

Keräävätkö ne vain englanninkielisiä tietoja?

**Esimerkki 5.1537**

Luonnollisen kielen käsittelyssä (NLP) kullekin kielelle ominaisen morfologisen monimutkaisuuden huomioon ottaminen voi olla tärkeää nykyisten menetelmien parantamiseksi tai mukauttamiseksi, sillä sanatasolle koodatun semanttisen ja kieliopillisen tiedon määrä voi vaihdella merkittävästi kielestä toiseen. Tämä tieto voisi olla hyödyllistä kielellisessä analyysissä ja mitattaessa eri sanamuotojen normalisointityökalujen vaikutusta kielestä riippuen.

**Tulos**

Mikä on tämän asiakirjan käytännön sovellus?

**Esimerkki 5.1538**

Suunnittelimme kolme arviointikokonaisuutta: (1) Perusjoukko (1264 näytettä), joka on peräisin simuloidusta datasta. (2) laajennettu joukko (1280 näytettä), joka muodostettiin lisäämällä perusjoukkoon kaksi jakauman ulkopuolista oiretta ja niitä vastaavat dialogisisällöt ja kyselyt ("verenvuoto" ja "kylmä", jotka eivät koskaan esiintyneet harjoitusaineistossa). (3) Reaalimaailman joukko (944 näytettä), joka on rajattu manuaalisesti reaalimaailman dialogien oireiden tarkistusosioista (noin 4 tuntia) ja merkitty arviointinäytteiksi.

**Tulos**

Miten he valitsevat testitapaukset pidättäytyneeseen testijoukkoonsa?

**Esimerkki 5.1539**

Pyrimme mahdollisuuksien mukaan rajoittamaan kunkin tekstin sisällön alkuperäisen teoksen tiivistelmään ja johtopäätökseen. Koska nämä osiot olivat kuitenkin monissa tapauksissa liian lyhyitä, otimme huomioon myös muita alkuperäisteosten osia, kuten johdannon tai keskusteluosion. Sen varmistamiseksi, että poimitut tekstiosuudet soveltuvat AV-tehtävään, kukin alkuperäisteos esikäsiteltiin manuaalisesti. Tarkemmin sanottuna poistimme taulukot, kaavat, lainaukset, sitaatit ja lauseet, jotka sisälsivät muuta kuin kielellistä sisältöä, kuten matemaattisia konstruktioita tai tutkijoiden, järjestelmien tai algoritmien nimiä.

**Tulos**

Mikä on itse koottu korpus?

**Esimerkki 5.1540**

Suoritamme myös muita CTC-kokeita 36-kerroksisella muuntajalla (kokonaisparametrit $\pm $120 miljoonaa). 36 kerroksen perustasolla on sama suorituskyky kuin 24 kerroksella, mutta lisäämällä ehdotetut menetelmät 36 kerroksen suorituskyky parani ja antoi parhaat tulokset.

**Tulos**

Kuinka monta kerrosta he käyttävät parhaiten toimivassa verkossaan?

**Esimerkki 5.1541**

Taulukon TABREF7 tulokset osoittavat, että SRCC suoriutuu tiedon louhinnasta paljon paremmin. Kahden asiakirjan sisältö sisältää saman ajatuksen, joka on ilmaistu termeillä eri järjestyksessä: John oli pyytänyt Marya naimisiin ennen kuin tämä lähti. On ilmeistä, että kosinin samankaltaisuus ei pysty tunnistamaan tätä yhteyttä, mutta SRCC on tunnistanut sen onnistuneesti ja tuottanut samankaltaisuusarvon -0,285714.

**Tulos**

Miten ne arvioivat tiedon louhinnan suorituskykyä?

**Esimerkki 5.1542**

Käytämme Fisherin espanjankielistä puhekorpusta BIBREF11, joka koostuu 819 puhelusta, joiden keskimääräinen kesto on 12 minuuttia, eli yhteensä 160 tuntia dataa.

**Tulos**

Mitä kieltä he tarkastelevat?

**Esimerkki 5.1543**

Kuten taulukosta käy ilmi, ehdotettu PS-rnn-elmo parantaa MAP-suorituskykyä merkittävästi verrattuna edelliseen parhaaseen malliin, CompClip-LM:ään (0,696-0,734 absoluuttista).

**Tulos**

Kuinka paljon parempi suorituskyky ehdotetulla mallilla on verrattuna vastausvalintamalleihin?

**Esimerkki 5.1544**

Cambridge Assessment English BIBREF27:n BULATS-testi (Business Language Testing Service) on monitasoinen tietokonepohjainen englannin kielen koe. Se koostuu luetun puheen ja vapaan puhumisen osista, joissa kokelas vastaa kehotuksiin. BULATSin puhutun kielen testissä on viisi osiota, joissa kaikissa on liiketoimintatilanteisiin soveltuvaa materiaalia. Tässä työssä BULATS-testin ei-äidinkielen puhetta käytetään sekä koulutus- että testiaineistona puhujan todentamisjärjestelmissä. Järjestelmien yleistettävyyden tutkimiseksi testiaineisto otetaan myös Cambridge Assessment English Linguaskill -verkkotestistä. BULATSin tavoin tämäkin testi on monitasoinen, ja sen muoto koostuu samoista viidestä osiosta kuin edellä kuvatun kaltaisesti, mutta siinä arvioidaan yleistä englannin kielen taitoa.

**Tulos**

Mitä standardoitua suurta puhujan todentamiskorpusta käytetään arvioinnissa?

**Esimerkki 5.1545**

Vuoropuhelujen kerääminen: Ennen virallisen tiedonkeruun aloittamista vaadimme työntekijöitä tekemään pienen määrän vuoropuheluja ja annoimme heille palautetta vuoropuhelun laadusta. Sitten hyvin koulutetut työntekijät paritettiin keskustelemaan annettujen tavoitteiden mukaisesti. Työntekijöitä pyydettiin myös merkitsemään sekä käyttäjän että järjestelmän tilat. 2. Dialogin merkintä: Käytimme joitakin sääntöjä merkitäksemme automaattisesti dialogin tekoja käyttäjän tilojen, järjestelmän tilojen ja dialogihistorian mukaan. Dialogitapahtumien ja -tilojen annotoinnin laadun arvioimiseksi kolme asiantuntijaa annotoi manuaalisesti 50 dialogin dialogitapahtumat ja -tilat. Tulokset osoittavat, että annotaatiomme ovat korkealaatuisia. Lopuksi jokainen dialogi sisältää strukturoidun tavoitteen, tehtävänkuvauksen, käyttäjän tilat, järjestelmän tilat, dialogitoiminnot ja lausumat.

**Tulos**

Miten korpus annotoitiin?

**Esimerkki 5.1546**

Koulutamme mallit Sentiment140- ja Amazon-tuotearvosteluilla. Molemmissa tietokokonaisuuksissa keskitytään lyhyen tekstin esittämiin tunnetiloihin. Kouluttaaksemme softmax-mallia jaamme tekstin sentimentin kahteen tunnetyyppiin, positiiviseen ja negatiiviseen. Tanh-mallin kouluttamista varten muunnamme positiivisen ja negatiivisen tunteen [-1.0, 1.0] jatkuvaksi tunnepisteytykseksi, kun taas 1.0 tarkoittaa positiivista ja päinvastoin.

**Tulos**

Koulutettiinko käyttöön otettu LSTM+CNN-malli annotoidulla datalla valvotusti?

**Esimerkki 5.1547**

Tulokset osoittavat, että CJFA-kooderi saavuttaa huomattavasti paremman puhelimen luokittelutarkkuuden kuin VAE-perusversio ja myös kuin CJFS-kooderi. Nämä tulokset toistuvat myös puhujantunnistustehtävissä.

**Tulos**

Kumpi kahdesta paperissa ehdotetusta lähestymistavasta toimi kokeissa paremmin?

**Esimerkki 5.1548**

Käytämme kahta suosittua tietopohjaa: WordNet BIBREF0 ja Freebase BIBREF1. Käytämme erityisesti WN18 (WordNetin osajoukko) BIBREF24 ja FB15K (Freebasen osajoukko) BIBREF2, koska niiden tekstikuvaukset ovat helposti julkisesti saatavilla. Taulukossa 1 luetellaan näiden kahden tietokokonaisuuden tilastotiedot.

**Tulos**

Mitä tietokokonaisuuksia käytetään tämän asiakirjan arvioinnissa?

**Esimerkki 5.1549**

Lähestymistapa ::: Menetelmä: Menetelmä: Koulutus ja testaus ::: Osa aiempien lähestymistapojen ongelmista näyttää johtuvan infiksien käytöstä. Tässä kokeessa vertaamme käännöksen BLEU-2-pistemääriä havaitaksemme erot esityksen tulkittavuudessa. Lähestymistapa ::: Menetelmä:: Harjoittelu ja testaus ::: Kokeessa 2: NykytilaTässä kokeessa verrataan verkkojamme viimeaikaisiin aiempiin töihin. Laskemme tietyn testipistemäärän yksinkertaisella "oikein vs. väärin" -menetelmällä. Vastaus lausekkeeseen on suoraan sidoksissa siihen, että kaikki käännöstermit ovat oikein, minkä vuoksi emme ota huomioon osittaista tarkkuutta. Vertailemme keskimääräisiä tarkkuuksia kolmen testikokeen aikana eri satunnaisesti poimituilla testijoukoilla kustakin MWP-tietokannasta.

**Tulos**

Miten tätä ongelmaa arvioidaan?

**Esimerkki 5.1550**

Kehitimme AdaBoost-pohjaisen luokittelijan, joka koostuu 200 matalasta päätöspuusta (syvyys 2), havaitsemaan uudet väärennetyt arvostelut.

**Tulos**

Millaista mallia he käyttävät havaitsemiseen?

**Esimerkki 5.1551**

Tämän tietokokonaisuuden tarkoituksena on auttaa kehittämään puolueellisten uutisten tunnistinta. Tietokokonaisuutta voidaan käyttää useilla eri tavoilla järjestelmän kehittämiseksi. Tunnistin voidaan esimerkiksi kouluttaa käyttämällä kustantajatason merkintöjä ja testata artikkelitason merkinnöillä. On myös mahdollista käyttää puolivalvottua oppimista ja käsitellä kustantajatason osaa valvomattomana tai käyttää vain artikkelitason osaa. Julkaisimme myös tutkimuksen raakadatan, jotta voidaan kehittää uusia mekanismeja artikkelitason merkintöjen päättämiseksi.

**Tulos**

Mitä esimerkkejä sovelluksista mainitaan?

**Esimerkki 5.1552**

Tätä tehtävää varten julkaistu harjoitusaineisto sisältää 22 lääkemerkintää, joihin viitataan nimellä Training-22, ja joissa on kultainen standardiannotaatio. Koska Training-22 on suhteellisen pieni tietokokonaisuus, käytämme lisäksi ulkoista tietokokonaisuutta, jossa on 180 annotoitua lääkemerkintää ja joka on nimeltään NLM-180 BIBREF5 (lisää myöhemmin).

**Tulos**

Mitä koulutustietoja he käyttivät?

**Esimerkki 5.1553**

Tarvitsemme suuren määrän dataa kokeillaksemme syväoppimisen huippualgoritmeja. Saatavilla on paljon avoimia tietokokonaisuuksia, jotka koskevat nimiä, paikkoja ja organisaatioita, mutta ei edellä olevassa tiivistelmässä määriteltyjä aiheita. Aiheiden määrittely ja päättely on myös yksilöllistä, eikä niiden määrittelyyn ole olemassa kiinteitä sääntöjä. Määritelmämme mukaan voimme kuitenkin käyttää wikipedian otsikoita kohdeaiheina. Englanninkielisessä wikipedia-aineistossa on yli 18 miljoonaa otsikkoa, jos otetaan huomioon kaikki tähänastiset versiot. Meidän oli siivottava otsikot ja poistettava roskanimikkeet, sillä wikipediaotsikko sisältää lähes kaikki päivittäin käyttämämme sanat. Joidenkin lisäpuhdistusten jälkeen jäljelle jäi 10 miljoonaa otsikkoa. Meillä on 15 miljoonaa englanninkielistä uutisartikkelia, jotka on julkaistu viimeisten neljän vuoden aikana. Lisäksi vähensimme artikkelien määrää poistamalla päällekkäiset ja lähes samanlaiset artikkelit. Käytimme esivalmennettuja doc2vec-malleja ja kosinin samankaltaisuutta havaitaksemme lähes samanlaiset uutisartikkelit.

**Tulos**

Kuinka suuri on heidän käyttämänsä aineisto?

**Esimerkki 5.1554**

Työmme liittyy myös koneoppimismallien tulkittavuuteen ja läpinäkyvyyteen BIBREF11, BIBREF35, BIBREF12, jossa ihmiset ovat yhä useammin mukana esimerkiksi mallin tulkittavuuden jälkiarvioinnissa.

**Tulos**

Miten lähestymistavan tulkittavuus osoitetaan?

**Esimerkki 5.1555**

Saavutimme ensimmäisen sijan 8 suunnan osalta: Kolme muuta suuntaa sijoittui toiseksi (joukkueiden mukaan), mukaan lukien liettualaiset suomen-, liettuan-, suomen- ja kazakstaninkieliset suuntia: liettualaiset suuntia: englanti, suomalaiset suuntia: englanti ja englantilaiset suuntia: kolme muuta suuntaa: liettualaiset suuntia: englanti, suomalaiset suuntia: englanti ja kazakstaninkieliset suuntia: englanti.

**Tulos**

Mihin kielisuuntiin WMT:n konekäännösjärjestelmiä arvioidaan?

**Esimerkki 5.1556**

Tämän luokittelutekniikan F1-pistemäärän lopulliset tulokset ovat 0,52 yhden ja 0,31 yhden ja usean käännöksen datan osalta.Vertaamme sekä yhden että usean käännöksen kokeilujen tuloksia testidatan tarkkuuteen BLEU-pistemäärän perusteella.

**Tulos**

mihin olemassa oleviin mittareihin niitä verrataan?

**Esimerkki 5.1557**

GPT-2 saa korkeimman pistemäärän ja $n$-grammi alhaisimman.

**Tulos**

Minkä mallin suorituskyky on paras?

**Esimerkki 5.1558**

Kiinnostava aiempi työ sosiaalisten normien rikkomisen kvantifioimiseksi on keskittynyt vahvasti tietoon BIBREF8 , BIBREF9 .

**Tulos**

Ehdotetaanko tässä asiakirjassa uutta tehtävää, jota muut voivat yrittää parantaa?

**Esimerkki 5.1559**

Arvioimme lähestymistapojamme kahdella julkisella kiinalaisella span-extraction machine reading comprehension -tietokannalla: CMRC 2018 (yksinkertaistettu kiina) BIBREF8 ja DRCD (perinteinen kiina) BIBREF9. Näiden kahden tietokokonaisuuden tilastot on lueteltu taulukossa TABREF29.

**Tulos**

Onko tämä span-pohjainen (uuttamispohjainen) laadunvarmistustehtävä?

**Esimerkki 5.1560**

Toisessa, jatkuvassa sanapussissa (CBOW), mallille annetaan sanasarja ilman keskimmäistä sanaa, ja se yrittää ennustaa tämän poisjätetyn sanan.

**Tulos**

Mitä sanojen upotuksia analysoidaan?

**Esimerkki 5.1561**

Lisäksi annotaatioiden luotettavuuden arvioimiseksi ja paremman arvioinnin mahdollistamiseksi testijoukon jokaiseen kysymykseen vastaa vähintään kaksi muuta asiantuntijaa.

**Tulos**

Ovatko vastaukset kaksinkertaisesti (eikä kolminkertaisesti) merkittyjä?

**Esimerkki 5.1562**

Lauseisiin perustuvan SMT:n pohjalta haemme naapureita, joilla on suuri paikallinen, lauseen alainen päällekkäisyys lähdelauseen kanssa. Mukautamme lähestymistapaamme hakemaan n-grammeja lauseiden sijasta. Esitämme jokaisen lauseen sen pelkistetyllä n-grammijoukolla. Jokaisesta INLINEFORM0 -muodossa olevasta n-grammista etsitään lähin n-grammi harjoitusjoukosta käyttämällä edellä määriteltyä IDF- samankaltaisuutta.

**Tulos**

Mistä he hakevat naapurien n-grammeja lähestymistavassaan?

**Esimerkki 5.1563**

POS-tunnisteiden merkitsemisessä havaittiin virheiden vähenemistä 0,71 % GSD:ssä, 0,81 % Sequoiassa, 0,7 % Spokenissa ja 0,28 % ParTUT:ssa. LAS-virheiden vähennykset ovat GSD:n osalta 2,96 %, Sequoian osalta 3,33 %, Spokenin osalta 1,70 % ja ParTUT:n osalta 1,65 %. XNLI-vertailussa CamemBERT parantaa suorituskykyä monikielisiin kielimalleihin verrattuna TRANSLATE-TRAIN-asetuksella (81,2 vs. 80,2 XLM:lle), kun se käyttää alle puolet vähemmän parametreja (110M vs. 250M). Sen suorituskyky jää kuitenkin edelleen jälkeen alkuperäisellä englanninkielisellä harjoitusjoukolla koulutetuista malleista TRANSLATE-TEST-asetuksessa (81,2 vs. 82,91 RoBERTa). Molemmat parannukset johtavat 2,36 pisteen lisäykseen F1-tulokseen verrattuna parhaaseen SEM-arkkitehtuuriin (BiLSTM-CRF), mikä tekee CamemBERT:stä huippuluokan mallin FTB:n kielentämisen alalla.

**Tulos**

Kuinka paljon paremmat CamemBERTin tulokset olivat kuin aiemmat tulokset näissä tehtävissä?

**Esimerkki 5.1564**

Yksikielisestä mallista tehdään kopiot kutakin kieltä varten ja lisätään kieltenvälisiä latentteja muuttujia (CLV) yksikielisten mallien yhdistämiseksi, jolloin ne kuvaavat kieltenvälisiä semanttisia roolimalleja. Konkreettisesti, kun harjoitellaan rinnakkaisia lauseita, aina kun argumenttien pääsanat ovat samansuuntaisia, lisäämme CLV:n kahden vastaavan roolimuuttujan vanhemmaksi.

**Tulos**

Mitä muita latentteja muuttujia mallissa käytetään?

**Esimerkki 5.1565**

Rikkomukset Haastavimmat ominaisuudet liittyvät kaikki rikkomuksiin. Heikko suorituskyky Infl/Agr Violations -ominaisuudessa, joka merkitsee morfologisia rikkomuksia (Hän pesi itsensä, Tämä on onnellinen), on erityisen silmiinpistävää, koska suhteellisen suuri osa (29 %) näistä lauseista on yksinkertaisia. Nämä mallit ovat todennäköisesti puutteellisia morfologisten piirteiden koodaamisessa, koska ne ovat sanatason malleja, eikä niillä ole suoraa pääsyä sanan alaiseen informaatioon, kuten taivutuspäätteisiin, mikä osoittaa, että näitä piirteitä on vaikea oppia tehokkaasti pelkästään leksikaalisten jakaumien perusteella.

**Tulos**

Onko kirjoittajilla hypoteesi siitä, miksi morfologista yhteisymmärrystä ei juuri opita millään mallilla?

**Esimerkki 5.1566**

Ikkunointi-menetelmästä huolimatta luokkajakauma on edelleen vinoutunut, ja tarkkuusmittarin olisi heijastettava aineistomme erityistä luokkajakaumaa. Sen vuoksi käytämme painottamatonta keskimääräistä palautusprosenttia (UAR), jota käytetään yleisesti tunteiden luokittelututkimuksissa. UAR on uudelleen painotettu tarkkuus, jossa molempien luokkien näytteitä painotetaan yhtä paljon yhteenlaskettuna. UAR simuloi siten tasaista luokkajakaumaa. Tavoitteen saavuttamiseksi luokittimemme koulutetaan asianmukaisesti painotetulla harjoitusaineistolla. Huomaa, että UAR:n sattumanvarainen suorituskyky on määritelmän mukaan 50 %, mikä tekee tuloksista vertailukelpoisempia eri aineistoissa.

**Tulos**

Mitä he käyttävät mittarina löytääkseen kuumia kohtia kokouksessa?

**Esimerkki 5.1567**

Kontekstit ovat joko kyseisestä tietokokonaisuudesta peräisin olevia todellisia konteksteja tai ne ovat Wikipedian kohtia, jotka on haettu TF-IDF BIBREF24 -toiminnolla HotpotQA-kysymyksen perusteella. Uusi, päällekkäinen joukko konteksteja muodostettiin jälleen Wikipediasta HotpotQA:n avulla käyttäen samaa menetelmää kuin kierroksella 1. Kierroksella 3 Wikipediasta peräisin olevien konteksteiden lisäksi otimme mukaan konteksteja myös seuraavilta aloilta: Uutiset (poimittu Common Crawl -palvelusta), kaunokirjallisuus (poimittu BIBREF27:stä ja BIBREF28:sta), muodollinen puhuttu teksti (poimittu Open American National Corpusin Manually Annotated Sub-Corpus (MASC) -palvelun tuomioistuinten ja presidentinvaalikeskustelujen pöytäkirjoista) ja kausaalinen tai menettelytapateksti, jossa kuvataan tapahtumien tai toimien sarjoja, poimittu WikiHow'sta. Lopuksi keräsimme myös annotaatioita käyttäen GLUE RTE -koulutusaineistossa olevia pidempiä konteksteja, jotka olivat peräisin RTE5-tietokannasta BIBREF29.

**Tulos**

Mitä tietolähteitä he käyttävät tietokokonaisuuksiensa luomiseen?

**Esimerkki 5.1568**

Käyttäjätutkimuksessamme verrataan kolmen skenaarion oikeellisuutta: jäsentäjän oikeellisuus - perustasomme on niiden esimerkkien prosenttiosuus, joissa semanttisen jäsentäjän palauttama ykköskysely oli oikea.käyttäjän oikeellisuus - niiden esimerkkien prosenttiosuus, joissa käyttäjä valitsi oikean kyselyn jäsentäjän tuottamasta seitsemästä parhaasta kyselystä.hybridi-korrektius - kahden edellisen skenaarion yhdistelmällä palautettujen kyselyjen oikeellisuus. Järjestelmä palauttaa käyttäjän oikeaksi merkitsemän kyselyn; jos käyttäjä merkitsee kaikki kyselyt vääriksi, järjestelmä palauttaa jäsentäjän parhaan ehdokkaan.

**Tulos**

Tehdäänkö käyttäjätutkimus, jossa näytetään NL-käyttöliittymä selityksen kanssa ja ilman selitystä?

**Esimerkki 5.1569**

Koulutamme kielimallit kahdella kielellä: Toinen malli on arvioitu WMT'18:n jakaman saksankielisen uutiskatsauksen perusteella, joka käsittää 260 miljoonaa lausetta tai 6 miljardia merkkiä. Toinen malli koulutetaan englanninkielisellä newscrawl-aineistolla, joka käsittää 193 miljoonaa lausetta tai 5 miljardia merkkiä. Opettelemme saksan- ja englanninkielisestä uutisdatasta 37 000 tyypin yhteisen BPE-sanaston (Byte-Pair-Encoding; Sennrich et al., 2016) ja koulutamme kielimallit tällä sanastolla. Tarkastelemme kahta vertailuarvoa: Useimmat kokeet suoritetaan WMT'18 English-German (en-de) -uutiskäännöstehtävässä, ja validoimme tuloksemme WMT'18 English-Turkish (en-tr) -uutistehtävässä. WMT'18 English-German -tehtävän harjoituskorpus koostuu kaikista saatavilla olevista bitext-korpuksista lukuun ottamatta ParaCrawl-korpusta, ja siitä poistetaan yli 250 merkkiä pidemmät lauseet sekä lauseparit, joiden lähde-kohde-pituus-suhde on yli 1,5. Tuloksena on 5,18 miljoonaa lauseparia. Tokenisoimme kaiken aineiston Mosesin tokenizerilla BIBREF8 ja käytämme yksikielisistä korpuksista opittua BPE-sanastoa.

**Tulos**

Mitä tietokokonaisuutta he käyttävät?

**Esimerkki 5.1570**

Toteutamme vertailua varten kolme perusmallia: Noun WordNet Semantic Text Exchange Model (NWN-STEM), General WordNet Semantic Text Exchange Model (GWN-STEM) ja Word2Vec Semantic Text Exchange Model (W2V-STEM).

**Tulos**

Mitkä ovat asiakirjassa mainitut perusmallit?

**Esimerkki 5.1571**

Kokonaisuudet muodostettiin yksinkertaisesti keskiarvottamalla yksittäisten mallien ennusteet.

**Tulos**

Miten heidän ensemble-menetelmänsä toimii?

**Esimerkki 5.1572**

Valvojan arvioinnin korpus sisältää 26972 lausetta. Yhteenvetotilastot sanojen määrästä lauseessa ovat: min:4 max:217 keskiarvo:15.5 STDEV:9.2 Q1:9 Q2:14 Q3:19.

**Tulos**

Mikä on lauseiden keskimääräinen pituus?

**Esimerkki 5.1573**

Tässä artikkelissa käytetyt Hausa-tiedot ovat osa LORELEI-kielipakettia. Se koostuu Broad Operational Language Translation (BOLT) -aineistosta, joka on kerätty uutissivustoilta, foorumeilta, verkkopäiväkirjoista, Wikipedia-artikkeleista ja Twitter-viesteistä. Käytämme 10k harjoitus- ja 1k testitapauksen jakoa. Tässä työssä käytetty Yorùbá NER -data on BIBREFin hiljattain julkaisema Global Voices -uutisartikkelien annotoitu korpus22. Aineisto koostuu 1 101 lauseesta (26k tokenia), jotka on jaettu 709 harjoituslauseeseen, 113 validointilauseeseen ja 279 testilauseeseen 65 %/10 %/25 %:n jakosuhteen perusteella.

**Tulos**

Kuinka paljon merkittyjä tietoja on saatavilla näistä kahdesta kielestä?

**Esimerkki 5.1574**

Aukkojen täyttäminen on tiedonlouhintatehtävä, josta on tullut suosittu viime vuosina BIBREF3 . Tehtävän tarkoituksena on poimia henkilöitä, organisaatioita tai geopoliittisia yksiköitä koskevia tietoja laajasta uutis-, verkko- ja keskustelufoorumidokumenttien kokoelmasta.

**Tulos**

Mikä on lähtö- ja saapumisaikojen täyttämisen tehtävä?

**Esimerkki 5.1575**

Vertaillaksemme aiempiin menetelmiin arvioimme malliamme AIDA-B:llä ja muilla tietokokonaisuuksilla. AIDA-CoNLL BIBREF14 on kommentoitu Reutersin uutisartikkeleihin. Se sisältää harjoitus- (AIDA-Train), validointi- (AIDA-A) ja testijoukot (AIDA-B).ACE2004 BIBREF15 on osajoukko ACE2004 Coreference -asiakirjoista.MSNBC BIBREF16 sisältää kymmenen uutiskategorian (politiikka, liike-elämä, urheilu jne.) kaksi tärkeintä juttua.)AQUAINT BIBREF17 on Xinhua News Service, New York Times ja Associated Press -lehtien uutiskorpus.WNED-CWEB BIBREF18 on satunnaisesti poimittu FACC1-annotoidusta ClueWeb 2012 -tietokannasta.WNED-WIKI BIBREF18 on poimittu Wikipedian sivuilta, joilla on alkuperäiset hyperlinkki-annotaatiot.

**Tulos**

Mitä tietokokonaisuuksia arvioinnissa käytetään?

**Esimerkki 5.1576**

menetelmämme voi silti parantaa BIBREF7-tarkkuutta 60,32 %:sta 60,34 %:iin Vaikka vain 57 % testikysymyksistä voi hyötyä peruskysymyksistä, menetelmämme voi silti parantaa BIBREF7-tarkkuutta 60,32 %:sta 60,34 %:iin,

**Tulos**

Millaista tarkkuutta he lähestyvät ehdotetulla menetelmällä?

**Esimerkki 5.1577**

Logician koulutetaan huomiopohjaisella sekvenssistä sekvenssiin -paradigmalla, jossa on kolme mekanismia: rajoitettu kopiointimekanismi kirjaimellisen rehellisyyden varmistamiseksi, kattavuusmekanismi ali- ja liiallisen uuttamisen ongelman lieventämiseksi ja riippuvuushuomiointimekanismi riippuvuustiedon huomioon ottamiseksi (gated dependency attention).

**Tulos**

Miten Logician eroaa perinteisistä seq2seq-malleista?

**Esimerkki 5.1578**

Testataksemme, voidaanko aihepiirimalleja käyttää runouden ajoitukseen tai tekijyyden määrittelyyn, suoritamme valvottuja luokittelukokeita Random Forest Ensemble -luokittelijoilla.

**Tulos**

Mitä algoritmia käytetään luokittelutehtävissä?

**Esimerkki 5.1579**

Sulauttaminen: Käytimme mallin eri variaatioita, joissa käytimme tyhjästä opittuja yksinkertaisia hakutaulukkomerkintöjä ja tehokkaita kontekstuaalisia merkintöjä, jotka ovat ELMo BIBREF11, BERT BIBREF16 ja ClinicalBERT BIBREF13 (koulutettu ja tekijöiden toimittama).

**Tulos**

Mitä sulautumia käytetään?

**Esimerkki 5.1580**

Haemme jokaisesta ehdokkaasta mainintajaksosta Freebase API:n mukaan 10 tärkeintä oliota. Tämän jälkeen luomme ristikon, jonka solmut vastaavat maininta-entiteettipareja, jotka on pisteytetty Freebase API:n pisteytyksen mukaan, ja särmät kuvaavat sitä, että yksikään entiteettien ja mainintojen yhteistoimitus ei voi sisältää päällekkäisiä ajanjaksoja. Otamme 10 parasta polkua ristikon läpi mahdollisiksi entiteettien yksiselitteisyyksiksi. Kullekin mahdollisuudelle luodaan $n$ -parasta parafraasia, joka sisältää entiteettien mainintojen välit. Lopulta tämä prosessi luo yhteensä $10n$ parafraasia.

**Tulos**

Kuinka monta rinnastusta luodaan kysymystä kohden?

**Esimerkki 5.1581**

Jakomenetelmät (satunnaisjaon lisäksi) ovat seuraavat:Output length: BIBREF2:n kuvaaman asetelman muunnelma, jossa harjoitusjoukko koostuu esimerkeistä, joiden ulostulon (sparql-kyselyn tai toimintosarjan) pituus on $\le \hspace{-2.5pt}. N$, kun taas testijoukko koostuu esimerkeistä, joiden tuloksen pituus on $> \hspace{-2.5pt}. N$. CFQ:n osalta käytämme $N = 7$ rajoituksia. Skannauksessa käytämme $N = 22$ toimintoja: Vaihtoehto yllä olevasta asetelmasta, jossa harjoitusjoukko koostuu esimerkeistä, joiden syötteen (kysymyksen tai käskyn) pituus on $\le N$, kun taas testijoukko koostuu esimerkeistä, joiden syötteen pituus on $> N$. CFQ:ssa käytämme $N=19$ kieliopin lehtiä. SCANissa käytetään $N=8$ merkkejä: BIBREF8:n kuvaaman asetelman muunnelma, jossa jako perustuu sellaisten esimerkkien klustereiden satunnaiseen määrittelyyn, joilla on sama ulostulokuvio (kysely- tai toimintosekvenssi). Kyselymallit määritetään anonymisoimalla entiteetit ja ominaisuudet; toimintojaksomallit yhdistävät primitiiviset toiminnot ja suunnat: Vaihtoehto edellisestä asetelmasta, jossa jako perustuu sellaisten esimerkkien klustereiden satunnaiseen määrittelyyn, joilla on sama syötemalli (kysymys tai käsky). Kysymysmallit määritetään anonymisoimalla entiteettien ja ominaisuuksien nimet; käskymallit yhdistävät verbit ja vaihdettavat parit vasen/oikea, ympäri/vastakkain, kaksi kertaa/kolme kertaa.

**Tulos**

Mitä muita lähestymistapoja on olemassa komposition yleistämisen vertailuarvojen luomiseen?

**Esimerkki 5.1582**

Suoritamme kaikki kokeilumme 635 tunnin audiodatalla seitsemällä intialaisella kielellä, jotka on kerätty $\textbf {All India Radio}$-uutiskanavalta. Keräsimme ja kuratoimme noin 635 tuntia audiodataa seitsemästä intialaisesta kielestä, nimittäin kannadasta, hindistä, telugusta, malayalamista, bengalista ja englannista. Keräsimme datan All India Radio -uutiskanavalta, jossa näyttelijä lukee uutisia noin 5-10 minuutin ajan. Kattaaksemme monia puhujia aineistoon, indeksoimme dataa vuodesta 2010 vuoteen 2019. Koska ääni on hyvin pitkä, jotta mitään syvää neuroverkkoa voitaisiin kouluttaa suoraan, segmentoimme äänileikkeet pienempiin palasiin Voice activity detectorin avulla. Koska äänileikkeisiin on uutisten aikana upotettu musiikkia, käytämme Inhouse-musiikin havaitsemismallia musiikin segmenttien poistamiseen tietokokonaisuudesta, jotta tietokokonaisuus olisi puhdas, ja tietokokonaisuutemme sisältää 635 tuntia puhdasta ääntä, joka on jaettu 520 tunnin harjoitusdataan, joka sisältää 165 000 lausumaa, ja 115 tunnin testausdataan, joka sisältää 35 000 lausumaa.

**Tulos**

Miten äänitiedot kerättiin?

**Esimerkki 5.1583**

Tärkeimmissä kokeissa käytetään lineaarista tukivektorikonetta (Joachims, 1998) murteiden luokitteluun CxG-piirteiden avulla. Tässä artikkelissa on käytetty tietoon perustuvaa kielikartoitusta englannin kielen kansallisten murteiden valitsemiseksi maailmanlaajuiseen murteiden tunnistamismalliin.

**Tulos**

Mitä vertailtavat mallit ennustavat?

**Esimerkki 5.1584**

Löysimme mielenkiintoisia malleja, jotka mallimme oppii ja jotka auttavat ymmärtämään näitä yksikielisiä voittoja. Toistuva malli on esimerkiksi se, että englannin kielen sanat, jotka käännetään kohdekielessä samaksi sanaksi tai semanttisesti läheisiksi sanoiksi, päätyvät muunnoksen jälkeen lähemmäksi toisiaan. Esimerkiksi englannin ja espanjan välisessä tapauksessa seuraavat parit olivat niiden parien joukossa, joiden samankaltaisuus kasvoi eniten muunnostamme soveltaen: kännykkä-puhelin, elokuva-elokuva, kirja-käsikirjoitus tai rytmi-sadenssi, jotka joko käännetään samaksi sanaksi espanjaksi (eli teléfono ja película kahdessa ensin mainitussa tapauksessa) tai jotka ovat jo hyvin lähellä toisiaan espanjankielisessä avaruudessa. Yleisemmin havaitsimme, että sanaparit, jotka liikkuvat eniten yhdessä, ovat yleensä semanttisesti hyvin samankaltaisia ja kuuluvat samaan alueeseen, esimerkiksi auto-polkupyörä, ooppera-kino tai lumi-jää.

**Tulos**

Miksi malli paranee myös yksikielisissä tiloissa?

**Esimerkki 5.1585**

MTMSN BIBREF4 on ensimmäinen ja toistaiseksi ainoa malli, jossa on erityisesti pyritty käsittelemään DROPin monialaisia kysymyksiä.

**Tulos**

Mikä on edellinen malli, jonka suunnittelussa on pyritty käsittelemään monialaisia kysymyksiä?

**Esimerkki 5.1586**

Ranskalaisen puheteknologian osalta neljä radio- ja televisiolähetyksiä sisältävää korpusta ovat yleisimmin käytettyjä: ESTER1 BIBREF13, ESTER2 BIBREF14, ETAPE BIBREF15 ja REPERE BIBREF16.

**Tulos**

Mitkä ovat neljä suurinta ranskankielisen lähetyksen korporaatiota?

**Esimerkki 5.1587**

Hajonnan vähentämiseksi ja tarkkuuden lisäämiseksi yhdistämme 10 CNN:ää ja 10 LSTM:ää pehmeän äänestyksen avulla. Yhdistetyillä malleilla on erilaiset satunnaispainojen alustukset, eri epookkien määrä (yhteensä 4-20), eri suodatinkokojen joukko (joko INLINEFORM0 , INLINEFORM1 tai INLINEFORM2 ) ja erilaiset upottamisen esivalmennusalgoritmit (joko Word2vec tai FastText).

**Tulos**

Kuinka monta CNN:ää ja LSTM:ää koottiin?

**Esimerkki 5.1588**

MATT-BiE-LSTM ja WATT-BiE-LSTM saavuttavat samankaltaisia tuloksia, kun niitä testataan sekä Vaderin että ihmisen kommentoimilla näytteillä, vaikka niiden tavat laskea huomio (painot ja vektorit) eroavat toisistaan siten, että WATT laskee huomiopainot ja senti-emoji-kuvion upotukset kunkin sanan perusteella, kun taas MATT saa senti-emoji-kuvion LSTM-koodaimen perusteella koko kontekstista ja laskee senti-emoji-kuvion upotuksen huomiopainot kaikkien sanojen osalta. Molemmat mallit ylittävät huipputason perusmallit, mukaan lukien ATT-E-LSTM Attention-based LSTM with emojis: Käytämme myös sana-emoji-kuviointia emoji-sanan huomion laskemiseen yhtälön EQREF20 ja EQREF21 mukaisesti, ja ainoa ero on, että korvaamme huomiosta johdetun senti-emoji-kuvioinnin fasttextin ennalta koulutetulla sana-emoji-kuvioinnilla, jota kutsutaan nimellä ATT-E-LSTM. LSTM, jossa on bi-sense-emoji-sulautus (ehdotettu): Kuten SECREF13 -jaksossa esiteltiin, ehdotamme kahta huomiota hyödyntävää LSTM-verkkoa, jotka perustuvat bi-sense-emoji-merkintöjen upottamiseen ja joita kutsutaan nimillä MATT-BiE-LSTM ja WATT-BiE-LSTM.

**Tulos**

Mitkä SOTA-mallit päihittävät?

**Esimerkki 5.1589**

Näytämme lähestymistapamme vastaamalla avoimen alan tyhjään-kenttään täytettäviin luonnollisen kielen kysymyksiin.

**Tulos**

Minkä tehtävän perusteella he arvioivat?

**Esimerkki 5.1590**

Tämän vuoksi tutkimme tapoja havaita epäilyttävät tilit tarkastelemalla niiden twiittejä ryhmissä (chunks). Hypoteesimme on, että epäilyttävillä tileillä on ainutlaatuinen malli twiittijaksojen lähettämisessä. Koska heidän tarkoituksenaan on johtaa harhaan, heidän tapaansa siirtyä twiittien sarjasta toiseen on piilotettu tunnusmerkki, jota heidän aikomuksensa vääristävät. Kun kyseessä on uutisia sisältävä Twitter-tili, luemme sen twiitit tilin aikajanalta. Sitten lajittelemme twiitit julkaisupäivän mukaan nousevaan järjestykseen ja jaamme ne $N$-kappaleisiin. Kukin palanen koostuu lajitellusta twiittien sarjasta, joka on merkitty vastaavan tilin tunnisteella.

**Tulos**

Miten tässä teoksessa määritellään "virka-apuraha"?

**Esimerkki 5.1591**

Ehdotamme BERT:iin perustuvaa luonnollisen kielen tuottamismallia, jossa käytetään hyväksi valmiiksi koulutettua kielimallia koodaus- ja dekoodausprosessissa, ja malli voidaan kouluttaa alusta loppuun ilman käsityönä tehtyjä piirteitä. Mallin koulutuksen aikana mallimme tavoite on kahden prosessin summa, joka koulutetaan yhdessä "opettaja-pakottavan" algoritmin avulla. Koulutuksen aikana syötämme perustotuuden yhteenvedon kullekin dekooderille ja minimoimme tavoitteen.$$$L\_{malli} = \hat{L}\_{dec} + \hat{L}\_{refine}$$ (Yht. 23)Testihetkellä jokaisella aika-askeleella valitsemme ennustetun sanan $\hat{y} = argmax\_{y^{\prime }} P(y^{\prime }|x)$ , käytämme palkkihakua luonnosyhteenvetojen tuottamiseen ja ahnetta hakua tarkennettujen yhteenvetojen tuottamiseen.

**Tulos**

Miten mallin eri osat koulutetaan? Koulutetaanko se alusta loppuun?

**Esimerkki 5.1592**

Ilman upotuspainoja mallimme vaatii 100k parametria, kun taas BIBREF8 vaatii 250k parametria.

**Tulos**

kuinka paljon parametrien ero heidän mallinsa ja aiempien menetelmien välillä oli?

**Tulos**

kuinka monta parametria heidän mallinsa käytti?

**Esimerkki 5.1593**

Taulukossa TABREF23 esitetään malleidemme ja eri perusmallien merkitsemättömät INLINEFORM0-pisteet.

**Tulos**

mitkä olivat arviointimittarit?

**Esimerkki 5.1594**

ACL Anthology Reference Corpus sisältää kanoniset 10 921 laskennallisen kielitieteen artikkelia, joista olen luonut 622 144 lausetta suodatettuani pois heikompilaatuiset lauseet.

**Tulos**

Mitä tietokokonaisuutta käytetään?

**Esimerkki 5.1595**

Jokaisessa koulutusvaiheessa mallia koulutetaan alkuperäisten esimerkkien ja tasapainoisesti merkittyjen vastakkaisten esimerkkien avulla. Vastakkaisten esimerkkien osuus on noin 10 prosenttia erästä. Koulutuksen aikana luomme vastakkaisia esimerkkejä, joiden kohteena on nykyinen malli, ja päivitämme mallin parametrit hybridierän avulla iteratiivisesti.

**Tulos**

Mikä on erottelija tässä generatiivisessa vastakkaisasetelmassa?

**Esimerkki 5.1596**

Valitsimme perustasoksi Conditional Copy (CC) -mallin, joka on Wisemanin paras malli.

**Tulos**

Mikä on vahva lähtötaso?

**Esimerkki 5.1597**

Koulutimme sanojen upotukset joko GloVe BIBREF11- tai SGNS BIBREF12 -ohjelmalla pienellä tai suurella korpuksella.

**Tulos**

Minkälaisia sanojen esitystapoja he arvioivat?

**Esimerkki 5.1598**

Koulutustietokantojen koko vaihtelee huomattavasti 469 viestistä 17 miljoonaan; ero on neljä suuruusluokkaa.

**Tulos**

Kuinka suuri tämä tietokokonaisuus ja luettelo ovat?

**Esimerkki 5.1599**

Kuten jäljempänä osoitamme, STransE toimii paremmin kuin SE- ja TransE-mallit ja muut huipputason linkkien ennustemallit kahdessa vakiomuotoisessa linkkien ennustamiseen tarkoitetussa tietokokonaisuudessa WN18 ja FB15k, joten se voi toimia uutena perustasona KB:n täydentämisessä.

**Tulos**

Mitä tietokokonaisuuksia käytetään mallin arvioinnissa?

**Esimerkki 5.1600**

Käytimme tallenteita sairaanhoitajan aloitteesta käydyistä puhelinkeskusteluista, joita käytiin sairaalasta kotiutumisen jälkeen etäseurannassa olevien sydämen vajaatoimintapotilaiden kanssa. Analysoidaksemme kysely-vastausparien kielellistä rakennetta koko 41 tunnin aineistossa otimme satunnaisotoksen 1 200 vuorosta koostuvasta siemenaineistosta ja luokittelimme ne manuaalisesti eri tyyppeihin, jotka on esitetty taulukossa TABREF14 yhdessä vastaavien esiintymistiheystilastojen kanssa.

**Tulos**

Mitä tietoja he käyttävät vuoropuhelua koskevan tietokokonaisuuden lähtökohtana?

**Esimerkki 5.1601**

Arviointimittarina käytettiin oikeiden "käännöksien" (vastaavuuksien) suhdelukua.

**Tulos**

Mitä arviointimittaria he käyttävät?

**Esimerkki 5.1602**

Käytämme päättelyssä säteenhakua, jonka säteen koko on 5 De-En-, En-Fr- ja En-Vi-käännöstehtävissä.

**Tulos**

Mitkä ovat kolme tärkeintä konekäännöstehtävää?

**Esimerkki 5.1603**

Käytämme väärennettyjen arvostelujen luomiseen Yelp Challenge -tietokokonaisuutta BIBREF2.

**Tulos**

Mitä tietokokonaisuutta he käyttävät lähtökohtana väärennettyjen arvostelujen luomisessa?

**Esimerkki 5.1604**

Tietokanta koostuu 198 112 uutisartikkelista.

**Tulos**

Kuinka monta artikkelia heillä oli?

**Esimerkki 5.1605**

Tavanomaisesti käytetään yhteisen tavoitteen tarkkuuden mittaria, kun malliamme verrataan aiempiin töihin.

**Tulos**

Mitä suorituskykymittareita käytetään?

**Esimerkki 5.1606**

Yllättäen parhaan laadunvarmistusmallin (bert-large-wwm) käyttäminen ei johda parhaaseen korrelaatioon ihmisen arvioiden kanssa. CNN/DM:ssä bert-large-wwm alittaa hieman bert-base- ja bert-large-mallit.

**Tulos**

Mitä malleja arvioidaan QAGS:n avulla?

**Esimerkki 5.1607**

Muissa tapauksissa malli osoittaa uskottavia signaaleja, jotka annotoija on jättänyt huomiotta ja joita voidaan pitää virheinä kultaisessa standardissa. Malli esimerkiksi huomaa helposti, että kysymysmerkit viittaavat solutionhood-suhteeseen, vaikka kommentoijat ovat ohittaneet ne ja merkinneet WH-sanat:. [RGB]230, 230, 230Mikä [RGB]230, 230, 230Edellinen [RGB]230, 230, 230Virginia [RGB]230, 230, 230Governor(s) [RGB]230, 230, 230, 230tehdä [RGB]230, 230, 230, 230sinä [RGB]230, 230, 230melkein [RGB]230, 230, 230, 230mielessä [RGB]230, 230, 230, 230ja [RGB]230, 230, 230ja [RGB]230, 230, 230miksi [RGB]12, 12, 12? $\xrightarrow[\text{pred:solutionhood}]{\text{gold:solutionhood}}$ [RGB]230, 230, 230Thomas [RGB]230, 230, 230Jefferson [RGB]183, 183, 183. Se tarttuu kuitenkin myös toistuvaan tapaan, joka esiintyy käyttöohjeissa, joissa lukijaan viittaava toisen persoonan pronomini on usein jonkin toimenpiteen hyötyjä, mikä edistää tarkoituksen lukemista ja auttaa erottamaan niin, vaikka annotoijat eivät pidä sitä signaalina.

**Tulos**

Missä ehdotettu metriikka eroaa jumanin tuomiosta?

**Esimerkki 5.1608**

INLINEFORM0 BIBREF2-harjoitusjoukossa ja kasvattaa solua INLINEFORM1 jokaista tässä ikkunassa esiintyvää INLINEFORM2-sanaa kohti (muodostaen INLINEFORM3-parin). LexVec mukauttaa PPMI-matriisin käyttämällä kontekstijakauman tasoitusta BIBREF3 . Vertailemme 1) n-grammaisten osasanojen käyttöä, kuten fastText, ja 2) Morfessor BIBREF11 -ohjelmalla tunnistettuja valvomattomia morfeemeja, jotta saamme selville, tarjoavatko kielellisesti motivoidummat osasanat etua pelkkiin n-grammeihin verrattuna.

**Tulos**

Minkälaisia alasanoja he sisällyttävät malliinsa?

**Esimerkki 5.1609**

Kanavan kovarianssimatriisin perusteella tapahtuvaan elektrodien välisten tilallisten yhteyksien purkamiseen käytetään CNN BIBREF19 -nimistä CNN:ää, joka on nelikerroksinen 2D-CNN, jossa on kaksi konvoluutiokerrosta ja kaksi täysin kytkettyä piilokerrosta. INLINEFORM0-ominaisuuskartta tietyllä CNN-kerroksella, jossa on tulo INLINEFORM1 , painomatriisi INLINEFORM2 ja bias INLINEFORM3, saadaan seuraavasti: INLINEFORM4 . Tällä ensimmäisellä hierarkiatasolla verkko koulutetaan vastaavilla merkinnöillä tavoitelähdöiksi optimoimalla risti-entropian kustannusfunktiota. Rinnakkain sovellamme nelikerroksista rekursiivista neuroverkkoa kanavien kovarianssimatriiseihin elektrodien piilotettujen ajallisten ominaisuuksien tutkimiseksi. Hyödynnämme nimittäin LSTM BIBREF20 -verkostoa, joka koostuu kahdesta täysin kytketystä piilokerroksesta, jotka on pinottu kahdella LSTM-kerroksella ja koulutettu samalla tavalla kuin CNN.

**Tulos**

Miten EEG-signaalin alueellinen aspekti laskettiin?

**Esimerkki 5.1610**

Monet käytännön valeuutisten havaitsemisalgoritmit käyttävät tilastollisten ominaisuuksiensa lisäksi eräänlaista semanttista sivutietoa, kuten sitä, onko tuotettu teksti asiallisesti oikein. Vaikka tilastollinen sivutieto olisi helppo sisällyttää hypoteesin testauksen kehykseen, on vielä ymmärrettävä, miten tällainen semanttinen tieto voidaan sisällyttää tilastolliseen päätöksentekoteorian kehykseen.

**Tulos**

Mitkä semanttiset piirteet auttavat havaitsemaan, onko teksti aito vai keksitty? of

**Esimerkki 5.1611**

DQN-pohjaisen agentin tulokset esitetään kuvassa: skenaarioiden vertailu. Kukin kuvaaja kuvaa kaikkien esitysmenetelmien keskimääräistä palkkiota (5 siemenen osalta). Voidaan nähdä, että NLP-edustus on muita menetelmiä parempi. NLP-edustukset pysyvät kestävinä ympäristön muutoksille sekä tilassa esiintyville tehtävän aiheuttamille häiriöille.

**Tulos**

Mitkä kokeiden tulokset viittaavat siihen, että luonnolliseen kieleen perustuvat agentit ovat kestävämpiä?

**Esimerkki 5.1612**

Ensinnäkin voimme havaita, että lopullinen malli "Ours with Mask and Ordered Triplets" päihittää Baseline- ja Ablation-mallit kaikilla mittareilla aiemmin nähdyissä ympäristöissä. Ero suorituskyvyssä on erityisen selvä Exact Match- ja Goal Match -mittareissa, joissa mallimme parantaa tarkkuutta 35 % ja 25 % verrattuna Baseline- ja Ablation-malleihin. Nämä tulokset viittaavat siihen, että käyttäytymiseen liittyvän navigointigraafin tarjoaminen mallille ja sen antaminen mallin käsitellä tätä tietoa tietopohjana alusta loppuun on hyödyllistä. Lopuksi on syytä huomata, että ehdottamamme malli (taulukon TABREF28 viimeinen rivi) päihittää kaikki muut mallit aiemmin nähdyissä ympäristöissä. Erityisesti saamme yli INLINEFORM0 lisäystä EM- ja GM-arvoissa mallimme ja kahden seuraavaksi parhaan mallin välillä.

**Tulos**

Kuinka paljon heidän mallinsa oli parempi kuin perusmalli?

**Esimerkki 5.1613**

Kvantitatiivisesti arvioidaksemme kaksisuuntaiseen rekursiiviseen lähestymistapaan perustuvan lähestymistapamme suorituskykyä käytämme METEOR-mittaria sen vankan suorituskyvyn vuoksi.

**Tulos**

mitä mittareita arvioinnissa käytettiin?

**Esimerkki 5.1614**

Tarkemmin sanottuna havaitsemme seuraavien tekijöiden vaikutuksen: (BIBREF11, BIBREF12, rekursiiviset BIBREF13 ja muuntopohjaiset lauseen koodaajat BIBREF14 kysymysten esittämisstrategioina; (ii) visuaalisten piirteiden louhintaan käytetyt erilaiset konvoluutiohermoverkot BIBREF15, BIBREF16, BIBREF17; ja (iii) vakioidut fuusiointistrategiat sekä kahden tärkeimmän huomiomekanismin BIBREF18 ja BIBREF19 merkitys.

**Tulos**

Millaisia kokeita tehdään?

**Esimerkki 5.1615**

Järjestelmän kuvaus ::: Arkkitehtuuri Prosessiin osallistuvat komponentit ovat: QnAMaker Portal: Tämä on graafinen käyttöliittymä (GUI) QnAMakerin käyttöä varten. QnaMaker Management API:t: Tätä käytetään kysymys-vastausparien (QA) poimimiseen puolistrukturoidusta sisällöstä. Azure Search Index: Tallentaa KB:n kysymyksineen ja vastauksineen indeksoitaviksi sarakkeiksi ja toimii siten hakukerroksena. QnaMaker WebApp: Toimii kerroksena botin, hallintasovellusten ja Azure Search Indexin välillä. Botti: Kutsuu WebAppia käyttäjän kyselyllä saadakseen tuloksia.

**Tulos**

Mistä komponenteista QnAMaker koostuu?

**Esimerkki 5.1616**

Style-perustason tarkkuus on 83 %, LDA-ominaisuuksien 89 % ja näiden kahden yhdistelmän 90 %. Harjoittelu täydellisillä runoilla kuitenkin laskee tämän tason 42-52 prosenttiin.

**Tulos**

Arvioidaanko LDA-analyysin tuloksia jollakin tavalla?

**Esimerkki 5.1617**

Joukkotyöntekijät sijaitsivat Yhdysvalloissa, ja heidät palkattiin BIBREF22-alustan kautta.

**Tulos**

Keitä ovat joukkorakentajat?

**Esimerkki 5.1618**

Arkkitehtuurimme arvioinnissa käytämme kahta tärkeintä aineistoa, jotka ovat vuoden 2014 i2b2-haasteen BIBREF2-tietoaineisto ja hoitotyön muistiinpanojen korpus BIBREF3 .

**Tulos**

Millä kahdella tietokokonaisuudella järjestelmä on testattu?

**Esimerkki 5.1619**

Tässä esitelty tehtävä oli helppo ja yksinkertainen analysoida, mutta jatkossa voidaan tehdä työtä luonnollisen kielen tehtävien parissa. Jos nämä ominaisuudet pitävät paikkansa, se saattaa viitata siihen, että NLP:n arvioinnissa olisi edistettävä uutta paradigmaa, jossa korostetaan suorituskykyä epätyypillisillä (mutta rakenteellisesti hyvillä) syötteillä harjoituksessa tyypillisesti esiintyvien tietojen lisäksi.

**Tulos**

Voidaanko tämän asiakirjan tulokset yleistää yleiskäyttöiseen tehtävään?

**Esimerkki 5.1620**

Jotta löydettäisiin kuhunkin tekstiin liittyvien skalaarien luokitteluun parhaiten soveltuva luokittelija, suoritamme arviointeja käyttämällä stokastista gradienttilaskeutumista, naive bayes-, päätöspuu- ja satunnaismetsäluokittimia.

**Tulos**

Mikä oli lähtötaso?

**Esimerkki 5.1621**

Perusmallit ja ehdotetut mallit laaditaan seuraavasti: LSTM tekstin upottamisen kanssa: CNN:iä ja LSTM:iä käytetään laajalti tekstisisällön koodaamiseen sentimenttianalyysia varten BIBREF:ssä45 , BIBREF:ssä46 ja monissa verkko-oppaissa. Tässä valitaan syötteeksi tavallinen LSTM, jossa on valmiiksi koulutettu sanojen upotus, ja lisätään LSTM-koodaajan päälle yksi täysin kytketty kerros, jossa on sigmoidiaktivointi (kuten kaikissa muissakin malleissa), jota kutsutaan T-LSTM:ksi. LSTM, jossa on emoji-merkki: Pidämme emojia yhtenä erityisenä sanana ja syötämme sekä esivalmennetun tekstin että emojien upotukset samaan LSTM-verkkoon eli E-LSTM:ään. Vastaavasti yhdistämme esivalmennetut kahden aistimuksen emoji-merkinnät yhdeksi erityiseksi sanaksi ja syötämme ne LSTM-verkkoon. Tätä mallia kutsutaan nimellä BiE-LSTM. Huomioon perustuva LSTM emojien kanssa: Käytämme myös sana-emoji-kuviointia emoji-sanan huomion laskemiseen yhtälön EQREF20 ja EQREF21 mukaisesti, ja ainoa ero on se, että korvaamme huomiosta johdetun senti-emoji-kuvioinnin fasttextin esivalmennetulla sana-emoji-kuvioinnilla, jota kutsutaan nimellä ATT-E-LSTM.

**Tulos**

Mikä on kokeilujen lähtökohta?

**Esimerkki 5.1622**

Reaaliaikaiset twiittipisteet laskettiin samalla tavalla kuin historialliset tiedot, ja ne laskettiin yhteen minuutin ajalta ja lähetettiin koneoppimismalliin yhdessä Bitcoinin edellisen minuutin hinnan ja liukuvan keskihinnan kanssa. Se ennusti seuraavan minuutin Bitcoin-hinnan annettujen tietojen perusteella. Todellisen hinnan saavuttua laskettiin RMS-arvo ja koneoppimismalli päivitti itseään ennustaakseen seuraavan arvon paremmalla ymmärryksellä.

**Tulos**

Mitä kokeellista arviointia käytetään?

**Esimerkki 5.1623**

Käytämme sanakohdistuksia, kuten muissakin annotaatioiden projisointityössä, projisoidaksemme AMR-kohdistukset kohdekieliin. Lähestymistapamme perustuu erääseen perusolettamukseen: jos lähdesana on sanavastaavuudeltaan kohdesanan mukainen ja se on AMR-vastaavuudeltaan AMR-solmun mukainen, myös kohdesana on vastaavuudeltaan AMR-solmun mukainen. Sanakohdistukset luotiin fast\_align BIBREF10 -ohjelmalla, kun taas AMR-kohdistukset luotiin JAMR BIBREF11 -ohjelmalla.

**Tulos**

Miten annotaatioprojektio tehdään, kun kielillä on erilainen sanajärjestys?

**Esimerkki 5.1624**

Suoritamme ensin kokeen BiLSTM:llä, BiLSTM-CNN:llä, BiLSTM-CRF:llä ja BiLSTM-CNN-CRF:llä käyttäen taulukossa TABREF30 mainittuja hyperparametreja.

**Tulos**

Mitä koneoppimismalleja ne tutkivat?

**Esimerkki 5.1625**

Otamme käyttöön Amazonin Mechanical Turk -palvelun avulla toteutetun BIBREF-annotaatiokoneiston5 ja annotoimme jokaisen predikaatin kahdella koulutetulla työntekijällä itsenäisesti, kun taas kolmas yhdistelee heidän annotaationsa lopulliseksi rooli- ja argumenttikokonaisuudeksi. Tässä yhdistämistehtävässä työntekijä validoi kysymykset, yhdistää, jakaa tai muuttaa saman roolin vastauksia ohjeiden mukaisesti ja poistaa turhat roolit valitsemalla luonnollisemmin muotoillut kysymykset.

**Tulos**

Mitä erilaista parannetussa annotaatioprotokollassa on?

**Esimerkki 5.1626**

ILP-pohjaisen lähestymistapamme tuottamia tiivistelmiä verrattiin vastaaviin manuaalisiin tiivistelmiin käyttämällä ROUGE BIBREF22 unigram-pistemäärää.

**Tulos**

Mitä arviointimittareita käytettiin tiivistämistehtävässä?

**Esimerkki 5.1627**

Espanjan kieli (SPA) on sitä vastoin morfologisesti rikas, ja siinä on paljon laajemmat verbiparadigmat kuin englannissa. Englannin tavoin se on suffiksoiva kieli, ja lisäksi se käyttää sisäisiä kantamuutoksia (esim. o $\rightarrow $ ue). Valitsimme zulun (ZUL), joka on bantoidinen kieli. Kahdesta ensimmäisestä kielestä poiketen se on vahvasti prefiksoiva. Toiseksi, HUN:lla esivalmennettu järjestelmä, joka toimi hyvin, viittaa jälleen kerran siihen, että lähdekielestä, jonka morfologia on agglutinatiivinen eikä fuusionaalinen, näyttää olevan myös hyötyä. TUR ja HUN saavat jälleen korkean tarkkuuden, mikä on lisäindikaattori hypoteesillemme, jonka mukaan lähdekieli, jossa on agglutinatiivinen morfologia, helpottaa taivutuksen oppimista toisessa kielessä.

**Tulos**

Käytetäänkö agglutinatiivisia kieliä sekä prefiksoivien että suffiksoivien kielten ennustamisessa?

**Esimerkki 5.1628**

Kokeissa tarkastellaan neljää tietoaineistoa, joista kaksi on äskettäin luotuja ja loput kaksi jo julkisia: CNN, TIME, 20 Newsgroups ja Reuters-21578. Koodi ja kaksi uutta tietoaineistoa ovat saatavilla osoitteessa github.com/baiyangwang/emgan. Kaikkien asiakirjojen esikäsittelyä varten muutimme kaikki merkit pieniksi kirjaimiksi, poistimme asiakirjojen kantoja ja ajoimme word2vec-mallin jokaiselle korpukselle saadaksemme sanojen upotukset, joiden koko on 300. Kaikissa myöhemmissä malleissa otamme huomioon vain yleisimmät INLINEFORM0-sanat kaikissa aineiston korpuksissa.

**Tulos**

Mitä korporaatioita he käyttävät?

**Esimerkki 5.1629**

Tutkiaksemme, voiko BERT-pohjainen mallimme siirtää tietoa kielen ulkopuolelta, pidämme kuvaominaisuuksia yksinkertaisina visuaalisina merkkeinä, jotka voidaan esittää mallille analogisesti tekstimerkkien upotusten kanssa. Jotta $o\_j$-vektorit (ulottuvuus $2048+4=2052$) olisivat vertailukelpoisia BERT-merkkiutumisten (ulottuvuus 768) kanssa, käytämme yksinkertaista lineaarista ristikkäismodaalista projektointikerrosta $W$, jonka ulottuvuudet ovat $2052\hspace{-1.00006pt}\times \hspace{-1.00006pt}768$. Kuvassa havaitut $N$ kohdealueet esitetään siten muodossa $X\_{img} = (W.o\_1,...,W.o\_N)$. Kun kuva on kuvattu BERT:n sulautusavaruuteen $W$:n avulla, muu malli näkee kuvan yksikköjen sarjana, jossa ei ole selvää merkintää siitä, onko kyseessä tekstin vai kuvan sulautus.

**Tulos**

Miten multimodaaliset representaatiot yhdistetään?

**Esimerkki 5.1630**

GP-GNN:t rakentavat ensin täysin yhdistetyn graafin $\mathcal {G} = (\mathcal {V}, \mathcal {E})$ , jossa $\mathcal {V}$ on olioiden joukko ja jokainen reuna $(v\_i, v\_j) \in \mathcal {E}, v\_i, v\_j \in \mathcal {V}$ vastaa tekstistä poimittua sekvenssiä $s = x\_0^{i,j}, x\_1^{i,j}, \dots , x\_{l-1}^{i,j}}$.

**Tulos**

Tässä asiakirjassa siis muutetaan jäsentymätön tekstisisääntulo parametreiksi, joita GNN:t voivat lukea?

**Esimerkki 5.1631**

Twiittien tilastollinen analyysi ::: Tutkimuksessa on tarkasteltu kolmea sanamerkkien muotoa eli unigram-, bigram- ja trigram-muotoa.

**Tulos**

Mitkä sanafrekvenssit heijastavat kirjoittajien mukaan twitterin käyttäjien psykologiaa?

**Esimerkki 5.1632**

Taulukossa TABREF23 esitetään Pearsonin, Spearmanin ja Kendallin korrelaatiot Rougen ja Seran sekä pyramidipisteiden välillä. Mielenkiintoista on havaita, että monilla Rouge-pisteiden muunnelmilla ei ole korkeita korrelaatioita ihmisten pyramidipisteiden kanssa. Alhaisimmat F-pistekorrelaatiot ovat Rouge-1:llä ja Rouge-L:llä (INLINEFORM0 =0,454). Rouge-1:n heikko korrelaatio osoittaa, että ehdokasyhteenvedon ja kultaisten yhteenvetojen välisten unigrammien yhteensovittaminen ei ole tarkka yhteenvedon laadun kvantifioinnissa.

**Tulos**

Mitä erilaisia korrelaatioita saadaan, kun käytetään ROUGE-pisteiden eri muunnelmia?

**Esimerkki 5.1633**

Kieliopin matemaattisen rakenteen tutkiminen on osoittanut, että lauseiden perustana eivät ole sanat vaan jotkin atomiset kieliopilliset tyypit, kuten substantiivi- ja lausetyyppi BIBREF23 , BIBREF24 , BIBREF25 . Transitiiviverbityyppi ei ole atomaattinen kieliopillinen tyyppi, vaan kahdesta substantiivityypistä ja yhdestä lausetyypistä koostuva komposiitti. Näin ollen erityisen mielenkiintoista tässä on se, että atomisella ei todellakaan tarkoiteta pienintä....

**Tulos**

Väittävätkö he, että kaikki sanat voidaan johtaa muista (alku)sanoista?

**Esimerkki 5.1634**

Tehtävän yksinkertaistamiseksi laadimme binäärisen luokittelun: ehdokkaat, joista on pidetty tai jotka on valittu valintalistalle, katsotaan kuuluviksi haluttavaan luokkaan ja muut kuuluviksi ei-haluttavaan luokkaan.

**Tulos**

Miten "palkattavuus" määritellään?

**Esimerkki 5.1635**

Tutkimme vastausten leksikaalista ja semanttista monimuotoisuutta tekemällä kolme analyysia. Ensinnäkin yhdistimme kaikki työntekijän vastaukset tiettyyn kysymykseen yhdeksi kysymystä vastaavaksi luetteloksi. Toiseksi verrattiin yksittäisten vastausten monimuotoisuutta kunkin kysymyksen kohdalla valvonta- ja AUI-järjestelmien välillä. Kysymyksen monimuotoisuuden mittaamiseksi laskimme vastausten määrän jaettuna kyseiseen kysymykseen annettujen yksittäisten vastausten määrällä. Kutsumme tätä arvoa vastaustiheydeksi.

**Tulos**

Miten leksikaalista monimuotoisuutta mitattiin?

**Esimerkki 5.1636**

Tulokset ja analyysi

**Tulos**

mitä kvantitatiivista analyysia tehdään?

**Esimerkki 5.1637**

Jaettua tehtävää varten tarjotaan tasapainoinen tietokokonaisuus, joka koostuu 2 396 ironisesta ja 2 396 ei-ironisesta twiitistä. Ironinen korpus muodostettiin keräämällä itse kommentoituja twiittejä, joissa oli hashtagit #ironia, #sarkasmi ja #ei. Tämän jälkeen twiitit puhdistettiin ja tarkistettiin ja merkittiin manuaalisesti käyttäen hienojakoista annotaatiojärjestelmää BIBREF3 .

**Tulos**

Mikä on tietokokonaisuuden koko?

**Esimerkki 5.1638**

Menetelmät ::: BIBREF15, BIBREF16. Jaamme ensin harjoituslauseparit kolmeen ryhmään kohde- ja lähteen pituuden suhteen mukaan (merkkeinä). Ihannetapauksessa haluamme ryhmän, jossa kohde on lyhyempi kuin lähde (lyhyt), ryhmän, jossa ne ovat samankokoisia (normaali), ja viimeisen ryhmän, jossa kohde on pidempi kuin lähde (pitkä). Käytännössä valitsemme kaksi kynnysarvoa $t\_\text{min}$ ja $t\_\text{max}$ pituussuhdejakauman mukaan. Kaikki lauseparit, joiden pituussuhde on välillä $t\_\text{min}$ ja $t\_\text{max}$, kuuluvat normaaliryhmään, ne, joiden pituussuhde on alle $t\_\text{min}$, lyhyisiin ja loput pitkiin. Harjoittelun aikana liitämme jokaiseen lähdelauseeseen pituusmerkin sen ryhmän mukaan ($<$lyhyt$>$, $<$normaali$>$ tai $<$pitkä$>$), jotta yksi verkko voi erottaa ryhmät toisistaan (ks. kuva KUVA 2). Päättelyhetkellä pituusmerkin avulla verkko saa aikaan käännöksen, joka kuuluu haluttuun pituusryhmään.

**Tulos**

Miten ne ehdollistavat tuotoksen tiettyyn kohde-lähde-luokkaan?

**Esimerkki 5.1639**

Vaikka jokainen kaupunkikuvaus on suhteellisen lyhyt, Calvinon kirjoitus on täynnä harvinaisia sanoja, monimutkaisia syntaktisia rakenteita ja kuvakieltä. Kunkin kaupungin olennaisten osatekijöiden vangitseminen yhteen vektoriin ei siis ole yhtä yksinkertaista kuin tavanomaisempien tekstimuotojen kohdalla. Toivomme kuitenkin, että miljardien sanojen tekstin avulla koulutettujen kielimallien avulla voidaan poimia näistä kuvauksista mielekästä semantiikkaa. Kokeilemme kolmea erilaista esivalmennettua representaatiota: ELMo BIBREF5 , BERT BIBREF6 ja GloVe BIBREF18 . Yhden kaupungin upottamisen tuottamiseksi laskemme TF-IDF-painotetun elementtiviisauden keskiarvon merkkitason representaatioista. Kaikissa esivalmennetuissa menetelmissä pienennämme lisäksi kaupunkikuvioiden ulottuvuuden 40:een PCA:n avulla, jotta ne olisivat yhteensopivampia klusterointialgoritmimme kanssa.

**Tulos**

Miten he mallintavat kaupunkikuvauksen käyttämällä upotuksia?

**Esimerkki 5.1640**

Teemme kokeita (§ SECREF5 ) kahdella englannin ja japanin välisellä käännöskorpuksella arvioidaksemme menetelmän hyödyllisyyttä käännöstarkkuuden parantamisessa ja harjoitteluun tarvittavan ajan lyhentämisessä.

**Tulos**

Mitä kielipareja he kokeilivat?

**Esimerkki 5.1641**

Kuvassa 1 esitetään yksinkertainen lineaarinen malli, jossa on sijoitusrajoitus.

**Tulos**

Mitkä ovat niiden perusmenetelmät?

**Esimerkki 5.1642**

Tietokanta sisältää noin 6 000 kolmikkoa, jotka koostuvat videoista, kysymyksistä ja vastauksista, jotka on kerätty manuaalisesti kuvankäsittelyohjelmiston opetusvideoista, joissa on puhuttuja selostuksia.

**Tulos**

Millaisia opetusvideoita tietokannassa on?

**Esimerkki 5.1643**

Vaikka neuraalinen syntaksista riippumaton järjestelmä on parempi L1-tiedoissa, molemmat syntaksiin perustuvat järjestelmät tuottavat parempia analyyseja L2-tiedoissa. Lisäksi, kuten eri jäsentäjien välinen vertailu osoittaa, mitä parempia jäsentämistuloksia saamme, sitä paremman suorituskyvyn saavutamme L2-tiedoissa. Tämä osoittaa, että syntaktinen jäsennys on tärkeää semanttisen rakentamisen kannalta oppijan kiinan kielessä. Analyysimme mukaan tärkein syy on se, että syntaksiin perustuva järjestelmä voi tuottaa oikeita syntaktisia analyysejä osittaisille kieliopillisille fragmenteille L2-teksteissä, mikä tarjoaa ratkaisevaa tietoa SRL:n kannalta.

**Tulos**

Vihjailevatko kirjoittajat, miksi syntaktinen jäsennys on niin tärkeää kieltenvälisten kielten semanttisessa roolimerkinnässä?

**Esimerkki 5.1644**

Nämä toimet koostuvat seuraavista osatekijöistä:[leftmargin=\*]Yhteisviittausten ratkaiseminen: Joskus on tarpeen käyttää rinnakkaisviittausten ratkaisutekniikoita tehokkaan haun mahdollistamiseksi, jotta voidaan tukea usean käännöksen vuorovaikutusta. Kyselyn luominen: Tämä komponentti luo kyselyn käyttäjän ja järjestelmän aiempien vuorovaikutusten perusteella. Hakumalli: Tämä on keskeinen luokittelukomponentti, joka hakee asiakirjoja tai kohtia suuresta kokoelmasta. Tulosten tuottaminen: Haetut asiakirjat voivat olla liian pitkiä esitettäväksi joidenkin käyttöliittymien avulla.

**Tulos**

Mitä eri moduuleja Macawissa on?

**Esimerkki 5.1645**

Teemme kokeita SQuAD-tietokannalla BIBREF3.

**Tulos**

Millaisilla tietokokonaisuuksilla kokeita tehdään?

**Esimerkki 5.1646**

Käytämme kahta lähdettä e-kirjojen merkintätiedoista: (i) toimittajan tunnisteita ja (ii) Amazonin hakusanoja. Toimittajatunnisteita varten keräämme 48 705 e-kirjan tiedot 13 kustantajalta, jotka ovat Kunstmann, Delius-Klasnig, VUR, HJR, Diogenes, Campus, Kiwi, Beltz, Chbeck, Rowohlt, Droemer, Fischer ja Neopubli. Amazonin hakusanoja varten keräämme 21 243 e-kirjan hakukyselylokit 12 kuukauden ajalta (eli marraskuusta 2017 lokakuuhun 2018).

**Tulos**

mitä tietokokonaisuutta käytettiin?

**Esimerkki 5.1647**

Mallinsimme sanojen määrän ja kahden käyttäjän sitoutumista kuvaavan mittarin (kokonaisarvosana, keskimääräinen kierrosten määrä) välistä suhdetta erillisillä lineaarisilla regressioilla.Tulokset osoittivat, että käyttäjät, jotka tuottivat keskimäärin enemmän sanoja sisältäviä lauseita, antoivat merkittävästi korkeampia arvosanoja ($\beta $=0.01, SE=0.002, t=4.79, p$<$0.001) (ks. kuvio 2) ja osallistuivat Gunrockiin huomattavasti useamman kierroksen ajan ($\beta $=1.85, SE=0.05, t=35.58, p$<$0.001) (ks. kuvio 2). Nämä tulokset voidaan tulkita todisteeksi Gunrockin kyvystä käsitellä monimutkaisia lauseita, joissa käyttäjät eivät ole sidottuja yksinkertaisiin vastauksiin tullakseen ymmärretyksi ja tunteakseen olevansa mukana keskustelussa - ja todisteeksi siitä, että yksilöt ovat tyytyväisempiä keskusteluun, kun he ottavat aktiivisemman roolin sen sijaan, että järjestelmä hallitsisi vuoropuhelua. Toisaalta toinen tulkinta on se, että puhelias käyttäjät saattavat nauttia keskustelusta botin kanssa yleensä ja siten antaa korkeampia arvosanoja yhdessä korkeamman keskimääräisen sanamäärän kanssa.

**Tulos**

Miten he suhteuttavat käyttäjien taustatietokyselyt käyttäjien tyytyväisyyteen?

**Esimerkki 5.1648**

On yleisesti tunnettu tosiasia, että NLP-sovellukset teollisuudessa joutuvat usein käsittelemään meluisia tietoja. Mahdollisia häiriöitä ovat muun muassa ei-kanoninen teksti, kuten kirjoitusvirheet, typografiset virheet, puhekieliset ilmaisut, lyhenteet, slangi, internet-jargon, hymiöt, upotetut metatiedot (kuten hashtagit, URL-osoitteet, maininnat), epätyypilliset syntaktiset rakenteet ja oikeinkirjoitusvariantit, kieliopillisesti virheellinen teksti sekä kahden tai useamman kielen sekoitus. Tällaiset meluiset tiedot ovat tyypillisiä käyttäjien tuottamalle tekstisisällölle, ja niitä esiintyy yleisesti esimerkiksi sosiaalisessa mediassa, keskusteluissa, verkkoarvosteluissa ja verkkofoorumeilla.

**Tulos**

Minkälaista kohinaa esiintyy tyypillisissä teollisuustiedoissa?

**Esimerkki 5.1649**

Tässä artikkelissa tarjoamme uudenlaisen reaaliaikaisen ja mukautuvan kryptovaluutan hinnan ennustamisalustan, joka perustuu Twitter-tunnelmiin. Integroiva ja modulaarinen alusta selviytyy kolmesta edellä mainitusta haasteesta monin tavoin. Ensinnäkin se tarjoaa Spark-pohjaisen arkkitehtuurin, joka käsittelee suuren määrän saapuvaa dataa pysyvällä ja vikasietoisella tavalla. Toiseksi ehdotettu alusta tarjoaa lähestymistavan, joka tukee VADERiin perustuvaa tunneanalyysia, joka pystyy vastaamaan suuriin määriin luonnollisen kielen käsittelykyselyjä reaaliajassa. Kolmanneksi alusta tukee verkko-oppimiseen perustuvaa ennakoivaa lähestymistapaa, jossa koneoppimismalli mukauttaa painojaan uusien hintojen ja tunteiden mukaan. Lopuksi, alusta on modulaarinen ja integroiva siinä mielessä, että se yhdistää nämä eri ratkaisut tarjotakseen uudenlaisen reaaliaikaisen työkalutuen bitcoinin hinnan ennustamiseen, joka on skaalautuvampi, tietorikkaampi ja ennakoivampi ja joka voi auttaa nopeuttamaan päätöksentekoa, paljastamaan uusia mahdollisuuksia ja tarjoamaan oikea-aikaisempia oivalluksia saatavilla olevan ja yhä suuremman taloudellisen tiedon määrän ja monipuolisuuden perusteella.

**Tulos**

Mitkä alustan osat ovat modulaarisia?

**Esimerkki 5.1650**

Käytämme valvomatonta sanojen segmentointimenetelmää latticelm, jolla voidaan segmentoida sanoja suoraan lausuttujen lauseiden puheentunnistustulosten BIBREF22 ristikoista.

**Tulos**

Mitä menetelmää he käyttävät sanojen segmentointiin?

**Esimerkki 5.1651**

Esimerkkinä INLINEFORM0-pisteytyksestä arvioimme ja vertailimme kahden eri SBD-järjestelmän suorituskykyä YouTube-videoiden joukossa moniviiteympäristössä. Ensimmäinen järjestelmä (S1) käyttää konvolutiivista neuroverkkoa määrittämään, vastaako liukuikkunan keskimmäinen sana SU-rajaa vai ei BIBREF30 . Toisessa lähestymistavassa (S2) sen sijaan otetaan käyttöön kaksisuuntainen rekursiivinen neuroverkkomalli, jossa on huomiomekanismi rajojen havaitsemiseen BIBREF31 .

**Tulos**

Mitä SBD-järjestelmiä he vertasivat?

**Esimerkki 5.1652**

Käytämme Gal & Ghahramani galin pudotustekniikkaa perustasona, koska se on kaikkein samankaltaisin pudotustekniikka kuin meidän lähestymistapamme, ja nimitämme sitä VBD:ksi (variational binary dropout).

**Tulos**

Mikä on binäärinen variaatiopudotus?

**Esimerkki 5.1653**

Laadun valvomiseksi varmistimme, että yksi kommentoija kommentoi enintään 120 uutisotsikkoa (tämä suojaa kommentoijia lukemasta liikaa uutisotsikoita ja hallitsemasta annotaatioita). Toiseksi annoimme tehtävään osallistua vain sellaisia annotoijia, jotka asuvat maantieteellisesti Yhdysvalloissa.Testasimme annotoijia 1100 dollarin testikysymysten joukolla ensimmäisessä vaiheessa (noin 10 prosenttia aineistosta) ja 500:lla toisessa vaiheessa. Annotoijien oli läpäistävä 95 prosenttia.

**Tulos**

Miten merkintöjen laatua mitataan?

**Esimerkki 5.1654**

Menetelmämme perustuu sosiaalipsykologiseen tutkimukseen, erityisesti moraalin perusteita koskevaan teoriaan BIBREF26. MFT pyrkii selittämään ihmisen moraalin rakennetta ja vaihtelua eri kulttuureissa, ja siinä ehdotetaan viittä moraalista perustaa: Huolehtiminen / vahingoittaminen, oikeudenmukaisuus / huijaaminen, uskollisuus / pettäminen, auktoriteetti / kumoaminen ja pyhyys / rappio. Kukin perusta tiivistyy positiiviseen ja negatiiviseen napaan, jolloin saadaan kymmenen hienojakoista moraalikategoriaa.

**Tulos**

Mitä hienojakoisia esimerkkejä moraalisesta ulottuvuudesta ne esittelevät?

**Esimerkki 5.1655**

Saadaksemme suuren määrän brasilialaisen musiikin sanoituksia loimme indeksoijan, joka navigoi Vagalume-sivustolla ja poimi kustakin musiikinlajista kaikki kappaleet kaikilta luetelluilta tekijöiltä.

**Tulos**

mikä on laulun sanoituksen lähde?

**Esimerkki 5.1656**

Logistinen regressiomalli, jossa käytetään merkkitason n-grammiominaisuuksia, esitetään vahvana vertailukohtana, koska se on osoittautunut erittäin tehokkaaksi.

**Tulos**

Mikä on heidän perustasonsa?

**Esimerkki 5.1657**

Aiomme soveltaa semanttista aukkotelineistöä uutisten tiivistämiseen. Voimme erityisesti merkitä kriittisiä entiteettejä, kuten henkilöiden tai paikkojen nimiä, varmistaaksemme, että ne on otettu oikein huomioon luodussa tiivistelmässä. Aiomme myös kerätä ihmisen ja ihmisen välisen dialogin tietokokonaisuuden, jossa on monipuolisempia ihmisen kirjoittamia tiivistelmiä.

**Tulos**

Odotetaanko, että puhujan rooli, semanttinen aukko ja dialogialueen annotaatiot löytyvät reaalimaailman tietokannoista?

**Esimerkki 5.1658**

Forexiin, erityisesti EUR:iin ja USD:hen, liittyvät twiitit hankittiin Twitterin hakusovellusliittymän kautta seuraavalla kyselyllä: "EURUSD", "USDEUR", "EUR" tai "USD". Kolmen vuoden aikana (tammikuusta 2014 joulukuuhun 2016) kerättiin lähes 15 miljoonaa twiittiä. Osajoukko niistä (44 000 twiittiä) leimattiin manuaalisesti asiantuntevien rahoitusalan opiskelijoiden toimesta.

**Tulos**

Kuinka monta twiittiä merkittiin manuaalisesti?

**Esimerkki 5.1659**

Kokoelmassa on lähes 1,1 miljoonaa lausetta. Erilaisia relaatiotyyppejä (ainutlaatuisia predikaatteja) on 119, ja niissä on muutamasta suhteesta muutamaan miljoonaan suhteeseen.

**Tulos**

Kuinka suuri on heidän tietokokonaisuutensa?

**Esimerkki 5.1660**

Kokeissamme käytämme syötteenä Stanford Sentiment Treebank -testijoukon BIBREF2 2210 tokenisoitua lausetta, jotka esikäsittelemme pienentämällä ne, kuten BIBREF8:ssa tehtiin.

**Tulos**

Mitä tietokokonaisuuksia käytetään arvioinnissa?

**Esimerkki 5.1661**

Tässä työssä koulutamme kaksikielisiä UWS-malleja, joissa käytetään uhanalaista kieltä Mboshi kohdekielenä ja erilaisia hyvin resursoituja kieliä yhdenmukaistettuna tietona. Tulokset osoittavat, että samankaltaiset kielet sijoittuvat paremmin segmentointisuorituskyvyn suhteen ja että yhdistämällä eri malleilla opittua tietoa segmentointi paranee entisestään. Tämä saattaa johtua erilaisista kieliriippuvaisista rakenteista, jotka saadaan selville käyttämällä useampaa kuin yhtä kieltä.

**Tulos**

Miten hyvin resursoitu kieli vaikuttaa tuotoksen laatuun?

**Esimerkki 5.1662**

Käytämme samaa perustasoa kuin lang2011unsupervised, jota on osoittautunut vaikeaksi päihittää. Tämä lähtötaso määrittää konstituentille semanttisen roolin sen syntaktisen funktion perusteella eli riippuvuussuhteen perusteella sen päähän.

**Tulos**

Mitä perusmallia käytetään?

**Esimerkki 5.1663**

Espanjan serigos2017using poimi argentiinalaisista sanomalehdistä anglismeja yhdistämällä sanakirjahakuja (TreeTaggerin ja NLTK-lemmatisaattorin avulla), suuraakkosellisten sanojen automaattista suodatusta ja manuaalista tarkastusta. Serigos2017applying-moduulissa lisättiin merkkien n-gram-moduuli, jolla arvioitiin todennäköisyys, että sana on englanninkielinen tai espanjankielinen. moreno2018configuracion käytti erilaisia hahmontäsmäytyssuodattimia ja sanakirjan etsintää poimiakseen anglismikadidaatteja yhdysvaltalaisista espanjankielisistä twiiteistä koostuvasta korpuksesta.

**Tulos**

Mainitaanko artikkelissa muita teoksia, joissa ehdotetaan menetelmiä espanjan kielen anglismien havaitsemiseksi?

**Esimerkki 5.1664**

Esimerkiksi alkuperäinen lause `Kävimme ostoksilla lauantaina' ja korjattu versio `Kävimme ostoksilla lauantaina' tuottavat seuraavan mallin: (VVD shop\_VV0 II, VVD shopping\_VVG II)Kun taustakorpuksesta on kerätty tilastot, virheet voidaan lisätä virheettömään tekstiin. Opitut kuviot käännetään nyt päinvastaisiksi ja etsitään syötelauseesta tuplan oikea puoli. Käytämme vain kuvioita, joiden frekvenssi on INLINEFORM0 , jolloin harjoitusaineistostamme saadaan yhteensä 35 625 kuviota.

**Tulos**

Mitä tekstikuvioita poimitaan?

**Esimerkki 5.1665**

Tässä artikkelissa analysoimme jatkuvien lauseiden automaattista korjaamista. Kehitämme kaksi menetelmää: ehdollisen satunnaiskentän mallin (roCRF) ja Seq2Seq-huomiointimallin (roS2S), ja osoitamme, että ne päihittävät mallit, jotka ovat peräisin sisartehtävistä, kuten välimerkkien palauttamisesta ja koko lauseen kielioppivirheiden korjaamisesta.

**Tulos**

Millaisia koneoppimismalleja he käyttävät korjatakseen loputetut lauseet?

**Esimerkki 5.1666**

Teemme myös kaksi inhimillistä arviointia arvioidaksemme, a) minkä tyyppistä tiivistelmää osallistujat suosivat (vertailemme uuttuvia ja abstrahoivia järjestelmiä) ja b) kuinka paljon asiakirjan keskeistä tietoa tiivistelmässä säilyy (pyydämme osallistujia vastaamaan asiakirjan sisältöön liittyviin kysymyksiin lukemalla järjestelmän tiivistelmiä).

**Tulos**

Käyttävätkö he ROUGEn lisäksi muita arviointimittareita?

**Esimerkki 5.1667**

Kieliopin induktiomenetelmä perustuu opetussuunnitelman oppimiseen BIBREF7 , jossa jäsentäjä oppii ensin jäsentämään yksinkertaisia lauseita ja sitten oppimaan monimutkaisempia lauseita. Induktiomenetelmä on iteratiivinen, puoliautomaattinen ja perustuu usein esiintyviin kuvioihin. Tekstistä indusoidaan kontekstivapaa kielioppi (CFG), jota edustavat useat kerrokset semanttisia merkintöjä.

**Tulos**

Miten he saivat aikaan CFG:n?

**Esimerkki 5.1668**

Käytämme BIBREF5-menetelmää neuraalisten espanjan ja englannin ST-mallien kouluttamiseen sekvenssistä sekvenssiin. Kultaisten teemamerkintöjen saaminen aineistoomme vaatisi huomattavaa manuaalista annotaatiota, joten sen sijaan käytämme 1K (train20h) harjoitusjoukon lausumien ihmiskäännöksiä NMF-teemamallin kouluttamiseen scikit-learn BIBREF14 -ohjelmalla.

**Tulos**

Mikä on mallin arkkitehtuuri?

**Esimerkki 5.1669**

Vaikka kirosanat ovat yleisiä twiiteissä, ne edustavat vain 1,15 prosenttia kaikista käytetyistä sanoista BIBREF21 . Sitä vastoin havaitsimme, että 5,72 prosenttia kaikista jengin jäsentilien julkaisemista sanoista oli luokiteltu kirosanoiksi, mikä on lähes viisi kertaa enemmän kuin keskimääräinen kirosanojen käyttö Twitterissä. Pilvet heijastavat myös sitä, että jengiläiset puhuvat usein huumeista ja rahasta termeillä kuten smoke, high, hit ja money, kun taas tavalliset käyttäjät eivät juuri puhu taloudesta ja huumeista. Huomasimme myös, että jengiläiset puhuvat aineellisista asioista termeillä, kuten got, money, make, real, need, kun taas tavalliset käyttäjät pyrkivät ilmaisemaan tunteitaan termeillä, kuten new, like, love, know, want, look, make, us.

**Tulos**

Mitä eroja kielenkäytössä on jengiläisten ja muun Twitter-väestön välillä?

**Esimerkki 5.1670**

Twitter tarjoaa hyvin dokumentoidun API:n, jonka avulla voi pyytää mitä tahansa tietoja twiiteistä, käyttäjistä ja heidän profiileistaan sekä nopeusrajoituksista. On olemassa erityyppinen API, jota kutsutaan Streaming API:ksi ja joka tarjoaa reaaliaikaisen twiittien virran. Tärkein ero tavalliseen API:hin on se, että yhteys pidetään yllä mahdollisimman pitkään ja twiitit lähetetään reaaliajassa asiakkaalle. Streaming API:ssa on kolme meitä kiinnostavaa päätepistettä: "sample", "filter" ja "firehose". Ensimmäinen tarjoaa näytteen (satunnainen osajoukko) koko twiittivirrasta. Toisen avulla voidaan vastaanottaa twiittejä, jotka vastaavat joitakin hakukriteerejä: ne vastaavat yhtä tai useampaa hakusanaa, ovat osajoukon käyttäjien tuottamia tai ovat peräisin tietystä maantieteellisestä sijainnista. Viimeinen tarjoaa koko twiittien joukon, vaikka se ei olekaan oletusarvoisesti käytettävissä. Meidän tapauksessamme yksinkertaisin lähestymistapa olisi käyttää "sample"-päätepistettä, mutta se tarjoaa twiittejä kaikilla mahdollisilla kielillä kaikkialta maailmasta, kun taas meitä kiinnostaa vain yksi kieli (venäjä). Tämän päätepisteen käyttöä varten toteutimme suodatuksen kielen perusteella. Suodatin on yksinkertainen: jos twiitti ei sisällä vähintään kolmen kyrillisen symbolin muodostamaa merkkijonoa, se katsotaan ei-venäläiseksi. Vaikka tällä lähestymistavalla säilytetään mongolian-, ukrainan- ja muiden slaavilaiskielten twiitit (koska niissä käytetään kyrillisiä aakkosia), väärien positiivisten tulosten kokonaismäärä on tässä tapauksessa häviävän pieni. Osoittaaksemme tämän teimme yksinkertaisen kokeen: satunnaisesta 200 twiitin otoksesta vain 5 twiittiä oli muulla kuin venäjän kielellä. Jotta emme luottaisi Twitterin kielentunnistukseen, päätimme käyttää tätä kielipohjaista suodatusmenetelmää. "Näyte"-päätepisteen kautta saatujen twiittien määrä ei kuitenkaan ollut tyydyttävä. Tämä johtuu luultavasti siitä, että "sample"-päätteestä lähetetään aina samaa sisältöä kaikille sen asiakkaille, ja pieni osa siitä on venäjänkielistä. Jotta venäjänkielisten twiittien louhinta voitaisiin pakottaa, valitsimme "filter"-päätteen, joka vaatii jonkinlaisen hakukyselyn. Rakensimme heuristisen kyselyn, joka sisälsi joitakin venäjänkielisiä apusanoja: konjunktioita, pronomineja ja prepositioita. Täydellinen luettelo on seuraava: russian я, у, к, в, по, на, ты, мы, до, на, она, он, и, да.

**Tulos**

Mitä Twitter-korpusta käytettiin sanavektorien kouluttamiseen?

**Esimerkki 5.1671**

Arvioidaksemme korpuksen hyödyllisyyttä SMT-tarkoituksiin käytimme sitä automaattisen kääntäjän kouluttamiseen Moses BIBREF8 -ohjelmalla.

**Tulos**

Mitä SMT-malleja he tarkastelivat?

**Esimerkki 5.1672**

Seuraavassa esitellään vaiheet, joita tarvitaan toisiinsa liittyvien tapahtumien tietämysgraafin kehittämiseksi. Kuvassa FIGREF2 esitetään koko prosessin yleiskatsaus. Tämä putki sisältää seuraavat päävaiheet, joita käsitellään yksityiskohtaisesti myöhemmin. (1) Twiittien kerääminen useiden uutiskanavien, kuten BBC:n ja CNN:n, Twitter-virrasta. (2) Taustatietomallista sopiminen. (3) Tapahtumien annotointi sisältää mahdollisesti kaksi osatehtävää (i) tapahtumien tunnistaminen ja (ii) tapahtumien luokittelu. (4) Entiteettien/suhteiden annotointi käsittää mahdollisesti joukon tehtäviä, joita ovat (i) entiteettien tunnistaminen, (ii) entiteettien linkittäminen, (iii) entiteettien disambiguointi, (iv) entiteettien semanttisten roolien merkitseminen ja (v) implisiittisten entiteettien päätteleminen. (5) Tapahtumien linkittäminen toisiinsa yli ajan ja median rajojen. (6) Tapahtumatietograafin julkaiseminen linkitetyn avoimen datan parhaiden käytäntöjen pohjalta.

**Tulos**

Mistä uutisorganisaatioista otsikot ovat peräisin?

**Esimerkki 5.1673**

Puheen sisältö saatiin YouTuben ASR-järjestelmän tarjoamista kuvateksteistä (CC), joita voidaan käyttää tekstityksinä.

**Tulos**

Mitä ASR-järjestelmää he käyttävät?

**Esimerkki 5.1674**

Jotta voimme mukauttaa FactorCell BIBREF4:n tarkoituksiimme, korvaamme käyttäjän upotukset matalaulotteisella kuvaesityksellä. Näin pystymme muokkaamaan jokaisen kyselyn täydennyksen yksilölliseksi tietylle kuvaesitykselle.

**Tulos**

Miten he täydentävät käyttäjän kyselyn etuliitteen, jonka ehtona on kuva?

**Esimerkki 5.1675**

Rakennamme ja testaamme MMT-mallejamme Multi30K-tietokannassa BIBREF21 . Jokainen Multi30K:n kuva sisältää yhden englanninkielisen (EN) kuvauksen, joka on otettu Flickr30K:sta BIBREF22 , sekä ihmisen tekemät käännökset saksaksi (DE), ranskaksi (FR) ja tšekiksi BIBREF23 , BIBREF24 , BIBREF25 . Tietokokonaisuus sisältää 29 000 instanssia harjoittelua varten, 1 014 instanssia kehittämistä varten ja 1 000 instanssia testausta varten. Kokeilemme vain saksan- ja ranskankielisiä aineistoja, sillä ne ovat kieliä, joista meillä on sisäistä asiantuntemusta esiteltävän tyyppistä analyysia varten. Virallisen Multi30K-testijoukon (test 2016) lisäksi käytämme myös viimeisimmän WMT-arviointikilpailun testijoukkoa, test 2018 BIBREF25 .

**Tulos**

Ilmoitetaanko tulokset vain englanninkielisestä aineistosta?

**Esimerkki 5.1676**

Weibo valvoo tiukasti BIBREF0-palvelussaan julkaistavaa sisältöä yhteistyössä vallanpitäjien kanssa. Zhu et al. zhu-etal:2013 mukaan Weibo käyttää sensuroitaviin viesteihin kohdistamiseen erilaisia strategioita, jotka vaihtelevat avainsanalistan suodattamisesta yksittäisten käyttäjien seurantaan. Kaikista lopulta sensuroiduista viesteistä lähes 30 prosenttia sensuroidaan 5-30 minuutin kuluessa ja lähes 90 prosenttia 24 tunnin kuluessa BIBREF1. Oletamme, että edelliset tapahtuvat automaattisesti, kun taas jälkimmäiset poistetaan ihmisen suorittaman sensuurin avulla.

**Tulos**

Tiedetäänkö, sensuroivatko Sina Weibo -viestit ihmiset vai jokin automaattinen luokittelija?

**Esimerkki 5.1677**

Taulukossa TABREF16 esitetään yleiskatsaus toimitetusta tietokokonaisuudesta. Huomaa, että käytettävissä oleva tietomäärä vaihtelee kielittäin.

**Tulos**

Onko tietokokonaisuus tasapainossa kaikkien neljän kielen kesken?

**Esimerkki 5.1678**

aika: vuosikymmen (1900-luvun ja 2010-luvun väliset luokat) ja vuosi, jotka edustavat aikaa, jolloin genrestä tuli valtavirta.

**Tulos**

Mitä vuosikymmeniä he tarkastelivat?

**Esimerkki 5.1679**

Ensimmäinen dataerä koostuu lumimyrskyihin, hurrikaaneihin ja maastopaloihin liittyvistä twiiteistä sillä rajoituksella, että niitä twiittaavat "vaikutusvaltaiset" twiittaajat, jotka määrittelemme henkilöiksi, joilla on varmasti luokiteltavissa oleva tunne kyseisestä aiheesta. Oletamme esimerkiksi, että kaikki Al Goren laatimat ilmastonmuutosta koskevat twiitit ovat positiivinen näyte, kun taas kaikki salaliittotili @ClimateHiJinxin twiitit ovat negatiivinen näyte. Seuraavissa menetelmissä tekemämme oletus (joka on vahvistettu kohtuulliseksi SECREF2 -jaksossa) on, että vaikutusvaltaisia twiittaajia voidaan käyttää twiittien merkitsemiseen massoittain, kun manuaalisesti merkittyjä twiittejä ei ole.

**Tulos**

Millä menetelmällä kompensoidaan vähäiset merkityt tiedot?

**Esimerkki 5.1680**

Korpuskokeilujen perusteella saamme hyviä tuloksia SarcasmCorpus-korpuksella, joka on ainoa korpus, joka sisältää Amazonin arvosteluja. Valitettavasti, kun koulutamme mallejamme ristikkäiskorporaatiossa tai kaikissa korpuksissa, tuloksemme heikkenevät dramaattisesti, erityisesti ristikkäiskorporaatiossa. Nämä tulokset tarkoittavat, että SarcasmCorpusin sarkasmi välittyy sellaisten piirteiden kautta, joita ei ole muissa korpuksissa.

**Tulos**

Millä aloilla sarkasmia välitetään eri tavoin?

**Esimerkki 5.1681**

Kaikista kolmesta esiharjoittelutehtävästä SR toimii hieman paremmin kuin kaksi muuta tehtävää (NSG ja MDG).

**Tulos**

Mikä kolmesta esivalmennustehtävästä on hyödyllisin?

**Esimerkki 5.1682**

Olemme keränneet tähän mennessä yli 3 934 610 miljoonaa twiittiä.

**Tulos**

Kuinka suuri tietokokonaisuus on?

**Esimerkki 5.1683**

tietylle kysymykselle vastaamisen kannalta merkitykselliset tosiseikat) merkitään harjoittelun aikana.

**Tulos**

Mitä tarkoittaa tosiasioiden valvonnan tukeminen?

**Esimerkki 5.1684**

Merkkijonojen ytimet ovat tapa käyttää tietoa merkkitasolla mittaamalla merkkijonojen samankaltaisuutta merkkien n-grammien avulla.

**Tulos**

Mikä on merkkijonon ydin?

**Esimerkki 5.1685**

BioASQ-järjestäjät toimittavat koulutus- ja testausdatan. Harjoitusaineisto koostuu kysymyksistä, kultaisista standardidokumenteista, pätkistä, käsitteistä ja ideaalivastauksista (joita emme käyttäneet tässä asiakirjassa, mutta käytimme viime vuonna BIBREF2:ta). Testidata on jaettu vaiheisiin A ja B. Vaiheen A tietokokonaisuus koostuu kysymyksistä, yksilöllisistä tunnuksista ja kysymystyypeistä. Vaiheen B tietokokonaisuus koostuu kysymyksistä, kultaisista standardidokumenteista, pätkistä, yksilöllisistä tunnuksista ja kysymystyypeistä.

**Tulos**

Mitä tietokokonaisuutta he käyttivät?

**Esimerkki 5.1686**

Vuoteen 2015 verrattuna meidän QA-SRL-kultamme saatiin 1593 annotaatiota, joista 989 oli ydin- ja 604 liitetietoa, kun taas QA-SRL-kulta tuotti 1315 annotaatiota, 979 ydin- ja 336 liitetietoa.

**Tulos**

Kuinka suuri tietokokonaisuus on?

**Tulos**

Kuinka paljon suurempi kattavuus on uudessa tietokokonaisuudessa?

**Esimerkki 5.1687**

Käsittelemme pääasiassa kahta seuraavaa upotusavaruuden rakennetta.Semanttinen samankaltaisuusrakenne: Semanttisesti samankaltaiset entiteetit ovat lähellä toisiaan upotusavaruudessa ja päinvastoin. Semanttinen suuntarakenne: Upotusavaruudessa on semanttisia suuntia, joiden avulla vain yksi semanttinen aspekti muuttuu, kun taas kaikki muut aspektit pysyvät samoina. Se voidaan tunnistaa vektorierolla, kuten kahden upotusvektorin välisellä vähennyksellä.

**Tulos**

Mitkä ovat upotusavaruuden oudot semanttiset rakenteet?

**Esimerkki 5.1688**

Esimerkiksi kysymys ja apulause korreloivat keskenään, koska englannin kielen päälauseen kysymykset edellyttävät subjekti-aux-inversiota ja monissa tapauksissa apulauseen do (Do lions meow?) lisäämistä. Ekspletiivit eli "tyhjät" argumentit ovat semanttisesti inerttejä argumentteja. Yleisimmät ekspletiivit englannissa ovat it ja there, vaikka kaikki näiden kohtien esiintymät eivät olekaan ekspletiivejä.

**Tulos**

Raportoidaanko tulokset vain englanninkielisistä tiedoista?

**Esimerkki 5.1689**

Taulukossa TABREF1 esitetään yhteenveto edellä mainittujen aiempien tutkimusten määrällisistä tuloksista. Taulukossa TABREF1 esitetään yhteenveto edellä mainittujen aiempien tutkimusten määrällisistä tuloksista.

**Tulos**

Mikä on uusimpien menetelmien ilmoittama tarkkuus?

**Esimerkki 5.1690**

Kaikki korporaatiot tarjoavat tietokokonaisuuksia/jakoja vastausten valintaa varten, kun taas vain (WikiQA, SQuAD) ja (WikiQA, SelQA) tarjoavat tietokokonaisuuksia vastausten poimimista ja vastausten käynnistämistä varten. SQuAD on kooltaan paljon suurempi, vaikka tämän korpuksen kysymykset ovat usein moninkertaisesti muotoiltuja. Päinvastoin, SQuADin keskimääräinen ehdokkaiden määrä kysymystä kohti ( INLINEFORM0 ) on pienin, koska SQuAD poimii vastausehdokkaat kappaleista, kun taas muut poimivat ne osioista tai tietolaatikoista, jotka koostuvat suuremmista yhteyksistä. Vaikka InfoboxQA on suurempi kuin WikiQA tai SelQA, InfoboxQA:n merkkityyppien määrä ( INLINEFORM1 ) on pienempi kuin näiden kahden, mikä johtuu infolaatikoiden toistuvasta luonteesta.Kaikkien korporaatioiden vastausehdokkaiden keskimääräiset pituudet ( INLINEFORM0 ) ovat samankaltaisia, paitsi InfoboxQA:ssa, jossa jokaista infolaatikon riviä pidetään ehdokkaana. SelQA:ssa ja SQuAD:ssa kysymysten keskimääräiset pituudet ( INLINEFORM1 ) ovat samankaltaisia, koska niiden merkintäjärjestelmät ovat samankaltaisia. Ei ole yllättävää, että WikiQA:n kysymysten keskimääräinen pituus on pienin, kun otetaan huomioon, että kysymykset on otettu hakukyselyistä. InfoboxQA:n keskimääräinen kysymyksen pituus on suhteellisen pieni, mikä johtuu siitä, että infoboxeista voidaan kysyä vain rajoitetusti tietoa. InfoboxQA:lla ja WikiQA:lla on vähiten kysymys-vastaus-sanojen päällekkäisyyksiä kysymyksissä ja vastauksissa ( INLINEFORM2 ja INLINEFORM3 taulukossa TABREF2 ). Päällekkäisten sanojen F1-pistemäärän ( INLINEFORM4 ) osalta SQuAD antaa vähiten päällekkäisyyksiä kysymys-vastaus -parien välillä, vaikka WikiQA on hyvin lähellä.

**Tulos**

Miten he analysoivat kontekstuaalisia samankaltaisuuksia eri tietokokonaisuuksien välillä?

**Esimerkki 5.1691**

Mallin kestävyyden arvioimiseksi laadimme testijoukon, joka koostuu "vastakkaisista" esimerkeistä eli häiriintyneistä esimerkeistä, jotka voivat mahdollisesti muuttaa perusmallin ennustetta. Tällaisia esimerkkejä voivat olla esimerkiksi lauseen parafraasointi, esimerkiksi leksikaaliset ja syntaktiset muutokset. Käytämme kahta kirjallisuudessa kuvattua lähestymistapaa: takaisinkääntämistä ja kohinaista sekvenssin automaattista kooderia. Huomattakoon, että nämä esimerkit muistuttavat mustan laatikon hyökkäyksiä, mutta niitä ei ole tarkoituksellisesti suunniteltu huijaamaan järjestelmää, ja siksi käytämme termiä "vastahyökkäys" laajasti. Käytämme näitä tekniikoita tuottaaksemme monia parafraaseja ja löydämme osajoukon lausumia, jotka ovat hyvin samankaltaisia kuin alkuperäinen testisarja, mutta joiden ennusteet ovat vääriä. Mittaamme mallin kestävyyttä tällaisia muutoksia vastaan.

**Tulos**

Miten kirjoittajat luovat vastakohtaisen testijoukon mallin kestävyyden mittaamiseksi?

**Esimerkki 5.1692**

CFQ sisältää suuruusluokaltaan eniten kyselymalleja ja myös huomattavasti enemmän kyselyitä ja kysymyksiä kuin muut tietokokonaisuudet.

**Tulos**

Miten kirjoittajat perustelevat, että esitetty kysymysvastaustietokanta on realistinen?

**Esimerkki 5.1693**

Arvioimme malliamme ja perusmalleja kolmella versiolla aineistosta. The first one (Inc) only considers the original data, containing naturally incorrect tweets, and achieves accuracy of 80$\%$ against BERT's 72$\%$. Toisessa versiossa (Corr) otetaan huomioon korjatut twiitit, ja sen tarkkuus on suurempi, koska se on vähemmän kohinainen.

**Tulos**

Pitäisikö heidän lähestymistapaansa soveltaa vain silloin, kun on kyse epätäydellisistä tiedoista?

**Esimerkki 5.1694**

SG-mallin keskimääräinen samankaltaisuuspistemäärä oli korkea, 0,650, ja seuraavaksi korkein oli CBoW, jonka keskimääräinen samankaltaisuuspistemäärä oli 0,632. Myös GloVe saavutti huomattavan keskimääräisen pistemäärän 0,591. SdfastTextin keskimääräinen samankaltaisuuspistemäärä on kuitenkin 0,388, ja sanaparia Microsoft-Bill Gates ei löydy SdfastTextin sanastosta. Lisäksi taulukossa TABREF78 on esitetty maiden ja niiden pääkaupunkien semanttisen sukulaisuuden keskimääräinen samankaltaisuuspistemäärä englanninkielisen käännöksen kanssa, jossa SG:n keskimääräinen pistemäärä on myös paras, 0,663, ja sen jälkeen CBoW:n pistemäärä on 0,611. GloVe tuottaa myös paremman semanttisen samankaltaisuuden (0,576) ja SdfastTextin (0,391).

**Tulos**

Miten ehdotetut sanojen upotukset vertautuvat Sindhi fastText -sanojen esityksiin?

**Esimerkki 5.1695**

Kullekin kieliparille tarkastellaan kahta eri mallia: perusmallia ja asiakirjamallia. Arvioimme niitä kolmella testijoukolla ja raportoimme BLEU- ja TER-pisteet. Kaikki kokeet suoritetaan 8 kertaa eri siemenillä, ja raportoimme kunkin kokeen keskiarvotulokset ja p-arvot.

**Tulos**

Mitä arviointimittareita he käyttivät?

**Esimerkki 5.1696**

Keräsimme Simple English Wikipediasta vapaasti saatavilla olevan yksinkertaistetun aineiston, jota on aiemmin käytetty monissa tekstin yksinkertaistamismenetelmissä BIBREF0 , BIBREF10 , BIBREF3 .

**Tulos**

Mihin kieleen tässä asiakirjassa keskitytään?

**Esimerkki 5.1697**

Luomme uuden manuaalisesti annotoidun multimodaalisen vihapuheaineiston, joka koostuu 150 000 dollarin twiiteistä, joista jokainen sisältää tekstiä ja kuvan.

**Tulos**

Kuinka suuri tietokokonaisuus on?

**Esimerkki 5.1698**

Kolme annotoijaa, A1-A3, merkitsevät jokaisen datasetin H twiitin humalaiseksi tai selväksi.

**Tulos**

Tarkistetaanko kaukovalvonnassa hankitut tiedot jossain vaiheessa ihmisten toimesta?

**Esimerkki 5.1699**

Ehdotetun menetelmän suorituskyvyn arvioimiseksi NLI-tehtävässä käytetään SNLI BIBREF22- ja MultiNLI BIBREF23 -tietoaineistoja. Käytämme Quora Question Pairs -tietokokonaisuutta BIBREF24 arvioidessamme menetelmämme suorituskykyä PI-tehtävässä. Tunteiden luokittelun suorituskyvyn arvioinnissa käytetään Stanford Sentiment Treebank (SST) BIBREF25 -tietokantaa.

**Tulos**

Mitä tietokokonaisuuksia käytettiin?

**Esimerkki 5.1700**

Tämän jälkeen lasketaan käsitteiden upotukset (GloVe) erikseen haastattelukuvauksille ja tutkintokuvauksille. Yksinkertaisin tapa tuottaa tekstin upotuksia termien upotusten perusteella on käyttää jonkinlaista termien upotusten aggregaatiota, kuten keskiarvoa. Tätä lähestymistapaa testattiin esimerkiksi BIBREF21:ssä ja BIBREF13:ssa. BIBREF22 laski termien upotusten painotetun keskiarvon rakentamalla häviöfunktion ja harjoittelemalla painotukset gradienttilaskeutumismenetelmällä.Käyntien lopulliset upotukset saadaan yhdistämällä erikseen haastattelulle ja lääkärintarkastukselle lasketut keskimääräiset upotukset, ks. kuva KUVA FIGREF4 .

**Tulos**

Millaisia sanasulkeumia he käyttävät edustamaan lääkärikäyntejä?

**Esimerkki 5.1701**

Tämän huomioimiseksi laskemme tasapainotetun tarkkuuden eli kunkin luokan kolmen tarkkuuden keskiarvon.

**Tulos**

Arvioidaanko mallin selityksen tulosta, mitä mittaria käytettiin?

**Esimerkki 5.1702**

Kun ennustetaan vihaa, iloa tai valenssia, niiden järjestelmien määrä, jotka johdonmukaisesti antavat korkeammat pisteet lauseille, joissa on naispuolisia substantiivilausekkeita (21-25), on selvästi suurempi kuin niiden järjestelmien määrä, jotka antavat korkeammat pisteet lauseille, joissa on miespuolisia substantiivilausekkeita (8-13). (Muistutettakoon, että korkeampi valenssi tarkoittaa positiivisempaa tunnetta.) Sitä vastoin pelkotehtävässä useimmat järjestelmät antoivat korkeammat pisteet lauseille, joissa oli miespuolisia substantiivilausekkeita (23), verrattuna niiden järjestelmien määrään, jotka antoivat korkeammat pisteet lauseille, joissa oli naispuolisia substantiivilausekkeita (12). Surua ennustettaessa niiden järjestelmien määrä, jotka antoivat korkeammat pisteet naispuolisten substantiivilauseiden lauseille (18), on lähellä niiden järjestelmien määrää, jotka antoivat korkeammat pisteet miespuolisten substantiivilauseiden lauseille (16). Nämä tulokset ovat linjassa joidenkin yleisten stereotypioiden kanssa, kuten että naiset ovat tunteikkaampia ja tilanteet, joissa miespuoliset toimijat ovat mukana, ovat pelokkaampia BIBREF27 Suurin osa järjestelmistä antoi korkeammat pisteet lauseille, joissa oli afroamerikkalaisia nimiä, vihaa, pelkoa ja surua ennustettaessa. Iloa ja valenssia koskevissa tehtävissä useimmat järjestelmät antoivat korkeammat pisteet lauseille, joissa oli eurooppalaisamerikkalaisia nimiä. Nämä suuntaukset heijastavat joitakin yleisiä stereotypioita, joiden mukaan afroamerikkalaiset liitetään kielteisempiin tunteisiin BIBREF28 .

**Tulos**

Mikä rotu ja sukupuoli saavat korkeamman tunteen voimakkuusennusteen?

**Esimerkki 5.1703**

Rinnakkainen skannauksen päättelySemiring-algoritmien kommutatiiviset ominaisuudet mahdollistavat joustavuuden siinä järjestyksessä, jossa laskemme $A(\ell )$. Dynaamisen ohjelmoinnin algoritmien tyypilliset toteutukset ovat järjestyksen pituuden suhteen sarjamuotoisia. Vektoroitu jäsennysLaskennan monimutkaisuus on vielä suurempi ongelma jäsennysalgoritmeille, joita ei voi yhtä helposti rinnakkaistaa. Semiring-matriisioperaatiotKaksi edellistä optimointia vähentävät suurimman osan semiring-matriisikertolaskennan kustannuksista. Erityistapauksessa $(\sum , \times )$ semiringin tapauksessa nämä voidaan laskea erittäin tehokkaasti käyttämällä matriisikertolaskentaa, joka on hyvin viritetty GPU-laitteistoon. Valitettavasti muiden semiringien, kuten login ja maxin, tapauksessa nämä operaatiot ovat joko hitaita tai hyvin muistitehottomia. Esimerkiksi matriiseille $T$ ja $U$, joiden koko on $N \times M$ ja $M \times O$, voimme lähettää $\otimes $:lla tensorin, jonka koko on $N \times M \times O$, ja sitten pienentää dim $M$:n $\bigoplus $:lla, mikä maksaa valtavasti muistia.

**Tulos**

Mitä lähtötasoja kokeissa käytetään?

**Esimerkki 5.1704**

Kutsumme tätä elinikäiseksi vuorovaikutteiseksi oppimiseksi ja päättelyksi (LiLi). Elinikäinen oppiminen näkyy siinä, että äskettäin hankitut tiedot säilytetään tietovarastossa ja niitä käytetään tulevien kyselyjen päättelyssä ja että päivitetyn tietovaraston lisäksi kertynyttä tietoa, mukaan lukien aiemmat päättelytulokset, hyödynnetään tulevan vuorovaikutuksen ja oppimisen ohjaamisessa.

**Tulos**

Millä tavoin LiLi jäljittelee sitä, miten ihmiset hankkivat tietoa ja tekevät päätelmiä vuorovaikutteisen keskustelun aikana?

**Esimerkki 5.1705**

Tavoitteenamme on irrottaa sisällön valinta dekooderista ottamalla käyttöön ylimääräinen sisällönvalitsin. Intuitiivisin tapa on kouluttaa sisällönvalitsin kohdentamaan jotakin heuristisesti poimittua sisältöä. Voimme esimerkiksi kouluttaa valitsijan valitsemaan päällekkäiset sanat lähteen ja kohteen välillä BIBREF6 , lauseet, joilla on korkeampi tf-idf-pistemäärä BIBREF20 , tai tunnistetut kuvakohteet, jotka esiintyvät kuvatekstissä BIBREF21 . INLINEFORM0 latenttina muuttujana: Toinen tapa on käsitellä INLINEFORM1:tä latenttina muuttujana ja kouluttaa valitsija ja generaattori yhdessä maksimoimalla marginaalisen datan todennäköisyys. Tällä tavoin valitsijalla on mahdollisuus tutkia automaattisesti optimaalisia valintastrategioita, jotka sopivat parhaiten vastaavaan generaattorikomponenttiin. Reinforce-select (RS) BIBREF24 , BIBREF9 käyttää vahvistusoppimista marginaalisen todennäköisyyden lähentämiseen. Ehdotamme Variational Reinforce-Select (VRS) -menetelmää, jossa sovelletaan variationaalista päättelyä BIBREF10 varianssin vähentämiseksi.

**Tulos**

Miten malli koulutetaan tekemään vain sisällön valinta?

**Esimerkki 5.1706**

Luokittelijoita arvioitaessa käytetään yleisesti tarkkuutta, täsmällisyyttä ja palautusta sekä Hammingin häviötä. On osoitettu, että laskettaessa tarkkuutta ja palautusta monimerkkiluokittelijoille voi olla edullista käyttää mikrokeskiarvoista tarkkuutta ja palautusta BIBREF6 . Mikrokeskiarvoistetun tarkkuuden kaavat ovat seuraavat: DISPLAYFORM0 DISPLAYFORM1

**Tulos**

mitä arviointimittareita käsitellään?

**Esimerkki 5.1707**

Yksikielisestä mallista tehdään kopiot kutakin kieltä varten ja lisätään kieltenvälisiä latentteja muuttujia (CLV) yksikielisten mallien yhdistämiseksi, jolloin ne kuvaavat kieltenvälisiä semanttisia roolimalleja. Konkreettisesti, kun harjoitellaan rinnakkaisia lauseita, aina kun argumenttien pääsanat ovat samansuuntaisia, lisäämme CLV:n kahden vastaavan roolimuuttujan vanhemmaksi.

**Tulos**

Lisäävätkö he Bayesin malliin yhden latentin muuttujan kutakin kieliparia varten?

**Esimerkki 5.1708**

Olemme samoilla linjoilla kuin äskettäisessä BIBREF-työssä16 , jossa ehdotetaan arvioinnin siirtämistä absoluuttisista arvoista tutkivampiin arviointeihin, joissa keskitytään sulautumien heikkouksiin ja vahvuuksiin eikä niinkään yleisiin pisteytyksiin. Yksi mittari voisi esimerkiksi olla sen tarkistaminen, ovatko kaikki sanat, joiden tiedetään kuuluvan samaan luokkaan, lähempänä toisiaan kuin eri luokkiin kuuluvat sanat, riippumatta todellisesta kosinuksesta.

**Tulos**

Mitä uusia mittareita ehdotetaan edistymisen seuraamiseksi?

**Esimerkki 5.1709**

Visual Question Generation (VQG) on toinen nouseva aihealue, jossa pyritään esittämään kysymyksiä kuvan perusteella. BIBREF10 ehdotti tämän innoittamana avointa VQG:tä, jonka tarkoituksena on luoda luonnollisia ja kiinnostavia kysymyksiä kuvasta.

**Tulos**

Tekevätkö he kyselytutkimuksen visuaalisen kysymyksenasettelun?

**Esimerkki 5.1710**

Tutkijat ovat kiinnittäneet yhä enemmän huomiota emojien laajamittaiseen käyttöön BIBREF4 , BIBREF5 , koska emojit välittävät hedelmällistä semanttista ja tunnepohjaista tietoa täydentämään visuaalisesti tekstimuotoista tietoa, mikä on erittäin hyödyllistä tekstiin sisältyvien tunnesignaalien ymmärtämisessä BIBREF6 . Aiemmassa kirjallisuudessa ei kuitenkaan ole otettu huomioon emojien kielellistä monimutkaisuutta ja monimuotoisuutta. Tämän vuoksi aiemmat emojien upotusmenetelmät eivät pysty käsittelemään tilannetta, jossa opittujen emojien upotusten semantiikka tai tunteet ovat ristiriidassa vastaavista yhteyksistä saatavan tiedon kanssa BIBREF5 , tai kun emojit välittävät useita semanttisia ja tunnetason merkityksiä, kuten ( ja ). Käytännössä emojit voivat joko tiivistää ja korostaa kontekstiensa alkuperäistä sävyä tai ilmaista monimutkaisempaa semantiikkaa, kuten ironiaa ja sarkasmia, yhdistämällä ne konteksteihin, joiden semantiikka tai tunteet ovat ristiriitaisia. Perinteinen emoji-analyysi voi poimia kustakin emojista vain yhden upotuksen, ja tällaiset upotukset hämmentävät seuraavaa tunneanalyysimallia, koska syötetekstien ja emojien tunnesignaalit ovat epäjohdonmukaisia.

**Tulos**

Mikä on motivaatio bi-sense embedding -koulutukselle?

**Esimerkki 5.1711**

Tulosten analysoimiseksi päätimme käyttää BIBREF10:n tarjoamaa testiä, joka koostuu 19 eri luokkaan jaetuista 19 791 $ analogioista: (8915 analogiaa) ja 13 "syntaktiseen" makroalueeseen (10876 analogiaa). Kaikki analogiat koostuvat kahdesta sanaparista, joilla on yhteinen suhde, joka on kaavamaisesti ilmaistu yhtälöllä: $a:a^{\*}=b:b^{\*}$ (esim. "mies : nainen = kuningas : kuningatar"); missä $b^{\*}$ on arvottava sana ("kuningatar"), $b$ on siihen kytketty sana ("kuningas"), $a$ on poistettavien komponenttien sana ("mies") ja $a^{\*}$ on lisättävien komponenttien sana ("nainen").

**Tulos**

Testataanko sanasulkeumia NLP-tehtävässä?

**Esimerkki 5.1712**

Oikeudenmukaisen vertailun varmistamiseksi koulutimme kutakin mallia samalla 10 miljoonan lauseen korpuksella, joka oli kerätty Wikipediasta.

**Tulos**

Millä aineistolla (aineistoilla) he laskevat sanojen upotukset?

**Esimerkki 5.1713**

Strategiapohjaiset menetelmät poikkeavat esivalmennusvaiheesta ja pyrkivät hyödyntämään esivalmennettuja malleja kohdetehtävän oppimisprosessin aikana. Näihin lähestymistapoihin kuuluvat hienosäätöaikataulut, joissa oppimisnopeuden ohjausta suunnitellaan tarkkaan optimointia varten, sijaistehtävät, joissa käytetään merkittyjä tietoja, jotta esivalmennettu malli sopisi paremmin kohdetiedon jakaumaan, ja tiedon tislausta koskevat lähestymistavat, joissa luovutaan esivalmistettujen parametrien avulla tapahtuvasta alustamisesta omaksumalla esivalmennettu malli opettajaverkoksi.

**Tulos**

Miten strategiapohjaiset menetelmät käsittelevät NLG:n esteitä?

**Esimerkki 5.1714**

Sekä POS-taggausta että riippuvuuksien jäsentämistä varten suoritamme kokeita Penn Treebankin Wall Street Journal (WSJ) -osuudella.

**Tulos**

Arvioidaanko ne vain englanninkielisillä tietokokonaisuuksilla?

**Tulos**

Millaisia tietokokonaisuuksia he arvioivat?

**Esimerkki 5.1715**

Seuraavaksi tarkastelemme seuraavaa ongelmaa: kun on annettu kaksi BabelNet-luokkaa $A$ ja $B$, ennustetaan, ovatko ne todennäköisesti käsitteellisiä naapureita niiden tekstikorpuksen lauseiden perusteella, joissa ne molemmat mainitaan. Tällaisen luokittelijan kouluttamiseksi käytämme harjoitusdatana SECREF8-jaksosta saatuja etävalvontalappuja. Kun tämä luokittelija on koulutettu, voimme käyttää sitä ennustamaan sellaisten luokkien käsitteellistä naapuruutta, joista tiedetään vain vähän tapauksia. Löytääksemme lauseet, joissa sekä $A$ että $B$ mainitaan, käytämme yksiselitteistä tekstikorpusta, jossa BabelNet-kategorioiden maininnat on nimenomaisesti merkitty. Tällainen yksiselitteinen korpus voidaan rakentaa automaattisesti esimerkiksi BIBREF30 mancini-etal-2017-embeddingin ehdottamien menetelmien kaltaisilla menetelmillä. Jokaisen ehdokasluokkaparin osalta haemme siis kaikki lauseet, joissa ne esiintyvät yhdessä. Seuraavaksi esitämme jokaisen poimitun lauseen vektorina.

**Tulos**

Miten he tunnistavat käsitteelliset naapurit?

**Esimerkki 5.1716**

Kuten taulukosta TABREF46 käy ilmi, konsensusfuusio parantaa NDCG:n pistemäärää noin 1,0:lla yhteisen mallin pistemäärään verrattuna, mutta antaa silti vertailukelpoiset pisteet muille mittareille. Kuten taulukosta TABREF46 käy ilmi, ensemble-malli näyttää ottavan parhaat tulokset kustakin mallista.

**Tulos**

Kumpi integraatiomenetelmä toimii paremmin ensemble- tai konsensusfuusio jaetuilla parametreilla?

**Esimerkki 5.1717**

Käytimme annotaatioiden tekemiseen Amazon Mechanical Turk (AMT) -palvelua ja rajasimme tehtävän koskemaan työntekijöitä viidessä englanninkielisessä maassa (Yhdysvallat, Yhdistynyt kuningaskunta, Kanada, Uusi-Seelanti ja Australia), joilla oli yli 1000 valmista HIT:tä ja vähintään 95 prosentin hyväksymisprosentti.

**Tulos**

Mitä joukkoistamisalustaa he käyttivät?

**Esimerkki 5.1718**

Ennustustehtävämme on siis suoraviivainen sanatason binääriluokittelutehtävä. Kehitämme seuraavat viisi piirreryhmää, jotka kuvaavat ominaisuuksia, jotka liittyvät siihen, miten sanaa käytetään selityksessä (ks. taulukko TABREF18 täydellisen luettelon osalta):[itemsep=0pt,leftmargin=\*,topsep=0pt]Sanan kontekstittomat ominaisuudet. Nämä ominaisuudet johdetaan suoraan sanasta, ja ne kuvaavat yleistä taipumusta siihen, että sana kaikuu selityksissä. sanan käyttö OP:ssa tai PC:ssä (kaksi ryhmää). Nämä piirteet kuvaavat sitä, miten sanaa käytetään OP:ssa tai PC:ssä. Näin ollen jokaiselle piirteelle on kaksi arvoa OP:lle ja PC:lle.Miten sana yhdistää OP:n ja PC:n. Nämä piirteet tarkastelevat sanan käytön eroa OP:ssa ja PC:ssä. Odotamme, että tämä ryhmä on tehtävämme kannalta tärkein. yleiset OP/PC-ominaisuudet. Nämä piirteet kuvaavat keskustelun yleisiä ominaisuuksia. Niitä voidaan käyttää luonnehtimaan kaikujen taustajakaumaa.Taulukossa TABREF18 esitetään lisäksi intuitio kunkin ominaisuuden sisällyttämiselle ja tiivistetyt $t$-testitulokset Bonferroni-korjauksen jälkeen. Erityisesti testataan, onko selityksissä kaikuuntuneilla sanoilla erilaiset ominaisuusarvot kuin niillä sanoilla, joita ei kaikuuntunut. Kaikkien sanojen tarkastelun lisäksi tarkastelemme myös erikseen stop-sanoja ja sisällön sanoja kuvion FIGREF8 valossa. Korostamme tässä muutamia havaintoja:[itemsep=0pt,leftmargin=\*,topsep=0pt]Vaikka odotamme, että monimutkaisempia sanoja (#merkkejä) kaikuisi useammin, näin ei kuitenkaan keskimäärin ole. Havaitsemme myös mielenkiintoisen esimerkin Simpsonin paradoksista Wordnetin syvyyttä BIBREF38 koskevissa tuloksissa: matalammat sanat kaikuvat todennäköisemmin kaikissa sanoissa, mutta syvemmät sanat kaikuvat todennäköisemmin sisältö- ja stop-sanoissa.

**Tulos**

Mitä ominaisuuksia ehdotetaan?

**Esimerkki 5.1719**

Sen sijaan analyysimme paljastaa, että tämä käyttäytyminen johtuu käännösprosessista itsestään johtuvista hienovaraisista artefakteista. Erityisesti osoitamme, että kunkin instanssin eri osien kääntäminen erikseen (esim. premissi ja hypoteesi NLI:ssä) voi muuttaa aineiston pinnallisia malleja (esim. niiden välisen leksikaalisen päällekkäisyyden astetta), mikä vaikuttaa vakavasti nykyisten mallien yleistämiskykyyn. Saatujen oivallusten perusteella parannamme XNLI:n nykytilaa ja osoitamme, että joitakin aiempia havaintoja on harkittava uudelleen tämän ilmiön valossa. Vertailuanalyysi koostuu pätevyystestistä, jossa arvioidaan kykyä ymmärtää antonyymisuhde ja suorittaa numeerista päättelyä, häiriötestistä, jossa arvioidaan kestävyyttä matalien mallien, kuten leksikaalisen päällekkäisyyden ja kieltosanojen esiintymisen, suhteen, ja kohinatestistä, jossa arvioidaan kestävyyttä kirjoitusvirheiden suhteen. Aiempien kokeiden tapaan raportoimme tulokset kunkin testisarjan parhaasta epookin tarkistuspisteestä.

**Tulos**

Mitkä ovat esimerkkejä näistä keinotekoisista toimista?

**Esimerkki 5.1720**

Kokeissamme babidatasetissa ehdotetut Memory Neural Networks (MemNNs) toimivat perustasona.

**Tulos**

Mitä sanatason ja merkkitason mallien perustasoja käytetään?

**Esimerkki 5.1721**

Verrattaessa yhtälöä DISPLAY\_FORM14 yhtälöön DISPLAY\_FORM22 voidaan havaita, että yhtälö DISPLAY\_FORM14 on itse asiassa $F1$:n pehmeä muoto, jossa käytetään jatkuvaa $p$:tä binäärisen $\mathbb {I}( p\_{i1}>0.5)$:n sijasta. Tämä ero ei ole suuri ongelma tasapainoisissa aineistoissa, mutta se on erittäin haitallinen, jos suuri osa harjoitusesimerkkejä on helposti negatiivisia: helposti negatiiviset esimerkit voivat helposti hallita harjoittelua, koska niiden todennäköisyydet voidaan melko helposti painaa nollaan. Samaan aikaan malli tuskin pystyy erottamaan vaikeasti negatiivisia esimerkkejä positiivisista, millä on valtava negatiivinen vaikutus lopulliseen F1-suorituskykyyn.Tämän ongelman ratkaisemiseksi ehdotamme, että pehmeä todennäköisyys $p$ kerrotaan pienenevällä kertoimella $(1-p)$, jolloin yhtälö DISPLAY\_FORM22 muuttuu seuraavaan muotoon: Voidaan ajatella, että $(1-p\_{i1})$ on kuhunkin esimerkkiin liitetty painoarvo, joka muuttuu harjoittelun edetessä. Kun $p\_{i1}$ muutetaan $(1-p\_{i1}) p\_{i1}$:ksi, tarkoituksena on alentaa helppojen esimerkkien painoa. Helppojen esimerkkien, joiden todennäköisyys lähestyy 0:a tai 1:tä, $(1-p\_{i1}) p\_{i1}$ saa mallin painottamaan niitä huomattavasti vähemmän. Kuvassa KUVA23 annetaan selitys derivaatan näkökulmasta: $\frac{(1-p)p}{1+(1-p)p}$:n derivaatta $p$:n suhteen lähestyy 0:aa heti, kun $p$ lähestyy 0:aa, mikä tarkoittaa, että malli kiinnittää vähemmän huomiota esimerkkeihin sen jälkeen, kun ne on luokiteltu oikein.Yhtälön Eq.DISPLAY\_MUOTO14 tarkempi tarkastelu paljastaa, että se itse asiassa jäljittelee ajatusta fokusmenetyksestä (focal loss, lyhyesti FL) BIBREF16 kohteiden havaitsemisessa näkökentässä. Fokaalihäviötä ehdotettiin yksivaiheista objektinilmaisinta varten, jotta voidaan käsitellä harjoittelun aikana esiintyvää etualan ja taustan välistä kompromissia. Se pienentää hyvin luokitelluille esimerkeille määritettyä tappiota lisäämällä $(1-p)^{\beta }$-kertoimen, jolloin lopullinen tappio on $(1-p)^{\beta }\log p$.

**Tulos**

Miten painotuksia säädetään dynaamisesti?

**Esimerkki 5.1722**

Suoritamme myös manuaalisen arvioinnin, joka osoittaa En-It-tehtävän kohdalla lievää laadun heikkenemistä, minkä vastineeksi keskimääräinen pituussuhde pienenee tilastollisesti merkittävästi, 1,05:stä 1,01:een.

**Tulos**

Suorittavatko he minkäänlaista ihmisarviointia?

**Esimerkki 5.1723**

MSA:n osalta hankimme diakritisoidun korpuksen, jota käytettiin RDI BIBREF7 -diakritisaattorin ja Farasa BIBREF31 -diakritisaattorin kouluttamiseen. Korpus sisältää 9,7 miljoonaa merkkiä, joissa on noin 194 000 ainutlaatuista pintamuotoa (lukuun ottamatta numeroita ja välimerkkejä). Testaukseen käytimme vapaasti saatavilla olevaa WikiNews-testijoukkoa BIBREF31, joka koostuu 70 MSA WikiNews-artikkelista (18 300 tokenia) ja kattaa tasaisesti eri genrejä, kuten politiikka, talous, terveys, tiede ja teknologia, urheilu, taide ja kulttuuri. CA:ta varten saimme kirjankustantajalta suuren kokoelman täysin diakritisoituja klassisia tekstejä (2,7 miljoonaa merkkiä), ja pidimme testausta varten poissa pienen 5 000 lauseen (noin 400 000 sanaa) osajoukon.

**Tulos**

mitä tietokokonaisuuksia käytettiin?

**Esimerkki 5.1724**

Sanakohdistusten lisääminen rinnakkaisiin lauseisiin johtaa pieniin, ei merkittäviin parannuksiin, vaikka lähdekielestä olisi saatavilla jonkin verran merkittyä tietoa. Tämä vaikeus osoittaa rinnakkaisten korpusten hyödyllisyys SRI:n kannalta voi johtua nykyisistä oletuksista, jotka koskevat roolien kohdistuksia ja joiden mukaan vain pieni osa rooleista on kohdistettu. Lisäanalyysit paljastavat, että pienten tietomäärien merkitsemisellä voidaan helposti saavuttaa parempi suorituskyky kuin lisäämällä suuria merkitsemättömiä tietokokonaisuuksia ja rinnakkaisia korpuksia.

**Tulos**

Parantaako rinnakkainen data kaiken kaikkiaan semanttisten roolien induktiota useilla kielillä?

**Esimerkki 5.1725**

Koulutus ja testaus suoritetaan vuorotellen: Kussakin epookissa LSTM-verkolle esitetään koulutusta varten ensin 1000 näytettä tietyllä kielellä, jotka on luotu tietyn diskreetin todennäköisyysjakauman mukaan, joka on tuettu suljetulle äärelliselle aikavälille. Tämän jälkeen jäädytämme kaikki painot mallissamme, luettelemme tyhjentävästi kaikki kielen jaksot niiden pituuksien mukaan ja määrittelemme ensimmäiset $k$ lyhimmät jaksot, joiden tulosteet malli tuottaa epätarkasti. kokeilimme 1, 2, 3 ja 36 piilotettua yksikköä $a^n b^n$:lle; 2, 3, 4 ja 36 piilotettua yksikköä $a^n b^n c^n$:lle; ja 3, 4, 5 ja 36 piilotettua yksikköä $a^n b^n c^n d^n$:lle. BIBREF7 , BIBREF12 , BIBREF9 ja monissa muissa tutkimuksissa omaksutun perinteisen lähestymistavan mukaisesti koulutamme neuroverkkomme seuraavasti. Kullakin ajanhetkellä esitämme mallillemme yhden syötemerkin ja pyydämme sitä sitten ennustamaan seuraavien mahdollisten merkkien joukon nykyisen merkin ja aiempien piilotettujen tilojen perusteella. Koska sanasto $\mathcal {V}^{(i)}$ on kooltaan $d$ , käytämme yhden pisteen esitystä tuloarvojen koodaamiseen; näin ollen kaikki tulovektorit ovat $d$ -ulotteisia binäärivektoreita. Lähtöarvot ovat kuitenkin $(d+1)$ -ulotteisia, koska ne voivat sisältää $\mathcal {V}^{(i)}$ -symbolien lisäksi lopetussymbolin $\dashv $ . Lähtöarvoja ei aina koodata yhden kuuman esityksen avulla, koska sarjan seuraavalle merkille voi olla useita vaihtoehtoja, joten käytämme sen sijaan $k$ -kuuman esityksen lähtöarvojen koodaamiseen. Tavoitteenamme on minimoida sekvenssiennusteiden keskimääräinen neliövirhe (MSE).

**Tulos**

Mitä koulutusasetuksia he kokeilivat?

**Esimerkki 5.1726**

Työn lisäsovelluksena olemme toteuttaneet valvotun luokittelutehtävän, jonka tarkoituksena on ennustaa vaaratapahtuman vahingollisuuden aste suoraan tekstin ja käsin koodattujen ominaisuuksien (esim. ulkoinen luokka, lääketieteen erikoisala, sijainti) perusteella. Tarkistimme myös, voiko valvomattoman sisällönohjautuvan klusterimerkintämme käyttäminen lisäominaisuuksina parantaa valvotun luokittelun suorituskykyä.

**Tulos**

Miten sisällön klustereita käytetään parantamaan vaaratilanteiden vakavuuden ennustamista?

**Esimerkki 5.1727**

Rakennamme docqan julkisesti saatavilla olevan huipputason kysymysten vastausjärjestelmän varaan. Järjestelmä laajentaa BiDAF BIBREF4:ää itsetarkkailulla ja suoriutuu hyvin asiakirjatason laadunvarmistuksesta.

**Tulos**

Mikä on kysymysten vastausalgoritmi?

**Esimerkki 5.1728**

Jaksossa SECREF16 kerromme ensin tarkemmin noudattamastamme koeasetelmasta. Kuten kohdassa SECREF3 selitetään, käytimme BabelNet BIBREF29 -vertailutaksonomiaa. BabelNet on laajamittainen täysimittainen taksonomia, joka koostuu heterogeenisistä lähteistä, kuten WordNet BIBREF36, Wikidata BIBREF37 ja WiBi BIBREF38 , joten se soveltuu hypoteesimme testaamiseen yleisessä ympäristössä. Testataksemme ehdotettua luokkainduktiomallia tarkastelemme kaikkia BabelNetin luokkia, joissa on alle 50 tunnettua tapausta. Tämä johtuu siitä, että käsitteellinen naapuruus on hyödyllinen lähinnä tapauksissa, joissa tunnettujen tapausten määrä on pieni. Kunkin kategorian osalta jaamme tunnettujen tapausten joukon 90 prosenttiin harjoittelua ja 10 prosenttiin testausta varten. Näiden luokkien ennakkotodennäköisyyden $\lambda \_A$ virittämiseksi pidämme 10 prosenttia koulutusjoukosta validointijoukkona.

**Tulos**

Mitä kokeita he tekevät osoittaakseen, että heidän lähestymistapansa johtaa tarkempiin aluepohjaisiin esityksiin?

**Esimerkki 5.1729**

Taulukon TABREF13 tuloksissa on kymmenen koetinta. ELMo-transformerin ja mSynC:n suorituskyky on jälleen samankaltainen, mutta mSynC suoriutuu hieman huonommin seitsemästä yhdeksästä tehtävästä. Odotetusti mSynC saavuttaa kappaletunnisteiden ennustamista koskevassa tehtävässä 96,9 $F\_1$ ja ELMo-transformer 92,2 $F\_1$, mikä osoittaa, että mSynC todellakin koodaa matalaa syntaksia. Kaiken kaikkiaan tulokset vahvistavat edelleen, että eksplisiittinen matala syntaksi ei tarjoa mitään etuja ELMo-transformeriin verrattuna.

**Tulos**

Mitä parannuksia nämä kaksi lähestymistapaa tuovat pelkkään ELMo-perusasetukseen verrattuna?

**Esimerkki 5.1730**

RNN-pohjainen NMT-malliAluksi esittelemme lyhyesti RNN-pohjaisen neuraalisen konekäännöksen (RNN-pohjainen NMT) mallin. Transformer-NMTR Viime aikoina Transformer-malli BIBREF4 on edistynyt merkittävästi konekääntämisessä. Tässä mallissa on monipäinen itsetarkkailukooderi ja monipäinen itsetarkkailun dekooderi.

**Tulos**

mihin NMT-malleihin niitä verrattiin?

**Esimerkki 5.1731**

Yhden merkin aihepiiriluokituskokeissa käytämme Switchboard Telephone Speech Corpus BIBREFiä21 , joka on kokoelma kaksipuolisia puhelinkeskusteluja. Lisäksi arvioimme aiheen tunnistussuorituskykyämme DARPA:n LORELEI-ohjelman (Low Resource Languages for Emergent Incidents) julkaisemilla kolmen kielen puhekorpuksilla.

**Tulos**

Mitä tietokokonaisuuksia käytetään järjestelmän suorituskyvyn arvioinnissa?

**Esimerkki 5.1732**

Tämän jälkeen poimimme nämä ominaisuudet kahdesta kontekstitekstilähteestä, erityisesti sen uutisartikkelin otsikosta, johon kommentti on lähetetty, ja kommentin lähettäneen käyttäjän näytön nimestä.

**Tulos**

Mitä asiayhteyttä he käyttävät?

**Esimerkki 5.1733**

Menetelmämme demonstroimiseksi mustassa laatikossa keskitämme kokeilumme Googlen konekäännösjärjestelmään (GMT), jota käytetään sen Cloud API:n kautta.

**Tulos**

Mitä neuraalista konekäännösjärjestelmää käytetään?

**Esimerkki 5.1734**

Perusmalli, INLINEFORM0 , koulutetaan 190 tunnin ( INLINEFORM1 100K instansseja) transkriptoidulla puheaineistolla. Sen jälkeen se valitsee osajoukon 1700 tunnin ( INLINEFORM2 1,1M instanssia) merkitsemättömästä tietokokonaisuudesta. Kyselemme valitun osajoukon etiketit ja otamme ne mukaan harjoitteluun. Oppimisnopeudet viritetään pienellä 2048 instanssin validointijoukolla. Koulutettua mallia testataan 156 tunnin ( INLINEFORM3 100K instansseja) testijoukolla, ja raportoimme CTC-tappion, merkkivirheprosentin (CER) ja sanavirheprosentin (WER).

**Tulos**

Mitä tietokokonaisuutta he käyttävät?

**Esimerkki 5.1735**

Korkea lähimpien naapureiden tarkkuus osoittaa, että syntaksitieto on onnistuttu tallentamaan upotuksiin.

**Tulos**

Tehdäänkö opittujen upotusten kvantitatiivinen laatuanalyysi?

**Esimerkki 5.1736**

Koska malleissamme ei käytetty RL-jalostusta, nämä tulokset merkitsevät 1,34 ja 1,12 BLEU-pisteen merkittävää parannusta BIBREF3:n vahvoihin perusmalleihin verrattuna. Myös perpleksisyyspisteet ovat paremmat. Google Production -tietokannassa mallimme saavutti 1,01 korkeamman testin BLEU-pistemäärän, vaikka sitä oli harjoiteltu vain yksi kuudesosa ajasta.

**Tulos**

Miten MOE-malli parantaa konekääntämistä SOTA-malliin verrattuna?

**Esimerkki 5.1737**

Järjestelmämme sijoittui 43 tiimin joukosta 421:nneksi virallisen F1-pistemäärän ollessa 0,2905 testijoukossa. Vaikka mallimme on validointijoukossa F1-pistemäärän osalta parempi kuin perusjoukko, havaitsemme kaikkien mittareiden osalta huomattavia pudotuksia testijoukkoon verrattuna, mikä osoittaa, että arkkitehtuuri ei näytä kykenevän yleistämään hyvin.

**Tulos**

Millaisia tuloksia he saivat testijoukossa?

**Esimerkki 5.1738**

Käytämme valmista jäsentäjää, tässä tapauksessa Stanford CoreNLP BIBREF11 , binääristen vaalipiiriparssien luomiseen.

**Tulos**

Miten he saavat analysoituja lähdelauseet?

**Esimerkki 5.1739**

Vaiheessa yksi sekoitimme ja jaoimme datan (1913 signaalia 14 yksilöltä) satunnaisesti harjoittelu- (80 %), kehitys- (10 %) ja testisarjoihin (10 %). Jotta PHASE-TWO:ssa voitaisiin tehdä oikeudenmukainen vertailu aiempiin menetelmiin, jotka on raportoitu samasta tietokokonaisuudesta, suoritamme ristiinvalidointikokeen, jossa jätetään yksi koehenkilö pois ja käytetään parhaita PHASE-ONE:ssa opittuja asetuksia.

**Tulos**

Kuinka monta elektrodia tutkittavaan käytettiin EEG-istunnoissa?

**Esimerkki 5.1740**

Maksimaalinen sovittaminen (MM) on yksi suosituimmista sanojen segmentoinnin perus- ja rakennealgoritmeista BIBREF19 . Se ei kuitenkaan ratkaise moniselitteisten sanojen ja tuntemattomien sanojen ongelmaa, joita ei ole sanakirjassa.

**Tulos**

Mitkä ovat nykyisten vietnaminkielisten sanojen segmentointijärjestelmien rajoitukset?

**Esimerkki 5.1741**

Tässä kokeilussa vertaamme Enitity-GCN:ää samaa tehtävää käsitteleviin viimeaikaisiin aiempiin töihin. Taulukossa 2 esitetään testi- ja kehitystulokset (jos niitä on) molemmille tietokokonaisuuden versioille. BIBREF0:sta luetellaan ihmissuoritukseen perustuva oraakkeli sekä kaksi tavallista luetun ymmärtämisen mallia, nimittäin BiDAF BIBREF3 ja FastQA BIBREF6 . Vertaamme myös Coref-GRU BIBREF12 , MHPGM BIBREF11 ja Weaver BIBREF10 . Lisäksi otamme mukaan tuloksia MHQA-GRN BIBREF23 , joka on peräisin hiljattain julkaistusta arXiv-preprintistä, jossa kuvataan samanaikaista työtä. Niissä koulutetaan yhdessä graafisia neuroverkkoja ja rekursiivisia koodaajia. Raportoimme kahden parhaan yksittäisen mallin ja ensemblen yksittäiset ajot peittämättömässä testijoukossa (muistutettakoon, että testijoukko ei ole julkisesti saatavilla, ja tehtävän järjestäjät raportoivat vain peittämättömät tulokset) sekä validointijoukon molemmat versiot.

**Tulos**

Mihin perustasoon he vertasivat Entity-GCN:ää?

**Esimerkki 5.1742**

Jos harjoittelemme skip-gram-mallin uudelleen ja päivitämme emoji-kartan säännöllisesti uusien escort-ilmoitusten perusteella, kun ihmiskauppiaat siirtyvät uusiin emojeihin, kartta voi yhdistää uudet emojit vanhoihin, mikä auttaa ihmiskaupan vastaisia asiantuntijoita laajentamaan ihmiskauppaa koskevien merkkien sanastoa. Tämä lähestymistapa toimii myös lyhenteiden ja tahallisten kirjoitusvirheiden osalta.

**Tulos**

Miten salakuljetuslippujen sanastoa laajennetaan?

**Esimerkki 5.1743**

Parannus on 3,5 F1-pistettä, kun vaikea osajoukko, johon on lisätty asiantuntijalausuntoja, sekoitetaan jäljelle jäävään joukkorekisteröintiin, mikä on paljon enemmän kuin silloin, kun siihen lisätään satunnainen joukko asiantuntijalausuntoja.

**Tulos**

Kuinka paljon laadukkaampi on tuloksena saatava kommentoitu tieto?

**Esimerkki 5.1744**

oss-analyysi. Perusteellisemman arvioinnin suorittamiseksi tutkimme edelleen mallin käyttäytymistä sekä rekonstruoinnin että uudelleenrakentamisen kannalta.

**Tulos**

Miten he arvioivat tuotetun tekstin laatua?

**Esimerkki 5.1745**

Tietokanta koostuu kyselyistä ja niitä vastaavista kuvahaun tuloksista. Kunkin kyselyn jokaiselle merkille annetaan kielimerkintä, joka perustuu Google Images -hakua tekevän käyttäjän asettamaan kotikieleen.

**Tulos**

Onko kuvissa monikielisiä vai yksikielisiä merkintöjä?

**Esimerkki 5.1746**

Vertailemme käännöslähestymistapojen suorituskykyä neljän mittarin perusteella:[align=left,leftmargin=0em,labelsep=0.4em,font=]Kuten BIBREF20:ssä, EM on 1, jos ennustettu suunnitelma vastaa täsmälleen pohjatotuutta, muuten se on 0.Tarkkuuden ja palautuksen harmoninen keskiarvo koko testijoukon BIBREF26 yli.Vähimmäismäärä lisäyksiä, poistoja tai vaihto-operaatioita, jotka tarvitaan ennustetun käyttäytymissarjan muuttamiseksi perustotuuden mukaiseksi BIBREF27 .GM on 1, jos ennustettu suunnitelma saavuttaa perustotuuden määränpään (vaikka koko käyttäytymissarja ei vastaisi täsmälleen perustotuutta). Muussa tapauksessa GM on 0.

**Tulos**

Mitä arviointimittareita käytetään?

**Esimerkki 5.1747**

Ensimmäinen tietokokonaisuus on laajennettu versio BIBREF9:n kommentoidusta Wikipedian keskusteluja koskevasta tietokokonaisuudesta. Tässä tietokokonaisuudessa käytetään tarkoin valvottuja joukkoresurssien antamia merkintöjä, jotka on suodatettu tiukasti sen varmistamiseksi, että keskustelut ovat sivistyneitä aina henkilökohtaisen hyökkäyksen hetkeen asti. Tämä on hyödyllinen ominaisuus mallianalyysin kannalta, ja siksi keskitymme tähän ensisijaisena tietokokonaisuutena. Olemme kuitenkin tietoisia siitä, että nämä tiukat merkinnät eivät ehkä täysin kuvaa sellaista käyttäytymistä, josta moderaattorit käytännössä välittävät. Siksi otamme käyttöön toissijaisen tietokokonaisuuden, joka on muodostettu ChangeMyView (CMV) -alipalvelusta, jossa ei käytetä post-hoc-merkintöjä. Sen sijaan ennustustehtävänä on ennustaa, joutuuko keskustelu tulevaisuudessa moderaattorin toimenpiteiden kohteeksi. Wikipedia-aineisto. BIBREF9:n `Conversations Gone Awry'-tietokanta koostuu 1 270 keskustelusta, jotka käytiin Wikipedian toimittajien välillä julkisesti saatavilla olevilla keskustelusivuilla. Keskustelut ovat peräisin WikiConv-tietoaineistosta BIBREF59 ja joukkoistajat ovat merkinneet ne joko sisältävän henkilökohtaisen hyökkäyksen sisältäpäin (eli yhden käyttäjän vihamielinen käytös keskustelussa, joka kohdistuu toiseen käyttäjään) tai pysyvän koko ajan kohteliaana. Redditin CMV-tiedot. CMV-aineisto on muodostettu Redditin API:n kautta kerätyistä keskusteluista. Toisin kuin Wikipediaan perustuvassa tietokannassa, vältämme nimenomaisesti jälkikäteisen merkinnän käyttöä. Sen sijaan käytämme merkkinä sitä, onko moderaattori poistanut kommentin keskustelusta säännön 2 (älä ole epäkohtelias tai vihamielinen muita käyttäjiä kohtaan) rikkomisen vuoksi.

**Tulos**

Mitä merkintöjä epäsosiaalisista tapahtumista on saatavilla tietokannoissa?

**Esimerkki 5.1748**

Tässä työssä esittelemme uuden loogisen päättelymoottorin nimeltä MonaLog, joka perustuu luonnolliseen logiikkaan ja vanBenthemEssays86:sta peräisin olevaan monotonisuutta koskevaan työhön. Koska logiikkamme operoi pintamuotojen yli, on suoraviivaista hybridisoida mallimme. Tutkimme MonaLogin käyttöä yhdessä BERT BIBREF20 -kielimallin kanssa, mukaan luettuna kompositionaalinen datan lisääminen, eli harjoitusjoukkojen esimerkkien sisältämien versioiden luominen uudelleen. Teemme kaksi kokeilua MonaLogin testaamiseksi. Ensin käytämme MonaLogia ratkaisemaan ongelmia yleisesti käytetyssä luonnollisen kielen päättelyaineistossa SICK BIBREF1 ja vertaamme tuloksiamme aiempiin järjestelmiin. Toiseksi testaamme MonaLogin tuottaman datan laatua. Tätä varten luomme lisää harjoitusaineistoa (lausepareja) SICKin harjoitusaineistosta järjestelmämme avulla ja suoritamme hienosäätöä BERT BIBREF20:llä, joka on muunnosarkkitehtuuriin BIBREF23 perustuva kielimalli, laajennetulla aineistolla.

**Tulos**

Miten he yhdistävät MonaLogin ja yhteistyöelimen?

**Esimerkki 5.1749**

Avoimen toimialueen chatbotit ovat yleisempiä dialogijärjestelmiä. Esimerkkinä tästä on BIBREF7 humeau2019real:n Poly-enkooderi. Se on parempi kuin Bi-enkooderi BIBREF8, BIBREF9 ja vastaa Cross-enkooderin BIBREF10, BIBREF11 suorituskykyä samalla kun laskenta-aika on kohtuullinen. Se suoriutuu erinomaisesti pareittaisia vertailuja sisältävistä myöhemmistä kielenymmärtämistehtävistä ja osoittaa huipputuloksia ConvAI2-haasteessa BIBREF12. Feed Yourself BIBREF13 on avoimen alueen dialogiagentti, jossa on itseään ruokkiva malli. Kun keskustelu sujuu hyvin, dialogista tulee osa harjoitusdataa, ja kun keskustelu ei suju, agentti pyytää palautetta. Kvmemnn BIBREF14 on avainarvomuistiverkko, jossa on tietopohja, joka käyttää avainarvojen hakumekanismia harjoitellakseen samanaikaisesti useilla aloilla. Käytämme kaikkia kolmea näistä malleista vertailukohtina. Vertailemme neljään dialogijärjestelmän perusmalliin: Kvmemnn, Feed Yourself, Poly-encoder ja BERT bi-ranker -perustason, joka on koulutettu Persona-Chat-tietokannalla käyttäen samoja koulutushyperparametreja (mukaan lukien oppimisnopeuden aikataulutus- ja pituuden rajausasetukset), jotka on kuvattu kohdassa SECREF20.

**Tulos**

Mitä perusmalleja käytetään?

**Esimerkki 5.1750**

Keskeinen etu on se, että RNN päättelee tilan latentin esityksen, jolloin ei tarvita tilamerkintöjä.

**Tulos**

Auttaako latentti vuoropuhelun tila heidän malliaan?

**Esimerkki 5.1751**

kun yhteinen CLWE-avaruus on luotu, tehtävänä on hakea kohdekielen käännöksiä lähdekielen sanojen testijoukosta hakea kohdekielen käännöksiä lähdekielen sanojen testijoukosta BLI on BLI-asetelmassamme ranking-tehtävä, jossa on vain yksi oikea käännös kullekin "kyselysanalle", MAP on yhtä suuri kuin keskimääräinen vastavuoroinen sijoitus (MRR).

**Tulos**

Miten BLI mittaa linjauksen laatua?

**Esimerkki 5.1752**

Alun perin siemensanakirjat kattoivat tyypillisesti useita tuhansia sanapareja BIBREF15 , BIBREF18 , BIBREF19 , mutta uudempi työ on osoittanut, että CLWE:t voidaan indusoida jopa heikommalla valvonnalla pienistä sanakirjoista, jotka kattavat useita satoja sanapareja BIBREF20 , identtisiä merkkijonoja BIBREF21 tai jopa vain yhteisiä numeroita BIBREF22 .

**Tulos**

Miten siemensanakirjat saadaan täysin valvomattomilla menetelmillä?

**Esimerkki 5.1753**

Bengali, gujarati, marathi, malayalam ja tamil ovat ensisijaiset lähdekielet, ja kääntäminen näistä kielistä hindiksi on lasten tehtävä.

**Tulos**

Kuinka monen kieliparin kohdalla he osoittavat, että avustavan kielen lauseiden esijärjestäminen auttaa käännöksen laatua?

**Esimerkki 5.1754**

Koulutustietokannan luomisessa noudatimme samanlaista lähestymistapaa kuin LASERissa. Tietokanta sisältää 6 kieltä: Kielet ovat englanti, espanja, saksa, hollanti, korea ja kiinan mandariini. Tietokokonaisuus luotiin käyttämällä Tatoeban ja OpenSubtitles BIBREF16:n tarjoamia käännöksiä.

**Tulos**

Mitä korpusta he käyttävät?

**Esimerkki 5.1755**

Vertaamme tuloksia uudempaan monikieliseen kielimalliin XLM BIBREF12 sekä uusimpaan CoNLL 2018 -tehtävän jaetun tehtävän tuloksiin, joissa on ennustettu tokenisointi ja segmentointi, artikkelin päivitetyssä versiossa. Ranskan kielellä ei ole tehty laajaa työtä, koska NER-korporaatioita on vain vähän saatavilla. Vertaamme malliamme vahvoihin peruslinjoihin, joihin BIBREF49 asettui, jossa koulutettiin sekä CRF- että BiLSTM-CRF-arkkitehtuurit FTB:llä ja parannettiin niitä heuristiikoilla ja esivalmistetuilla sanojen upotuksilla. TRANSLATE-TRAIN-asetelmassa ilmoitamme aiemmasta kirjallisuudesta saadut parhaat tulokset yhdessä omamme kanssa. BiLSTM-max on paras malli alkuperäisessä XNLI-julkaisussa, mBERT, joka on raportoitu ranskaksi BIBREFissä52, ja XLM (MLM+TLM) on paras esitelty malli BIBREFissä50.

**Tulos**

Mikä on tekniikan nykytila?

**Esimerkki 5.1756**

Korpus koostuu 2 346 kliinisen muistiinpanon kokoelmasta (vastaanottokertomukset, potilaskertomukset ja kotiutuskertomukset), joissa on yhteensä 2 372 323 tokenia (keskimäärin 1 011 tokenia per muistiinpano). Kaikki muistiinpanot kirjoitettiin englanniksi ja poimittiin 183 psykoosipotilaan sähköisestä potilastietojärjestelmästä McLean Psychiatric Hospitalista Belmontissa, MA:ssa, ja kaikilla potilailla oli historiassaan vähintään yksi 30 päivän takaisinottotapaus.

**Tulos**

Mitä tietokokonaisuutta käytetään?

**Esimerkki 5.1757**

Havaittujen tarkistusten arviointiin käytetään tarkkuutta, palautusta ja F-mittaa.

**Tulos**

Mitä mittareita käytetään tarkistusten havaitsemisen arvioinnissa?

**Esimerkki 5.1758**

Toisaalta Go-Explore Seq2Seq osoittaa lupaavia tuloksia, sillä se ratkaisee lähes puolet näkymättömistä peleistä. Kuva FIGREF62 (liitteessä SECREF60) osoittaa, että suurin osa hävityistä peleistä kuuluu vaikeimpaan joukkoon, jossa pelin voittaminen edellyttää hyvin pitkää toimintasarjaa. Nämä tulokset osoittavat sekä Seq2Seq-mallin kouluttamisen suhteellisen tehokkuuden Go-Explore-pelien liikeradoilla että sen, että tarvitaan lisäponnistuksia sellaisten vahvistusoppimisalgoritmien suunnittelemiseksi, jotka yleistyvät tehokkaasti näkymättömiin peleihin.

**Tulos**

Miten kirjoittajat osoittavat, että heidän oppimansa toimintatapa yleistyy olemassa olevia ratkaisuja paremmin näkymättömiin peleihin?

**Esimerkki 5.1759**

Lause-korpus (S) koostuu 80 000 lauseesta ja 280 gigatavusta tavallista tekstiä, jotka on poimittu BIBREF6 aristo2016:combiningin käyttämiltä verkkosivuilta. Käytämme BIBREF6 aristo2016:combiningin tekstikorpusta (S) rakentaaksemme tuple KB:n. Otamme 200 parasta osumaa, ajamme Open IE v4:n ja yhdistämme tuloksena saadut tuplat kaikkiin $a \ in A$ ja kaikkiin kysymyksiin $Q\_\mathit {tr}$:ssa luodaksemme tuple KB:n (T).

**Tulos**

Mikä oli tekstilähde, johon OpenIE:tä sovellettiin?

**Esimerkki 5.1760**

Työssämme käytetään neljää tietokokonaisuutta, joita ovat CoNLL 2003 German BIBREF9 , CoNLL 2002 Spanish BIBREF10 , OntoNotes 4 BIBREF11 ja Weibo NER BIBREF12 . Taulukossa TABREF22 esitetään kiinalaisen OntoNotes 4.0:n tulokset.

**Tulos**

Millä kielillä he työskentelevät?

**Esimerkki 5.1761**

Algoritmi "Online-oppiminen" osoittaa, miten CNN-mallimme voidaan kouluttaa puhtaasti verkossa. Aluksi alustetaan mallin parametrit $\theta \_0$ (rivi 1), joka voi olla muista katastrofitapahtumista koulutettu malli tai se voidaan alustaa satunnaisesti tyhjästä.Kun uusi erä merkittyjä twiittejä $B\_t= \lbrace \mathbf {s}\_1 \ldots \mathbf {s}\_n \rbrace $ saapuu, lasketaan ensin yhtälön 11 mukainen log-loss (ristiinentropia) $B\_t$:lle suhteessa nykyisiin parametreihin $\theta \_t$ (rivi 2a). Sitten laskemme takaisinkulkeutumisen avulla häviön gradientit $f^{\prime }(\theta \_{t})$ nykyisten parametrien suhteen (rivi 2b). Lopuksi päivitämme parametrit oppimisnopeudella $\eta \_t$ ja gradienttien keskiarvolla (rivi 2c). Käytämme gradienttien keskiarvoa, jotta voimme käsitellä erikokoisia minieriä. Huomaa, että otamme huomioon vain nykyisen minierän saadaksemme päivitetyn mallin.

**Tulos**

Mitä uutta tässä stokastisessa gradienttilaskeutumisalgoritmissa on?

**Esimerkki 5.1762**

Koska olemme kiinnostuneita esityksemme nollakuvauskyvystä, koulutimme sentimenttianalyysimallimme vain englanninkielisellä IMDB Large Movie Review -tietokannalla ja testasimme sitä kiinalaisella ChnSentiCorp-tietokannalla ja saksankielisillä SB-10K BIBREF24 , BIBREF25 . Luonnollisen kielen päättelytehtävä koostuu kahdesta lauseesta; premissi ja hypoteesi, jotka ovat joko ristiriitoja, seurauksia tai neutraaleja. NLI-tehtävän oppiminen edellyttää tiettyä vivahteikasta kielen ymmärtämistä. Siksi on kiinnostavaa, pystyykö UG-WGAN ymmärtämään tarvittavat kielelliset piirteet.

**Tulos**

Kokeilivatko he matematiikassa muita tehtäviä kuin sanaongelmia?

**Esimerkki 5.1763**

Käytämme yksikielisenä perusmallina garg2012unsupervised Bayes-mallia. Semanttiset roolit ovat predikaattikohtaisia.

**Tulos**

Mistä yksittäinen malli koostuu?

**Esimerkki 5.1764**

Koko tietokokonaisuus koostuu 81 826 näytteestä, jotka englannin äidinkieliset puhujat ovat kommentoineet. Niistä 80 prosenttia käytetään harjoitusjoukkona. Niistä 10 prosenttia käytetään validointijoukkona ja loput testijoukkona. Testijoukosta valitaan 3000 kovaa näytettä.

**Tulos**

Mitä eroa on täydellisellä testisarjalla ja kovalla testisarjalla?

**Esimerkki 5.1765**

N-grammit, POS, hierarkkiset ominaisuudet: sanasäkkimalli, joka sisältää sekä merkityt että merkitsemättömät unigrammit ja bigammit. CNN: Kim BIBREF28 osoitti lähes huippuluokan suorituskyvyn useissa lauseiden luokittelutehtävissä (mukaan lukien TREC-kysymysten luokittelu) käyttämällä esivalmennettuja sanojen upotuksia BIBREF40 CNN-mallin ominaisuuksien poimijoina.

**Tulos**

Mihin aiempiin menetelmiin heidän malliaan verrataan?

**Esimerkki 5.1766**

Loimme tämän CORD-19-NER-tietokokonaisuuden, jossa on kattava nimettyjen entiteettien annotaatio CORD-19-korpuksesta (2020-03-13). Korpus on luotu CORD-19-korpuksen (2020-03-13) 29 500 asiakirjasta.

**Tulos**

Mikä on tämän tietokokonaisuuden koko?

**Esimerkki 5.1767**

Taulukossa 2 luetellaan kaikkien menetelmien tarkkuudet kahdella luokitteluarkkitehtuurilla. Tulokset osoittavat, että eri luokittimien arkkitehtuurien eri tietokokonaisuuksissa ehdollisen BERT-yhteenliittymän lisääminen parantaa mallin suorituskykyä eniten. BERT voi myös lisätä lauseita jossain määrin, mutta ei yhtä paljon kuin ehdollinen BERT.

**Tulos**

Ilmoittavatko kirjoittajat ehdollisen bertin suorituskyvyn tehtävissä, joissa tietoja ei ole lisätty?

**Esimerkki 5.1768**

Kuvakokoelmaa tarkastelemalla voi olla vaikea havaita kuvioita. Toinen menetelmä on merkitä kaikki kuvaukset puhekielitiedoilla, jolloin voidaan esimerkiksi nähdä, mitä adjektiiveja käytetään yleisimmin tietyistä substantiiveista. Eräs menetelmä, jota lukijat saattavat pitää erityisen hyödyllisenä, on hyödyntää Flickr30K Entities BIBREF8 -tietokannan rakennetta. Tämän jälkeen sovelsin Louvainin klusterointia BIBREF9 coreference-graafiin, minkä tuloksena saatiin samanlaisiin entiteetteihin viittaavien ilmausten klustereita.

**Tulos**

Millä menetelmillä voidaan löytää esimerkkejä ennakkoluuloista ja perusteettomista päätelmistä??

**Esimerkki 5.1769**

Määritämme tehokkuuden ja tarkkuuden välisen kompromissin verrattuna kahteen sääntöpohjaiseen perusratkaisuun: Unif ja Stopword.

**Tulos**

Mitkä ovat käytetyt perustasot?

**Esimerkki 5.1770**

Jopa niiden tutkijoiden keskuudessa, jotka tunsivat edellä esitetyt määritelmät, yksimielisyys oli edelleen vähäistä (Krippendorffin INLINEFORM0 ).

**Tulos**

Miten luotettavuutta mitattiin?

**Esimerkki 5.1771**

Käytämme myös neuroverkon viimeistä piilokerrosta väitteen tehtäväkohtaisena upottamisena yhdessä verkkotodisteiden kanssa.

**Tulos**

Mitä tietoja käytetään tehtäväkohtaisten sulautusten rakentamiseen?

**Esimerkki 5.1772**

vastaamaan, kuinka loukkaavana he pitivät twiittiä 6-pisteisellä Likertin asteikolla 1:stä (ei lainkaan loukkaava) 6:een (erittäin loukkaava). Jos osallistujat vastasivat 4 tai korkeampi, heillä oli mahdollisuus ilmoittaa, mitkä sanat olivat heidän mielestään loukkaavia.

**Tulos**

Tarkoitettiinko loukkaavuuden asteella sitä, kuinka yleisesti loukkaava teksti oli vai kuinka henkilökohtaisesti loukkaava se oli kommentoijalle?

**Esimerkki 5.1773**

Jotta voimme arvioida tämän ja muiden mallien ennustuskykyä, tarvitsemme jonkin menetelmän, jolla voimme vertailla malleja. Tätä tarkoitusta varten käytämme ROC-käyriä (receiver operating characteristic curves), joka on visuaalinen esitys ennustuskyvystä. ROC-käyrillä verrataan mallin ennusteiden todellisten positiivisten ennusteiden (TPR) ja väärien positiivisten ennusteiden (FPR) osuutta eri kynnystasoilla. Käyrän alle jäävä pinta-ala (AUC) (0:n ja 1:n välillä) on numeerinen mittari, jossa mitä suurempi AUC on, sitä paremmin malli toimii. Ristiinvalidoimme mallimme jakamalla korpuksen ensin satunnaisesti koulutusjoukkoon (95 % korpuksesta) ja testijoukkoon (5 % korpuksesta). Sovitamme sitten mallin harjoitusjoukkoon ja käytämme sitä ennustamaan testijoukon asiakirjojen vastauksia. Toistamme tämän prosessin 100 kertaa. Näistä ennusteista saadaan kynnyskeskiarvoinen ROC-käyrä BIBREF13 , joka on esitetty kuvassa 3 . Taulukossa 1 esitetään kunkin tarkastellun mallin AUC-arvo.

**Tulos**

Miten suorituskykyä mitataan?

**Esimerkki 5.1774**

Tulkittavuuden vuoksi projisoimme myös upotusten ja PCA-mallien kertoimet takaisin dummy-muuttuja-avaruuteen.

**Tulos**

Miten niiden tulkita kertoimia?

**Esimerkki 5.1775**

Näissä nauhoituksissa on taustahälyä (kuten ostoskeskuksessa on odotettavissa) ja eri kieliä (intialaiset puhuvat englantia ja hindiä sekaisin).

**Tulos**

Mitä kieliä otetaan huomioon?

**Esimerkki 5.1776**

Yksinkertainen, mutta toistuva tekstimalli on tunnettu sanojen yhteisesiintymisverkko. Valinnaisten tekstin esikäsittelyvaiheiden jälkeen yhteisesiintymisverkossa jokaisesta eri sanasta tulee solmu, ja särmät muodostetaan yhteisesiintymisen avulla halutussa ikkunassa. Yleinen strategia yhdistää vain vierekkäiset sanat ns. sanojen vierekkäisyysverkoissa. Vaikka yhteisesiintymisesitys tuottaa hyviä tuloksia luokittelutilanteissa, mallissa ei oteta huomioon joitakin tärkeitä ominaisuuksia. Esimerkiksi pitkän matkan syntaktiset yhteydet, vaikka ne ovatkin harvinaisempia kuin vierekkäiset syntaktiset suhteet, saatetaan jättää huomiotta yksinkertaisessa sanojen vierekkäisyyttä koskevassa lähestymistavassa BIBREF12.

**Tulos**

Mistä johtuu, että perinteiset esiintymisverkot eivät onnistu luomaan yhteyksiä samankaltaisten sanojen välille aina, kun ne esiintyvät tekstissä kaukana toisistaan?

**Esimerkki 5.1777**

Sekä ääni että transkriptio on poistettu (poistamalla tunnistetiedot) digitaalisilla nollilla ja [de-identified]-tunnisteilla.

**Tulos**

Ovatko tiedot tunnistamattomia?

**Esimerkki 5.1778**

Keskitymme seuraaviin NLP-tutkimuksen näkökohtiin: koko, demografiset tiedot, tutkimusalat, vaikutus ja viittausten ja demografisten ominaisuuksien (ikä ja sukupuoli) välinen korrelaatio.

**Tulos**

Mitä NLP-tutkimuksen osa-aluetta tarkastellaan?

**Esimerkki 5.1779**

Kolmannen persoonan monikossa olevalla pronominilla `they' ei ole sukupuolta englannissa ja useimmissa muissa kielissä (toisin kuin kolmannen persoonan yksikössä olevilla pronomineilla `he', `she' ja `it'). Jos nämä lauseet käännetään ranskaksi, ensimmäisessä lauseessa oleva `they' pitäisi kääntää `elles', joka viittaa Janeen ja Susaniin, ja toisessa lauseessa oleva `they' pitäisi kääntää `ils', joka viittaa Frediin ja Georgeen. Pari esimerkkiä: Saksan kielessä sana `sie' toimii sekä muodollisena toisen persoonan prounominina (aina isolla alkukirjaimella), kolmannen persoonan feminiinisenä yksikössä että kolmannen persoonan monikossa.

**Tulos**

Mitä kieltä he tutkivat?

**Esimerkki 5.1780**

Tässä luvussa esitellään todennäköisyysmallimme, joka johtaa avainsanan odotuksen ja kouluttaa kohdemallin samanaikaisesti.

**Tulos**

Millaisia luokittelijoita käytetään?

**Esimerkki 5.1781**

Yksi puhuja kertoi 2000 lausetta, mikä kesti useita päiviä.

**Tulos**

Kuinka monta kommentoijaa osallistui?

**Esimerkki 5.1782**

Ensimmäinen perustasomme on ROUGE-L BIBREF1 , koska se on yleisimmin käytetty metriikka pakkaustehtävissä. Vertaamme sitä toutanova2016datasetista löytyviin parhaisiin n-grammi-overlap-mittareihin; lisäksi vertaamme sitä negatiiviseen LM-ristientropiaan eli log-todennäköisyyteen, joka on normalisoitu vain lauseen pituuden mukaan. Seuraavana vertailukohtana on perpleksisyys, joka vastaa eksponentoitua risti-entropiaa: Sen suosion vuoksi teimme alustavia kokeita myös BLEU BIBREF17 -menetelmällä.

**Tulos**

Onko ROUGE heidän ainoa perustasonsa?

**Esimerkki 5.1783**

Testaamme ehdotettua lähestymistapaamme sarkasmin tai ironian binääriseen luokitteluun seitsemällä vertailutietoaineistolla, jotka on haettu eri medialähteistä. Seuraavassa kuvataan kukin tietokokonaisuus, ja taulukossa TABREF1 on yhteenveto. Twitter: Käytämme SemEval 2018 Task 3, Irony Detection in English Tweets BIBREF18 -tehtävää varten toimitettua Twitter-tietokokonaisuutta. Lopuksi käytämme BIBREF20 -tietokokonaisuutta, joka keräsi käyttäjän itsensä kommentoiman korpuksen twiittejä, joissa oli #sarcasm-hashtag. Reddit: BIBREF21 keräsi SARC-korpuksen, joka koostuu 600 000 sarkastisesta kommentista Redditissä. Käytämme pääjoukkoa SARC 2.0 ja poliittista osajoukkoa SARC 2.0 pol. Verkkodialogit: Käytämme Sarcasm Corpus V1:tä (SC-V1) ja Sarcasm Corpus V2:tä (SC-V2), jotka ovat Internet Argument Corpusin (IAC) osajoukkoja. Verrattuna muihin valikoimassamme oleviin tietokokonaisuuksiin nämä eroavat lähinnä tekstin pituuden ja rakenteen monimutkaisuuden BIBREF22 osalta. Jotta kutakin tietokokonaisuutta voidaan täydentää ulkoisilla tiedoilla, suodatamme ensin kieliarvoitusjärjestelmien avulla pois twiitit, jotka eivät ole englanninkielisiä.

**Tulos**

Arvioivatko he vain englannin kieltä?

**Esimerkki 5.1784**

Käytämme kokeissamme TB-Dense- ja MATRES-ohjelmia, ja taulukossa TABREF33 on lyhyt yhteenveto datatilastoista.

**Tulos**

Mitä tietokokonaisuuksia tässä työssä käytettiin?

**Esimerkki 5.1785**

Valitsemalla englannin (En) kääntökieleksi suoritamme identtisten englanninkielisten segmenttien kääntökohdistukset Europarlin Fr-En- ja En-De-rinnakkaiskorpuksissa BIBREF18 , jolloin muodostuu moniparalliskorpus Fr-En-De. Tämän jälkeen kukin Fr\*-De- ja Fr-De\*-pseudorinnakkaiskorpuksista muodostetaan moniparallisaineistosta soveltamalla edellisessä jaksossa kuvattua pivot-kielipohjaista käännöstä. Automaattisessa kääntämisessä käytetään valmiiksi koulutettua ja julkisesti julkaistua NMT-mallia En $\rightarrow $ De:lle ja koulutetaan toinen NMT-malli En $\rightarrow $ Fr:lle WMT'15 En-Fr-paralleelikorpuksen BIBREF19 avulla.

**Tulos**

Mistä synteettiset tiedot kerätään?

**Esimerkki 5.1786**

Näiden kokoonpanojen aikaansaamiseksi TIMIT-tiedot jaettiin. Kuvassa FIGREF12 esitetään tietojen jakaminen kahdeksaan osajoukkoon (A-H). TIMIT-tietokokonaisuus sisältää 462 puhujan puhetta harjoittelussa ja 168 puhujan puhetta testijoukossa, ja jokaisella puhujalla on 8 lausumaa. TIMIT-koulutus- ja testijoukko on jaettu kahdeksaan lohkoon, joissa kussakin lohkossa on kaksi satunnaisesti valittua lausumaa kutakin puhujaa kohti. Näin ollen kukin lohko A,B,C,D sisältää dataa 462 puhujalta, joiden 924 lausumaa on otettu koulutusjoukoista, ja kukin lohko E,F,G,H sisältää puhetta 168:lta testijoukon puhujalta, joiden 336 lausumaa on otettu. Testidata on loput eli lohkot (D+H). Tehtävässä b sulautusten ja luokittelijoiden koulutuksessa käytetään lohkoja (A+B+E+F) ja (C+G) ja testissä taas lohkoja (D+H). Tehtävä c pitää molemmat erillään: upotukset koulutetaan lohkoissa (A+B+C+D), luokittelijat lohkoissa (E+G) ja testit suoritetaan lohkoissa (F+H). Huomaa, että H on osa kaikkia tehtäviä ja että tehtävä c on huomattavasti helpompi, koska erotettavien puhujien määrä on vain 168, vaikka koulutusolosuhteet ovat vaikeammat.

**Tulos**

Mitä TIMIT-tietoaineistoja käytetään testauksessa?

**Esimerkki 5.1787**

Formalisoimme tämän intuition NextSum-nimisessä mallissa, joka valitsee seuraavan yhteenvetolauseen lähdetekstin ominaisuuksien lisäksi myös yhteenvedon aiemmin valittujen lauseiden perusteella.

**Tulos**

Miten nextsum toimii?

**Esimerkki 5.1788**

Olemme havainneet, että epookkien määrä on tärkeä parametri, ja sen lisääminen johtaa tuloksiin, joissa kaksi huonointa malliamme ovat lähes yhtä hyviä tai jopa parempia kuin muut.

**Tulos**

Miten eri parametriasetukset vaikuttavat tuloksena olevan mallin suorituskykyyn ja semanttiseen kapasiteettiin?

**Esimerkki 5.1789**

Esikäsitellyssä aineistossa on 8757 uutistietuetta.

**Tulos**

Kuinka suuri tietokokonaisuus on?

**Esimerkki 5.1790**

Tätä varten ehdotetaan menetelmää, jolla käytetyn tietokokonaisuuden (DSTC2) tavoitetunnisteet muunnetaan tunnisteiksi, joiden käyttäytyminen voidaan jäljitellä käyttöönoton aikana.

**Tulos**

mitä korpusta käytetään käyttäytymisen oppimiseen?

**Esimerkki 5.1791**

Kohteen tyypin löytämiseksi SimplerVoice rakentaa ensin ontologiapohjaisen tietopuun. Ehdotamme, että kohdeobjektiin sopivien verbien luomiseen käytetään kahta menetelmää: heuristiikkapohjaista ja n-grammimallia.

**Tulos**

Mitä mallia he käyttävät avainviestien tuottamiseen?

**Esimerkki 5.1792**

Kouluttaaksemme ja arvioidaksemme avoimen alueen faktatietokannan laadunvarmistusjärjestelmää reaalimaailman kysymyksiin, luomme uuden kiinalaisen laadunvarmistustietokannan nimeltä WebQA.

**Tulos**

Mitä kieliä he kokeilevat?

**Esimerkki 5.1793**

Ehdotamme uudenlaista suhdesuositusongelmaa (RSR). DSN-verkkojen vastavuoroisen suosittelun ongelmasta poiketen RSR-tehtävämme toimii tavallisissa sosiaalisissa verkostoissa (RSN) ja arvioi pitkäaikaisten ja vakavien suhteiden yhteensopivuutta sosiaalisten viestien, kuten twiittien, perusteella.

**Tulos**

Ovatko muut ihmiset työskennelleet tämän tehtävän parissa?

**Esimerkki 5.1794**

UTD:n tarkoituksena on tunnistaa ja klusteroida automaattisesti toistuvia termejä (esim. sanoja tai lauseita) puheesta. Kiertääksemme tyhjentävän DTW-pohjaisen haun, jota rajoittaa INLINEFORM0 -aika BIBREF6 , hyödynnämme Zero Resource Toolkitin (ZRTools) BIBREF7 skaalautuvaa UTD-kehystä, joka mahdollistaa haun INLINEFORM1 -ajassa.

**Tulos**

Miten sanan kaltaisten yksiköiden tai foneemien kaltaisten yksiköiden sanasto löydetään automaattisesti?

**Esimerkki 5.1795**

Järjestelmät koulutettiin kaikilla WMT 2016 -tapahtumasta saatavilla olevilla rinnakkaisilla tiedoilla. Uutiskommenttikorpuksen, Euroopan parlamentin istuntoselostusten ja yhteisen ryömintäkorpuksen yhteenlaskettu määrä on 3,7 miljoonaa lausetta ja noin 90 miljoonaa sanaa.

**Tulos**

Mitä tietokokonaisuutta he käyttävät?

**Esimerkki 5.1796**

Moniluokkaista luokittelua varten käytettiin One-Vs-The-Rest-strategiaa, ja raportoin F-pisteet, mikro-F-, makro-F- ja painotetut F-pisteet käyttäen 10-kertaista ristiinvalidointia.

**Tulos**

Mitä mittareita otetaan huomioon?

**Esimerkki 5.1797**

Argus-tietoaineisto AI2-8grade/CK12-tietoaineisto Pidämme tätä aineistoa alustavana, koska sitä ei ole tarkistanut ihminen ja koska monet hypoteesit ovat ilmeisesti todistamattomia keräämiemme todisteiden perusteella (eli teoreettinen huipputarkkuus on paljon pienempi kuin 1,0). MCTest-tietokanta Käytämme RTE-arviointiin tietokannan virallista laajennusta, jossa kysymykset ja vastaukset yhdistetään tekstin avulla.

**Tulos**

mitä tietokokonaisuuksia he käyttivät?

**Esimerkki 5.1798**

Arvioimme monikieliseen esivalmennukseen perustuvaa siirtomenetelmäämme useita vahvoja perusratkaisuja vastaan kahdella julkisella tietokokonaisuudella, Europarl BIBREF31 ja MultiUN BIBREF32, jotka sisältävät useita rinnakkaisia arviointitietoja nollapisteen suorituskyvyn arvioimiseksi.

**Tulos**

millä tietokokonaisuuksilla he tekivät kokeita?

**Esimerkki 5.1799**

Käytimme Levenshteinin etäisyysmittaria BIBREF8 , joka on toteutettu Apache Lucene -kirjastossa BIBREF9 . Toinen yksinkertainen lähestymistapa on edellä mainittu diakriittinen vaihtaminen, joka on termi, jonka otamme tässä käyttöön viitatakseen BIBREF4:n työn innoittamaan ratkaisuun. Eräs lupaava menetelmä, joka on peräisin englannin kielen oppijoiden tekstien korjaamista käsittelevästä työstä BIBREF11 , perustuu käsitteeseen, jonka mukaan oikeinkirjoitusvirhettä lähimpänä oleva korjaus valitaan jonkin etäisyyskäsitteen mukaan. Tässä Levenshteinin etäisyyttä käytetään painotettuna summana sanavektoreiden väliseen kosinietäisyyteen. ELMo-lisäyksellä varustettu LSTM on kaksisuuntainen.

**Tulos**

Mitä menetelmiä PIEWi testaa?

**Esimerkki 5.1800**

Kokeissa, joissa käytetään rinnakkaista dataa ulkomaisten parametrien alustamiseen, käytämme samoja tietokokonaisuuksia kuin BIBREF6:n työssä. Erityisesti käytämme United Nations Parallel Corpus BIBREF18 -korpusta en-ru, en-ar, en-zh ja en-fr -kieleen. Keräämme en-hi-rinnakkaisdataa IIT Bombayn korpuksesta BIBREF19 ja en-vi-dataa OpenSubtitles 2018:sta.

**Tulos**

Mitä tietokokonaisuuksia käytetään arvioinnissa?

**Esimerkki 5.1801**

Rakensimme vihamielisen puheen havaitsemista meemeistä varten tietokokonaisuuden, jossa oli 5 020 kuvaa, jotka oli heikosti merkitty vihamielisiksi tai ei-vihamielisiksi meemeiksi niiden lähteen mukaan.

**Tulos**

Miten kukin tietokokonaisuuden instanssi on kommentoitu?

**Esimerkki 5.1802**

Kunkin testijoukon osalta käytämme vastaavia harjoituskysymyksiä $Q\_\mathit {tr}$ hakeaksemme alaan liittyviä lauseita S:stä. Tarkemmin sanottuna käytämme jokaisen Q\_\mathit {tr}$:n monivalintakysymyksen $(q,A) \ ja jokaisen A$:n vaihtoehdon $a \a$ kohdalla kaikkia $q$:n ja $a$:n sisältämiä ei-stop-sanamerkkejä ElasticSearchin kyselynä S:lle. Otamme 200 parasta osumaa, suoritamme Open IE v4:n ja yhdistämme tuloksena saadut tuplat kaikkien $a \ A$:n ja kaikkien $Q\_\mathit {tr}$:n kysymysten yli luodaksemme tuplan KB (T).

**Tulos**

Mitä menetelmää käytettiin OpenIE-uutteiden tuottamiseen?

**Esimerkki 5.1803**

Arvioimme kysymysten tuottamisessa yleisesti käytetyillä mittareilla BIBREF13: BLEU-1 (B1), BLEU-2 (B2), BLEU-3 (B3), BLEU-4 (B4) BIBREF17, METEOR (MET) BIBREF18 ja ROUGE-L (R-L) BIBREF19. Käytämme Chen2015MicrosoftCC:n julkaisemaa arviointiskriptiä.

**Tulos**

Mitä mittareita he käyttävät?

**Esimerkki 5.1804**

Huomaa, että vain noin 12 prosentissa (7 kärkeä 59:stä) eniten siteeratuista tutkimusaloista naiset saivat keskimäärin enemmän viittauksia kuin miehet. Näihin kuuluvat: sentimenttianalyysi, tiedonlouhinta, asiakirjojen tiivistäminen, puhuttu dialogi, monikielisyys (tutkimus), dialogi, järjestelmät, kielen tuottaminen. (Huomaa tietysti, että osa 59:stä alueesta, kuten otsikkotermien bigrammien avulla arvioituna, on päällekkäisiä. Emme myöskään sisällyttäneet edellä olevaan luetteloon suursanastoa, koska keskiarvojen ero on hyvin pieni ja koska se ei ole varsinainen tutkimusalue). Siten viittauskuilu on yleistä suurimmalla osalla NLP:n korkean viittausmäärän alueista.

**Tulos**

Millä NLP-alueella on korkein keskimääräinen siteeraus naiskirjoittajalle?

**Esimerkki 5.1805**

Käytimme vain 300 ulottuvuuden malleja ja Word2Vec-, Wang2Vec- ja FastText-malleissa skip-grammiarkkitehtuurin malleja.

**Tulos**

mitä sanojen upottamistekniikoita he kokeilivat?

**Esimerkki 5.1806**

Vertaamme ehdotettuja malleja olemassa oleviin sanojen upottamismenetelmiin. Nämä ovat: Bernoulli embeddings (b-emb) BIBREF1 , continuous bag-of-words (CBOW) BIBREF5 , Distributed Memory version of Paragraph Vector (PV-DM) BIBREF11 ja Global Vectors (GloVe) BIBREF6.

**Tulos**

Mitä sanasulkeumia he testaavat?

**Esimerkki 5.1807**

Nämä ominaisuudet saadaan yleensä kouluttamalla syvää neuroverkkoa yhdessä useilla kielillä, joista on saatavilla merkittyjä tietoja. Verkon alimmat kerrokset ovat yleensä yhteisiä kaikille harjoituskielille. Tämän jälkeen verkko jakautuu erillisiin osiin kutakin kieltä varten tai sillä on yksi yhteinen ulostulo. Viimeisen ulostulokerroksen kohteina ovat puhelimien merkinnät tai HMM-tilat. Viimeisen jaetun kerroksen dimensiot ovat usein pienempiä kuin tulokerroksen, ja siksi sitä kutsutaan "pullonkaulaksi".

**Tulos**

Mitkä ovat pullonkaulaominaisuuksia?

**Esimerkki 5.1808**

Korpus koostuu 53 asiakirjasta, joissa on keskimäärin 156,1 lausetta asiakirjaa kohti ja joissa jokaisessa on keskimäärin 19,55 merkkiä.

**Tulos**

Kuinka monta asiakirjaa uudessa korpuksessa on?

**Esimerkki 5.1809**

Tässä jaksossa raportoidaan kahdella aineistolla suoritettujen kokeiden tulokset, joilla arvioidaan wDTW:n ja wTED:n suorituskykyä muihin perusmenetelmiin verrattuna. Merkitsemme seuraavia etäisyys- ja samankaltaisuusmittoja. WMD: kohdassa SECREF1 esitelty Word Mover's Distance. VSM: jaksossa UID12 esitelty samankaltaisuusmitta. PV-DTW: PV-DTW on sama kuin algoritmi SECREF21 , paitsi että kahden kappaleen välinen etäisyys ei perustu algoritmiin SECREF20, vaan se lasketaan INLINEFORM0:n avulla, jossa INLINEFORM1 on kappaleen INLINEFORM2 PV-sisällyttäminen. PV-TED: PV-TED on sama kuin algoritmi SECREF23 , paitsi että kahden kappaleen välinen etäisyys ei perustu algoritmiin SECREF20 vaan lasketaan INLINEFORM0:n avulla.

**Tulos**

Mitkä ovat uusimmat mallit?

**Esimerkki 5.1810**

Olemme osoittaneet, että muuntajapohjaiset kooderit yhdessä GRU-mallien kanssa tarjoavat parhaan suorituskyvyn kysymysten esittämisessä. Erityisesti osoitimme, että esivalmennettujen tekstiedustusten käyttö parantaa suorituskykyä johdonmukaisesti useilla hyperparametrikokoonpanoilla. Olemme myös osoittaneet, että ulkoisen datan avulla hienosäädetyn objektinilmaisimen käyttö parantaa tarkkuutta huomattavasti. Lisäksi olemme osoittaneet, että huomiomekanismit ovat ensiarvoisen tärkeitä huippusuorituskykyisten verkkojen oppimisessa, kun ne mahdollistavat kysymystietoisten kuvarepresentaatioiden tuottamisen, jotka kykenevät koodaamaan avaruudellisia suhteita. On käynyt selväksi, että Top-Down on suositeltavin huomiomenetelmä, kun otetaan huomioon sen tulokset ReLU-aktivoinnin kanssa.

**Tulos**

Mitkä osatekijät on määritelty VQA-mallien koulutuksen keskeisiksi osatekijöiksi?

**Esimerkki 5.1811**

He käyttivät 20 miljoonan sanan dataa, joka oli satunnaisesti poimittu CoNLL 2017 Shared Task - Automatically Annotated Raw Texts and Word Embeddings BIBREF8 -tehtävässä julkaistusta raakatekstistä, joka on yhdistelmä Wikipedian tyhjennystä ja yleistä indeksointia. Vertasimme esimerkiksi ELMoForManyLangsin latvian kielen mallia malliin, jonka koulutimme täydellisellä (wikidump + common crawl) latvian kielen korpuksella, jossa on noin 280 miljoonaa merkkiä.

**Tulos**

Kuinka suuret ovat näiden ELMo-versioiden harjoitusjoukot verrattuna aiempiin versioihin?

**Esimerkki 5.1812**

Menetelmämme avulla merkitsimme 10 000 arabiankielisen twiitin aineiston loukkaavuuden perusteella, ja loukkaavien twiittien osuus oli noin 19 prosenttia twiiteistä. Lisäksi merkitsimme twiitit mauttomiksi tai vihapuheiksi.

**Tulos**

Kuinka monta twiittiä tietokannassa on?

**Esimerkki 5.1813**

Huomasimme, että kaikki testaamamme tekijät voivat vaikuttaa laadullisesti siihen, miten malli yleistyy kysymyksenmuodostustehtävässä. Näitä tekijöitä ovat toistuvan yksikön tyyppi, huomion tyyppi ja peräkkäisen vs. puupohjaisen mallin rakenteen valinta.

**Tulos**

Mitä arkkitehtonisia tekijöitä tutkittiin?

**Esimerkki 5.1814**

Vertaamme ehdottamaamme diskreettiä CVAE:tä (DCVAE) kaksivaiheisella otantamenetelmällä kolmeen luokkaan kuuluviin vasteiden muodostamismalleihin: Perusasetukset: CVAE BIBREF14: Mukautamme alkuperäisen työn, joka koskee monikierroksista keskustelua, yhden kierroksen asetelmaa varten. Oikeudenmukaisen vertailun vuoksi käytämme samoja avainsanoja, joita käytimme verkon esivalmennuksessa, tietoon perustuvina ominaisuuksina tässä mallissa. muut parannetut koodaaja-dekooderimallit: Hierarchical Gated Fusion Unit (HGFU) BIBREF12, joka sisällyttää dekooderiin pistemäisen keskinäisen informaation (PMI) avulla poimitun vihjesanan mielekkäiden vastausten luomiseksi; Mechanism-Aware Neural Machine (MANM) BIBREF13, joka ottaa käyttöön latentteja upotuksia, jotta voidaan luoda useita erilaisia vastauksia.

**Tulos**

Millaisia muita sukupolvimalleja kokeissa käytetään?

**Esimerkki 5.1815**

(ehdotettu) Bi-LSTM/CRF + Bi-CharLSTM ja modaliteettihuomio (W+C): käyttää modaliteettihuomiota yhdistämään sana- ja merkkikokoelmia. (ehdotettu) Bi-LSTM/CRF + Bi-CharLSTM + Inception (W+C+V): ottaa syötteenä InceptionNetistä poimittuja visuaalisia konteksteja, jotka on yhdistetty sana- ja merkkikokoelmavektoreihin.(ehdotettu) Bi-LSTM/CRF + Bi-CharLSTM + Inception with modality attention (W+C+V): käyttää modaliteettihuomiota yhdistämään sana-, merkki- ja visuaaliset upotukset syötteenä entiteetti-LSTM:lle.

**Tulos**

Oppiiko heidän NER-mallinsa NER:ää sekä tekstistä että kuvista?

**Esimerkki 5.1816**

Koska muuntosääntöjämme on tulkittavissa, tunnistamme puutteet molemmissa resursseissa ja käytämme kumpaakin toistensa validoimiseksi. Pystyimme löytämään tiettyjä tapauksia, joissa UniMorph-merkintöjä oli sovellettu virheellisesti, sekä erityisiä tapauksia, joissa molemmissa resursseissa oli kieltenvälisiä epäjohdonmukaisuuksia.

**Tulos**

Etsitäänkö UniMorphissa eri kielten annotaatioiden välisiä epäjohdonmukaisuuksia?

**Esimerkki 5.1817**

Tässä tutkimuksessa kokeillaan neljää erilaista luokittelumallia, mukaan lukien logistinen regressio (LR), rekursiivinen neuroverkko (RNN) BIBREF35, konvoluutio-neuroverkko (CNN) BIBREF36 ja Google BERT BIBREF37.

**Tulos**

Mitä asiakirjaluokittelijoita he kokeilevat?

**Esimerkki 5.1818**

Käytämme 16 erilaista tietokokonaisuutta, jotka ovat peräisin useista suosituista BIBREFissä20 käytetyistä arvostelukorpuksista. Nämä tietokokonaisuudet koostuvat 14 tuotearvostelutietokokonaisuudesta ja kahdesta elokuva-arvostelutietokokonaisuudesta. Käytämme CoNLL 2000 BIBREF22 -jaksomerkintätietokokonaisuutta sekä POS Tagging- että Chunking-tehtävissä.

**Tulos**

Mitä tietokokonaisuutta he käyttivät?

**Esimerkki 5.1819**

. Vaikka automaattiset arviointimittarit, kuten ROUGE, mittaavat koneellisten ja inhimillisten tiivistelmien välistä leksikaalista samankaltaisuutta, ihmiset voivat mitata paremmin sitä, kuinka johdonmukainen ja luettava tiivistelmä on. Arviointitutkimuksessamme tutkitaan, lisääkö PG-net-mallin virittäminen tiivistelmän johdonmukaisuutta pyytämällä arvioijia valitsemaan, mistä kolmesta saman asiakirjan tiivistelmästä he pitävät eniten: CNN/DM:llä koulutetusta PG-net-mallista, opiskelijoiden pohdinnoilla koulutetusta mallista ja lopuksi CNN/DM:llä koulutetusta ja opiskelijoiden pohdinnoilla viritetystä mallista. Laitoksestamme rekrytoitiin 20 arvioijaa, joita pyydettiin tekemään kukin 20 annotaatiota. Yhteenvedot esitetään arvioijille satunnaisessa järjestyksessä. Arvioijia pyydetään sitten valitsemaan tiivistelmä, joka on heidän mielestään luettavin ja johdonmukaisin.

**Tulos**

Keitä käytettiin ihmisarvioijina?

**Esimerkki 5.1820**

Taulukossa 1 on esimerkki vaikeista ja helpoista esimerkeistä vaikeuden määritelmämme mukaisesti. Alleviivattu teksti kuvaa toimialan asiantuntijoiden antamaa (yksimielistä) viitemerkintää. Vaikeissa esimerkeissä joukko työntekijöitä merkitsi tekstin, joka erosi näistä referenssimerkinnöistä, kun taas helpoissa tapauksissa he toistivat ne kohtuullisen tarkasti. Vaikeat lauseet ovat yleensä rakenteeltaan monimutkaisia ja niissä on jargonia.

**Tulos**

Onko instanssi lause vai IE-tuple?

**Esimerkki 5.1821**

Kommentoija äänestää yhtä seitsemästä tunteesta eli Ekmanin kuutta perustunnetta BIBREF1 sekä neutraalia.

**Tulos**

Mitä merkintöjä tietokokonaisuudella on?

**Esimerkki 5.1822**

Määritimme palkkion arvoksi 1, jos tehtävä on suoritettu onnistuneesti, ja 0, jos se ei ole suoritettu onnistuneesti. $0.95$:n alennusta käytettiin kannustamaan järjestelmää suorittamaan dialogit nopeammin kuin hitaammin, jolloin palautus oli 0 epäonnistuneista dialogeista ja $G = 0.95^{T-1}$ onnistuneista dialogeista, missä $T$ on järjestelmän vuorojen määrä dialogissa. 0.95^{T-1} palkinto 0.95^{T-1}

**Tulos**

Mikä on vahvistusoppimisen sovelluksen palkitsemismalli?

**Esimerkki 5.1823**

Kokeiluaineisto on peräisin Microsoft Research (MSR) -yritykseltä. Se sisältää kolme aluetta: elokuva, taksi ja ravintola. Taulukossa TABREF11 ilmoitetaan dialogien kokonaismäärä alueittain ja koulutus-/validointi-/testijakauma. Joka käänteessä sekä käyttäjän että agentin tekoja kommentoidaan, mutta käytämme kokeessamme vain agenttipuolta kohteina. Teot on järjestetty tietokokonaisuudessa (jokainen tulostettu lause vastaa yhtä tekoa).

**Tulos**

Mitä tietokokonaisuuksia käytetään mallien harjoittelussa/testauksessa?

**Esimerkki 5.1824**

KryptoOracle on rakennettu Apachen ekosysteemiin ja käyttää Apache Sparkia. Sparkin tietorakenteet perustuvat resilientteihin hajautettuihin datajoukkoihin (RDD), jotka ovat vain lukukelpoisia moniulotteisia datajoukkoja, jotka voidaan hajauttaa koneklusteriin ja jotka ovat vikasietoisia. Spark RDD:llä on luontainen kyky toipua itsestään, koska se tallentaa kaikki suoritusvaiheet lineage-graafiin. Jos järjestelmässä ilmenee vikoja, Spark tekee uudelleen kaikki aiemmat suoritukset rakennetusta DAG:sta ja palautuu edelliseen vakaaseen tilaan mistä tahansa viasta, kuten muistin ylikuormittumisesta. Spark RDD:t ovat KryptoOracin ytimessä ja helpottavat siksi sen toipumista vioista. Lisäksi muistin ylikuormituksen tai järjestelmän kaatumisen kaltaiset viat saattavat vaatia koko järjestelmän kovaa uudelleenkäynnistystä. Koska Apache Hivessä on kuitenkin monistetut kopiot RDD-tietueista ja koneoppimismallin aiempi tila on tallennettu, KryptoOracle voi helposti palautua edelliseen vakaaseen tilaan.

**Tulos**

Miten arkkitehtuuri on vikasietoinen?

**Esimerkki 5.1825**

Kehitystietojen merkitsemätöntä joukkoa käytettiin sekä UBM:n että i-vektorin poimijan koulutuksessa.

**Tulos**

Valitaanko kiinteästä SRE-koulutusjoukosta validointijoukko?

**Esimerkki 5.1826**

Tässä tehtävässä käytetty aineisto on peräisin Switchboardin englanninkielisestä keskustelukorpuksesta BIBREF31 .

**Tulos**

Mitä tietokokonaisuutta he käyttävät?

**Esimerkki 5.1827**

Tässä artikkelissa esitellään kuvassa 1 esitetty koneellisen luetun ymmärtämisen ekstraktio- ja synteesi-kehys, jossa vastaus syntetisoidaan ekstraktiotulosten perusteella. Rakennamme todisteiden louhintamallin, jolla ennustetaan tärkeimpiä osa-alueita tekstikappaleista todisteiksi, ja sitten kehitämme vastauksen synteesimallin, joka ottaa todisteet lisäominaisuuksina yhdessä kysymyksen ja tekstikappaleen kanssa lopullisen vastauksen tarkentamiseksi.

**Tulos**

Mitä puitteita he ehdottavat tässä asiakirjassa?

**Tulos**

Mitkä kaksi osatekijää sisältyvät heidän ehdottamaansa kehykseen?

**Esimerkki 5.1828**

Jokaista syötteen Open IE -tuplan $(S; P; O\_1; O\_2 \ldots )$ kohdetta $O\_i$ varten lisätään tähän taulukkoon kolmikko $(S; P; O\_i)$.

**Tulos**

Ovatko OpenIE-uutteet kaikki kolmosia?

**Esimerkki 5.1829**

Testaamme merkki- ja sanatason variantteja ennustamalla hashtageja pidättäytyneelle testiviestien joukolle.

**Tulos**

Millä muilla tehtävillä he testaavat menetelmäänsä?

**Esimerkki 5.1830**

Moduulit: Moduulit ovat erilaisten mallien perusrakennuspalikoita. LeafNATSissa tarjoamme valmiita moduuleja, joilla voidaan rakentaa NATS:lle RNN-pohjaisia sekvenssistä sekvenssiin (Seq2Seq) perustuvia malleja, esimerkiksi osoitinverkkogeneraattoriverkko BIBREF1 . Näihin moduuleihin kuuluvat muun muassa upottaja, RNN-kooderi, huomio BIBREF24 , ajallinen huomio BIBREF6 , huomio dekooderissa BIBREF2. Käytämme näitä perusmoduuleja myös kootaksemme osoitingeneraattorin dekooderimoduulin ja vastaavat palkkihakualgoritmit. Upotuskooderin avulla voidaan myös toteuttaa upotus-painojen jakomekanismi BIBREF2 .

**Tulos**

Mitä malleja työkalupakki sisältää?

**Esimerkki 5.1831**

Järjestelmämme oppii Perceptron-malleja BIBREF37 käyttäen Apache OpenNLP -projektin tarjoamaa koneoppimiskoneistoa ja omia räätälöityjä (paikallisia ja klusterointi-) ominaisuuksia. Paikalliset piirteet muodostavat perusjärjestelmämme, jonka päälle lisätään klusterointipiirteet.

**Tulos**

mitkä ovat peruslinjat?

**Esimerkki 5.1832**

Facebook-sivujen lopullinen kokoelma tässä artikkelissa kuvattuja kokeita varten on seuraava: The Guardian, Cartoon Network, Cooking Light, Home Cooking Adventure, Justin Bieber, Nickelodeon, Paavo, Disney.

**Tulos**

Mitä Facebook-sivuja he katsoivat?

**Esimerkki 5.1833**

Käytämme kolmea luovaa englanninkielistä tietokokonaisuutta, joilla on erilaiset kielelliset ominaisuudet: (1) 740 klassisen ja nykyaikaisen englanninkielisen runon korpus, (2) 14950 metaforalauseen korpus, joka on haettu metaforatietokannan verkkosivulta, ja (3) 1500 laulun sanoituksen korpus, joka koostuu eri genreistä. Ensin esivalmennamme generaattorimme Gutenbergin tietokokonaisuudella BIBREF24 20 epookin ajan ja sitten hienosäädämme BIBREF19 niitä kohdetietokantojamme varten kielimallinnustavoitteella.

**Tulos**

Mitä tietokokonaisuuksia käytetään?

**Esimerkki 5.1834**

Ehdotimme yhteisesti oppivia NLP-malleja, joissa käytetään perusyksikköinä konvoluutiohermoverkkoa (convolutional neural network, CNN) BIBREF8 ja kaksisuuntaista pitkää lyhytkestoista muistia (bi-directional long short-term memory, BiLSTM) BIBREF9, BIBREF10. Mallimme voivat automaattisesti poimia seksuaalista häirintää koskevista tarinoista keskeiset elementit ja samalla luokitella tarinat eri ulottuvuuksiin. Ehdotetut mallit päihittivät yhden tehtävän mallit ja saavuttivat korkeamman tarkkuuden kuin aiemmin raportoidut tarkkuudet ahdistelulomakkeiden luokittelussa BIBREF6.

**Tulos**

Mitä mallia he käyttivät?

**Esimerkki 5.1835**

Merkintöihin osallistui kaikkiaan noin 500 eri työntekijää. noin 500

**Tulos**

Kuinka monta henkilöä osallistui heidän taulukoista tekstiksi -mallien arviointitutkimukseensa?

**Esimerkki 5.1836**

Taulukossa TABREF31 esitetään ehdotettujen mallien F-tulos, ROC-käyrän alapuolinen alue (AUC) ja keskimääräinen tarkkuus (ACC), kun käytettävissä on erilaisia syötteitä. Kuvassa FIGREF32 esitetään eri mallien Precision vs Recall -diagrammi ja ROC-käyrä (jossa esitetään True Positive Rate vs False Positive Rate).

**Tulos**

Mitä mittareita käytetään tulosten vertailemiseksi?

**Esimerkki 5.1837**

Turisti: Matkailija: En voi mennä enää suoraan eteenpäin.Opas: Selvä: Opas: Joo. Katsoin väärää pankkiaTuristi: Matkailija: Ilmoitan, kun olen takaisin Brooks Brothersin ja pankin luona.Matkailija: OPAS: käänny oikealle, kun pankki on vasemmalla puolellasiTuristi: käänny oikealle, kun pankki on vasemmalla puolellasiTuristi: käänny oikealle, kun pankki on vasemmalla puolellasi: TOIMINTA:ETEENPÄIN TOIMINTA:ETEENPÄIN TOIMINTA:KÄÄNTÄÄNOIKEALLETuristi: Oikealle kääntyminen pankin kohdalla.turisti: ACTION:FORWARD ACTION:FORWARDTourist: Turisti: En voi mennä tuohon suuntaan: ACTION:TURNLEFTTuristi: Turisti: Pankki on edessäni oikeallaTuristi: KÄÄNTYMINEN: Käänny oikealle: ACTION:FORWARD ACTION:FORWARD ACTION:TURNLEFTGuide: käänny ympäri tuossa risteyksessäTuristi: Matkailija: Voin mennä vain vasemmalle tai takaisin sitä tietä, jota juuri tulin.Matkailija: Turisti: Voin mennä vain vasemmalle tai takaisin sitä tietä, jota juuri tulin: Näetkö kulmilla olevia kauppoja?Opas: Opas: Käänny vasemmalle, käänny vasemmalle, käänny vasemmalle, käänny vasemmalle, käänny vasemmalle, käänny vasemmalle, käänny vasemmalle, käänny vasemmalle, käänny vasemmalle: Turisti: Jos olet pankin kulmassa, mene kadun yliTuristi: Jos olet pankin kulmassa, mene kadun yliTuristi: Jos olet pankin kulmassa, mene kadun yli: Turisti: Olen takaisin siinä, mistä aloitin, kaupan ja pankin luona.Turisti: Opas: Mene takaisin: KÄÄNNY OIKEALLE.

**Tulos**

Millä kielellä agentit puhuvat?

**Esimerkki 5.1838**

Vertailukohteiden arviointi ::: luokittelumallitSVM: MLP: Monikerroksinen perceptron, jonka syötteenä on USE embeddings BIBREF4.FastText: Pinnallinen neuroverkko, joka keskiarvoistaa n-grammien upotukset BIBREF5.CNN: Convolutional neural network with non-static word embeddings initialized with GloVe BIBREF6.BERT: Neuroverkko, joka on koulutettu ennustamaan tekstissä esiintyviä väistyneitä sanoja ja joka on sitten hienosäädetty aineistollamme BIBREF1.Platforms: Alustat: Tehtäväkeskeisten agenttien kehittämiseen on olemassa useita alustoja. Tarkastelemme Googlen DialogFlowta ja Rasa NLU:ta spacy-sklearnilla.

**Tulos**

Mitkä luokittelijat arvioidaan?

**Esimerkki 5.1839**

Hypoteesi 4 (H4) : Ehdotettu mallipohjainen synteesimalli on parempi kuin pelkkä sanojen korvausmalli. Taulukko TABREF13 tukee H4:ää osoittamalla, että ehdotettu synteesimalli on parempi kuin WordNet-perusmalli koulutuksessa (rivit 7, 8 ja 19, 20) lukuun ottamatta Stat2016:ta ja virityksessä (10, 11 ja 22, 23) kaikilla kursseilla. Se osoittaa myös, että vaikka synteettisen datan lisääminen perustasosta ei aina auta, synteettisen datan lisääminen mallimallista auttaa parantamaan sekä koulutus- että viritysprosessia. Sekä CS- että ENGR-kursseilla synteettisen datan virittäminen parantaa kaikkia ROUGE-pistemääriä verrattuna virittämiseen vain alkuperäisellä datalla. (rivit 9 ja 11). Stat2015-kurssin osalta R-1 ja R-$L$ paranivat, kun taas R-2 heikkeni. Stat2016:n osalta R-2 ja R-$L$ paranivat ja R-1 heikkeni (rivit 21 ja 23). Harjoittelu sekä opiskelijoiden pohdintatiedoilla että synteettisillä tiedoilla verrattuna harjoitteluun vain opiskelijoiden pohdintatiedoilla tuottaa samanlaisia parannuksia, mikä tukee H3:a (rivit 6, 8 ja 18, 20).

**Tulos**

Onko mallipohjainen malli realistinen?

**Esimerkki 5.1840**

Samoin kuin BIBREF9:ssä, kokeilemme SVM:ää RBF-ytimellä, jonka piirteet edustavat (1) argumenttipuun yksinkertaisia ominaisuuksia ja (2) väitteen kielellisiä ominaisuuksia.

**Tulos**

Mitä malleja, jotka perustuvat ainoastaan väittämäkohtaisiin kielellisiin piirteisiin, käytetään vertailukohtina?

**Esimerkki 5.1841**

Kun tarkastellaan nykyistä WinoGrande-listaa, näyttää siltä, että tekniikan edellinen taso perustuu RoBERTa BIBREF2:een, jota voidaan luonnehtia pelkän koodaimen muuntimen arkkitehtuuriksi.

**Tulos**

Mikä on tekniikan aiempi taso?

**Esimerkki 5.1842**

Jotta oppisimme paremman graafisen esityksen usean osapuolen dialogeista, otamme käyttöön dialogit, joissa on 8-15 lausetta ja 3-7 puhujaa. Tehtävän yksinkertaistamiseksi suodatamme dialogit, joissa on pitkiä lauseita (yli 20 sanaa). Lopulta saamme 52 053 dialogia ja 460 358 lausumaa.

**Tulos**

Kuinka suuri ehdotettu tietokokonaisuus on?

**Esimerkki 5.1843**

Kokeissamme vertailimme noin 60 perusmittaria, jotka voidaan jakaa seuraavasti:MT-mittaritBLEU, ROUGE, METEOR, TERpBLEU:n muunnokset: BLEU\_1gram, BLEU\_2gram, BLEU\_3gram, BLEU\_4gram ja seitsemän tasoitusmenetelmää NLTK:n BIBREF32:sta.TERp:n väliosat, jotka ovat BIBREF18:n innoittamia: esim. lisäysten, poistojen, siirtojen jne. määrä.Luettavuusmetriikat ja muut lausetason ominaisuudet: BIBREF28 , kuten sanojen lukumäärää, sanojen pituuksia, kielimallin todennäköisyyttä ja INLINEFORM0 -grammitaajuutta koskevat tilastot, kielimallin todennäköisyyttä ja INLINEFORM0 -grammitaajuutta koskevat metriikat.Muihin piirteisiin perustuvat metriikat: frekvenssitaulukon sijainti, konkreettisuus sellaisena kuin se on poimittu BIBREF33 'n BIBREF33 -luettelosta, sanojen kielimallin todennäköisyys käyttäen BIBREF:n34 konvoluutio-sekvenssi-sekvenssimallia, vertailumenetelmät käyttäen valmiiksi koulutettuja fastText-sanojen upotuksia BIBREF35 tai Skip-thought-lauseiden upotuksia BIBREF36 .

**Tulos**

mitä lähestymistapoja verrataan?

**Esimerkki 5.1844**

Yhdistämme yhteistyö- ja e-yhteistyöelimet a) yhdistämällä niiden tuotokset (AVG) tai b) yhdistämällä kokonaisuuden ja sen nimen vinoviivalla (CONCAT), esim. seuraavasti: Jean\_Marais / Jean Mara ##on.

**Tulos**

Mitkä ovat kaksi tapaa koota yhteen yhteistyö- ja e-yhteistyömuodot?

**Esimerkki 5.1845**

Sovelsimme malliamme myös MCTest-tietokantaan, jossa koneiden on vastattava fiktiivisiä tarinoita koskeviin monivalintakysymyksiin.

**Tulos**

Kokeilevatko he ehdotettua mallia muilla kuin MovieQA:n aineistoilla?

**Esimerkki 5.1846**

Tässä työssä tutkimme menetelmämme soveltuvuutta kolmeen suosittuun arkkitehtuuriin: LSTM-pohjaiseen, CNN-pohjaiseen ja muuntajaan perustuvaan arkkitehtuuriin. Taulukossa TABREF46 esitetään menetelmämme suorituskyky eri sekvenssimallinnusarkkitehtuureilla. Taulukosta nähdään ensinnäkin, että LSTM-pohjainen arkkitehtuuri suoriutui paremmin kuin CNN- ja transformer-pohjaiset arkkitehtuurit.

**Tulos**

Mihin sekvenssimalliarkkitehtuureihin tämä menetelmä voidaan siirtää?

**Esimerkki 5.1847**

Annotaatiojärjestelmä kehitettiin kuvaamaan kolmea luokkahuonekeskustelun näkökohtaa, joiden on kirjallisuudessa katsottu olevan tärkeitä keskustelun laadun ja oppimismahdollisuuksien kannalta: argumentointi (prosessi, jossa idean tueksi esitetään systemaattisia perusteluja), spesifisyys (laatu, joka kuuluu tai liittyy yksiselitteisesti tiettyyn aiheeseen) ja tietämysalue (asiantuntemuksen alue, jota keskustelun sisältö edustaa).

**Tulos**

miten he mittaavat keskustelun laatua?

**Esimerkki 5.1848**

Taulukossa TABREF32 esitetään mallien perpleksisyys kussakin testijoukossa. Odotetusti kukin malli suoriutuu paremmin testijoukossa, johon se on koulutettu. Kun mallia sovelletaan toiseen testijoukkoon, kummankin mallin perpleksiteetti kasvaa. Leipzigin mallilla näyttää kuitenkin olevan enemmän vaikeuksia yleistää: sen perpleksiteetti lähes kaksinkertaistuu SwissCrawl-testijoukossa ja nousee kaksikymmentä kertaa yhdistetyssä testijoukossa.

**Tulos**

Miten kielen mallintamista arvioidaan?

**Esimerkki 5.1849**

Lisäksi käytämme myös kahta sen vastakohtajoukkoa, nimittäin AddSent ja AddOneSent BIBREF6 , arvioidaksemme MRC-mallien kestävyyttä kohinan suhteen. Vastakohtajoukkojen kohdat sisältävät harhaanjohtavia lauseita, joiden tarkoituksena on häiritä MRC-malleja. Emme arvioi KAR:n suorituskykyä ainoastaan kehitysjoukolla ja testijoukolla, vaan teemme sen myös vastakkaisilla joukoilla.

**Tulos**

Miten kirjoittajat tutkivat, onko malli kestävä kohinan suhteen vai ei?

**Esimerkki 5.1850**

Edellä esitetystä musiikin genrejen kuvauksesta käy ilmi, että supergenrejä ja johdannaislinjoja BIBREF19, BIBREF20 on rajallinen määrä, kuten kuviossa KUVIO 1 esitetään.Laskennallisesta näkökulmasta genret ovat luokkia, ja vaikka niitä voidaan käsitellä koneoppimisalgoritmeilla, ne eivät sisällä tietoa niiden välisistä suhteista. Jotta genrejen väliset suhteet voitaisiin virallistaa laskennallisia tarkoituksia varten, määrittelemme jatkuvan genreasteikon, joka ulottuu kokeellisimmasta ja introvertoiduimmasta supergenrestä euforisoivimpaan ja osallistavimpaan genreen. Valitsimme Wikipediasta 77 genreä, jotka mainitsimme lihavoituna edellisessä kappaleessa, ja pyysimme kahta riippumatonta arvioijaa lukemaan genrejen Wikipedian sivut, kuuntelemaan genren näytteitä tai artisteja (jos he eivät tienneet niitä jo ennestään) ja sitten kommentoimaan seuraavia ulottuvuuksia:

**Tulos**

Kuinka monesta genrestä he keräsivät?

**Esimerkki 5.1851**

Tietokanta sisältää yhteensä 9710 kohtaa, joissa on keskimäärin 6,24 lausetta kohtaa kohti, 16,16 sanaa lausetta kohti ja joiden keskipituus on 86 sanaa.

**Tulos**

Mikä on tämän tietokokonaisuuden koko?

**Esimerkki 5.1852**

Yksittäisten lauseiden esittämiseksi lisäämme ulkoisia [cls]-merkkejä kunkin lauseen alkuun, ja jokainen [cls]-symboli kerää sitä edeltävän lauseen piirteitä. Käytämme myös intervallin segmenttien upotuksia erottaaksemme useita lauseita asiakirjan sisällä. Tällä tavoin asiakirjan representaatiot opitaan hierarkkisesti, jolloin alemmat Transformer-kerrokset edustavat vierekkäisiä lauseita, kun taas ylemmät kerrokset yhdessä itsehuomautuksen kanssa edustavat usean lauseen diskurssia. Alkuperäisessä Bert-mallissa sijaintikoodausten enimmäispituus on 512. Ylitämme tämän rajoituksen lisäämällä lisää sijaintikoodauksia, jotka alustetaan satunnaisesti ja joita hienosäädetään koodaajan muiden parametrien avulla.

**Tulos**

Mitä uutta heidän asiakirjatason koodaajassaan on?

**Esimerkki 5.1853**

Osallistujat nauhoittavat äänileikkeitä lukemalla lahjoitetuista lauseista.

**Tulos**

Mitä aloja korpus kattaa?

**Esimerkki 5.1854**

Laskemme lauseen todennäköisyyden pitkä-lyhytkestoisen muistin (LSTM, korkereiter1997long) LM:llä eli erityisellä RNN LM:llä, joka on koulutettu suurella korpuksella. Koulutamme LSTM LM:t englanninkielisellä Gigaword-korpuksella BIBREF15 , joka koostuu uutisdatasta.

**Tulos**

mitä kielimalleja ne käyttävät?

**Esimerkki 5.1855**

Torch-Struct-kirjaston rakenne noudattaa sekä TensorFlow'n että PyTorch BIBREF29:n käyttämää jakeluliitäntää.

**Tulos**

Antaako API mahdollisuuden liittää malleja, jotka on kirjoitettu jollakin muulla syväoppimiskehyksellä?

**Esimerkki 5.1856**

Käytämme luokittelukokeissamme BIBREF4:n käyttöön ottamaa MH17 Twitter -tietokokonaisuutta, joka on kerätty tutkimaan MH17-lento-onnettomuutta koskevan (epä)tiedon kulkua Twitterissä.

**Tulos**

Mitä tietokokonaisuutta tässä tutkimuksessa käytetään?

**Esimerkki 5.1857**

Keräämme syötteeksi $37 263$ käsitejoukkoja, joista kukin sisältää kolmesta viiteen yhteistä käsitettä. Nämä käsitejoukot on poimittu useista suurista kuva-/videotekstien korpuksista siten, että niiden sisältämät käsitteet esiintyvät todennäköisemmin yhdessä luonnollisissa kohtauksissa. Tehtävässämme odotettujen käsitejoukkojen oletetaan esiintyvän luonnollisissa, jokapäiväisessä elämässä esiintyvissä kohtauksissa. Kuvien/videoiden kuvateksteissä olevilla käsitteillä, jotka yleensä kuvaavat jokapäiväisen elämämme kohtauksia, on siis haluttu ominaisuus. Keräämme siksi suuren määrän kuvatekstilauseet eri tietokannoista, kuten VATEX BIBREF4, LSMDC BIBREF12, ActivityNet BIBREF13 ja SNLI BIBREF15, ja muodostamme yhteensä 1 040 330 lausetta.

**Tulos**

Mistä konseptisarjat tulevat?

**Esimerkki 5.1858**

Rekrytoimme 102 osallistujaparia Amazon Mechanical Turkista ja jaoimme satunnaisesti puhujan ja kuuntelijan roolit.

**Tulos**

Tehtiinkö tämä koe laboratoriossa?

**Esimerkki 5.1859**

Kunkin mallin osalta tarkasteltiin sanatason perpleksiteettiä, R@3 seuraavan sanan ennustamisessa, latenssia (ms/q) ja energiankulutusta (mJ/q). Tutkiaksemme perpleksiteetin ja palautuksen välistä suhdetta keräsimme yksittäisiä perpleksiteetti- ja palautustilastoja jokaisesta testisarjan lauseesta.

**Tulos**

Mitä näkökohtia on vertailtu eri kielimallien välillä?

**Esimerkki 5.1860**

Siemenleksikko koostuu positiivisista ja negatiivisista predikaateista. Jos uutetun tapahtuman predikaatti on siemenleksikossa eikä siihen liity monimutkaisia ilmiöitä, kuten negaatiota, annamme tapahtumalle vastaavan polariteettipisteytyksen ($+1$ positiivisille tapahtumille ja $-1$ negatiivisille tapahtumille). Se on

**Tulos**

Mikä on siementen sanasto?

**Esimerkki 5.1861**

Kieliopillisten virheiden korjaaminen on haastava tehtävä, koska virhetyypit vaihtelevat ja virheet ovat syntaktisesti ja semanttisesti riippuvaisia ympäröivästä kontekstista. Useimmissa kielioppivirheiden korjausjärjestelmissä käytetään luokittelua ja sääntöpohjaisia lähestymistapoja tiettyjen virhetyyppien korjaamiseen. Nämä järjestelmät käyttävät kuitenkin useita kielellisiä vihjeitä ominaisuuksina. Tavanomaiset kielianalyysityökalut, kuten POS-taggerit (part-of-speech) ja jäsentimet, on usein koulutettu hyvin muodostetulle tekstille, ja ne toimivat huonosti epäkieliopillisessa tekstissä. Tämä aiheuttaa lisää virheitä ja rajoittaa sääntöpohjaisten ja luokitteluun perustuvien lähestymistapojen suorituskykyä GEC:ssä. Tämän seurauksena lausepohjainen tilastollinen konekääntäminen (SMT) on saavuttanut suosiota, koska se pystyy oppimaan tekstin muunnoksia virheellisestä tekstistä oikeaksi tekstiksi virheiden korjaamista rinnakkaisista korpuksista ilman kielellistä lisätietoa. Me mallinnamme GEC-järjestelmämme fraasipohjaisen SMT-lähestymistavan pohjalta. Perinteiset fraasipohjaiset SMT-järjestelmät käsittelevät kuitenkin sanoja ja fraaseja erillisinä kokonaisuuksina. Hyödynnämme jatkuvan tilan esittämistä lisäämällä kaksi neuroverkkokomponenttia, joiden on osoitettu parantavan SMT-järjestelmiä BIBREF3 , BIBREF4 . NNJM:n kouluttamiseen käytämme julkisesti saatavilla olevaa toteutusta, Neural Probabilistic Language Model (NPLM) BIBREF14 . Mosesin uusin versio voi sisällyttää NPLM:n avulla koulutetun NNJM:n ominaisuudeksi dekoodauksen aikana. NNGLM:n tavoin käytämme NNJM:n kouluttamiseen käännösmallin kouluttamiseen käytettyä rinnakkaistekstiä. Lähdekonteksti-ikkunan koko on 5 ja kohdekonteksti-ikkunan koko 4. Valitsemme lähdekontekstin sanastoon 16 000 yleisintä sanaa lähdepuolelta. Kohdekontekstin sanastoksi ja tulostussanastoksi asetetaan 32 000 yleisintä sanaa. Käytämme yhtä piilokerrosta nopeuttaaksemme harjoittelua ja dekoodausta, kun syötteen upotusulottuvuus on 192 ja piilokerroksen solmut 512.

**Tulos**

Käyttävätkö ne neuroverkkomalleissaan esivalmennettuja sanarepresentaatioita?

**Esimerkki 5.1862**

Suoritamme kokeita SQuAD-tietokannan BIBREF1 asiakirjatason muunnoksilla.

**Tulos**

Millä aineistoilla tätä menetelmää on arvioitu?

**Esimerkki 5.1863**

Valvojan arvioinnin korpus sisältää 26972 lausetta.

**Tulos**

Mikä on todellisen tietokokonaisuuden koko?

**Esimerkki 5.1864**

Vaiheessa 1 (jota kutsutaan myös "etsintävaiheeksi") algoritmi tutkii tilaavaruutta pitämällä kirjaa aiemmin käytetyistä tiloista ylläpitämällä arkistoa. Tämän vaiheen aikana algoritmi ei jatka etsintää tyhjästä, vaan alkaa tutkia arkiston lupaavia tiloja löytääkseen tehokkaita ratoja.

**Tulos**

Miten on kehityskaari miten palkkiot uutetaan?

**Esimerkki 5.1865**

Käytämme kokeissamme julkisesti saatavilla olevaa tietokokonaisuutta KVRET BIBREF5. Tämä tietokokonaisuus sisältää vuoropuheluja kolmelta alueelta: kalenteri, sää ja navigointi (POI), mikä soveltuu eri alojen vuoropuhelukokeiluihin.

**Tulos**

Mitä alueita he tutkivat?

**Esimerkki 5.1866**

Tässä jaksossa tarkastelemme kahdenlaisia relaatioiden luokittelutehtäviä: (1) relaatioiden ennustaminen ja (2) relaatioiden poimiminen. Toivomme voivamme suunnitella yksinkertaisen ja selkeän koeasetelman, jonka avulla voimme tehdä virheanalyysin relaatioiden ennustamista varten. Siksi tarkastelemme tyypillistä menetelmää TransE BIBREF3:n kohteena ja FB15K BIBREF3:n tietokokonaisuutena. Relaatioiden louhintaa varten tarkastelemme valvottua relaatioiden louhinta-asetusta ja TACRED-tietokokonaisuutta BIBREF10 . Aiheen mallina käytämme TACRED-tietokannan parasta mallia - sijaintitietoista hermosekvenssimallia.

**Tulos**

Mitä kilpailukykyisiä relaatioluokitusmalleja ne testaavat?

**Esimerkki 5.1867**

Jokainen twiitti tunnistettiin NLTK TweetTokenizer -ohjelmalla ja luokiteltiin yhdeksi kymmenestä mahdollisesta tilistä, jolta se saattoi olla peräisin. Tilit valittiin sen perusteella, mistä aiheista kunkin tiedetään yleensä twiittaavan.

**Tulos**

Mitä tekstin luokittelutehtävää tarkastellaan?

**Esimerkki 5.1868**

Kokeemme tehdään todellisella uhanalaisella kielellä, Mboshi (Bantu C25), jota puhutaan Kongo-Brazzavillessa, käyttäen BIBREF17:n kaksikielistä ranska-mboshi 5K -korpusta. Mboshi-kielen osalta tarkastelemme aakkosellista esitystapaa, jossa ei ole ääntämistietoa. Ranskan puolella otetaan huomioon vain oletusarvoinen segmentointi sanoihin.

**Tulos**

Mitä tietokokonaisuutta käytetään tutkimuksessa?

**Esimerkki 5.1869**

Tässä mallissa käytetään tiettyä RNN-solua tallentamaan vain olennaiset tiedot annetusta kysymyksestä. Dynaaminen muisti tallentaa tietoa $T$:ssä olevista olioista. Tärkein panoksemme on $s\_t^T q$-termin lisääminen porttifunktioon.

**Tulos**

Miten malli tunnistaa entiteetit ja niiden suhteen vastauksiin päättelyhetkellä, kun vastaukset eivät ole saatavilla?

**Esimerkki 5.1870**

Koe-erässä 4 järjestelmämme (nimeltään FACTOIDS) saavutti korkeimman recall-pistemäärän 0,7033, mutta alhaisen tarkkuuden 0,1119, mikä jättää avoimeksi kysymyksen siitä, miten olisimme voineet tasapainottaa näitä kahta mittaria paremmin.

**Tulos**

Mikä oli heidän korkein muistipistemääränsä?

**Esimerkki 5.1871**

Käytämme ylimääräisiä INLINEFORM0-harjoitusartikkeleita, jotka on merkitty kustantajan mukaan, ohjaamattomana aineistona BERT-mallin harjoitteluun. Tutkimme ensin esivalmennuksen vaikutusta BERT-BASE:n suorituskykyyn. Samalla tietokoneella mallin hienosäätö pienellä harjoitusjoukolla kesti vain noin 35 minuuttia sekvenssin pituuden 100 osalta.

**Tulos**

Miten nämä kaksi eri mallia koulutetaan?

**Esimerkki 5.1872**

Arvioidaksemme näytteenottostrategiaamme vertaamme sitä esivalmistettuun malliin ja hienosäädettyyn malliin 16 eri korpuksen avulla. Tulokset osoittavat, että lähestymistapamme on kaikissa korporaatioissa parempi kuin perusmenetelmämme: se saavuttaa keskimäärin 1,6 % alhaisemman puhelinvirheprosentin.

**Tulos**

Kuinka paljon ne ovat keskimäärin parempia kuin monikielinen perusmalli 16:ssa vähäisten resurssien tehtävässä?

**Esimerkki 5.1873**

Vertaamme lähestymistapaamme useisiin menetelmiin BIBREF1 , BIBREF31 , BIBREF11 , BIBREF8 , BIBREF10 , BIBREF39 kahdessa eri toimialueiden välisessä ympäristössä. Giménez-Pérez et al. BIBREF10 raportoi merkkijonoytimiä käyttäen paremmasta suorituskyvystä kuin SST BIBREF31 ja KE-Meta BIBREF11 monilähdealueilla. Lisäksi vertaamme lähestymistapaamme SFA BIBREF1:een, CORAL BIBREF8:aan ja TR-TrAdaBoost BIBREF39:ään yhden lähteen alueella. Transduktiiviset merkkijonojen ytimet. Esittelemme yksinkertaisen ja suoraviivaisen lähestymistavan merkkijonoille soveltuvan transduktiivisen samankaltaisuusmitan tuottamiseksi. Transduktiivinen ydinluokittimemme (TKC) koostuu kahdesta oppimisiteraatiosta.

**Tulos**

Mitä koneoppimisalgoritmeja käytetään?

**Esimerkki 5.1874**

Ensimmäinen, jäljempänä seq2seq, on tavallinen RNN-pohjainen neuraalinen konekäännösjärjestelmä, jossa kiinnitetään huomiota BIBREF0 . Tässä perustasossa ei käytetä jäsenneltyä dataa. Monilähdejärjestelmät parantavat huomattavasti molempia perustasoja, jopa 1,5 BLEU:n parannukset seq2seq-perustasoon verrattuna ja jopa 1,1 BLEU:n parannukset parse2seq-perustasoon verrattuna.

**Tulos**

Mikä on heidän mallinsa suorituskyvyn lasku, kun ei ole analysoitua syötettä?

**Esimerkki 5.1875**

Arvioimme muunnettuja upotusmalleja ensin monikielisillä leksikaalisen samankaltaisuuden ja sukulaisuuden tehtävillä, jotta voimme varmistaa, että sanamerkityksen induktioprosessi ei haittaa upotusten yleistä suorituskykyä. Sen jälkeen testaamme merkityssisällön sulautuksia WSD-tehtävässä.

**Tulos**

Suoritettiinko ulkopuolinen arviointi?

**Esimerkki 5.1876**

Yksi monista muodoista, joihin FHIR voidaan sarjallistaa, on RDF. Koska RDF on kuitenkin suunniteltu abstraktiksi tietomalliksi ja FHIR on suunniteltu terveydenhuollon operatiiviseen käyttöön, mallien välillä voi olla pieni epäsuhta. Tämä ilmenee kahdella tavalla: Ensinnäkin RDF:ssä esitetään tosiseikkoja koskevia lausumia, kun taas FHIR:ssä esitetään tapahtumia koskevia tietueita. Toiseksi RDF:n on tarkoitus olla monotoninen, mikä tarkoittaa, että uudet tosiasiat eivät voi mitätöidä aiempia tosiasioita.

**Tulos**

Miten FHIR ja RDF yhdistetään?

**Esimerkki 5.1877**

Base koulutetaan vain puhtaalla datalla, kun taas Base+Noise koulutetaan sekä puhtaalla että meluisella datalla ilman kohinan käsittelyä. Global-CM käyttää kohinan mallintamiseen globaalia sekoitusmatriisia kaikille meluisille tapauksille, kuten BIBREF3:ssa ehdotetaan ja SECREF3-jaksossa esitellään. Samaa arkkitehtuuria käytetään Global-ID-CM:ssä, mutta sekoitusmatriisi alustetaan identiteettimatriisilla (kaavan DISPLAY\_FORM5 sijaan) ja sitä mukautetaan vain harjoittelun aikana.Klusteripohjaiset mallit, joita ehdotamme jaksossa SECREF4, ovat Brown-CM ja K-Means-CM. Kokeilimme klusterien lukumäärää 5, 10, 25 ja 50. Mallit, jotka valitsevat vain suurimmat ryhmät $G$, on merkitty \*-Freq-merkinnällä, ja ne valitsevat joko 30 tai 50 % klustereista. Interpolointimalleilla on jälkiliite \*-IP, jossa $\lambda \in \lbrace 0,3, 0,5, 0,7\rbrace $ . Molempien yhdistelmää kutsutaan nimellä \*-Freq-IP. Kuten kaikkien muidenkin hyperparametrien osalta, valinta tehtiin kehityssarjasta. toteutimme Cleaning BIBREF15- ja Dynamic-CM BIBREF14-mallit. Molempia malleja ei ollut kehitetty sekvenssien merkintätehtäviä varten, joten niitä oli mukautettava. Cleaning-mallin osalta noudatimme BIBREF3:n ohjeita. Dynamic-CM-mallin sulauttamis- ja ennustuskomponentit korvattiin perusmallimme mukaisesti. Tiheän kerroksen tulosta käytettiin syötteenä dynaamisen matriisin generoinnissa. Kokeilimme niiden ehdottaman jälkihäviön kanssa ja ilman sitä.

**Tulos**

Mitä käytetään perustasona?

**Esimerkki 5.1878**

Recurrent Neural Network (RNN) -kielimallit koulutetaan tuottamaan tulevan sanan todennäköisyysjakauma kontekstin perusteella ilman, että kontekstin rakennetta esitetään eksplisiittisesti BIBREF9, BIBREF10. ActionLSTMmallintaa lauseen linearisoidun sulkeistetun puurakenteen oppimalla ennustamaan seuraavaa toimintoa, jota tarvitaan lauseen rakenteen jäsentämiseen BIBREF16. Generatiiviset rekursiiviset neuroverkkokieliopit (RNNG)mallintavat yhdessä sanasekvenssin sekä taustalla olevan syntaktisen rakenteen BIBREF18.

**Tulos**

Mitkä ovat perusmallit?

**Esimerkki 5.1879**

Arvioimme menetelmäämme IWSLT16:n saksan ja englannin (DE-EN) välisessä käännöksessä molempiin suuntiin, WMT15:n englannin ja saksan välisessä käännöksessä (EN-DE) molempiin suuntiin ja NIST:n kiinan ja englannin välisessä käännöksessä (ZH$\rightarrow $EN).

**Tulos**

Mitä korpuksia käytetään?

**Esimerkki 5.1880**

Ilman ELMo:ta (sama asetus kuin taulukon 3 4. rivi ), data-asetuksemme ovat samat kuin qin-EtAl:2017:Long, jonka suorituskyky oli huippuluokkaa ja jota verrataan suoraan. Huomaamme, että suorituskykymme on parempi myös ilman esivalmennettua ELMo-kooderia, mikä johtuu pääasiassa paremmista lausepariedustuksistamme.

**Tulos**

Miksi heidän mallinsa toimii paremmin kuin aiemmat mallit?

**Esimerkki 5.1881**

cQA-foorumien automatisointi voidaan jakaa kolmeen tehtävään: kysymyksen ja kommentin relevanssi (tehtävä A), kysymyksen ja kysymyksen relevanssi (tehtävä B) ja kysymyksen ja ulkoisen kommentin relevanssi (tehtävä C). Meidän cQA-tehtävissämme objektiparit ovat (kysymys, kysymys) tai (kysymys, kommentti), ja suhde on relevantti/relevantti. Tehtävässä C alkuperäisen kysymyksen (oriQ) ja ulkoisen kommentin (relC) lisäksi annetaan myös kysymys, jota relC kommentoi (relQ). Tämän lisätiedon sisällyttämiseksi tarkastelemme monitehtäväistä oppimiskehystä, joka oppii yhdessä ennustamaan kolmen parin (oriQ/relQ, oriQ/relC, relQ/relC) suhteet.

**Tulos**

Mitä täydentäviä tehtäviä käytetään monitehtäväoppimisessa?

**Esimerkki 5.1882**

Ehdotetun luokittelumallin validoimiseksi todellisissa joukkoistamisympäristöissä Amazon Mechanical Turk (AMT) -palvelua käytettiin useiden kommentoijien merkintöjen saamiseksi kahdesta suositusta tietokokonaisuudesta: Reuters-21578 BIBREF30 ja LabelMe BIBREF31 . Jotta ehdotettu malli voitaisiin ensin validoida luokitusongelmiin hieman kontrolloidummassa ympäristössä, käytettiin tunnettua 20-Newsgroups-vertailukorpusta BIBREF29 simuloimalla useita annotoijia, joilla oli eritasoista asiantuntemusta.

**Tulos**

mitä tietokokonaisuuksia käytettiin?

**Esimerkki 5.1883**

Näissä malleissa määritellään parametrisoidut samankaltaisuuspisteytysfunktiot $: Q\times T\rightarrow \mathbb {R}$ , jossa $Q$ on luonnollisen kielen kysymysten joukko ja $T$ on loogisten muotojen parafraasien joukko.

**Tulos**

Ottaako neuraalinen pisteytysfunktio syötteenä sekä kysymyksen että loogisen muodon?

**Esimerkki 5.1884**

Ensinnäkin monet uutisartikkelit alkavat toimittajien nimillä, mediatoimistoilla, päivämäärillä tai muulla sisällön kannalta epäolennaisella sisällöllä, esimerkiksi "New York (CNN) -", "Jones Smith, 10. toukokuuta 2018:". Siksi käytämme yksinkertaisia säännöllisiä lausekkeita poistaaksemme nämä etuliitteet. Toiseksi, varmistaaksemme, että tiivistelmä on tiivis ja että artikkeli sisältää riittävästi keskeistä tietoa, säilytämme vain artikkelit, joiden kolme ensimmäistä lausetta sisältävät 10-150 sanaa ja loput 150-1200 sanaa ja jotka sisältävät yhteensä vähintään kuusi lausetta. Tällä tavoin suodatamme pois i) artikkelit, joiden sisältö on liian pitkä, jotta muistin kulutusta voidaan vähentää; ii) hyvin lyhyet johtavat lauseet, joissa on vain vähän tietoa ja jotka eivät todennäköisesti ole hyvä tiivistelmä. Kannustaaksemme mallia tuottamaan abstrakteja tiivistelmiä poistamme myös artikkelit, joissa jokin kolmesta ylimmästä lauseesta toistuu täsmälleen samassa kohdassa artikkelin loppuosassa.Kolmanneksi pyrimme poistamaan artikkelit, joiden kolme ylimpää lausetta eivät välttämättä muodosta relevanttia tiivistelmää.

**Tulos**

Mistä tietojen puhdistus- ja suodatusprosessi koostuu?

**Esimerkki 5.1885**

Menetelmät ::: Baselines ::: Tämä on hyvin samanlainen kuin Sum-QE, mutta nyt $\mathcal {E}$ on pino BiGRU:ita, joilla on oma huomio BIBREF21, BERT-instanssin sijaan. Lopullinen yhteenvetorepresentaatio ($h$) on tuloksena saatujen kontekstitietoisten merkkien upotusten summa ($h = \sum \_i a\_i h\_i$), joka on painotettu niiden itsetarkkailupisteillä ($a\_i$). Meillä on jälleen kolme makua: yksi yksitehtäväinen (BiGRU-ATT-S-1) ja kaksi monitehtäväistä (BiGRU-ATT-M-1 ja BiGRU-ATT-M-5).Menetelmät ::: Baselines ::: ROUGE: Tämä perustaso on ROUGE-versio, joka toimii parhaiten kussakin tietokokonaisuudessa BIBREF13:n tarkastelemien versioiden joukossa. Vaikka ROUGE keskittyy vertais- ja viitetiedostojen välisiin pintapuolisiin yhtäläisyyksiin, odotamme, että pitkiin $n$-grammeihin tai pisimpiin yhteisiin osasekvensseihin perustuvilla ROUGE-versioilla voidaan jossain määrin ottaa huomioon sellaiset ominaisuudet kuin kieliopillisuus, viittausten selkeys ja johdonmukaisuus: Baselines ::: Kielimalli (LM): Vertaisvertailun yhteenvedon osalta kohtuullinen arvio $\mathcal {Q}1$ (kieliopillisuudesta) on esivalmennetun kielimallin palauttama perpleksisyys. Kokeilemme esivalmennetulla GPT-2-mallilla BIBREF22 ja todennäköisyysestimaateilla, jotka BERT voi tuottaa kullekin merkille, kun merkkiä käsitellään peitettynä (BERT-FR-LM). Koska yhteenvedon kieliopillisuutta voi pilata vain muutama huono merkki, laskemme perpleksisyyden ottamalla huomioon vain $k$ huonoimmat (pienimmän LM-todennäköisyyden omaavat) merkit vertaisyhteenvedosta, jossa $k$ on viritetty hyperparametri. methods ::: Baselines ::: Seuraavan lauseen ennustaminen: BERT-koulutus perustuu kahteen tehtävään: peitettyjen merkkien ennustamiseen ja seuraavan lauseen ennustamiseen. Jälkimmäinen näyttää olevan linjassa $\mathcal {Q}3$ (Referential Clarity), $\mathcal {Q}4$ (Focus) ja $\mathcal {Q}5$ (Structure & Coherence) määritelmien kanssa. Intuitiivisesti, kun lause seuraa toista suurella todennäköisyydellä, siihen pitäisi sisältyä selkeitä viittausilmaisuja ja sen pitäisi säilyttää tekstin fokus ja paikallinen koherenssi. Siksi käytämme esivalmennettua BERT-mallia (BERT-FR-NS) laskeaksemme kunkin yhteenvedon lausetason monitulkintaisuuden: missä $p(s\_i|s\_{i-1})$ on todennäköisyys, jonka BERT antaa lauseiden $\left< s\_{i-1}, s \right>$ järjestykselle, ja $n$ on lauseiden lukumäärä vertaisyhteenvedossa.

**Tulos**

Mitä yksinkertaisempia malleja he tarkastelevat?

**Esimerkki 5.1886**

Tässä työssä ehdotamme vastaamattomien kysymysten luomista muokkaamalla vastauksen sisältävää kysymystä ja ehdollistamalla vastauksen sisältävän kappaleen. Näin luodut vastaamattomat kysymykset ovat leksikaalisesti samankaltaisempia ja asiaankuuluvampia asiayhteyden kannalta. Lisäksi käyttämällä vastattavaa kysymystä prototyyppinä ja sen vastausväliä uskottavana vastauksena, luodut esimerkit voivat tarjota erottelevamman harjoittelusignaalin kysymyksiin vastaamisen mallille.

**Tulos**

Edellyttääkö heidän lähestymistapansa tietokokonaisuutta, jossa vastaamattomat kysymykset on yhdistetty vastaaviin kysymyksiin, joihin voidaan vastata?

**Esimerkki 5.1887**

Arviointimittari. Koska ongelmamme on epätasapainoinen, käytämme arviointimittarina F1-pistemäärää. Merkintämenetelmää käytettäessä keskiarvoistamme sellaisten sanojen merkinnät, joilla on sama kantasanaversio, jotta saamme yhden ennusteen kantasanalle. Perustason määrittämiseksi tarkastelemme satunnaista menetelmää, joka ennustaa positiivisen merkinnän 0,15 todennäköisyydellä (positiivisten tapausten perusprosentti).

**Tulos**

Mitä mittareita käytetään tämän tehtävän arvioinnissa?

**Esimerkki 5.1888**

Alkuperäisen WikiQA:n tulokset osoittavat, että kaikki kolme siirto-oppimismenetelmää eivät ainoastaan paranna tuloksia, vaan myös heikentävät F1-tulosta. Tämä johtuu siitä, että muut tietokokonaisuudet eivät pysty lisäämään uutta tietoa alkuperäiseen tietokokonaisuuteen tai ne ilmeisesti lisäävät turhaa tietoa, joka ei ole samanlaista kuin kohdetietokokonaisuus.

**Tulos**

Onko siirtäminen haittaa suorituskykyä on korporaatioiden eivät liity?

**Esimerkki 5.1889**

Syötämme nämä esitykset klusterointialgoritmiin, joka tuottaa tarkalleen yksitoista viiden kaupungin klusteria, ja arvioimme niitä sekä Calvinon alkuperäisiä merkintöjä että joukkoruokiteltuja ihmisten tekemiä arvioita vastaan. Vaikka yleinen korrelaatio Calvinon merkintöjen kanssa on alhainen, sekä tietokoneet että ihmiset pystyvät tunnistamaan luotettavasti joitakin konkreettisiin kohteisiin liittyviä temaattisia ryhmiä.

**Tulos**

Miten ne saavat inhimillisiä tuomioita?

**Esimerkki 5.1890**

Kun S-V-O on luotu, Text2Visual tarjoaa käyttäjille visuaalisia komponentteja, jotka välittävät S-V-O-tekstin merkitykset.

**Tulos**

Sisältyykö heidän ratkaisuunsa kuvien ja tekstin yhdistäminen?

**Esimerkki 5.1891**

GhostVLAD on laajennus NetVLAD-lähestymistavasta, jota käsittelimme edellisessä jaksossa. GhostVLAD toimii täsmälleen samalla tavalla kuin NetVLAD, paitsi että se lisää Ghost-klustereita NetVLAD-klustereiden rinnalle. Nyt meillä on siis K:n klusterin sijasta K+G:n määrä klustereita. Ghost-klusterit lisätään, jotta kaikki meluisa tai epäolennainen sisältö voidaan kuvata haamuklustereiksi, eikä niitä oteta mukaan ominaisuuksien yhdistämisvaiheeseen, kuten kuvassa 1 (oikealla) on esitetty.

**Tulos**

Mikä on GhostVLAD-lähestymistapa?

**Esimerkki 5.1892**

Esittelemme kaksi käyttötapausta, joiden pohjalta Seshatia kehitettiin: kliiniset haastattelut ja päiväkohtaiset lapsikohtaiset nauhoitukset.

**Tulos**

Kokeilivatko he työkalua?

**Esimerkki 5.1893**

käytämme terveen järjen mukaista määritelmää rasistisesta kielestä, johon kuuluvat kaikki negatiiviset lausahdukset, negatiiviset yleistykset ja loukkaukset, jotka koskevat etnistä alkuperää, kansallisuutta, uskontoa ja kulttuuria.

**Tulos**

miten he kysyivät, oliko twiitti rasistinen?

**Esimerkki 5.1894**

Tämä työ tarjoaa myös uuden INLINEFORM0-tietokannan, joka koostuu vapaamuotoisista luonnollisen kielen ohjeista ja korkean tason navigointisuunnitelmista. Tämä tietokokonaisuus kerättiin Mechanical Turkin kautta käyttäen 100 simuloitua ympäristöä ja vastaavaa topologista karttaa, ja tietojemme mukaan se on ensimmäinen laatuaan käyttäytymisnavigointia varten.

**Tulos**

Käyttivätkö kirjoittajat joukkoistamisalustaa?

**Esimerkki 5.1895**

Matkustamon AV-tietoaineistomme sisältää 30 tuntia multimodaalista dataa, joka on kerätty 30 matkustajalta (15 naiselta ja 15 mieheltä) 20 kyydissä/sessiossa. Matkustajien 10 erilaista aikomusta on tunnistettu ja kommentoitu seuraavasti: Muun muassa seuraavat: Aseta/muuta määränpäätä, Aseta/muuta reittiä (mukaan luettuna käännöksen mukaan annetut ohjeet), Aja nopeammin, Aja hitaammin, Pysähdy, Pysäköi, Pysähdy sivuun, Pysähdy pois, Avaa ovi ja Muut (musiikin/radion kytkeminen päälle/pois, ikkunan/ tavaratilan avaaminen/sulkeminen, ilmastointilaitteen/lämpötilan muuttaminen, kartan näyttäminen jne.). Asiaankuuluvat lähtö- ja saapumisajat yksilöidään ja merkitään seuraavasti: Sijainti, sijainti/suunta, kohde, aikaohjaus, henkilö, ele/katse (tämä, tuo, tuolla jne.) ja ei mitään. Lausetason aikomustyyppien ja niiden paikkojen lisäksi myös sanatason aikomusavainsanat merkitään aikomukseksi.

**Tulos**

Mitä aikomuksia paperi tutkii?

**Esimerkki 5.1896**

Rekrytoimme 176 AMT-työntekijää osallistumaan käsitteellistämistehtävään.

**Tulos**

Mikä oli työntekijöille annettu tehtävä?

**Tulos**

Mitä joukkoistamisalustaa käytettiin?

**Esimerkki 5.1897**

Päätelmä

**Tulos**

Mitä on kolmiomittaus?

**Esimerkki 5.1898**

Tarjoamalla ohjatulle useita dialogivaihtoehtoja (vapaan tekstin lisäksi) ohjasimme keskustelua ja pystyimme ottamaan käyttöön toimia, jotka muuttavat järjestelmän sisäistä tilaa. Tämä tarjoaa useita etuja: Ohjattu vuoropuhelu mahdollistaa asetettujen menettelyjen oppimisen ja vähentää sitä tietomäärää, joka tarvitaan vuoropuhelun hallintaan tarkoitetun koneoppimismallin lähentymiseen.Useiden vuoropuheluvaihtoehtojen tarjoaminen ohjatulle lisää vuorovaikutuksen vauhtia ja mahdollistaa monimutkaisempien skenaarioiden ymmärtämisen ja navigoinnin. Dialogin rakenne: Otimme käyttöön strukturoidut dialogit FSM-koneen (Finite State Machine) avulla, joka hallitsee nykyistä dialogin tilaa ja tarjoaa ohjatulle useita sopivia ja asiaankuuluvia tilasiirtymiä (toimia) vuorovaikutuksen pisteen, maailman tilan ja historian mukaan. Dialogitilojen, siirtymien ja lausumien graafi ladataan, kun järjestelmä käynnistetään, ja jokaisella keskusteluhuoneella on oma dialogitilansa, joka muuttuu toimien kautta. FSM:n tarjoamien siirtymien lisäksi käytettävissä on aina myös muita kiinteitä dialogivaihtoehtoja, kuten "Odota, 2 sekuntia", "Okay" tai "Anteeksi, voisitko toistaa sen?", jotka ovat oikoteitä yleisesti käytetyille dialogitoimille, sekä mahdollisuus kirjoittaa viesti vapaasti. Vuoropuhelussa on useita polkuja päästä samoihin tiloihin, joissa operaattorin hallinnan tai sitoutumisen taso vaihtelee, mikä rikastuttaa keskustelujen heterogeenisuutta.

**Tulos**

Mitä tarkoitetaan puoliksi ohjatulla vuoropuhelulla, mikä osa vuoropuhelusta on ohjattua?

**Esimerkki 5.1899**

Esikoulutettu englanninkielinen NER-malli Rakennamme englanninkielisen NER-järjestelmän BIBREF:n7 mukaisesti. Järjestelmä käyttää kaksisuuntaista LSTM-mallia merkkitason kielimallina, joka ottaa kontekstitietoa sanojen upottamista varten. Englannin kielen esivalmennetussa NER-järjestelmässä käytämme Flairin oletusarvoista NER-mallia.

**Tulos**

Mitä valmiiksi koulutettua englanninkielistä NER-mallia he käyttävät?

**Esimerkki 5.1900**

Huomaamme, että visuaalinen huomio on hyvin harva, sillä vain yhteen lähdekoodaukseen kiinnitetään huomiota (suurin visuaalinen huomio lähdekoodausten osalta koko testijoukossa on keskimäärin 0,99 ja keskihajonta 0,015), mikä rajoittaa modulaation käyttöä. Molemmissa tapauksissa havaitsemme, että huomion visuaalinen komponentti ei ole oppinut mitään vaihtelua lähdekoodausten yli, mikä taas viittaa siihen, että visuaaliset upotukset eivät sovellu parantamaan token-tason erottelukykyä ennustamisen aikana. Huomaamme tämän olevan johdonmukaista eripituisten lauseiden kohdalla.

**Tulos**

Mikä on niiden huomion jakautumisen analyysin tulos?

**Esimerkki 5.1901**

Tuloksena saatu 46 asiakirjan joukko muodostaa perusjoukkomme. Huomaa, että näiden asiakirjojen koko vaihtelee suuresti, joten tuloksena oleva korpus on monipuolinen, mutta ei tasapainoinen koon suhteen (taulukko TABREF43).

**Tulos**

Kuinka suuria PIE-tietoaineistoja saadaan sanakirjoista?

**Esimerkki 5.1902**

Tutkimme molempia menetelmiä joko erikseen tai yhdistettynä kahdella käännössuunnalla (En-It ja En-De), joissa kohteen pituus on keskimäärin pidempi kuin lähteen pituus. En-It, En-De molempiin suuntiin.

**Tulos**

Mihin kieliin he keskittyvät?

**Esimerkki 5.1903**

Tässä työssä esiteltiin MVCNN, uusi CNN-arkkitehtuuri lauseiden luokitteluun. Siinä yhdistyvät monikanavainen alustaminen - käytetään erilaisia versioita esivalmennetuista sanojen upotuksista - ja vaihtelevan kokoiset suodattimet - monisanaisia lauseita koskevat piirteet uutetaan vaihtelevan kokoisilla konvoluutiosuodattimilla.

**Tulos**

Millainen MVCNN on verrattuna CNN:ään?

**Esimerkki 5.1904**

Tehtävä: Kun annetaan kuvateksti ja kuvapari (jos käytetään), tavoitteena on merkitä jokainen kuvatekstin merkki BIO-kaavalla (B: alku, I: sisällä, O: ulkopuolella) BIBREF27 .

**Tulos**

Voivatko SnapCaptionsin nimetyt yksiköt olla epäsovinnaisia?

**Esimerkki 5.1905**

Videokysymyksiin vastaaminen on uusi tehtävä, ja tietojemme mukaan mitään mallia ei ole suunniteltu erityisesti tätä tehtävää varten. Ensimmäisenä askeleena kohti tämän ongelman ratkaisemista arvioimme muihin laadunvarmistustehtäviin, kuten lausetason ennustustehtävään ja kahteen segmenttien hakutehtävään, kehitettyjen uusimpien mallien suorituskykyä. Perustasot ::: Baseline1: Baseline1 ennustaa (alkavan lauseen indeksi, päättyvän lauseen indeksi), kun kyseessä on transkriptio (lauseiden sarja) ja kysymys. Malli perustuu RaSor BIBREF13 -malliin, joka on kehitetty SQuAD QA -tehtävää BIBREF6 varten. RaSor ketjuttaa aloittavan ja päättyvän sanan upotusvektorit edustamaan jänneväliä. Tämän ajatuksen mukaisesti Baseline1 edustaa lauseiden jaksoa ketjuttamalla aloitus- ja lopetuslauseiden vektorit. Kuvion FIGREF15 vasemmanpuoleinen kaavio havainnollistaa Baseline1-mallia.Model. Malli ottaa vastaan kaksi syötettä, transkripti, $\lbrace s\_1, s\_2, ... s\_n\rbrace $, jossa $s\_i$ on yksittäisiä lauseita, ja kysymys, $q$. Tuloksena saadaan span-pisteet, $y$, eli pisteet kaikista mahdollisista span-pisteistä. GLoVe BIBREF14:ää käytetään transkriptin ja kysymysten sanojen esittämiseen. Käytämme kahta bi-LSTM:ää BIBREF15 koodaamaan transkriptiota.missä n on lauseiden lukumäärä . Kulkutason koodauksen tuloksena $p$ on vektorijakso $p\_i$, joka edustaa kunkin lauseen latenttia merkitystä. Sitten malli yhdistää jokaisen lauseen upotusparin ($p\_i$, $p\_j$) luodakseen span embeddingin.jossa [$\cdot $,$\cdot $] tarkoittaa ketjuttamista. Lopuksi käytämme yksikerroksista feed forward -verkkoa laskemaan kunkin spanin ja kysymyksen välisen pistemäärän.Koulutuksessa käytämme ristiinentropiaa tavoitefunktiona. Testauksessa vastaukseksi valitaan korkeimman pistemäärän saanut span. Peruslinjat ::: Baseline2: Tarkastelemme myös yksinkertaisempaa tehtävää, kun ongelmamme muotoillaan hakutehtäväksi. Tarkemmin sanottuna annoimme mallille pelkän transkriptin lisäksi myös segmentointitiedot, jotka luotiin tiedonkeruulausekkeen aikana (ks. kohta SECREF3). Huomaa, että jokainen segmentti vastaa vastausehdokasta. Tehtävänä on sitten valita paras segmentti annettua kyselyä varten. Tämä tehtävä on helpompi kuin Baseline1:n tehtävä, koska segmentointitiedot annetaan mallille. Toisin kuin Baseline1, se ei kuitenkaan pysty palauttamaan vastausskaalaa eri rakeisuuksilla. Baseline2 perustuu Attentive LSTM BIBREF17 -malliin, joka on kehitetty InsuranceQA-tehtävää varten. Kuvan FIGREF15 oikeanpuoleinen kaavio havainnollistaa Baseline2-mallia.Model. Kaksi syötettä, $s$ ja $q$, edustavat segmenttitekstiä ja kysymystä. Malli koodaa ensin nämä kaksi syötettä.$h^s$ painotetaan sitten uudelleen huomiopainojen avulla.missä $\odot$ tarkoittaa elementtiviisasta kertolaskuoperaatiota. Lopullinen pistemäärä lasketaan käyttämällä yksikerroksista feed-forward-verkkoa.Koulutuksen aikana malli tarvitsee negatiivisia näytteitä. Kunkin positiivisen esimerkin (kysymys, perustotuussegmentti) osalta kaikkia muita saman transkriptin segmenttejä käytetään negatiivisina näytteinä. Tavoitefunktiona käytetään ristikkäistä entropiaa. Peruslinjat ::: Baseline3: Laskemme segmentin ja kysymyksen upotusten väliset kosinuskohtaiset samankaltaisuudet laskemalla segmentin ja kysymyksen upotusten väliset kosinuskohtaiset samankaltaisuudet. Tässä tehtävässä haluamme kuitenkin testata segmenttien hakutarkkuutta, kunhan ensin haemme oikean videon 76 videon joukosta. Ensin luodaan TF-IDF-merkinnät koko videon transkripteille ja kysymyksille. Seuraavassa vaiheessa haetaan ne videot, joiden kosini-etäisyys videokirjoitusten ja kysymyksen välillä on pienin. Tämän jälkeen suodatamme ja tallennamme kymmenen parasta videota, mikä vähentää seuraavassa vaiheessa tarvittavien laskutoimitusten määrää. Lopuksi laskemme kysymyksen ja niiden segmenttien väliset kosinusetäisyydet, jotka kuuluvat suodatettuihin 10 parhaaseen videoon, ja merkitsemme kysymyksen oikeaksi, jos se löytyy näistä videoista. Vaikka tehtävä on laskennallisesti vähemmän kallis kuin edellinen perustaso, emme opi segmenttien esityksiä, koska tämä tehtävä on yksinkertainen hakutehtävä, joka perustuu TF-IDF-merkintöihin. Kaksi ensimmäistä syötettä ovat kysymys q ja videon transkriptio v, jotka on koodattu niiden TF-IDF-vektoreilla: BIBREF18: Tämän jälkeen suodatetaan 10 parasta videokirjoitusta (76:sta), joilla on pienin kosinusetäisyys, ja lasketaan niiden segmenttien TF-IDF-vektorit, Stop10n, jossa n = 10. Tämän jälkeen lasketaan TF-IDF-vektorit. Toistamme prosessin vastaaville segmenteille: valitsemme segmentin, jonka kosinietäisyys kyselyyn on pienin.

**Tulos**

Mitä perusalgoritmeja esiteltiin?

**Esimerkki 5.1906**

Sovellamme Hania tällaisiin keskusteluihin peräkkäin syöttämällä jokaisen käyttäjän vuoron Hanille sitä mukaa, kun ne tapahtuvat, jotta voidaan määrittää, pitäisikö keskustelun eskaloitua. Jos näin on, käyttäjä siirretään live-chatin edustajalle keskustelun jatkamista varten. Käyttäjän on odotettava, että inhimillinen edustaja tarkastelee IVA-keskusteluhistoriaa ja jatkaa epäonnistunutta tehtävää.

**Tulos**

Miten he keräävät ihmisten arvioita?

**Esimerkki 5.1907**

Latasimme 76 videota kuvankäsittelyohjelmasta kertovalta opetusohjelmasivustolta.

**Tulos**

Mikä on kolmikoiden lähde?

**Esimerkki 5.1908**

Tässä yhteydessä ehdotamme vankkaa monikielistä tunteiden analysointimenetelmää, jota testattiin kahdeksalla eri kielellä: Espanjaa, englantia, italiaa, arabiaa, saksaa, portugalia, venäjää ja ruotsia.

**Tulos**

Mitä kahdeksaa kieltä raportoidaan?

**Esimerkki 5.1909**

CRF-moduuli saavutti parhaan tuloksen thaimaalaisen lauseen segmentointitehtävässä BIBREF8 , joten otamme Bi-LSTM-CRF-mallin perustasoksi.

**Tulos**

Mitä syväoppimisarkkitehtuuria he käyttävät lauseiden segmentointiin?

**Esimerkki 5.1910**

Testit ja kultainen standardi aineisto sisäistä arviointia vartenKäyttäen kultaista standardi aineistoa (kuvattu alla), suoritimme kolmenlaisia testejä:Luokanjäsenyystestit: kahden samaan semanttiseen luokkaan kuuluvan jäsenen (esim. "Vuoden kuukaudet", "Portugalin kaupungit", "hymiöt") upotusten pitäisi olla lähellä toisiaan, koska niiden oletetaan esiintyvän pääosin samoissa konteksteissa.Luokan erottelukykytesti: tämä on edellisen luokanjäsenyystestin käänteisluku. Eri luokkiin kuuluvien elementtien upotusten pitäisi olla erilaisia, koska eri luokkiin kuuluvien sanojen oletetaan esiintyvän huomattavasti erilaisissa yhteyksissä. sanojen vastaavuustestaus: synonyymejä, antonyymejä, lyhenteitä (esim. "porque" lyhennettynä "pq") ja osittaisia viittauksia (esim. "p") vastaavat upotukset. "slb ja benfica") pitäisi olla lähes yhtä suuria, koska molempia vaihtoehtoja on tarkoitus käyttää vaihdettavissa kaikissa yhteyksissä (joko säilyttäen tai kääntäen merkityksen).Näin ollen testeissämme kaksi sanaa katsotaan:erillisiksi, jos vastaavien upotusten kosinus on pienempi kuin 0,70 (tai 0,80).kuuluviksi samaan luokkaan, jos niiden upotusten kosinus on suurempi kuin 0,70 (tai 0,80).samanarvoisiksi, jos upotusten kosinus on suurempi kuin 0,85 (tai 0,95).

**Tulos**

Mitä sisäisiä arviointimittareita käytetään?

**Esimerkki 5.1911**

Laskimme Fliessin INLINEFORM0:n koejoukosta viidelle annotoijalle 21 twiitistä. INLINEFORM1 on .83 kerrokselle A (OFF vs. NOT), mikä osoittaa suurta yksimielisyyttä. Mitä tulee normalisointiin ja anonymisointiin, käyttäjän metatietoja tai Twitter-tunnuksia ei ole tallennettu, ja URL-osoitteet ja Twitter-maininnat on korvattu paikannimillä.

**Tulos**

Kuinka monta kommentoijaa osallistui?

**Esimerkki 5.1912**

Vihamieliset meemit haettiin Google Imagesista lataustyökalun avulla. Käytimme seuraavia kyselyjä kerätäksemme yhteensä 1695 vihameemiä: rasistinen meemi (643 meemiä), juutalainen meemi (551 meemiä) ja muslimimeemi (501 meemiä). Muut kuin vihamemit saatiin Reddit Memes Dataset -aineistosta .

**Tulos**

Mikä on meemien lähde?

**Esimerkki 5.1913**

Käytämme ensin ennusteita, jotka perustuvat kirjallisuudessa aiemmin ehdotettuihin sääntöihin: sanan pituus, foneemien määrä, tavujen määrä, aakkosjärjestys ja taajuus. Keräämme kaikki binomit, mutta teemme ennusteita vain binomeille, jotka esiintyvät yhteensä vähintään 30 kertaa, jaoteltuna subredditin mukaan. Mikään näistä piirteistä ei kuitenkaan näytä olevan erityisen ennustava kaikkialla (taulukko TABREF15). Yksinkertainen lineaarinen regressiomalli ennustaa lähellä satunnaisuutta, mikä vahvistaa näyttöä siitä, että nämä jäädytettyjä binomeja koskevat klassiset säännöt eivät ennusta yleisiä binomeja.

**Tulos**

Miten aiemmin ehdotetut säännöt toimivat hyvin laajassa mittakaavassa?

**Esimerkki 5.1914**

Koska näiden sarjojen odotetaan heijastavan sosiaalisia normeja, niistä käytetään jäljempänä nimitystä "Dos" ja "Don'ts". Vastaavasti jotkut negatiivisista sanoista kuvaavat vain sopimatonta käytöstä, kuten herjaus tai väärinkäytös, kun taas toiset ovat todellisia rikoksia, kuten murha.

**Tulos**

Raportoidaanko tulokset vain englanninkielisistä tiedoista?

**Esimerkki 5.1915**

Keskustelupohjaiset ominaisuudetNämä ominaisuudet pyrkivät hyödyntämään tietokokonaisuuden erityispiirteitä, sillä tietokokonaisuus on puurakenteinen ja kuvastaa keskusteluketjua.Tekstin samankaltaisuus lähdetwiittauksen kanssa: Tekstin samankaltaisuus vastatun twiitin kanssa: twiitin ja keskusteluketjun edellisen twiitin samankaltaisuuden aste (twiitti on vastaus kyseiseen twiittiin).Twiitin syvyys: syvyysarvo saadaan laskemalla solmuja lähteistä (juurista) kuhunkin twiittiin niiden hierarkiassa.

**Tulos**

Mitä keskustelupohjaisia ominaisuuksia käytetään?

**Esimerkki 5.1916**

WN18RR-tietokanta koostuu kahdenlaisista suhteista: symmetrisistä suhteista, kuten $\\_similar\\_to$, jotka yhdistävät luokkaan (b) kuuluvia entiteettejä, ja muista suhteista, kuten $\\_hypernym$ ja $\\_member\\_meronym$, jotka yhdistävät luokkaan (a) kuuluvia entiteettejä. Itse asiassa RotatE pystyy mallintamaan luokkaan (b) kuuluvia entiteettejä erittäin hyvin BIBREF7. HAKE saa kuitenkin 0,021 korkeamman MRR-arvon, 2,4 % korkeamman H@1-arvon ja 2,4 % korkeamman H@3-arvon RotatE:hen verrattuna. FB15k-237-tietokannassa on monimutkaisempia relaatiotyyppejä ja vähemmän entiteettejä kuin WN18RR:ssä ja YAGO3-10:ssä. Vaikka FB15k-237:ssä on suhteita, jotka kuvastavat hierarkiaa, on myös paljon suhteita, kuten "/location/location/time\_zones" ja "/film/film/prequel", jotka eivät johda hierarkiaan. Tämän tietokokonaisuuden ominaispiirteet selittävät sen, miksi ehdotetut mallit eivät ylitä aiempaa huipputekniikkaa yhtä paljon kuin WN18RR- ja YAGO3-10-tietokokonaisuuksien mallit. AGO3-10-tietokanta sisältää entiteettejä, joilla on korkea relaatiokohtainen indegree BIBREF18. Esimerkiksi linkkien ennustustehtävässä $(?, hasGender, male)$ on yli 1000 oikeaa vastausta, mikä tekee tehtävästä haastavan. Onneksi voimme pitää "urosta" hierarkian korkeamman tason entiteettinä ja ennustettuja pääentiteettejä alemman tason entiteetteinä. Näin ollen YAGO3-10 on tietokokonaisuus, jolla on selvästi semanttinen hierarkiaominaisuus, ja voimme olettaa, että ehdottamamme mallit pystyvät toimimaan hyvin tässä tietokokonaisuudessa. Taulukko TABREF19 vahvistaa odotuksemme. Sekä ModE että HAKE ylittävät merkittävästi aiemmat huipputason mallit. Erityisesti HAKE saa 0,050 korkeamman MRR-arvon, 6,0 % korkeamman H@1-arvon ja 4,6 % korkeamman H@3-arvon kuin RotatE.

**Tulos**

Kuinka hyvin HAKE-malli toimii nykyisiä menetelmiä paremmin?

**Esimerkki 5.1917**

Huomioon perustuva sekvenssistä sekvenssiin -malli, jossa emoji-vektori on lisäsyötteenä, kuten MojiTalk BIBREF16:ssä kuvataan. CVAE.RNN-pohjainen ehdollinen varioiva autokooderi dialogivastausten tuottamiseen BIBREF16, joka käyttää monimuuttujaista Gaussin latenttia muuttujaa vastauksen mallintamiseen ja liittää sen koodaajan viimeiseen piilotetun tilan kanssa dekooderin alkutilaksi. Koulutuksen aikana käytetään KL-hehkutusta, varhaisen pysäytyksen strategiaa ja sanapussin apuhäviötä. Käytämme BIBREFin julkaisemaa toteutusta16.

**Tulos**

Mitä muita perusmuotoja kuin vakiomuuntajia käytetään kokeissa?

**Esimerkki 5.1918**

Käytämme kokeissamme julkisesti saatavilla olevaa tietokokonaisuutta KVRET BIBREF5. Tämä tietokokonaisuus sisältää vuoropuheluja kolmelta alueelta: kalenteri, sää ja navigointi (POI), mikä soveltuu eri alojen vuoropuhelukokeiluihin.

**Tulos**

Mitä monialuetietokokonaisuutta käytetään?

**Esimerkki 5.1919**

Poimimme viisi pinta- ja leksikaalista piirrettä, nimittäin sekvenssin pituuden merkkien lukumääränä, sanan keskimääräisen pituuden, tyypin ja merkin suhteen sekä leksikaalisen ja merkkien suhteen (adjektiivien, verbien, substantiivien ja adverbien suhde merkkeihin). Poistimme kaksi piirrettä, jotka käyttävät opittua esitystä: Ensinnäkin saamme lauseen upotusominaisuuden, joka muodostetaan keskiarvottamalla syötetyn lauseen sanojen upotukset. Toiseksi louhimme fastText-edustuksen käyttämällä fastText-kirjastoa samoilla parametreilla kuin Joulin et al. joulin2016bag raportoi.

**Tulos**

Eroavatko ne toisistaan oivallukset, joissa on kyse opituista tai kehitetyistä representaatioista?

**Esimerkki 5.1920**

Käyttäjän aikomusten ja tunteiden ymmärtäminen on äärimmäisen tärkeää, jotta nykyiset älykkäät chatbotit voivat vastata asianmukaisesti ihmisten pyyntöihin. Nykyiset järjestelmät eivät kuitenkaan pysty toimimaan parhaalla mahdollisella tavalla, kun niille esitetään epätäydellisiä tietoja eli lauseita, joista puuttuu sanoja tai joissa on virheellisiä sanoja. Tämä skenaario on todennäköinen, kun otetaan huomioon kirjoittamisessa tehdyt inhimilliset virheet. Itse asiassa on melko naiivia olettaa, että käyttäjät kirjoittavat aina kieliopillisesti täysin oikeita lauseita.

**Tulos**

Miten kirjoittajat määrittelevät tai havainnollistavat "virheelliset sanat"?

**Esimerkki 5.1921**

Adaptiivinen monioppimisjärjestelmä perustuu vahvistusoppimisen (RL) paradigmaan. Kuvassa FIGREF18 havainnollistetaan koko oppimisprosessi. Monikurssioppimisjärjestelmä ajoitetaan sen mukaan, miten malli on suoriutunut validointijoukosta, ja ajoitusmekanismi toimii politiikkana $\pi$, joka on vuorovaikutuksessa vuoropuhelumallin kanssa oppimistilan $s$ hankkimiseksi. Monikurrikulaarisen oppimismekanismin palkkio $m\_t$ osoittaa, kuinka hyvin nykyinen dialogimalli suoriutuu. Positiivinen palkkio on odotettavissa, jos monikurrikulaarinen aikataulutustoimi $a\_t$ parantaa mallin suorituskykyä, ja nykyinen harjoitusnäytteiden pienoiserä otetaan aikataulutustoimen $a\_t$ avulla.

**Tulos**

Miten kehys valitsee automaattisesti erilaisia opetussuunnitelmia kehittyvän oppimisprosessin aikana neuraalisen dialogin luomisen mallin oppimistilanteen mukaan?

**Esimerkki 5.1922**

Parhaiden tulosten (BLSTM-CNN-CRF) perusteella suoritetaan virheanalyysi, joka perustuu viiteen virhetyyppiin (ei poimintaa, ei merkintöjä, väärä alue, väärä tunniste, väärä alue ja tunniste), samalla tavalla kuin BIBREF10:ssä, mutta analysoimme sekä kultaiset että ennustetut tunnisteet (lisätietoja kuvissa 1 ja 2).

**Tulos**

Millaisia virheitä BLSTM-CNN-CRF-järjestelmä tuotti?

**Esimerkki 5.1923**

Ensimmäisessä tehtävässä käytämme SST-2 (Stanford Sentiment Treebank, versio 2) -tietokokonaisuutta BIBREF25 sentimenttianalyysikokeilujen tekemiseen. SST-2 on lauseiden binääriluokittelutietokanta, jossa on train/dev/test-jako ja kahdenlaiset lauseiden merkinnät, positiiviset ja negatiiviset. Tässä kokeilemme Snips-tietokokonaisuutta BIBREF26 , jota käytetään laajasti SLU-tutkimuksessa. Tämä tietokokonaisuus sisältää testattuja puhuttuja lauseita (tekstiä), jotka on luokiteltu johonkin seitsemästä tarkoituksesta.

**Tulos**

Mitä reaalimaailman tietokokonaisuuksia he käyttivät?

**Esimerkki 5.1924**

Kuvion Y-akseli on agentin onnistumisprosentti (mitattuna taitoon johtaneiden dialogien määränä jaettuna dialogien kokonaismäärällä) ja X-akseli on oppimisvaiheiden määrä.

**Tulos**

Miten he mittasivat tehokkuutta?

**Esimerkki 5.1925**

Tutkimme kolmea eri epäjohdonmukaisuuden näkökohtaa ja suunnittelimme mittareita niiden mittaamiseksi. Sanatason ominaisuus, jossa käytetään tf-idf BIBREF22 -merkintää, on lisätty kestävyyden lisäämiseksi.

**Tulos**

Mitä ominaisuuksia ne poimivat?

**Esimerkki 5.1926**

Itsensä huomioiminen kysymyksessä. Koska kysymykseen on integroitu aiempia lausumia, mallin on yhdistettävä aiemmin mainittu käsite suoraan nykyiseen kysymykseen. Tämä on hyödyllistä käsitteiden siirtämisen ja ydinviittausten ratkaisemisen kannalta. Käytämme siis kysymyksen itsehuomiota.

**Tulos**

Sisältyykö malliin coreference ja entailment?

**Esimerkki 5.1927**

Tämän vuoksi emme väitä, että tätä tietokokonaisuutta voidaan pitää perustotuutena.

**Tulos**

Miten väärennettyjen uutisten pohjatotuus muodostetaan?

**Esimerkki 5.1928**

Tämän väitteen tueksi mittaamme järjestelmämme suorituskykyä eri alojen tietokokonaisuuksissa. Arvioinnit tehdään ohjattujen tuomareiden toimesta, jotka ymmärtävät tietopohjan ja arvioivat sitten käyttäjän kyselyjen relevanssia QA-parien (binäärilappujen) kannalta. Kutakin kysely-QA-paria arvioi kaksi tuomaria. Suodatamme pois tiedot, joiden kohdalla tuomarit eivät ole samaa mieltä etiketistä. Chit-chat itsessään voidaan katsoa toimialueeksi. Näin ollen arvioimme suorituskykyä tietyllä KB:lla sekä chit-chat-tietojen kanssa että ilman niitä (taulukon TABREF19 kaksi viimeistä riviä) sekä suorituskykyä pelkillä chit-chat-tiedoilla (taulukon TABREF19 toinen rivi).

**Tulos**

Mitä kokeita kirjoittajat esittävät validoidakseen järjestelmänsä?

**Esimerkki 5.1929**

Parhaiten hahmoyksikön osalta toimi RNN-muunnin, jossa oli ylimääräinen huomiomoduuli, jonka CER-arvo oli 7,8 % ja WER-arvo 17,6 %. Alasanayksiköiden osalta klassinen RNN-transducer, RNN-transducer, johon on liitetty huomio, ja yhteinen CTC-attention -moduuli osoittavat vertailukelpoista suorituskykyä alasanojen virhemäärän ja WER:n osalta, joista ensimmäinen on hieman parempi WER:n osalta (17,4 %) ja jälkimmäisen alhaisempi virheprosentti alasanojen osalta (14,5 %).

**Tulos**

Minkä mallin merkkivirheprosentti on pienin ja minkä sanavirheprosentti on pienin?

**Esimerkki 5.1930**

Käytämme MS COCO-, Bing- ja Flickr-tietoaineistoja BIBREF26:sta kysymysten luomiseksi käytettävän mallin kouluttamiseen. Nämä tietokokonaisuudet sisältävät kuvia koskevia luonnollisia kysymyksiä, joiden tarkoituksena on saada lisätietoja kuvasta. Kuten kuvasta FIGREF8 voidaan nähdä, kysymyksiin ei voida vastata vain katsomalla kuvaa. Kukin lähde sisältää 5 000 kuvaa, joissa on 5 kysymystä kuvaa kohti, joten yhteensä 15 000 kuvaa ja 75 000 kysymystä.

**Tulos**

Kuinka monta kysymystä kuvaa kohti on keskimäärin saatavilla tietokannassa?

**Esimerkki 5.1931**

Laskimme annotoijien välisen sopimuksen Krippendorffin Alpha BIBREF23 -menetelmällä, jossa otetaan huomioon eri annotoijaparit ja tyhjät arvot. Jotta voisimme myös tarkentaa kategorioittain erityistä yhteisymmärrystä, laskimme kunkin luokan keskinäiset F-pisteet. Tätä mittaria käytetään tavallisesti arvioitaessa järjestelmän suorituskykyä luokittain gold standard -aineistolla, mutta sitä voitaisiin soveltaa myös annotaatiopareihin vaihtamalla kahden annotaattorin roolia luokittelijan ja perustotuuden välillä. Vaikka sekä Relevance- että Subject-luokitusten annotointiprosentit ovat $0.71$ ja $0.70$, niiden yhteisymmärryspisteet ovat vain kohtuulliset, $\alpha =0.27$ ja $\alpha =0.29$. Stance- ja Sentiment-luokkien, joihin sisältyy enemmän kategorioita kuin kahteen edelliseen, yhteisymmärrysprosentti on molemmilla 0,54 $. Myös niiden yksimielisyyspisteet ovat kohtuulliset: $\alpha =0,35$ ja $\alpha =0,34$. Keskinäiset F-pisteet osoittavat huomattavia eroja kategorioiden välisessä yhteisymmärryksessä, jossa useimmin annotoidut kategoriat tuottavat yleensä korkeamman tuloksen. Tämä pätee luokkiin Relevant (0,81$), Vaccine (0,79$) ja Positive (0,64$). Negatiivinen kategoria tuottaa yhteisen F-pistemäärän 0,42 $, joka on korkeampi kuin useammin annotoidut kategoriat Neutraali (0,23 $) ja Ei selvä (0,31 $). Havaitsimme, että nämä kategoriat sekoittuvat usein keskenään. Kun näiden kahden luokan annotaatiot on yhdistetty, kantasopimus kasvaa arvoon $\alpha =0.43$.

**Tulos**

Mikä on heidän kommentoidun tietokokonaisuutensa yhteisymmärryspistemäärä?

**Esimerkki 5.1932**

Taulukosta TABREF14 käy ilmi mallimme (monitehtävämuuntaja) tehokkuus verrattuna perusmuuntajaan BIBREF8 .

**Tulos**

mitkä ovat peruslinjat?

**Esimerkki 5.1933**

Taulukossa 1 esitetään kaikkien mallien tulokset WikiLarge-tietokannassa. Näemme, että menetelmämme (NMT+synteettinen) saavuttaa korkeamman BLEU-arvon, alhaisemman FKGL-arvon ja korkean SARI-arvon verrattuna muihin malleihin, lukuun ottamatta Dressiä FKGL:n osalta ja SBMT-SARIa SARI:n osalta. Se vahvisti, että synteettisen datan sisällyttäminen harjoitteluun on erittäin tehokasta ja parantaa perus-NMF-malliamme 2,11 BLEU:lla, 1,7 FKGL:llä ja 1,07 SARI:llä. WikiSmall-tietokannan tulokset esitetään taulukossa 2. NMT:hen verrattuna huomattavia parannuksia (6,37 BLEU) saadaan lisäämällä yksinkertaistettu harjoitusaineisto, jossa on synteettisiä tavallisia lauseita.

**Tulos**

Kuinka paljon heidän mallinsa parani?

**Esimerkki 5.1934**

Span-ilmaisin. Otamme käyttöön monikierroksisen vastausmoduulin span-ilmaisimessa BIBREF1 . Muodollisesti aika-askeleella INLINEFORM0 alueella INLINEFORM1 tila on määritelty INLINEFORM2 . Alkutila INLINEFORM3 on yhteenveto INLINEFORM4 : INLINEFORM5 , jossa INLINEFORM6 . Tässä INLINEFORM7 lasketaan edellisestä tilasta INLINEFORM8 ja muistista INLINEFORM9 : INLINEFORM10 ja INLINEFORM11 . Lopuksi käytetään bilineaarista funktiota, jolla etsitään vastausvälien alku- ja loppupiste kussakin päättelyvaiheessa INLINEFORM12 : DISPLAYFORM0 DISPLAYFORM1 Lopullinen ennuste on kunkin aika-askeleen keskiarvo: INLINEFORM0 . Sovellamme satunnaista pudotusta askeleen tasolla jokaisessa aika-askeleessa harjoittelun aikana, kuten BIBREF1:ssä on tehty.

**Tulos**

Mikä on jännevälien ilmaisimen arkkitehtuuri?

**Esimerkki 5.1935**

Keskitymme tässä Europarl-alueeseen, josta meillä on runsaasti tietoja useilla kielillä, ja käytämme alan sisäisenä harjoitusaineistona Europarl-korpusta BIBREF5 kahta käännössuuntaa varten: Englanti INLINEFORM0 saksa ja englanti INLINEFORM1 ranska. Kun mittaamme out-of-domain-suorituskykyä, käytämme WMT newstest 2014. tai ranskan kielessä käytämme näytteitä News-Commentary-11- ja Wikipedia-korpuksista WMT 2014:n jaetusta käännöstehtävästä sekä korpuksista Multi-UN BIBREF9 ja EU-Bookshop BIBREF10. Saksan osalta käytämme näytteitä News-Commentary-11:stä, Rapidista, Common-Crawlista (WMT 2017) ja Multi-UN:sta (ks. taulukko TABREF5 ).

**Tulos**

mitä tietokokonaisuutta käytetään?

**Esimerkki 5.1936**

Tuloksemme tasoittavat tietä useiden uusien tutkimussuuntien tutkimukselle. Vaikka testasimme virtuaalisten reunojen tehokkuutta tietyssä tekstiluokittelutehtävässä, voisimme laajentaa tätä lähestymistapaa yleisiin luokittelutehtäviin. Sulauttamistekniikoiden järjestelmällinen vertailu voitaisiin tehdä myös muiden viimeaikaisten tekniikoiden BIBREF54, BIBREF55 sisällyttämiseksi. Voisimme myös tunnistaa muita asiaankuuluvia tekniikoita virtuaalisten reunojen luomiseksi, jolloin menetelmää voitaisiin käyttää muissakin verkottuneissa järjestelmissä kuin teksteissä. Verkko voitaisiin esimerkiksi rikastaa graafien sulauttamistekniikoilla saaduilla sulautuksilla. Yksinkertaisemmassa lähestymistavassa voitaisiin myös ottaa huomioon linkkien ennustaminen BIBREF56 virtuaalisten reunojen luomiseksi. Toinen mielenkiintoinen tutkimusryhmä koskee rinnakkaisesiintymisen ja virtuaalisten reunojen erottamista toisistaan, mahdollisesti luomalla uusia verkkomittauksia, joissa otetaan huomioon heterogeeniset linkit.

**Tulos**

Mitä muita luonnollisen prosessoinnin tehtäviä voitaisiin kirjoittajien mielestä tutkia käyttämällä sanojen upotuksia?

**Esimerkki 5.1937**

Tietokokonaisuus jaettiin ajallisesti harjoittelu-, validointi- ja testikokonaisuuksiin, ja hyperparametrit optimoitiin siten, että AUC-ROC-arvo maksimoitiin validointikokonaisuudessa.

**Tulos**

Mitä arviointimittareita käytettiin?

**Esimerkki 5.1938**

Sen sijaan meidän mallimme on koulutettu vain 76 000 kysymyksellä harjoitusjoukossa.

**Tulos**

Kokeilevatko kirjoittajat mallia myös muilla aineistoilla?

**Esimerkki 5.1939**

Käytämme mittauksina vakiomittauksia Precision, Recall ja F1 maininnan tasolla (Micro) ja asiakirjan tasolla (Macro). Näemme, että WW:n keskimääräinen linkitystarkkuus (Micro) on alhaisempi kuin TAC2010:n, ja NCEL päihittää kaikki perusmenetelmät sekä helpoissa että vaikeissa tapauksissa.

**Tulos**

Miten he varmistavat yleistämiskyvyn?

**Esimerkki 5.1940**

Monet keskittyvät sanan vektoriesityksen antamiseen, mutta yhä useammat antavat vektoriesityksen koko lauseesta tai tekstistä. Tämän työn jatkoksi pyrimme kehittämään järjestelmän, joka voi toimia verkossa ja joka suorittaa tekstin klusterointia tekstien upotusten perusteella yksi kerrallaan.

**Tulos**

Miten saapuvaa väittämää käytetään samankaltaisten faktatarkastettujen väittämien hakemiseen?

**Esimerkki 5.1941**

Koska BERT on jo saavuttanut huippuluokan suorituskyvyn kysymysten vastaamisessa, tässä jaksossa vertaamme ehdotettua malliamme huippuluokan kysymysten vastaamismalleihin (eli QANet BIBREF39) ja BERT-Base BIBREF26:een. BERT-mallilla on kaksi versiota: BERT-Base ja BERT-Large, mutta laskentaresurssien puutteen vuoksi voimme verrata vain BERT-Base-mallia BERT-Large-mallin sijaan.

**Tulos**

Mitkä ovat vahvat perusmallit tietyissä tehtävissä?

**Esimerkki 5.1942**

Kokeissamme käytetään E2E NLG -haastetietokokonaisuutta BIBREF21, joka on 50 000 ravintola-alan instanssin joukkoistettu tietokokonaisuus. Mallimme koulutetaan virallisella koulutusjoukolla ja verifioidaan virallisella testausjoukolla.

**Tulos**

Mitä tietokokonaisuuksia he käyttivät?

**Esimerkki 5.1943**

Arvioimme ehdotettuja malleja kolmella keskustelutietokokonaisuudella, kuten MojiTalk BIBREF16, PersonaChat BIBREF11 ja Empathetic-Dialogues BIBREF26.

**Tulos**

Mitä kolmea keskustelutietoaineistoa käytetään arvioinnissa?

**Esimerkki 5.1944**

Ehdotetun interaktiivisen MRC:n (iMRC) keskeinen idea on rajoittaa asiakirjakontekstia, jota malli havaitsee kerrallaan. Tarkemmin sanottuna jaamme tukevan asiakirjan sen sisältämiin lauseisiin ja pidättelemme nämä lauseet mallilta. Kun mallille annetaan kysymys, sen on annettava komentoja, joilla se voi tarkkailla pidätetyn joukon lauseita; varustamme mallit toiminnoilla, kuten Ctrl+F (etsi merkki) ja stop, joilla se voi etsiä osittain tarkkailtuja asiakirjoja. Malli hakee iteratiivisesti ja ehdollistaa jokaisen komennon syötetyn kysymyksen ja aiemmin havaitsemiensa lauseiden perusteella. Tehtävämme edellyttää siis, että mallit "ruokkivat itseään" sen sijaan, että ne syöttäisivät niille tietoa lusikalla. Tämä tekee MRC:stä peräkkäisen päätöksenteko-ongelman, johon voidaan soveltaa vahvistusoppimista (RL).

**Tulos**

Miten he kouluttavat malleja tässä asetelmassa?

**Tulos**

Tarjotaanko malleja koulutettaessa päätöksentekosekvenssejä valvontaan?

**Esimerkki 5.1945**

Molempien tehtävien syöttöaineisto koostuu 36:sta propagandistisesta ja 12:sta ei-propagandistisesta uutistoimituksesta kerätyistä vapaatekstimuotoisista uutisartikkeleista, jotka ammattimaiset kommentoijat ovat kommentoineet.

**Tulos**

Mitä tietokokonaisuutta käytettiin?

**Esimerkki 5.1946**

Tämän ansiosta Macaw voi tukea multimodaalista vuorovaikutusta, kuten tekstiä, puhetta, kuvaa, klikkausta jne.

**Tulos**

Mitä toimintatapoja Macaw tukee?

**Esimerkki 5.1947**

Tietojemme mukaan kaikki nykyiset järjestelmät käsittelevät tätä tehtävää kahden erillisen osatehtävän, eli tapahtumien louhinnan ja ajallisten suhteiden luokittelun, muodostamana putkistona, ja ne myös olettavat, että kultaiset tapahtumat on annettu, kun relaatioluokittelijaa koulutetaan BIBREF0, BIBREF1, BIBREF2, BIBREF3, BIBREF4 ja BIBREF5. Tarkemmin sanottuna he rakensivat päästä päähän -järjestelmiä, jotka ensin poimivat tapahtumat ja sitten ennustavat niiden väliset ajalliset suhteet (kuva FIGREF1). Näissä putkimalleissa tapahtumien poimintavirheet leviävät suhteiden luokitteluvaiheeseen, eikä niitä voida korjata jälkikäteen. Ensimmäinen panoksemme on ehdotus yhteiseksi malliksi, joka poimii sekä tapahtumia että ajallisia suhteita samanaikaisesti (ks. kuva FIGREF1).

**Tulos**

Onko tämä ensimmäinen julkaisu, jossa ehdotetaan yhteistä mallia tapahtumien ja ajallisten suhteiden erottamiseen?

**Esimerkki 5.1948**

Käytimme SemEval 2010 -tehtävän 8 BIBREF8 relaatioluokitteluaineistoa. Se koostuu lauseista, jotka on manuaalisesti merkitty 19 relaatiolla (9 suunnattua relaatiota ja yksi keinotekoinen luokka Other). Harjoitusjoukkona on käytetty 8 000 lausetta ja testijoukkona 2 717 lausetta.

**Tulos**

Millä aineistolla he kouluttavat mallinsa?

**Esimerkki 5.1949**

Akkadin kieli on jaettu aineistossa kuuteen murteeseen: Vanha babylonialainen, keskibabylonialainen perifeerinen, standardibabylonialainen, uusbabylonialainen, myöhäisbabylonialainen ja uusassyrialainen BIBREF14 .

**Tulos**

Mikä kieli on jaettu kuuteen murteeseen paperissa mainitussa tehtävässä?

**Esimerkki 5.1950**

Vaikka täydellisiä vastineita ei ole olemassa, päätimme valita parhaaksi vastineeksi Quoran kaksoiskysymyksiä koskevan tietokokonaisuuden BIBREF22. Tutkiaksemme sulautuksia laskimme euklidisen etäisyyden kahden kysymyksen välillä käyttämällä erilaisia sulautuksia, jotta voimme tutkia semanttisesti samankaltaisten ja erilaisten kysymysten välistä etäisyyttä.

**Tulos**

Mitä olemassa olevaa korpusta käytetään vertailukohtana näissä kokeissa?

**Esimerkki 5.1951**

Klusterointituloksen mittaamiseen käytetään kahta mittaria, tarkkuutta (ACC) ja normalisoitua keskinäistä informaatiota (NMI) BIBREF38 , BIBREF48 .

**Tulos**

Mitä arviointimittareita käytettiin?

**Esimerkki 5.1952**

Esitämme mielipidesuositusten lopulliset tulokset, joissa vertaamme ehdotettua malliamme seuraaviin uusimpiin perusjärjestelmiin: RS-Average on laajalti hyväksytty perusjärjestelmä (esim, Yelp.com), joka käyttää keskiarvotettuja arvostelupisteitä lopullisena pistemääränä.RS-Linear arvioi arvostelupistemäärän, jonka käyttäjä antaisi INLINEFORM0 BIBREF49 , jossa INLINEFORM1 ja INLINEFORM2 ovat käyttäjän INLINEFORM3 ja tuotteen INLINEFORM4 koulutuspoikkeamat.RS-Item soveltaa INLINEFORM0 NN:ää arvostelupistemäärän BIBREF50 arvioimiseksi. Valitsemme INLINEFORM1:n välisen kosinin samankaltaisuuden mittaamaan tuotteen välistä etäisyyttä.RS-MF on uusin suosittelumalli, joka käyttää matriisifaktorointia ennustamaan luokituspistemäärää BIBREF8 , BIBREF41 , BIBREF25 .Sum-Opinosis käyttää graafipohjaista kehystä tuottamaan abstrakteja yhteenvetoja, kun otetaan huomioon redundantit mielipiteet BIBREF51 .Sum-LSTM-Att on uusin neurologinen abstrahoiva tiivistelmäohjelma, joka käyttää tarkkaavaisuuteen perustuvaa neuromallia yhdistämään tietoa useista tekstilähteistä ja tuottamaan tiivistelmiä LSTM-purkua käyttäen BIBREF44 , BIBREF3 Kaikki perusmallit ovat yhden tehtävän malleja, joissa ei oteta huomioon arvosanan ja tiivistelmän ennustamista yhdessä. Tulokset esitetään taulukossa TABREF46 .

**Tulos**

Mitkä ovat perustasot?

**Esimerkki 5.1953**

Arvioimme ehdotettua mallia käyttämällä julkisesti saatavilla olevaa LibriSpeech ASR-korpusta BIBREF23. LibriSpeech-tietokanta koostuu 970 tunnista äänitietoja ja niitä vastaavista tekstitranskripteistä (noin 10 miljoonaa sanamerkkiä) sekä lisäksi 800 miljoonan sanamerkin pelkkä tekstiaineisto.

**Tulos**

Kuinka suuri LibriSpeech-tietokanta on?

**Esimerkki 5.1954**

Multimodaalinen esitys.Yhdistimme teksti- ja kuvaesitykset kahdella yksinkertaisella tavalla. Ensimmäinen tapa on tekstin ja kuvan yhdistäminen (concat). Ennen ketjuttamista sovellimme L2-normalisointia kuhunkin modaliteettiin. Toinen menetelmä on oppia yhteinen tila kahdelle modaliteetille ennen ketjuttamista (projektio).Kummankin modaliteetin projektio oppii $d$-ulottuvuusavaruuden niin, että $h\_{1}, h\_{2} \ in \mathbb {R}^{d}$. Kun multimodaalinen esitys on tuotettu ($h\_{m}$) vasemman ja oikean parin osalta, vektorit liitetään suoraan regressiokerroksiin. Projektiot opitaan päästä päähän regressiokerrosten kanssa ja MSE:tä käytetään tappiofunktiona.

**Tulos**

Mitä multimodaalisia representaatioita kokeissa käytetään?

**Esimerkki 5.1955**

Yleisessä muodossaan tyyppiloginen kielioppi koostuu seuraavista osista:

**Tulos**

Hyväksyykö Grail Prolog-syötteet?

**Esimerkki 5.1956**

Seuraavaksi pyrimme selvittämään, pystyykö valmiiksi koulutettu moniyhteenliitäntäinen yhteistyöjärjestelmä ratkaisemaan RC-tehtäviä nollakuvausasetelmassa.

**Tulos**

Mitä mallia käytetään lähtötasona?

**Esimerkki 5.1957**

Testataksemme tietokokonaisuutemme vaikeusastetta tarkistimme enemmistöluokkatunnisteen ja viiden eri lähestymistapoja käyttävän uusimman NLI-mallin tarkkuuden: BiMPM (Bilateral Multi-Perspective Matching Model; BIBREF31 , BIBREF31 ), ESIM (Enhanced Sequential Inference Model; BIBREF32 , BIBREF32 ), Decomposable Attention Model BIBREF33 , KIM (Knowledge-based Inference Model; BIBREF34 , BIBREF34 ) ja BERT (Bidirectional Encoder Representations from Transformers model; BIBREF35 , BIBREF35 ). BERT:n osalta tarkistimme Wikipedian ja BookCorpusin avulla esivalmennetun ja SNLI:llä ja MultiNLI:llä koulutetun mallin suorituskyvyn.

**Tulos**

Mitä NLI-malleja he analysoivat?

**Esimerkki 5.1958**

Arvioimme CBA:ta, LSA:ta, DCA:ta ja GMMv2b:tä käyttäen MOS-luonnollisuustuomioita (Mean Opinion Score), jotka on tuottanut joukko arvioijia. Pisteet vaihtelevat välillä 1-5, ja 5 tarkoittaa "täysin luonnollista puhetta".

**Tulos**

Miten he vertaavat erilaisia mekanismeja luonnollisuuden kannalta?

**Esimerkki 5.1959**

Taulukossa TABREF1 esitetään yhteenveto edellä mainittujen aiempien tutkimusten määrällisistä tuloksista. Se sisältää mallipohjan, ominaisuuksien kuvauksen, luokitellut kielet ja käytetyn tietokokonaisuuden sekä saavutetun tarkkuuden. Taulukossa luetellaan myös ehdottamiemme mallien kokonaistulokset (ylhäällä). Eri kirjoittajien käyttämät kielet ja niiden lyhenteet ovat englanti (En), espanja (Es), ranska (Fr), saksa (De), venäjä (Ru), italia (It), bengali (Ben), hindi (Hi) ja telegu (Tel).

**Tulos**

Verrataanko suorituskykyä perusmalliin?

**Esimerkki 5.1960**

Tässä työssä pidämme sanoja, jotka voivat osoittaa naapurisanojen ominaisuudet, kontekstuaalisina avainsanoina ja kehitämme lähestymistavan, jolla automaattisesti poimituista kontekstuaalisista avainsanoista voidaan luoda piirteitä.

**Tulos**

Mitä kontekstuaalisia ominaisuuksia käytetään?

**Esimerkki 5.1961**

Tärkein varoitus tämän esivalmennuksen yhteydessä on kuitenkin se, että lähdekooderi koulutetaan käyttämään englanninkielistä dekooderia, kun taas kohdekooderi koulutetaan käyttämään englanninkielisen kooderin - ei lähdekooderin - tuotoksia.

**Tulos**

Käytetäänkö pivot-kieltä kokeissa englanniksi vai jollakin muulla kielellä?

**Esimerkki 5.1962**

Toisin kuin nykyiset pullonkaulat, tässä työssä kohteena on kolme erityyppistä sosiaalista verkostoa (Formspring: kysymys- ja vastausfoorumi, Twitter: mikrobloggaus ja Wikipedia: yhteistoiminnallinen tietovarasto) kolmea verkkokiusaamisen aihetta (henkilökohtainen hyökkäys, rasismi ja seksismi) varten ilman eksplisiittistä ominaisuuksien suunnittelua kehittämällä syväoppimiseen perustuvia malleja yhdessä siirto-oppimisen kanssa.

**Tulos**

Mitä kyberbulling-aiheita he käsittelivät?

**Esimerkki 5.1963**

Tietokokonaisuutemme kattaa 218 videota NALCS:stä ja 103 LMS:stä, eli yhteensä 321 videota viikolta 1 viikolle 9 kevään 2017 turnaussarjassa kummastakin turnauksesta.

**Tulos**

Kuinka suuri esitelty tietokokonaisuus oli?

**Esimerkki 5.1964**

Enemmistö: teksti valitsee suurimman kokoisen etiketin. ESA: BIBREF0:ssa ehdotettu dataton luokittelija. Se kartoittaa sanat (tekstissä ja etikettien nimissä) Wikipedia-artikkelien otsikkoavaruuteen ja vertaa sitten tekstiä etikettien nimiin. Tämä menetelmä ei ole riippuvainen junasta. toteutimme ESA:n 08/01/2019 Wikipedia-dumppiin perustuen. Siinä on noin 6,1M sanaa ja 5,9M artikkelia. Word2Vec BIBREF23: Sekä tekstin että etikettien representaatiot ovat sanojen upotusten lisäys elementtiviisaasti. Tämän jälkeen kosinin samankaltaisuus määrittää etiketit. Tämäkään menetelmä ei perustu trainiin. Binary-BERT: Hienosäädämme BERT:n harjoituksessa, jolloin saadaan binäärinen luokittelija, joka päättää, onko kyseessä seuraamus vai ei; sitten testaamme sitä testissä - valitsemme yhden merkin skenaarioissa merkinnän, jolla on suurin todennäköisyys, ja usean merkin tapauksissa valitsemme kaikki merkinnät, joiden perusteella tehdään "seuraamus"-päätös.

**Tulos**

Mitkä ovat niiden perusmallit?

**Esimerkki 5.1965**

RoBERTa-perusversiomme suoriutuu HotpotQA:sta melko hyvin (77,0 F1), vaikka jokainen kappale käsitellään erikseen, mikä estää kappaleiden välisen päättelyn. Hienosäädämme esivalmennetun mallin, joka ottaa kysymyksen ja useita kappaleita ja ennustaa vastauksen samalla tavalla kuin BIBREF21:n yhden askeleen QA-malli. Käytämme $\textsc {RoBERTa}\_{\textsc {LARGE}}}$ BIBREF24:ää esivalmennettuna alustuksena.

**Tulos**

Mikä on se vahva lähtökohta, jonka tämä työ ylittää?

**Esimerkki 5.1966**

Tietokanta koostuu 90 annotoidusta asiakirjasta, joissa kussakin on 5 viittausta, jotka on luokiteltu 1-5, jossa 1 on vähiten merkityksellinen ja 5 on merkityksellisin, eli yhteensä 450 annotoitua viittausta.

**Tulos**

Mikä on tämän rakennetun korpuksen koko?

**Esimerkki 5.1967**

Kokeissa verrataan GluonCV/NLP:n ja muiden avoimen lähdekoodin toteutusten välisiä mallisuorituksia Caffe-, Caffe2-, Theano- ja TensorFlow-malleilla, mukaan lukien ResNet BIBREF8 ja MobileNet BIBREF9 kuvien luokittelua varten (ImageNet), Faster R-CNN BIBREF10 objektien havaitsemiseen (COCO), Mask R-CNN BIBREF11 instanssien segmentointiin, Simple Pose BIBREF12 asennon estimointiin (COCO), textCNN BIBREF13 tunteiden analysointiin (TREC) ja BERT BIBREF14 kysymyksiin vastaamiseen (SQuAD 1).1), tunteiden analysointiin (SST-2), luonnollisen kielen päättelyyn (MNLI-m) ja parafraasointiin (MRPC).

**Tulos**

Kokeilevatko he työkalupakkeja?

**Esimerkki 5.1968**

Lause todistetaan konstruktiivisesti valitsemalla monipäisen itsetarkkailukerroksen parametrit siten, että se toimii konvoluutiokerroksen tavoin.

**Tulos**

Miten he todistavat, että usean pään itsetarkkailu on vähintään yhtä tehokasta kuin konvoluutiokerros?

**Esimerkki 5.1969**

Teemme laajan joukon kokeita kaksikielisten sanakirjojen induktiota ja yksikielistä ja monikielistä sanojen samankaltaisuutta koskevilla vakiomittareilla sekä ulkoisella tehtävällä, joka on monikielisten hypernymien löytäminen. Sekä yksikielisissä että monikielisissä asetuksissa voimme havaita, että mallimme ovat yleisesti ottaen parempia kuin vastaavat perusmallit. Kuten taulukosta 1 käy ilmi, tarkennusmenetelmämme parantaa jatkuvasti perusmalleja (eli VecMapia ja MUSEa) kaikissa kielipareissa ja kaikilla mittareilla. Ennen kaikkea mallikohtaisissa vertailuissa havaitaan, että ehdotetut muutokset sekä VecMapiin että MUSEen parantavat niiden laatua johdonmukaisesti useimpien mittareiden ja datakokoonpanojen osalta.

**Tulos**

Mitä tehtäviä tämä menetelmä on parantanut?

**Esimerkki 5.1970**

Valvomattomassa skenaariossa samankaltaisuus lasketaan tuotettujen $h\_{L}$- ja $h\_{R}$-lauseen/kuvan esitysten kosinuksena.

**Tulos**

Miten ne laskevat esitysten välisen samankaltaisuuden?

**Esimerkki 5.1971**

Tässä jaksossa tarkastelemme kolmea eri luokkaa tietojen lisäämiseen liittyviä tekniikoita, joita voidaan soveltaa tehokkaasti videotietoaineistoihin. Koulutusdatan koon ja monimuotoisuuden lisäämiseksi voimme luoda uusia äänitiedostoja asettamalla kunkin alkuperäisen äänitiedoston päälle uusia meluisia äänitiedostoja aika-alueella. Monipuolisten meluisten äänitiedostojen saamiseksi käytämme AudioSet-tietokokonaisuutta, joka koostuu 632 äänitapahtumaluokasta ja kokoelmasta, joka sisältää yli 2 miljoonaa manuaalisesti kommentoitua 10 sekunnin mittaista äänileikettä YouTube-videoista BIBREF28. Harkitsemme taajuus- ja aikapeittotekniikoiden - joiden on osoitettu parantavan huomattavasti end-to-end ASR-mallien suorituskykyä BIBREF23 - soveltamista hybridijärjestelmiimme. Vastaavasti niitä voidaan soveltaa verkossa LF-MMI-koulutuksen jokaisen epookin aikana ilman uudelleensuuntausta. luvussa SECREF5 esitetyn äänen segmentoinnin jälkeen), ja laskemme sen log mel -spektrogrammin, jonka ulottuvuus on $\nu $ ja aika-askeleet $\tau $:Taajuusmaskurointia sovelletaan $m\_F$ kertaa, ja joka kerta maskoidaan taajuuskaistat $[f\_0$, $f\_0+ f)$, missä $f$ on näytteistetty alueelta $[0, F]$ ja $f\_0$ on näytteistetty alueelta $[0, \nu - f)$.Aikapeittoa sovelletaan valinnaisesti $m\_T$ kertaa, ja joka kerta peitetään aika-askeleet $[t\_0$, $t\_0+ t)$, jossa $t$ on näytteistetty alueelta $[0, T]$ ja $t\_0$ on näytteistetty alueelta $[0, \tau - t)$. Tietojen lisääminen ::: Sekä nopeus- että tilavuushäiriö jäljittelevät spektrin keskisiirtymiä BIBREF18, BIBREF19. Koulutusdatan nopeushäiriöiden suorittamiseksi tuotetaan kustakin äänitteestä kolme versiota nopeuskertoimilla $0.9$, $1.0$ ja $1.1$. Harjoitusdatan koko kolminkertaistuu näin ollen. Äänenvoimakkuuden häirintää varten jokainen ääni skaalataan satunnaismuuttujalla, joka on poimittu tasaisesta jakaumasta $[0.125, 2]$.

**Tulos**

Mitkä ovat parhaat kielensisäiset tiedonlisäysmenetelmät?

**Esimerkki 5.1972**

Se kehitettiin tekemällä kysely Followerwonk-verkkopalvelun sovellusliittymästä sijaintineutraaleilla alkusanoilla, joita jengiläisten tiedetään käyttävän Twitter-profiileissaan eri puolilla Yhdysvaltoja.

**Tulos**

Raportoivatko kirjoittajat vain englanninkielisistä tietokokonaisuuksista?

**Esimerkki 5.1973**

Tietokanta koostuu yli 0,3 miljoonasta tietueesta, ja se on asetettu saataville tulevaa tutkimusta varten.

**Tulos**

Mikä on tietokokonaisuuden koko?

**Esimerkki 5.1974**

Tietokokonaisuus MSParS on julkaistu NLPCC 2019 -arviointitehtävässä. Koko tietokokonaisuus koostuu 81 826 näytteestä, jotka ovat englannin äidinkielisten puhujien kommentoimia. Niistä 80 prosenttia käytetään harjoitusjoukkona. Niistä 10 prosenttia käytetään validointijoukkona ja loput testijoukkona. Testijoukosta valitaan 3000 kovaa näytettä. Tämän tietokokonaisuuden mittari on täsmällinen vastaavuustarkkuus sekä koko testijoukon että kovan testijoukon osalta. Kukin näyte koostuu kysymyksestä, loogisesta muodosta, parametreista (entiteetti/arvo/tyyppi) ja kysymystyypistä, kuten taulukko TABREF3 osoittaa.

**Tulos**

Tarjoavatko harjoitusaineistot loogisen lomakevalvonnan?

**Esimerkki 5.1975**

Keräämme ihmisten arvioita Amazon Mechanical Turk -palvelussa ParlAI BIBREF18 -palvelun kautta.

**Tulos**

Käyttävätkö ne joukkoistamista ihmisarvioiden keräämiseen?

**Esimerkki 5.1976**

Sanaleikit ovat sanaleikkivitsejä, joissa yksi merkki (esim. sana tai lause) viittaa kahteen tai useampaan merkitykseen hyödyntämällä polysemiaa, homonymiaa tai fonologista samankaltaisuutta toisen merkin kanssa tarkoituksenmukaisesti humoristisen tai retorisen vaikutuksen aikaansaamiseksi BIBREF3 .

**Tulos**

Mitä ovat sanaleikit?

**Esimerkki 5.1977**

Oletetaan, että asiakirjoissa on keskimäärin INLINEFORM0 mainintaa, ja näistä globaaleista malleista NCEL:llä on yllättäen pienin aikakompleksisuus INLINEFORM1, koska se ottaa huomioon vain vierekkäiset maininnat, jossa INLINEFORM2 on ala-GCN-kerrosten määrä, joka osoittaa iteraatiot konvergenssiin asti.

**Tulos**

Käytetäänkö vain vierekkäisten yksiköiden mainintoja vai käytetäänkö joissakin tapauksissa useampia mainintoja (vierekkäisten lisäksi)?

**Esimerkki 5.1978**

Tässä työssä käytämme esivalmennusta kunkin tehtävän merkitsemättömiin tietoihin ja osoitamme, että se voi parantaa luokittelujärjestelmien suorituskykyä.

**Tulos**

Mihin MVCNN liittyy?

**Esimerkki 5.1979**

Voimme siis asettaa rajoituksen, jonka mukaan mallimme esitys syötteen syntaksista ei voi sisältää tätä kontekstista riippumatonta tietoa. Tämä säännönmukaistaminen on ehdottomasti parempi vaihtoehto kuin se, että sanojen merkityksen kaikkien näkökohtien sallitaan siirtyä syötteen syntaksirepresentaatioon. Ilman tällaista rajoitusta kaikille syötteille voitaisiin periaatteessa antaa omat syntaktiset luokkansa.

**Tulos**

Tarkoittaako rajoitettujen hermoyksiköiden olemassaolo sitä, että sanojen merkitykset ovat kiinteitä eri yhteyksissä?

**Esimerkki 5.1980**

Varmistaaksemme ehdotetun mallin tehokkuuden suoritamme useita kokeita kolmella kiinalaisella koneellisella luetun ymmärtämisen tietokokonaisuudella, nimittäin CMRC-2017 BIBREF17 , People's Daily (PD) ja Children Fairy Tales (CFT) BIBREF2. Lisäksi käytämme myös Children's Book Test (CBT) -tietokokonaisuutta BIBREF1 testataksemme yleistämiskykyä monikielisessä tapauksessa.

**Tulos**

mitä julkisia tietokokonaisuuksia käytettiin?

**Esimerkki 5.1981**

Jotta voimme testata tämän uuden arkkitehtuurin tehokkuutta verkkokeskustelujen suistumisen ennustamisessa, kehitämme ja levitämme kaksi uutta tietokokonaisuutta. Ensimmäinen kolminkertaistaa laajasti kuratoidun "Conversations Gone Awry" -tietokannan BIBREF9, jossa Wikipedian keskustelusivun keskustelut on merkitty joukolla sen mukaan, johtavatko ne lopulta henkilökohtaisiin hyökkäyksiin; toinen perustuu suositun ChangeMyView-alipalvelun moderoinnin "in-the-wild" -tilanteeseen, jossa pyritään ennustamaan, joutuuko keskustelu myöhemmin moderaattorin toimenpiteisiin "epäkohteliaan tai vihamielisen" käyttäytymisen vuoksi.

**Tulos**

Mihin kahteen tietokokonaisuuteen mallia sovelletaan?

**Esimerkki 5.1982**

Työmme tärkein anti on siis uusi rinnakkainen data-teksti NLG-korpus, joka (1) on pikemminkin keskustelunomainen kuin tiedonhaku- tai kysymysvastaustyyppinen ja soveltuu siten paremmin avoimen alueen dialogijärjestelmään, (2) edustaa uutta, tutkimatonta aluetta, jolla on kuitenkin erinomaiset mahdollisuudet sovelluksiin keskustelevissa asiamiehissä, ja (3) sisältää korkealaatuisia, manuaalisesti puhdistettuja, ihmisten tuottamia lausumia.

**Tulos**

Miten kirjoittajat varmistivat, että korpus on puhdas, vaikka se on joukkorahoitettu?

**Esimerkki 5.1983**

Arvioidaksemme järjestelmällisesti EHR:ien vapaasta tekstistä poimittujen kliinisten tunnearvojen merkitystä, rakennamme ensin perusmallin käyttämällä strukturoituja ominaisuuksia, jotka ovat samankaltaisia kuin aiemmissa tutkimuksissa, jotka koskevat takaisinottoriskin ennustamista BIBREF6.

**Tulos**

Vertailevatko ne aiempia malleja?

**Esimerkki 5.1984**

Olemme toteuttaneet seuraavat rajapinnat Macaw'ta varten:[leftmargin=\*]File IO: Tämä rajapinta on suunniteltu kokeellisiin tarkoituksiin, kuten keskustelevan hakutekniikan suorituskyvyn arvioimiseen tietokokonaisuudessa, jossa on useita kyselyjä. Tämä ei ole vuorovaikutteinen käyttöliittymä.Standard IO: Tämä vuorovaikutteinen komentorivikäyttöliittymä on suunniteltu kehitystarkoituksiin, jotta voidaan olla vuorovaikutuksessa järjestelmän kanssa, tarkastella lokitietoja ja debugata tai parantaa järjestelmää.Telegram: Tämä interaktiivinen käyttöliittymä on suunniteltu vuorovaikutukseen todellisten käyttäjien kanssa (ks. KUVA 4). Telegram on suosittu pikaviestipalvelu, jonka asiakaspuolen koodi on avointa lähdekoodia. Olemme toteuttaneet Telegram-botin, jota voidaan käyttää eri laitteilla (henkilökohtaiset tietokoneet, tabletit ja matkapuhelimet) ja eri käyttöjärjestelmillä (Android, iOS, Linux, Mac OS ja Windows). Tämä käyttöliittymä mahdollistaa multimodaalisen vuorovaikutuksen (teksti, puhe, klikkaus, kuva). Sitä voidaan käyttää myös pelkkään puheeseen perustuvaan vuorovaikutukseen. Puheentunnistuksessa ja puheen tuottamisessa Macaw tukeutuu online-rajapintoihin, esimerkiksi Google Cloudin ja Microsoft Azuren tarjoamiin palveluihin. Lisäksi Telegramissa on useita suosittuja ryhmiä ja kanavia, mikä mahdollistaa sosiaalisten verkostojen integroinnin edelleen keskustelujärjestelmiin. Katso esimerkiksi Naserin ja Zamanin tutkimus uutisten suosiosta Telegramissa BIBREF12.

**Tulos**

Mikä käyttöliittymä Macawilla on tällä hetkellä?

**Esimerkki 5.1985**

holistic: Perusmalli, joka kartoittaa holistisen kuvan ominaisuuden ja LSTM-koodatun kysymysominaisuuden yhteiseen avaruuteen ja suorittaa elementtiviisaasti kertolaskun niiden välillä.TraAtt: Perinteinen huomiomalli, WTL-mallin BIBREF9 toteutus, jossa käytetään samoja $3\times 3$ alueita kuin SalAtt-mallissa.RegAtt: Alueiden huomiomalli, joka käyttää uutta huomiomenetelmäämme, joka on samanlainen kuin SalAtt-malli, mutta jossa ei ole alueiden esivalintaa.ConAtt: Konvoluutiomalli, jossa SalAtt-mallissa käytetty BiLSTM korvataan painoja jakavalla lineaarisella kartoituksella, joka on toteutettu konvoluutiokerroksella.Lisäksi vertaamme SalAtt-malliamme suosittuihin perusmalleihin i.iBOWIMG BIBREF4 , VQA BIBREF1 , ja uusimpia huomiointimalleja eli WTL BIBREF9 , NMN BIBREF21 , SAN BIBREF14 , AMA BIBREF33 , FDA BIBREF34 , D-NMN BIBREF35 , DMN+ BIBREF8 kahdessa COCO-VQA:n tehtävässä.

**Tulos**

Mihin aiempiin julkaisuihin tämä työ vertaa tuloksiaan?

**Esimerkki 5.1986**

Numeerisesti mallimme parantaa DAR-tarkkuutta verrattuna Bi-LSTM-CRF:ään 2,1 %:lla ja 0,8 %:lla SwDA:ssa ja MRDA:ssa.

**Tulos**

Kuinka paljon paremmat ne ovat kuin uusimmat ratkaisut SWDA:ssa ja MRDA:ssa?

**Esimerkki 5.1987**

Koska yksi tavoitteistamme on verrata automaattisessa käännöksessä käytettävien ydinviittausketjujen ominaisuuksia lähdetekstien ja ihmisen tekemien viittausten ominaisuuksiin, käytämme tietoja ParCorFull-korpuksesta, joka on englannin ja saksan välinen korpus, johon on merkitty täydelliset ydinviittausketjut BIBREF46.

**Tulos**

Mitä kieliä uutis- ja TED-tietoaineistoissa esiintyy?

**Esimerkki 5.1988**

1) Tyypilliset menetelmät. Perusmenetelmiksi valitaan kolme tyypillistä tietämysgraafin sulauttamismenetelmää: TransE, TransR ja TransH. 2) Polkupohjaiset menetelmät. Vertaamme menetelmäämme kahteen tyypilliseen polkupohjaiseen malliin, joihin kuuluvat PTransE ja ALL-PATHS BIBREF18. 3) Attribuutteihin perustuvat menetelmät. Useita uusimpia attribuutteihin perustuvia menetelmiä, kuten R-GCN BIBREF24 ja KR-EAR BIBREF26 , käytetään vertaamaan menetelmiimme kolmessa todellisessa tietokokonaisuudessa.

**Tulos**

Mitä seitsemää uusinta menetelmää käytetään vertailussa?

**Esimerkki 5.1989**

Lisäksi ehdotamme uutta vakiomuotoista kokeiluprotokollaa IBM-UB-1-tietokannalle BIBREF25 (SECREF50 ), jotta lähestymistapojen vertailu olisi tulevaisuudessa helpompaa. IAM-OnDB-tietokanta BIBREF42 on luultavasti käytetyin arviointitietokanta online-käsialan tunnistukseen. Se koostuu 298 523 merkistä 86 272 sanatapauksessa 221 kirjoittajan kirjoittamasta 11 059 sanasta koostuvasta sanakirjasta. Käytämme IAM-OnDB-tietokannan vakiomuotoista erottelua: yksi harjoitussarja, kaksi validointisarjaa ja testisarja, jotka sisältävät 5 363, 1 438, 1 518 ja 3 859 kirjoitettua riviä. Viritämme dekooderin painot käyttäen validointijoukkoa, jossa on 1 438 nimikettä, ja ilmoitamme virheprosentit testijoukon osalta. Esitämme arvioinnin tuotantojärjestelmästämme, joka on koulutettu omilla tietokokonaisuuksillamme ja jota on sovellettu useisiin kirjallisuudesta saataviin julkisesti saatavilla oleviin vertailutietokokonaisuuksiin. Huomattakoon, että kaikissa tässä jaksossa esitetyissä kokeissa arvioimme nykyistä järjestelmäämme ilman mitään virityksiä, jotka on tehty kyseisiä tehtäviä varten.ICDAR-2013 Competition for Online Handwriting Chinese Character Recognition BIBREF45 -kilpailussa esiteltiin tietokokonaisuus yleisimpien kiinalaisten merkkien luokittelua varten. Taulukossa TABREF56 ilmoitetaan virhetasot verrattuna kilpailussa julkaistuihin tuloksiin ja muiden tekemiin uudempiin töihin. ICFHR2018 Competition on Vietnamese Online Handwritten Text Recognition using VNOnDB BIBREF50 -kilpailussa tuotantojärjestelmäämme arvioitiin muita järjestelmiä vastaan. Kilpailussa käytetty järjestelmä on se, joka on raportoitu ja kuvattu tässä asiakirjassa. Lisenssirajoitusten vuoksi emme voineet tehdä kokeita kilpailun harjoitusaineistolla tai erityisiä virityksiä kilpailua varten, mikä ei koskenut muita tässä mainittuja järjestelmiä.

**Tulos**

Mitä tietokokonaisuuksia he käyttivät?

**Esimerkki 5.1990**

Kirjallisuudessa on jo ehdotettu erilaisia NMT:n monikielisiä laajennuksia. Kirjojen BIBREF18 ja BIBREF19 kirjoittajat soveltavat monitehtäväoppimista kouluttaakseen malleja useille kielille. Zoph ja Knight BIBREF20 ehdottavat monilähdemallia, ja BIBREF21 esittelee merkkitason koodaimen, joka on yhteinen useille lähdekielille. Omassa asetuksessamme noudatamme Johnson et al. BIBREF22 ehdottamaa pääideaa. Kyseisen artikkelin kirjoittajat ehdottavat yksinkertaista lisäystä lähdekielen puolella olevalla kielilipukkeella (ks. kuva 2 ), jolla osoitetaan kohdekieli, joka dekooderin on tuotettava. Tämä lippu liitetään tiheään vektoriesitykseen, ja sitä voidaan käyttää käynnistämään valitun kielen tuottaminen. Artikkelin kirjoittajat väittävät, että malli mahdollistaa siirto-oppimisen ja tukee kääntämistä sellaisten kielten välillä, jotka eivät ole eksplisiittisesti käytettävissä koulutuksessa.

**Tulos**

Mitä neuraalisia konekäännösmalleja voidaan oppia siirto-oppimisen kannalta?

**Esimerkki 5.1991**

Ensimmäisessä tehtävässä ihmisarvioijia pyydetään arvioimaan näitä tuotoksia kolmesta ensimmäisestä näkökulmasta, jotka ovat merkityksellisyys, houkuttelevuus ja kielen sujuvuus, Likertin asteikolla 1-10 (kokonaislukuarvot). Houkuttelevuuden osalta kommentoijilta kysytään, kuinka houkuttelevia otsikot ovat.

**Tulos**

Miten vetovoimapisteet mitataan?

**Esimerkki 5.1992**

Saaduista tuloksista käy ilmi, että muuntajamalli saavuttaa korkeamman BLEU-pistemäärän kuin sekä Attention-kooderi-dekooderi että sekvenssi-sekvenssi-malli.

**Tulos**

Millaisia tuloksia he saivat verrattuna uusimpaan tekniikkaan?

**Esimerkki 5.1993**

Kuten kirjallisuudessa BIBREF4 , BIBREF8 on tavallista, käytämme 300-ulotteisia vektoreita upotuksissa, ja kaikki sanavektorit normalisoidaan yksikköpituisiksi ennen arviointia.

**Tulos**

Millaisia ulottuvuuksia sanojen upotukset tuottavat käyttämällä faktorointia?

**Esimerkki 5.1994**

Sanasto muodostetaan koko datan perusteella, joten se sisältää sekä harjoitus- että testausjoukon sanaston.

**Tulos**

Onko sanasto sama kaikissa kielissä?

**Esimerkki 5.1995**

Kultaiseksi standardiksi sentimenttileksikoksi valitsimme manuaalisesti luodun sanaston tšekin BIBREF11 , saksan BIBREF12 , ranskan BIBREF13 , makedonian BIBREF14 ja espanjan BIBREF15 .

**Tulos**

mihin tunnelähteisiin niitä verrataan?

**Esimerkki 5.1996**

Tavanomaisen SMT:n valvotun uudelleenjärjestelyn innoittamana ehdotamme tässä artikkelissa valvottuun tarkkaavaisuuteen perustuvaa NMT-mallia (Supervised Attention based NMT, SA-NMT). Perinteisen SMT:n tapaan käytämme ensin valmiita kohdistimia (GIZA++ BIBREF3 tai fast\_align BIBREF4 jne.) saadaksemme kaksikielisen harjoituskorpuksen kohdistuksen etukäteen. Sitten, kun tätä kohdistustulosta pidetään tarkkaavuuden valvontana, opimme yhdessä tarkkaavuutta ja käännöstä, molemmat valvotusti. Koska tavanomaiset kohdistimet tuottavat laadukkaampia kohdistuksia, odotetaan, että valvotun huomion NMT:n kohdistus paranee, mikä johtaa parempaan käännössuorituskykyyn.

**Tulos**

Mitä perinteisiä linjausmalleja he käyttävät ohjeena?

**Esimerkki 5.1997**

Yhteenvetojen arvioinnissa käytämme vakiomittausta ROGUE. Vertailun vuoksi aiempiin AMR-pohjaisiin tiivistämismenetelmiin ilmoitamme ROGUE-1:n Recall-, Precision- ja INLINEFORM0-pisteet.

**Tulos**

Mitä arviointimenetelmiä käytetään?

**Esimerkki 5.1998**

Wen et al. wensclstm15 ehdottivat hiljattain RNN-pohjaista lähestymistapaa, joka päihitti aiemmat menetelmät useilla mittareilla. Tuotetut lauseet eivät kuitenkaan useinkaan sisältäneet kaikkia haluttuja attribuutteja.

**Tulos**

Miten RNN-pohjaisissa generointimalleissa menetetään jonkin verran tietoa?

**Esimerkki 5.1999**

Indusoituja latentteja rakenteita sisältävien koodaajien on osoitettu hyödyttävän useita tehtäviä, kuten asiakirjojen luokittelua, luonnollisen kielen päättelyä BIBREF12, BIBREF13 ja konekääntämistä BIBREF11.

**Tulos**

Onko näyttöä siitä, että latentteja rakenteita sisältävät kooderit toimivat hyvin muissa tehtävissä?

**Esimerkki 5.2000**

VG-harjoitusjoukkomme koostuu 85 prosentista datasta: 16k kuvaa ja 740k vastaavaa aluekuvausta.

**Tulos**

Kuinka suuria tietoja tämä tutkimus tarjoaa?

**Esimerkki 5.2001**

Hyödynsimme olemassa olevaa, annotoitua Twitter-tietokokonaisuutta, joka on rakennettu masennukseen liittyvien oireiden hierarkkisen mallin BIBREF12 , BIBREF13 perusteella. Tietokanta sisältää 9 473 annotaatiota 9 300 twiitille. Kukin twiitti on annotoitu siten, että siinä ei ole todisteita masennuksesta (esim. "Kansalaiset pelkäävät taloudellista masennusta") tai että siinä on todisteita masennuksesta (esim. "masentunut pettymyksestä").

**Tulos**

Arvioidaanko ne vain englanninkielisillä tietokokonaisuuksilla?

**Esimerkki 5.2002**

Käytämme kahta tietokokonaisuutta, NELL-Onea ja Wiki-Onea, jotka on muodostettu BIBREF-tietokannassa11 .

**Tulos**

Mitä tietokokonaisuuksia käytetään lähestymistavan arvioinnissa?

**Esimerkki 5.2003**

Recurrent +ELMo -mallimme käyttää BIBREF9:n kielimallia tarjoamaan kontekstualisoituja upotuksia edellä esitetylle perusmallille, kuten kirjoittajat suosittelevat.OpenAI GPT -mallimme hienosäätää 12-kerroksista 768-ulotteista 768-ulotteista yksisuuntaista muuntajaa BIBREF:stä27 , joka on esivalmennettu kielimallina Books-korpuksella BIBREF36 .

**Tulos**

käyttivätkö he muita esivalmisteltuja kielimalleja kuin bertiä?

**Esimerkki 5.2004**

Vastausten valinnan tulosten raportoinnissa käytetään MAP- ja MRR-arvoja, mutta vastausten käynnistämistehtävä arvioidaan F1-pistemäärän avulla. MULT-menetelmän tulos raportoidaan taulukossa. 1.

**Tulos**

Onko tarkkuus ainoa mittari, jota he käyttävät järjestelmien vertailussa?

**Esimerkki 5.2005**

Kun annetaan monivalintakysymys $qa$, jossa on kysymysteksti $q$ ja vastausvaihtoehdot A= $\lbrace a\_i\rbrace $ , valitsemme merkityksellisimmät tuplat $T$:stä ja $S$:sta seuraavasti. Valinta Tuple KB:stä: Käytämme käänteistä indeksiä löytääksemme ne 1000 tuplea, joissa on eniten päällekkäisiä merkkejä kysymyksen merkkien $tok(qa).$ kanssa.

**Tulos**

Käytetäänkö kokonaisuuksien yhdistämisprosessia?

**Esimerkki 5.2006**

Aineistomme on kehitetty indeksoimalla ja esikäsittelemällä OSG-verkkofoorumi. Foorumilla on paljon erilaisia ryhmiä, kuten masennus, ahdistus, stressi, parisuhde, syöpä, sukupuolitauti jne.

**Tulos**

Miten he saivat OSG-tietokannan?

**Esimerkki 5.2007**

Arvioimme ehdotettuja malleja neljällä eri NLG-alueella: ravintolan ja hotellin löytäminen, kannettavan tietokoneen ostaminen ja television ostaminen. Ravintola- ja hotellitietokannat kerättiin BIBREF4 -tietokannassa, kun taas kannettavan tietokoneen ja television tietokannat on julkaistu BIBREF22 -tietokannassa, jossa on paljon suurempi syöttöavaruus, mutta vain yksi harjoitusesimerkki kutakin DA:ta varten, joten järjestelmän on opittava käsitteiden osittainen toteuttaminen ja kyettävä yhdistämään ne uudelleen ja soveltamaan niitä näkymättömiin DA:iin.

**Tulos**

Arvioidaanko mallia NLG- vai dialogitietoaineistoilla?

**Esimerkki 5.2008**

Tietokannan rakentaminen: keräsimme Pekingin matkailutietoja verkosta, mukaan lukien hotelli-, nähtävyys- ja ravintola-alueita (jäljempänä näitä kolmea aluetta kutsutaan HAR-aluetiedoiksi). Sen jälkeen käytimme HAR-verkkotunnusten yksiköiden metrotietoja metrotietokannan rakentamiseen. Tavoitteiden luominen: tietokannan perusteella suunniteltiin monialueen tavoitegeneraattori. Verkkotunnusten välinen suhde on kuvattu kahdella tavalla. Yksi on kahden lähellä toisiaan sijaitsevan tavoitteen rajoittaminen. Toinen tapa on käyttää taksia tai metroa kahden HAR-alueilla sijaitsevan, kontekstissa mainitun kohteen välillä kulkemiseen. Vuoropuhelujen kerääminen: Ennen virallisen tiedonkeruun aloittamista vaadimme työntekijöitä tekemään pienen määrän vuoropuheluja ja annoimme heille palautetta vuoropuhelun laadusta. Sitten hyvin koulutetut työntekijät paritettiin keskustelemaan annettujen tavoitteiden mukaisesti. Vuoropuhelujen merkinnät: Käytimme joitakin sääntöjä, joiden avulla merkitsimme automaattisesti vuoropuhelun tekoja käyttäjän tilojen, järjestelmän tilojen ja vuoropuheluhistorian mukaan.

**Tulos**

Miten aineisto kerättiin?

**Esimerkki 5.2009**

NER-kokeiden tulokset raportoidaan erikseen ehdotetun järjestelmän kolmelle eri osalle. Tulokset osoittavat, että ensimmäisessä tapauksessa F1-tulos on harjoitusjoukossa aina yli 97 prosenttia, ja maksimiarvo on 99,65 prosenttia. Molemmilla testijoukoilla suorituskyky laskee ja vaihtelee välillä 94-97 %. UGC:n tapauksessa F1-pistemääriä vertaamalla voidaan havaita, että suorituskyky laskee merkittävästi.

**Tulos**

Millaisia tuloksia he ovat saaneet entiteettien tunnistustehtävässä?

**Esimerkki 5.2010**

Vertaamme lähestymistapaamme kahteen muuhun, joista ensimmäisessä käytetään lähes samaa twiittidataa kuin koulutuksessa, ja toisessa CrowdFlower-tietokokonaisuutta, johon on merkitty tunteita. Ensimmäisessä Wang et al. BIBREF21 latasivat yli 5 miljoonaa twiittiä, jotka sisälsivät yhden 131:stä emotionaalisesta hashtagista, jotka perustuvat Parrottin kolmitasoiseen tunteiden luokitteluun seitsemään kategoriaan: ilo, suru, viha, rakkaus, pelko, kiitollisuus, yllätys.

**Tulos**

Mitä tietoja kokeissa käytetään?

**Esimerkki 5.2011**

Ryhmittelemme mallit sen perusteella, mitä tavoitetehtävää ne optimoivat. Uskomme, että tämä työ voi auttaa ymmärtämään olemassa olevaa kirjallisuutta. Uskomme, että näiden mallien suorituskyky on hyvin riippuvainen siitä, mitä tavoitetehtävää se optimoi - vierekkäisen sanan (twiitin sisäiset suhteet), vierekkäisen twiitin (twiittien väliset suhteet), itse twiitin (autokooderi), strukturoitujen resurssien, kuten parafraasitietokantojen, mallintaminen ja heikko valvonta.

**Tulos**

Miten ne edistävät kirjallisuuden ymmärtämistä osana tavoitetehtäväänsä?

**Esimerkki 5.2012**

Kokeilujemme perusmalli on selitetty Alec Go:n artikkelissa [1]. Malli käyttää kokeessaan Naive Bayes-, SVM- ja Maximum Entropy -luokittelijoita.

**Tulos**

Mihin aiemmin ehdotettuihin menetelmiin tätä menetelmää verrataan?

**Esimerkki 5.2013**

N-GrAM sijoittui ensimmäiselle sijalle kaikissa muissa tapauksissa paitsi kielten moninaisuutta koskevassa tehtävässä. Tässä tapauksessa perusjärjestelmä oli parhaalla sijalla, ja meidän järjestelmämme sijoittui toiseksi pienellä marginaalilla. Järjestelmämme päihitti perustason huomattavasti yhteisessä tehtävässä, sillä perustason tulokset olivat huomattavasti alhaisemmat sukupuolitehtävässä kuin kielilajitehtävässä.

**Tulos**

Missä tehtävässä malli menestyy huonoimmin?

**Esimerkki 5.2014**

Peruskonseptina on käyttää puhelimia erottelevaa mallia tuottamaan kehystason foneettisia piirteitä ja sitten käyttää näitä piirteitä parantamaan RNN LID-järjestelmiä, jotka on alun perin rakennettu raakojen akustisten piirteiden avulla. Ensimmäinen vaihe on siis piirteiden yhdistäminen, ja foneettista piirrettä käytetään aputietona akustisen RNN LID:n apuna. Tätä parannetaan edelleen, sillä lisätutkimuksissa havaittiin, että yksinkertaisempi malli, jossa RNN LID-syötteenä käytetään vain äänteellistä piirrettä, tuottaa vielä parempaa suorituskykyä. Kutsumme tätä foneettisiin piirteisiin perustuvaa RNN-mallia foneettiseksi temporaaliseksi neuraaliseksi LID-lähestymistavaksi eli PTN LID:ksi. Yksinkertaisemman mallirakenteen lisäksi PTN tarjoaa syvällisemmän näkemyksen LID-tehtävään löytämällä uudelleen foneettisen temporaalisen ominaisuuden arvon kielen erottelussa.

**Tulos**

Mikä on tutkielman tärkein anti?

**Esimerkki 5.2015**

Arvioimme SECREF5:ssä kuvattua sekvenssistä sekvenssiin - QRNN-arkkitehtuuria haastavassa neuraalisessa konekäännöstehtävässä, IWSLT-saksan ja englannin käännöksessä, jossa käytetään täysin merkkitason segmentointia.

**Tulos**

Mitä kielipareja käytetään konekääntämisessä?

**Esimerkki 5.2016**

käytämme CORD-19-tietokokonaisuutta BIBREF2, joka sisältää yli 45 000 tieteellistä artikkelia, joista yli 33 000 on kokotekstejä COVID-19:stä, SARS-CoV-2:sta ja niihin liittyvistä koronaviruksista. Kehitämme lauseiden luokittelumenetelmiä kaikkien COVID-19:n radiologisia löydöksiä kuvaavien lauseiden tunnistamiseksi. Teemme kokeita menetelmämme tehokkuuden todentamiseksi. Menetelmämme löytää CORD-19-tietokannasta onnistuneesti joukon kliinisiä löydöksiä, jotka liittyvät läheisesti COVID-19:ään.

**Tulos**

Kuinka laaja on COVID-19-kirjallisuuden kokoelma?

**Esimerkki 5.2017**

Tietokanta koostuu 5000 bahasa indonesian kielisestä arvostelusta.

**Tulos**

Sisältyykö tietokantaan muita kuin englanninkielisiä arvioita?

**Esimerkki 5.2018**

Jotta voidaan mallintaa vastausta, joka on kokoelma span-merkkejä, usean span-merkinnän pää käyttää $\mathtt {BIO}$-merkintäformaattia BIBREF8: $\mathtt {B}$ merkitään span-merkinnän alkuun, $\mathtt {I}$ merkitään span-merkinnän sisälle ja $\mathtt {O}$ merkitään merkkeihin, jotka eivät sisälly span-merkintään. Tällä tavoin saamme sarjan kappaleita, jotka voidaan purkaa lopulliseksi vastaukseksi - kokoelmaksi spaneja.

**Tulos**

Miten he käyttävät sekvenssin merkitsemistä vastatakseen monitulkintakysymyksiin?

**Esimerkki 5.2019**

Wikilinkit voidaan nähdä laajamittaisena, luonnollisesti esiintyvänä, joukkoresursoituna tietokokonaisuutena, jossa tuhannet inhimilliset kommentoijat tarjoavat perustotuuksia kiinnostavista maininnoista. Tämä tarkoittaa, että tietokokonaisuus sisältää monenlaista kohinaa, joka johtuu erityisesti epäjohdonmukaisista yhteyksistä. Valmistelemme tietokokonaisuutemme Wikilinksin paikallisen kontekstin versiosta ja ratkaisemme ground-truth-linkit käyttämällä Wikipedia-dumppia huhtikuulta 2016. Käytämme sivu- ja redirect-taulukoita resoluutioon ja pidämme tietokannan pageid-saraketta Wikipedian entiteettien yksilöllisenä tunnisteena. Hylkäämme maininnat, joiden perustotuutta ei pystytty selvittämään (vain 3 % maininnoista).

**Tulos**

Miten laadunvalvonta suoritettiin, jotta teksti on meluisaa mutta huomautukset ovat tarkkoja?

**Esimerkki 5.2020**

Tämä toiminta-avaruus on suuruusluokkaa $A=\mathcal {O}(|V| \times |O|^2)$, jossa $V$ on toimintaverbien määrä ja $O$ on niiden erillisten objektien määrä maailmassa, joiden kanssa agentti voi olla vuorovaikutuksessa. Koska tämä on liian suuri tila, jota RL-agentti ei voi tehokkaasti tutkia, tietämysgraafia käytetään tämän tilan karsimiseen asettamalla toiminnot paremmuusjärjestykseen niiden esiintymisen perusteella nykyisessä tietämysgraafissa ja graafin objektien välisten suhteiden perusteella, kuten BIBREF7

**Tulos**

Miten aluetiedon siirto esitetään tietämysgraafina?

**Esimerkki 5.2021**

Aiempien töiden BIBREF56 , BIBREF57 , BIBREF17 mukaisesti käytämme MNCut-spektristä klusterointialgoritmia BIBREF58 , jota voidaan soveltaa laajalti samankaltaisiin NLP-tehtäviin, joissa on kyse suuriulotteisista ominaisuusavaruuksista BIBREF59 , BIBREF60 , BIBREF18 .

**Tulos**

Mitä klusterointialgoritmia käytetään VerbNet-erikoistuneiden esitysten päällä?

**Esimerkki 5.2022**

Siksi ehdotamme tässä asiakirjassa monimodaalista STT:tä, jossa on kaksi komponenttia: (i) englanninkielinen ASR-järjestelmä, joka on koulutettu How2-tietokannalla, ja (ii) muuntajaan perustuva BIBREF0-perusteinen visuaalisesti perustuva MMT-järjestelmä.

**Tulos**

Mitä tietokokonaisuutta tässä työssä käytettiin?

**Esimerkki 5.2023**

Diskurssiominaisuuksien johtamiseksi rakennetaan entiteettiruudukko syöttämällä asiakirja NLP-putken läpi, jotta voidaan tunnistaa erottuvat entiteetit. Kaksi erilaista diskurssipiirrettä luodaan täyttämällä olioruudukko joko i) kieliopillisilla suhteilla (GR) tai ii) RST-diskurssisuhteilla (RST). Kun piirre merkitään sellaisenaan INLINEFORM4 , muodostetaan poolausvektori INLINEFORM5 char-bigrammeja varten ja ketjutetaan INLINEFORM6 INLINEFORM7:ään ennen kuin tuloksena oleva vektori syötetään softmax-kerrokselle.

**Tulos**

Miten diskurssiominaisuudet sisällytetään malliin?

**Esimerkki 5.2024**

Verrattuna yhden tehtävän perustasoon suorituskyky parani vähäisten resurssien En-De-tehtävässä ja oli vertailukelpoinen runsaasti resursseja sisältävässä En-Fr-tehtävässä.

**Tulos**

Kuinka suuria kielteisiä vaikutuksia ehdotetuilla tekniikoilla on paljon resursseja vaativiin tehtäviin?

**Esimerkki 5.2025**

Mitä tulee nopeuteen, järjestelmämme pystyi lemmatisoimaan 7,4 miljoonaa sanaa kannettavalla tietokoneella lähes kahdessa minuutissa verrattuna MADAMIRA-järjestelmän 2,5 tuntiin, eli 75 kertaa nopeammin.

**Tulos**

Miten nopeus mitattiin?

**Esimerkki 5.2026**

Vertasimme monikierroksista emotionaalisesti sitouttavaa dialogimalliamme (MEED) kahteen perusmalliin - vaniljaiseen sekvenssistä sekvenssiin -malliin (S2S) ja HRAN:iin.

**Tulos**

Mitä kahta perusmallia käytetään?

**Esimerkki 5.2027**

Kokeilemme pienillä feed-forward-verkoilla neljää erilaista NLP-tehtävää varten: kielen tunnistaminen, puhekielisten osien merkitseminen, sanojen segmentointi ja tilastollisen konekääntämisen esijärjestäminen.

**Tulos**

Millaisissa NLP-tehtävissä kirjoittajat arvioivat feed-forward-verkkoja?

**Esimerkki 5.2028**

Kokeet suoritettiin useilla erityyppisillä luokittelijoilla: Gaussian Naive Bayes, Multinomial Naive Bayes, Decision Trees, Random Forests ja Maximum Entropy -luokitin.

**Tulos**

Mitä mallia he kouluttavat?

**Esimerkki 5.2029**

Kuten kuvasta 1 käy ilmi, kahden tutkimusartikkelin samankaltaisuuden arvioimiseksi tehokkaampi strategia vertailee ja yhdenmukaistaa (paikallisen painotuksen avulla) yksittäisiä tärkeitä sanoja (avainsanoja) tiivistelmäparin sisällä, kun taas muiden sanojen (esim. lopetussanojen) sisältämä informaatio, joka on yleensä vähemmän merkityksellistä, voidaan tehokkaasti jättää huomiotta (painottaa pienemmäksi). Ehdotamme, että opitaan semanttisesti tietoinen Network Embedding (NE), joka sisältää sanatason kohdistamisominaisuuksia, jotka on abstrahoitu tekstisekvensseistä, jotka liittyvät vertikaalipareihin. Kun annetaan lauseiden pari, mallimme yhdenmukaistaa ensin jokaisen sanan toisessa lauseessa toisen lauseen avainsanojen kanssa (joita painotetaan adaptiivisesti ylöspäin huomiomekanismin avulla), jolloin saadaan joukko hienojakoisia vastaavuusvektoreita.

**Tulos**

Mitä tekstisekvenssejä kuhunkin kärkeen liittyy?

**Esimerkki 5.2030**

Ennen SAOKE-tietokokonaisuutta julkaistiin OIE-algoritmien arvioimiseksi BIBREF20:ssä OIE-tehtäviä varten annotoitu tietokokonaisuus, joka sisälsi 3 200 lausetta kahdelta alalta. BIBREF20:n mukaan tietokokonaisuus oli "13 kertaa suurempi kuin edellinen suurin annotoitu Open IE -korpus".

**Tulos**

Mikä on edellisen suurimman OpenIE-tietokannan koko?

**Esimerkki 5.2031**

Analysoimalla suuria määriä vuoropuhelun tekosekvenssejä, jotka korreloivat tiettyihin lopputuloksiin, voidaan johtaa erilaisia sääntöjä, esimerkiksi: "Tietojen pyytämisen jatkaminen keskustelun loppuvaiheessa johtaa usein asiakkaan tyytymättömyyteen." Tämä voidaan sitten kodifioida automatisoituja järjestelmiä varten laadituiksi parhaiden käytäntöjen mallisäännöiksi, kuten "Tietopyyntö olisi esitettävä keskustelun alkuvaiheessa, ja vastauksen, informatiivisen lausunnon tai anteeksipyynnön olisi seurattava keskustelun loppupuolella.". Analyysimme auttaa rajaamaan, miten tiettyjen vuoropuhelulakien käyttö voi todennäköisesti johtaa erilaisiin lopputuloksiin. Havaitsemamme painotukset vaihtelevat tarjotun ymmärryksen määrän suhteen: esimerkiksi lisäavun tarjoaminen keskustelun lopussa tai asiakkaan kiittäminen tuottaa enemmän tyytyväisiä asiakkaita ja enemmän ratkaistuja ongelmia (suhdeluvut ovat yli 6:1). Jotkin tulokset ovat kuitenkin paljon hienovaraisempia: esimerkiksi kyllä-ei-kysymysten esittäminen keskustelun alkuvaiheessa on vahvasti yhteydessä ongelmanratkaisuun (suhde 3:1), mutta niiden esittäminen keskustelun lopussa on yhtä vahvasti yhteydessä tyytymättömiin asiakkaisiin. Kun keskustelun keskivaiheilla annetaan yksityiskohtaisia vastauksia, jotka eivät ole pelkkiä myöntäviä, kieltäviä tai vastauksen kuittaavia vastauksia (esim. vastaus (muu)), saadaan tyytyväisiä asiakkaita, jotka eivät ole turhautuneita. Vastaavasti tietojen pyytäminen keskustelun loppupuolella (mikä viittaa siihen, että lisätietoja tarvitaan vielä keskustelun päättyessä) johtaa tyytymättömiin ja tyytymättömiin asiakkaisiin, ja suhde on vähintään 4:1.

**Tulos**

Mitä malleja ja sääntöjä johdetaan?

**Esimerkki 5.2032**

Samaan aikaan tietty tieto taivutetuista sanamuodoista kontekstissa voi olla hyödyllistä, mutta se häviää lemmatisoinnin aikana, ja tämä johtaa pisteiden laskuun. Tämä tarkoittaa, että lemmatisointi tuo mukanaan sekä etuja että haittoja WSD:lle ELMon avulla.

**Tulos**

Mainitsevatko kirjoittajat mitään haittapuolia siitä, että syötteet lemmatoidaan ennen ELMon harjoittelua?

**Esimerkki 5.2033**

Klusterointituloksen arvioinnissa käytetään Calinski-Harabaszin BIBREF-pistemäärää42 , joka tunnetaan myös nimellä varianssisuhdekriteeri. CH-pistemäärä määritellään klusterin sisäisen hajonnan ja klusterin välisen hajonnan välisenä suhteena. CH-pistemäärän vaihteluväli on $[0, +\infty ]$, ja korkeampi CH-pistemäärä vastaa parempaa klusterointia. Visuaalista validointia varten piirrämme ja tarkastelemme myös opittujen representaatioiden t-Distributed Stochastic Neighbor Embedding (t-SNE) BIBREF52- ja Uniform Manifold Approximation and Projection (UMAP) BIBREF53 -karttoja.

**Tulos**

Miten he arvioivat menetelmäänsä?

**Esimerkki 5.2034**

Tulokset on esitetty kuvassa FIGREF87 eri aiheiden lukumäärille, ja niistä nähdään, että ehdotettu malli on parempi kuin kaikki perusmallit, ja svi-versio suoriutuu niistä parhaiten. Jotta voitaisiin arvioida stokastisen variationaalisen päättelyn (svi) laskennallisia etuja eräalgoritmiin verrattuna, log marginaalinen todennäköisyys (tai log evidenssi) piirrettiin iteraatioiden lukumäärän funktiona. Kuvassa FIGREF88 esitetään tämä vertailu. Ei ole yllättävää, että svi-versio konvergoituu paljon nopeammin suurempiin log marginaalisen todennäköisyyden arvoihin verrattuna eräversioon, mikä kuvastaa svi-algoritmin tehokkuutta.

**Tulos**

Mitkä ovat ehdotetun mallin edut?

**Esimerkki 5.2035**

Rakennamme Twitter-tilien tietokannan, joka perustuu kahteen aiemmissa töissä kommentoituun luetteloon. Muiden kuin tosiasioihin perustuvien tilien osalta käytämme BIBREF1:n 180 Twitter-tilin luetteloa. Toisaalta faktatilien osalta käytämme BIBREF19:n luetteloa, jossa on 32 muuta Twitter-tiliä, joita riippumattomat kolmannet osapuolet pitävät luotettavina.

**Tulos**

Miten he hankkivat tietokokonaisuuden?

**Esimerkki 5.2036**

Arvioidaksemme ehdotettua mallia käyttäytymiseen liittyvillä tiedoilla käytämme pariterapiakorpusta (CoupTher) BIBREF21 ja syöpäparien vuorovaikutustietoaineistoa (Cancer) BIBREF22. Nämä ovat kohdennettuja olosuhteita, joissa käyttäytymisohjattu kielimalli voi tarjota paremman suorituskyvyn. Hyödynnämme Couple's Therapy Corpus -korpusta toimialan sisäisenä kokeellisena korpuksena, koska myös käyttäytymisen luokittelumallimme on koulutettu sillä.

**Tulos**

Millä tietokokonaisuudella malli on koulutettu?

**Esimerkki 5.2037**

Tässä jaksossa kuvataan ensin perusmenetelmä, joka on saanut vaikutteita BIBREF12:n ja BIBREF13:n "align to segment" -menetelmästä. Sen jälkeen ehdotamme kahta laajennusta, jotka antavat mallille segmentointiprosessin kannalta merkityksellisen signaalin, jotta voidaan siirtyä segmentoinnin ja kohdistamisen yhteiseen oppimiseen.

**Tulos**

Ilmoitetaanko asiakirjassa vain linjausta koskeva perustaso?

**Esimerkki 5.2038**

Aivan kuten sec:johdannossa esitettiin, samankaltaisuuden laskeminen täsmällisesti on hankalaa, koska siihen liittyy INLINEFORM0-laskenta. Näin ollen tarkastelemme monte-carlo-approksimaatiota: DISPLAYFORM0jossa INLINEFORM0 on luettelo INLINEFORM1:stä poimituista oliopareista. Käytämme peräkkäistä näytteenottoa saadaksemme INLINEFORM6 , mikä tarkoittaa, että otamme ensin INLINEFORM7:n INLINEFORM8:n perusteella INLINEFORM9:stä ja sitten INLINEFORM10:n INLINEFORM11:n ja INLINEFORM12:n perusteella INLINEFORM13:sta.

**Tulos**

Mitä otantamenetelmää he käyttävät lähestyäkseen olioparien ehdollisten todennäköisyysjakaumien samankaltaisuutta?

**Esimerkki 5.2039**

Valitsimme seuraavat kolme mallia perusmalleiksi: K-means on tunnettu tietojen klusterointialgoritmi, toteutamme algoritmin sklearn-työkalupakilla ja edustamme dokumentit käyttämällä TF-IDF:llä painotettua sanasäkkiä (bag-of-words).LEM BIBREF13 on bayesiläinen mallinnusmenetelmä avoimen toimialueen tapahtumien louhintaan. Se käsittelee tapahtumaa latenttina muuttujana ja mallintaa tapahtuman syntymisen sen yksittäisten tapahtumaelementtien yhteisjakaumana. Toteutamme algoritmin oletuskonfiguraatiolla.DPEMM BIBREF14 on ei-parametrinen seosmalli tapahtumien louhintaan. Se korjaa LEM:n rajoituksen, jonka mukaan tapahtumien lukumäärä on tiedettävä etukäteen. Toteutamme mallin oletuskonfiguraatiolla.

**Tulos**

Mitä peruslähestymistapoja tämä lähestymistapa päihittää?

**Esimerkki 5.2040**

Word2vec-esitys on paljon parempi, kehittyneempi ja uudempi tekniikka, joka toimii kartoittamalla sanat 300-ulotteiseen vektoriesitykseen.

**Tulos**

Mikä on sulautumien ulottuvuus?

**Esimerkki 5.2041**

Käytämme kokeissamme kolmea tietokokonaisuutta: IWSLT14 saksa-englanti, turkki-englanti ja WMT14 englanti-saksa.

**Tulos**

mitä tietokokonaisuuksia käytettiin?

**Tulos**

mitä kielipareja tutkitaan?

**Esimerkki 5.2042**

Tässä artikkelissa ehdotamme parannettua RNN-T-mallia, jossa on kielellinen ennakkoasenne, ongelman ratkaisemiseksi. Malli koulutetaan ennustamaan kielitunnuksia sekä alasanoja. Varmistaaksemme, että malli voi oppia CS-tietoa, lisäämme kielitunnukset CS-kohtaan transkriptiossa, kuten kuvassa 1 on esitetty.

**Tulos**

Miten he saavat kielelliset identiteetit?

**Esimerkki 5.2043**

Taulukon TABREF31 tulokset voivat antaa käsityksen siitä, että INLINEFORM0 on vain skaalattu INLINEFORM1 . Vaikka on totta, että ne osoittavat lineaarista korrelaatiota, INLINEFORM2 voi tuottaa erilaisen järjestelmäsijoituksen kuin INLINEFORM3, kun otetaan huomioon sen noudattama integraalinen moniviittausperiaate. Se, mitä pidämme INLINEFORM4:ssä kaikkein kannattavimpana, on kuitenkin sen suorittama kaikkien käytettävissä olevien viitteiden kaksinkertainen huomioon ottaminen. Ensinnäkin INLINEFORM5:n rakentaminen, jotta saadaan kattavampi viite, jota vasten arvioidaan, ja sitten INLINEFORM6:n laskeminen, joka skaalaa tuloksen riippuen viitteiden välisestä yhteisymmärryksestä. Osoitimme, että INLINEFORM0 on kattava mittari, joka ei ainoastaan arvioi järjestelmän suorituskykyä kaikkia viitteitä vastaan, vaan ottaa huomioon myös niiden välisen sopimuksen.

**Tulos**

Mikä tekee siitä luotettavamman mittarin?

**Esimerkki 5.2044**

Rakennamme kunkin asiakirjan upotukset käyttämällä MLDoc-ohjelmalla hienosäädettyjä BERT-malleja. Kunkin asiakirjan sulautuksen keskiarvot kerätään yhteen, jotta jokaisesta asiakirjasta saadaan yksi vektori. Tämän jälkeen lasketaan englanninkielisen asiakirjan ja sen käännöksen upotusten välinen kosinin samankaltaisuus. Taulukossa TABREF21 havaitaan, että kosinin samankaltaisuuden mediaani kasvaa dramaattisesti vastakohtaisen harjoittelun myötä, mikä viittaa siihen, että upotuksista tuli kielestä riippumattomampia.

**Tulos**

Miten ne kvantifioivat asiakirjan upotusten ja käännöksen välisen vastaavuuden?

**Esimerkki 5.2045**

Sitten näytämme twiitin tekstin ja kuvan ja pyydämme heitä luokittelemaan sen johonkin kuudesta kategoriasta: Ei hyökkäyksiä mitään yhteisöä vastaan, rasistisia, seksistisiä, homofobisia, uskontoon perustuvia hyökkäyksiä tai hyökkäyksiä muita yhteisöjä vastaan.

**Tulos**

Mitä merkintöjä tietokokonaisuudessa on saatavilla - onko tweat käyttänyt vihapuheita vai ei?

**Esimerkki 5.2046**

Kohdetiedoissa koodi on kirjoitettu Python-ohjelmointikielellä.

**Tulos**

Mikä ohjelmointikieli on kohdekieli?

**Esimerkki 5.2047**

Vertaamme suorituskykyä toistuvien entiteettiverkkojen (EntNet) BIBREF17 -malliin. RelNet-mallin keskimääräinen virhe on 0,285 % kaikissa tehtävissä, mikä on parempi kuin EntNet-mallin BIBREF17 tulokset. RelNet-malli pystyy saavuttamaan 0 prosentin testivirheen 11 tehtävässä, kun taas EntNet-malli saavuttaa 0 prosentin virheen 7 tehtävässä.

**Tulos**

Mihin menetelmiin RelNetiä verrataan?

**Esimerkki 5.2048**

Arvioimme mallejamme kolmella eri tietokokonaisuudella: CSAT-tietokokonaisuus CSAT-ennustusta varten, joka koostuu puhutuista transkripteistä (automaattinen ASR:n avulla).20 uutisryhmää aiheiden tunnistustehtävää varten, joka koostuu kirjoitetusta tekstistä;Fisher Phase 1 -korpus aiheiden tunnistustehtävää varten, joka koostuu puhutuista transkripteistä (manuaalinen);

**Tulos**

Mitä tietokokonaisuuksia he käyttivät arvioinnissa?

**Esimerkki 5.2049**

Osoitimme kuitenkin myös, että tuotetuissa virheissä oli eroja kunkin menetelmän osalta ja että sanatasolla oli erilainen vaikutus lähestymistavasta tai yksiköistä riippuen. Tulevassa työssä keskitytäänkin analysoimaan näiden järjestelmien ortografisia tuotoksia kahdella tavalla: 1) tutkitaan päästä päähän -menetelmien tuottamia virheitä ja tutkitaan useita lähestymistapoja ranskan kielessä tehtyjen yleisten virheiden korjaamiseksi ja 2) verrataan päästä päähän -menetelmiä SLU-kontekstissa ja arvioidaan osittain oikein tuotettujen sanojen semanttista arvoa.

**Tulos**

Mihin tulevassa työssä keskitytään?

**Esimerkki 5.2050**

Käytämme Nguyenin aiemmin käyttämiä laajamittaisia antonyymi- ja synonyymipareja:16. Alun perin aineistoparit kerättiin WordNet BIBREF9- ja Wordnik-tietokannoista.

**Tulos**

Mitä tietokokonaisuutta he käyttävät menetelmänsä arvioinnissa?

**Esimerkki 5.2051**

Tietokanta: Kokeilemme kahta laajalti käytettyä tehtävää englannin ja japanin kieliparin välillä: KFTT BIBREF12 ja BTEC BIBREF13 . KFTT on kokoelma Kioton kaupunkia käsitteleviä Wikipedia-artikkeleita ja BTEC on matkakeskustelukorpus. BTEC on helpompi käännöstehtävä kuin KFTT, koska KFTT kattaa laajemman alan, siinä on enemmän harvinaisia sanoja ja suhteellisen pitkiä lauseita.

**Tulos**

Mitä tietokokonaisuuksia käytettiin?

**Esimerkki 5.2052**

Alustava havaintomme osoittaa, että on mahdollista rakentaa geometrinen avaruusprojektio upotusavaruuksien välille, jotta voidaan oppia rajoitetuista annotoiduista kielellisistä resursseista ja auttaa kieltenvälisessä NE-tunnistuksessa.

**Tulos**

Mikä on heidän mallinsa?

**Esimerkki 5.2053**

Tiedot ovat peräisin ADAPT:n ja Microsoftin yhteisestä hankkeesta.

**Tulos**

mitä tietokokonaisuutta käytettiin?

**Esimerkki 5.2054**

Ehdotettu järjestelmä koostuu kaksisuuntaisesta pitkäaikaisesta lyhytkestoisesta muistista (BiLSTM, Bi-directional Long Short-Term Memory) BIBREF16, kaksitasoisesta tarkkaavaisuusmekanismista BIBREF29, BIBREF30 ja jaetusta esityksestä tunteiden ja tunteiden analysointitehtäviä varten. BiLSTM koodaa kunkin sanan sanan edustuksen. Tämä esitys on yhteinen tunne- ja tunneanalyysin osajärjestelmien välillä. Kukin jaettu esitys syötetään sitten molempien osajärjestelmien ensisijaiseen huomiomekanismiin. Ensisijainen huomiomekanismi löytää kullekin sanalle parhaan edustuksen kutakin tehtävää varten. Toissijainen huomiomekanismi toimii ensisijaisen huomion lisäksi poimimalla parhaan lause-esityksen keskittymällä kuhunkin tehtävään sopivaan kontekstiin. Lopuksi molempien tehtävien representaatiot syötetään kahdelle eri feed-forward-neuraaliverkolle, jotka tuottavat kaksi tulosta - toisen tunneanalyysia ja toisen tunneanalyysia varten. Kukin komponentti selitetään seuraavissa alaluvuissa.

**Tulos**

Miten monitehtäväinen työskentely toteutetaan?

**Esimerkki 5.2055**

Ensimmäisessä skenaariossa dekooderi varustetaan ylimääräisellä morfologiataulukolla, joka sisältää kohdepuolen affiksit.

**Tulos**

Minkälaista morfologista tietoa "morfologinen taulukko" sisältää?

**Esimerkki 5.2056**

Ensimmäisessä vaiheessa tarinasta poimitaan tietämysgraafi, joka kuvaa paikkoja, hahmoja, esineitä ja näiden kokonaisuuksien välisiä suhteita. Esittelemme kaksi tekniikkaa. Ensimmäisessä käytetään neuraalista kysymys-vastaus-tekniikkaa suhteiden poimimiseen tarinatekstistä. Toisessa menetelmässä, joka toimii perustasona, käytetään OpenIE5 -tekniikkaa, joka on yleisesti käytetty sääntöpohjainen tiedonlouhintatekniikka. Yksinkertaisuuden vuoksi tarkastelimme ensisijaisesti sijainti-sijainti- ja sijainti-hahmo/objekti-suhteita, joita kuvassa FIGREF4 edustavat "next to"- ja "has"-reunat. Sen sijaan ehdotamme uutta menetelmää, jossa hyödynnetään kontekstiin perustuviin kysymys-vastaus-tehtäviin koulutettuja malleja entiteettien louhintaan ilman tehtävästä riippuvia tietoja tai hienosäätöä. Menetelmämme, jonka nimi on AskBERT, hyödyntää kysymys-vastausmallia ALBERT BIBREF15. AskBERT koostuu kahdesta päävaiheesta, jotka on esitetty kuvassa KUVA 7: vertexien louhinta ja graafin rakentaminen. Ensimmäisessä vaiheessa tarinasta poimitaan joukko entiteettejä - graafin kärkipisteet. Tarkoituksena on poimia tietoa erityisesti hahmoista, paikoista ja esineistä. Tämä tehdään esittämällä QA-mallille kysymyksiä, kuten "Kuka on tarinan hahmo?". BIBREF16 on osoittanut, että QA-mallille annettujen kysymysten muotoilu on tärkeää, ja tämä muodostaa perustan sille, miten muotoilemme kysymyksemme - kysymykset esitetään niin, että ne todennäköisemmin antavat yhden vastauksen, esimerkiksi kysymällä "Missä on tarinan paikka?" eikä "Missä ovat tarinan paikat?". Huomasimme erityisesti, että pronominin valinta voi olla ratkaisevaa: "Missä on tarinan paikka?" tuotti johdonmukaisemman tuloksen kuin "Mikä on tarinan paikka?". ALBERT QA on koulutettu tuottamaan myös erityinen <$no$-$answer$>-merkki, kun se ei löydä vastausta kysymykseen tarinasta. Menetelmämme hyödyntää tätä pyytämällä iteratiivisesti QA-mallilta kysymyksen ja peittämällä pois edellisessä vaiheessa tulostetun todennäköisimmän vastauksen. Tämä prosessi jatkuu, kunnes <$no$-$answer$>-merkistä tulee todennäköisin vastaus. Ensimmäisessä vaiheessa tarinasta poimitaan joukko entiteettejä - graafin kärkipisteitä. Haluamme poimia tietoa erityisesti hahmoista, paikoista ja esineistä. Tämä tehdään esittämällä QA-mallille kysymyksiä, kuten "Kuka on tarinan hahmo?". BIBREF16 on osoittanut, että QA-mallille annettujen kysymysten muotoilu on tärkeää, ja tämä muodostaa perustan sille, miten muotoilemme kysymyksemme - kysymykset esitetään niin, että ne todennäköisemmin antavat yhden vastauksen, esimerkiksi kysymällä "Missä on tarinan paikka?" eikä "Missä ovat tarinan paikat?". Huomasimme erityisesti, että pronominin valinta voi olla ratkaisevaa: "Missä on tarinan paikka?" tuotti johdonmukaisemman tuloksen kuin "Mikä on tarinan paikka?". ALBERT QA on koulutettu tuottamaan myös erityinen <$no$-$answer$>-merkki, kun se ei löydä vastausta kysymykseen tarinasta. Menetelmämme hyödyntää tätä pyytämällä iteratiivisesti QA-mallilta kysymyksen ja peittämällä pois edellisessä vaiheessa tulostetun todennäköisimmän vastauksen. Tätä prosessia jatketaan, kunnes <$no$-$answer$>-merkistä tulee todennäköisin vastaus.Seuraava vaihe on graafin rakentaminen. Tyypilliset interaktiiviset fiktiomaailmat ovat yleensä rakenteeltaan puita, eli ei syklejä paitsi paikkojen välillä. Tätä tosiasiaa hyväksikäyttäen käytämme lähestymistapaa, jossa graafi rakennetaan kärkijoukosta yksi relaatio - tai reuna - kerrallaan. Jälleen kerran käytämme koko tarinan juonta kontekstina ja kysymme ALBERT-QA-mallilta satunnaista lähtöpaikkaa $x$ aiemmin poimitusta kärkipistejoukosta ja kysymme kysymykset "Missä paikassa voin vierailla $x$:stä?" ja "Kuka/mikä on $x$:ssä?". Menetelmä näiden kysymysten muotoilussa noudattaa samaa menetelmää, joka kuvattiin pisteiden poiminnan yhteydessä. QA-mallin antama vastaus sovitetaan verteksijoukkoon valitsemalla verteksi $u$, joka sisältää parhaan päällekkäisyyden vastauksen kanssa. Pisteiden väliset suhteet lisätään laskemalla suhteiden todennäköisyys QA-mallin antaman vastauksen tulostodennäköisyyksien perusteella. Vertailimme ehdotettua AskBERT-menetelmää ei-neuraaliseen, sääntöpohjaiseen lähestymistapaan. Tämä lähestymistapa perustuu OpenIE5:n poimimaan tietoon, jota seuraa jälkikäsittely, kuten nimettyjen yksiköiden tunnistaminen ja puheosien merkitseminen. OpenIE5:ssä yhdistetään useita huippuluokan ideoita useista olemassa olevista julkaisuista BIBREF17, BIBREF18 ja BIBREF19 tehokkaan tiedonlouhintatyökalun luomiseksi. OpenIE5 tuottaa tietystä lauseesta useita kolmioita muodossa $\langle entity, relation, entity\rangle $ tiiviinä esityksenä lauseesta, joista jokaiselle on annettu luottamuspisteet. Näihin kolmioihin merkitään toisinaan myös sijaintitietoja, jotka osoittavat, että kolmio on tapahtunut tietyssä paikassa.Kuten neuraalisessa AskBERT-mallissa, pyrimme poimimaan tietoa paikoista, hahmoista ja esineistä. Koko tarinan juoni siirretään OpenIE5 -malliin, ja saamme joukon kolmioita. Tripeleiden sijainti-merkintöjä käytetään luomaan joukko sijainteja. Merkitsemme, mitkä tarinan lauseet sisältävät nämä paikat. Tämän jälkeen käytetään POS-tunnistusta, joka perustuu substantiivilauseiden merkitsemiseen, yhdessä NER:n kanssa kolmoiskappaleiden joukon suodattamiseksi edelleen - tunnistetaan tarinan hahmojen ja esineiden joukko.Graafi rakennetaan yhdistämällä kolmoiskappaleiden joukko sen perusteella, mihin paikkaan ne kuuluvat. Vaikka jotkin lauseet sisältävät hyvin selkeää paikkatietoa, jonka OpenIE5 voi merkitä kolmioihin, suurin osa lauseista ei sisällä sitä. Sen vuoksi oletamme, että sijainti pysyy samana kaikissa kolmioissa, jotka on poimittu sellaisten lauseiden väliltä, joissa sijainti mainitaan nimenomaisesti. Jos esimerkiksi tarinan 1. lauseessa on $paikka A$ ja 5. lauseessa $paikka B$, kaikkien lauseissa 1-4 kuvattujen tapahtumien katsotaan tapahtuvan $paikassa A$. Näissä tapahtumissa mainitut entiteetit liittyvät graafissa paikkaan $paikka A$.

**Tulos**

Miten tiedot poimitaan?

**Esimerkki 5.2057**

Teimme kokeita Amazonin arvostelutietokannalla BIBREF9, joka on monialaisen tunneanalyysin vertailutietokanta. Tämä tietokokonaisuus sisältää Amazonin tuotearvosteluja neljältä eri tuotealueelta: Kirjat (B), DVD (D), elektroniikka (E) ja keittiökoneet (K). Jokaiseen arvosteluun on alun perin liitetty 1-5 tähden luokitus, ja se on koodattu 5 000-ulotteisiksi ominaisvektoreiksi, jotka koostuvat sanasäkkipohjaisista unigrammi- ja bigrammivektoreista.Koe ::: Dataset and Task Design :::: Tästä tietokokonaisuudesta muodostimme 12 binääriluokkaista, eri osa-alueiden välistä tunneanalyysitehtävää: B$\rightarrow $D, B$\rightarrow $E, B$\rightarrow $K, D$\rightarrow $B, D$\rightarrow $E, D$\rightarrow $K, D$\rightarrow $K, E$\rightarrow $B, E$\rightarrow $D, E$\rightarrow $K, K$\rightarrow $B, K$\rightarrow $D, K$\rightarrow $E. Aiempien töiden asetusten mukaisesti käsittelimme arvostelua luokkana `1', jos se oli luokiteltu enintään 3 tähteen, ja luokkana `2', jos se oli luokiteltu 4 tai 5 tähteen. Kunkin tehtävän osalta $\mathcal {D}\_S$ koostui 1000 esimerkistä kustakin luokasta, ja $\mathcal {D}\_T$ koostui 1500 esimerkistä luokasta `1' ja 500 esimerkistä luokasta `2'. Koska on järkevää olettaa, että $\mathcal {D}\_T$ voi paljastaa kohdealueen tietojen jakauman, säädimme kohdealueen testitietokannan niin, että sen luokkasuhde on sama kuin $\mathcal {D}\_T$:n luokkasuhde. Käyttämällä samaa etiketinjakomekanismia tutkittiin myös mallin suorituskykyä eriasteisilla $\rm {P}(\rm {Y})$-siirtymillä, joita arvioitiin $\rm {P}\_S(\rm {Y}=i)/\rm {P}\_T(\rm {Y}=i), \kaikkien i=1, \cdots , L$ maksimiarvon perusteella. Lisätietoja tämän tutkimuksen tehtäväsuunnittelusta on liitteessä C. Koe ::: Dataset and Task Design ::: Rakensimme lisäksi 12 moniluokkaista, eri alojen välistä tunteiden luokittelutehtävää. Tehtävät suunniteltiin erottamaan 1 tai 2 tähteä (luokka 1) saaneet arvostelut 4 tähteä (luokka 2) ja 5 tähteä (luokka 3) saaneista arvosteluista. Kunkin tehtävän osalta $\mathcal {D}\_S$ sisälsi 1000 esimerkkiä kustakin luokasta, ja $\mathcal {D}\_T$ koostui 500 esimerkistä luokasta 1, 1500 esimerkistä luokasta 2 ja 1000 esimerkistä luokasta 3. Vastaavasti kontrolloimme myös kohdealueen testausaineistoa siten, että sillä oli sama luokkasuhde kuin $\mathcal {D}\_T$:lla.

**Tulos**

Mitä tunneanalyysitehtäviä käsitellään?

**Esimerkki 5.2058**

Mallimme tarkkuus on 7,8 % korkeampi kuin LSVM:n saavuttama paras tulos. Tulokset osoittavat, että tämä malli on 3,1 % parempi kuin uusimmat perusmallit, kuten hybridi-CNN BIBREF15 ja LSTM ja huomio BIBREF16, validointijoukossa ja 1 % testijoukossa.

**Tulos**

Mihin uusimpiin menetelmiin kirjoittajat vertaavat työtään?

**Esimerkki 5.2059**

Arvioinnissamme on 19 erilaista algoritmia, joilla suositellaan tunnisteita e-kirjoille. Ryhmittelimme ne i) suosioon perustuviin, ii) samankaltaisuuteen perustuviin (eli sisältötietoa hyödyntäviin) ja iii) hybridilähestymistapoihin.

**Tulos**

mitä algoritmeja he käyttivät?

**Esimerkki 5.2060**

uusi DMN+-malli ei edellytä, että tukevat tosiasiat Lisäksi otamme käyttöön uuden syöttömoduulin kuvien esittämistä varten.

**Tulos**

Mitä parannuksia he tekivät DMN:lle?

**Esimerkki 5.2061**

LUWAK on toteutettu puhtaalla JavaScript-koodilla, ja se käyttää selaimen LocalStoragea.

**Tulos**

Millä ohjelmointikielellä työkalu on kirjoitettu?

**Esimerkki 5.2062**

Olemme arvioineet kehystämme käyttämällä DBpedia BIBREF0 -tietokannasta peräisin olevaa synteettistä kirjainjoukkoa (S-Lite) ja todellista kirjainjoukkoa (R-Lite).

**Tulos**

Mitä KB:tä käytetään tässä työssä?

**Esimerkki 5.2063**

CoVoST perustuu tällä hetkellä tähän versioon ja kattaa seuraavat 11 kieltä: Ranska, saksa, hollanti, venäjä, espanja, italia, turkki, persia, ruotsi, mongoli ja kiina.

**Tulos**

Onko arabia yksi CoVostin 11 kielestä?

**Esimerkki 5.2064**

Käyttämällä hienojakoista, leksikaalisiin piirteisiin perustuvaa porttimekanismia voimme tarkasti hallita sanatason ja merkkitason välisiä tietovirtoja. Intuitiivisesti muotoilu on seuraava: INLINEFORM0jossa INLINEFORM0 on elementtiviisas kertolaskuoperaattori. kun portilla on korkea arvo, enemmän tietoa virtaa sanatason esityksestä; muussa tapauksessa char-taso ottaa hallitsevan aseman. Se on käytännöllinen todellisissa skenaarioissa. Esimerkiksi tuntemattomien substantiiviyksiköiden kohdalla portit pyrkivät painottumaan char-tason esitykseen, jotta voidaan huolehtia rikkaammasta morfologisesta rakenteesta.

**Tulos**

Miten gatint-mekanismi yhdistää sana- ja merkkitiedot?

**Esimerkki 5.2065**

Seuraavaksi analysoimme, miksi ne eivät suoriudu hyvin tässä tehtävässä ja näillä tiedoilla:[noitemsep,leftmargin=\*]Kohinainen data. Tämän tehtävän suurimpana haasteena on subjektiivisesta arvostelusta johtuva annotaatioiden välinen ristiriita. Vaikka tämä vaikuttaa myös pelkkää tekstiä käyttävään havaitsemiseen, sen vaikutukset ovat suuremmat monimutkaisemmissa tehtävissä, kuten kuvien tai multimodaalisen havaitsemisen yhteydessä. multimodaalisten suhteiden monimutkaisuus ja monimuotoisuus. Vihapuheen multimodaalisissa julkaisuissa käytetään paljon taustatietoa, minkä vuoksi niissä käytettyjen visuaalisten ja tekstuaalisten elementtien väliset suhteet ovat hyvin monimutkaisia ja monimuotoisia, minkä vuoksi neuraaliverkon on vaikea oppia niitä.Pieni joukko multimodaalisia esimerkkejä. Kuvassa FIGREF5 on esitetty joitakin haastavia multimodaalisia vihapuhe-esimerkkejä, jotka pyrimme havaitsemaan. Mutta vaikka olemme keränneet suuren 150 000 dollarin twiittien tietokokonaisuuden, siinä oleva multimodaalisen vihan osajoukko on edelleen liian pieni oppiaksemme monimutkaisia multimodaalisia suhteita, joita tarvitaan multimodaalisen vihan tunnistamiseen.

**Tulos**

Mikä on kirjoittajan mielipide siitä, miksi nykyiset multimodaaliset mallit eivät voi päihittää pelkkää tekstiä analysoivia malleja?

**Esimerkki 5.2066**

Erilaisten empiiristen tulosten tutkiminen voi auttaa meitä löytämään joitakin mielenkiintoisia ilmiöitä: Esikoulutuksen etu vähenee vähitellen merkityn datan määrän kasvaessa BIBREF14, BIBREF17, BIBREF18.Kiinteät esitykset tuottavat parempia tuloksia kuin hienosäätö joissakin tapauksissa BIBREF24.Kaiken kaikkiaan Seq2Seq-kooderin esikouluttaminen päihittää dekooderin esikouluttamisen BIBREF24, BIBREF17, BIBREF15, BIBREF16.

**Tulos**

Mitä kokeellisia ilmiöitä esitellään?

**Esimerkki 5.2067**

Tutkimme ensin kaikkien NMT-Fake-arvostelujen yleistä havaitsemista (1 006 väärennettyä arvostelua ja 994 aitoa arvostelua).

**Tulos**

Kuinka monta arvostelua (sekä luotuja että oikeita) he arvioivat Amazon Mechanical Turkissa?

**Esimerkki 5.2068**

PERPLEXITY

**Tulos**

Otetaanko asiakirjassa huomioon perpleksisyyden käyttö tekstin poikkeavuuksien tunnistamiseksi?

**Esimerkki 5.2069**

Menetelmiemme automaattiseen arviointiin ehdotamme, että käytämme laajalti käytettyä kuvan/videon tekstittämiseen tarkoitettua metriikkaa. Tämä johtuu siitä, että ehdotettua CommonGen-tehtävää voidaan pitää myös kuvatekstitehtävänä, jossa kontekstina ovat epätäydelliset kohtaukset, joissa on annetut käsitejoukot. Siksi valitsemme päämittareiksi BLEU-3/4 BIBREF28, ROUGE-2/L BIBREF29, CIDEr BIBREF30 ja SPICE BIBREF31. Näiden klassisten mittareiden lisäksi käytämme myös uutta sulauttamiseen perustuvaa mittaria nimeltä BERTScore BIBREF32. Vertailun selkeyttämiseksi näytämme BERTScore-tulosten delta-arvon vähentämällä tuloksen, joka saadaan, kun pelkkiä syöttökäsitejoukkoja käytetään kohdelauseina, nimellä $\triangle $BERTS.

**Tulos**

Mitä automaattisia mittareita käytetään tässä tehtävässä?

**Esimerkki 5.2070**

Syvät konvoluutiohermoverkot (CNN), joissa on 2D-konvoluutiot ja pienet ytimet BIBREF1, ovat saavuttaneet huippuluokan tuloksia useissa puheentunnistustehtävissä BIBREF2, BIBREF3, BIBREF4, BIBREF5, BIBREF6. r Mallit: Perus-CNN-mallimme BIBREF21 koostuu 15 konvoluutiokerroksesta ja yhdestä täysin yhdistetystä kerroksesta.

**Tulos**

Verrataanko mallia uusimpiin malleihin näillä tietokokonaisuuksilla?

**Esimerkki 5.2071**

Teemme suoran vertailun asiantuntijoiden ja joukko-osallistujien välillä 1000 termiryhmän osalta, jotka perustuvat annotaatioiden kokonaismäärään (200 termiryhmää, joissa on yhteensä 2 annotaatiota, 200 termiryhmää, joissa on yhteensä 3 annotaatiota, ja niin edelleen termiryhmiin, joissa on yhteensä 6 annotaatiota). Arvioijille annettiin yhteenveto termiryhmälle saaduista annotaatioista seuraavassa muodossa: Termiryhmä "epätasa-arvo epätasa-arvo" sai annotaatioita seuraavasti: 50,0 % surua, 33,33 % inhoa, 16,67 % vihaa. Sitten heitä pyydettiin arvioimaan asteikolla 1-5, kuinka pätevinä näitä merkintöjä pidettiin.

**Tulos**

Miten he vertailevat sanakirjoja?

**Esimerkki 5.2072**

Vertailemme kokeissamme kahdeksaa eri mallia. Neljässä niistä on projisoitu kerros (ks. kuva FIGREF2), kun taas muissa ei ole, ja tämä on ainoa ero näiden kahden malliryhmän välillä. Kokeissa on siis mukana neljä mallia (projisoitu kerros tai ei). LastStateRNN on klassinen RNN-malli, jossa viimeinen tila kulkee MLP:n läpi ja LR-kerros arvioi vastaavan todennäköisyyden. AvgRNN-mallissa sen sijaan tarkastelemme kaikkien soluista lähtevien tilojen keskimääräistä vektoria. AttentionRNN-malli on se, joka on esitetty BIBREF9:ssä. Lisäksi esitellään häirintäkielen havaitsemista varten MultiAttentionRNN-malli, joka sisältää yhden huomion sijasta neljä huomiota, yhden kutakin luokkaa varten.

**Tulos**

Mitä eri variaatioita tarkkaavaisuuteen perustuvasta lähestymistavasta tutkitaan?

**Esimerkki 5.2073**

Arvioimme malliamme sekä englannin- että kiinankielisellä segmentoinnilla. Molemmissa kielissä käytimme standarditietokantoja sanojen segmentointiin ja kielten mallintamiseen. EnglantiBrent-korpus on standardikorpus, jota käytetään lasten kielen omaksumisen tilastollisessa mallintamisessa BIBREF15 , BIBREF16 Käytämme yleisesti käytettyä versiota PTB:stä, jonka on laatinut BIBREF17 Koska kiinalainen ortografia ei merkitse sanojen välisiä välejä, sanojen rajoja on pyritty useaan otteeseen merkitsemään. Arvioinnissa käytetään kahta korpusta, jotka on segmentoitu manuaalisesti eri segmentointistandardien mukaisesti. Pekingin yliopiston korpus oli yksi korpuksista, joita käytettiin kansainvälisessä kiinalaisen sanan segmentointikilpailussa BIBREF18 . Käytämme Penn Chinese Treebank Version 5.1 BIBREF19 .

**Tulos**

Mitä tietokokonaisuutta käytetään?

**Esimerkki 5.2074**

Jotta tällaisia olemassa olevia tietolähteitä voitaisiin hyödyntää paremmin, ehdotamme huomioiviin osoitinverkkoihin perustuvaa end-to-end (E2E) -mallia, joka voidaan kouluttaa end-to-end ihmisen IE-tehtävien tulo/lähtö -pareilla ilman, että tarvitaan merkkitason merkintöjä. Koska mallimme ei tarvitse token-tason merkintöjä, luomme jokaisesta aineistosta E2E-version ilman token-tason merkintöjä pilkkomalla BIO-merkityt sanat ja käyttämällä merkintöjä poimittavina kenttinä.

**Tulos**

Olettavatko he rangaistustasoista valvontaa?

**Esimerkki 5.2075**

Ennen tämän NLP-työkalupaketin suunnittelua teimme kyselyn insinöörien keskuudessa ja määrittelimme kolme tyypillistä persoonaa.

**Tulos**

Miten kirjoittajat todistavat väitteen, jonka mukaan monille insinööreille on suuri työläs valita useiden kehysten, mallien ja optimointitekniikoiden välillä?

**Esimerkki 5.2076**

Kaikki tutkimamme menetelmät parantavat sukupuolittuneiden sanojen prosenttiosuutta, miesten prosenttiosuutta ja F1:tä Transformer-sukupolven perusmalliin verrattuna, mutta kaikkien menetelmien yhdistäminen yhteen - ALL-malli - on mielestämme edullisin.

**Tulos**

Mihin perustasoon kokeellisia tuloksia verrataan?

**Esimerkki 5.2077**

Käsittelemme kahta keskeistä mallia sekvenssien merkintäongelmien ratkaisemiseksi ja kuvaamme kummankin mallin harjoittelua yhden mallin monikielisessä ympäristössä: (1) Meta-LSTM BIBREF0 , joka on erittäin vahva perusmalli tehtäviämme varten, ja (2) monikielinen BERT-pohjainen malli BIBREF1 .

**Tulos**

Mikä on monikielinen perustaso?

**Esimerkki 5.2078**

Keräsimme 14 vapaaehtoiselta, 5 italialaiselta ja 9 muulta kuin italialaiselta kuuntelijalta yhteensä 657 arviota, jotka jakautuivat 24 videoleikkeen ja kolmen testausolosuhteen kesken.

**Tulos**

Kuinka monta henkilöä työskentelee subjektiivisen arvioinnin parissa?

**Esimerkki 5.2079**

Arvioimme lähestymistapaamme harjoittelemalla automaattista täydennysjärjestelmää 500 000 satunnaisesti poimitulla lauseella Yelp-arvosteluista BIBREF6 (katso lisätietoja liitteestä). Mittaamme viestintäjärjestelmän $(q\_{\alpha },p\_{\beta })$ tehokkuutta merkkien säilymisasteella, jota mitataan avainsanoissa säilytettävien merkkien osuutena. Järjestelmän tarkkuutta mitataan ahnaasti dekoodaamalla mallin tuottamien lauseiden osuutena, joka vastaa täsmälleen kohdelausetta.

**Tulos**

Miten malleja arvioidaan tässä ihmisen ja koneen välisessä viestintäpelissä?

**Esimerkki 5.2080**

Arvioimme mallejamme kahdessa tyypillisessä tehtävässä: tekstin luokittelussa ja tekstin semanttisessa yhteensovittamisessa.

**Tulos**

Mitä tehtäviä he kokeilevat?

**Esimerkki 5.2081**

Tässä julkaisussa keskitytään sellaisen tiedon esittämiseen, joka on arvokasta näissä tehtävissä, mutta joka on tähän asti jäänyt suurelta osin huomiotta yksinkertaistettuun kieleen keskittyvissä koneoppimismenetelmissä, erityisesti tekstin rakenteeseen (esim. kappaleet, rivit), typografiaan (esim. kirjasintyyppi, kirjasintyyli) ja kuvaan (sisältö, sijainti ja mitat) liittyvässä tiedossa. Tieto merkin kirjasintyypistä ja kirjasintyylistä (esim. kursivointi, lihavointi) ja sen sijainnista fyysisellä sivulla (vain PDF-tiedostojen osalta) määritettiin määritteinä merkkikerroksen merkkielementtien attribuuteiksi (ks. esimerkki kuvassa FIGREF34).

**Tulos**

Mitä typografiaa koskevia tietoja korpus sisältää?

**Esimerkki 5.2082**

Tässä työssä esittelemme menetelmän, jonka avulla VQA-algoritmit pystyvät tuottamaan ihmiselle tulkinnanvaraisia huomiokarttoja, jotka tehokkaasti maadoittavat vastauksen merkityksellisiin kuva-alueisiin. Saavutamme tämän hyödyntämällä Visual Genome -tietokannassa saatavilla olevia aluekuvauksia ja objektien annotaatioita ja käyttämällä niitä automaattisesti tarkkaavaisuuskarttojen rakentamiseen, joita voidaan käyttää tarkkaavaisuuden valvonnassa sen sijaan, että ihmisannotoijien olisi annettava manuaalisesti maadoitusmerkinnät.

**Tulos**

Miten he saavat aluekuvaukset ja kohteiden merkinnät?

**Esimerkki 5.2083**

Upotusviisaan konvoluution tarkoituksena on soveltaa konvoluutiosuodatinta w $\in \mathbb {R}^{\mathcal {\\\}k\*d}$ $k$ sanojen upotuksista koostuvaan ikkunaan uuden ominaisuuden tuottamiseksi, eli $k$ sanojen paikallisen kontekstin tiivistämiseksi. Soveltamalla konvoluutiosuodatinta kaikkiin lauseen mahdollisiin ikkunoihin luodaan ominaisuuskartta $c$.

**Tulos**

Mikä tukee väitettä, jonka mukaan CNN:n lisääminen toistuviin yksiköihin parantaa mallin kykyä ottaa huomioon paikallinen konteksti ja vähentää epäselvyyksiä?

**Esimerkki 5.2084**

Kuten taulukosta TABREF12 käy ilmi, T-T-malli on huomattavasti parempi kuin LSTM-pohjainen RNN-T-perusversio.

**Tulos**

Mikä oli edellinen uusin malli?

**Esimerkki 5.2085**

Katsotaanpa tarkemmin, mitä tarkoitamme kahden jäsennyspuun välisellä yksimielisyydellä. Jos englanninkielisessä lauseessa on kaksi sanaa, joita merkitään i:llä ja i':llä ja jotka vastaavat sanoja j ja j' samansuuntaisessa hindinkielisessä lauseessa, englanninkielisen jäsennyspuun i:n ja i':n välisen riippuvuusreunan voidaan olettaa vastaavan j:n ja j':n välistä reunaa hindinkielisessä jäsennyspuussa ja päinvastoin. Englanninkielisen jäsennyspuun reunaa vastaavaa reunaa tai polkua hindinkielisessä jäsennyspuussa kutsutaan englanninkielisen reunan projisoinniksi tai projisoiduksi poluksi hindinkieliseen jäsennyspuuhun, ja vastaavasti on olemassa projisoituja polkuja hindistä englantiin. Kaksoispurkualgoritmi pyrkii tuomaan kahden kielen jäsennyspuut yhteen rajoitteidensa avulla.

**Tulos**

Miten jäsennyspuiden välisen yhteisymmärryksen varmistaminen toimii eri kielillä?

**Esimerkki 5.2086**

Aloitamme tutkimalla muutaman merkittävän keskustelusivuston sisältöä: idebate.com, debatewise.org, procon.org.

**Tulos**

Mitä keskustelusivustoja he katsoivat?

**Esimerkki 5.2087**

Sen vuoksi käytimme tietojen keräämiseen CrowdFlower (CF) -joukkoistamisalustaa.

**Tulos**

Miten nämä tiedot kerättiin?

**Esimerkki 5.2088**

Yleisesti ottaen molemmat NN-mallit päihittävät CAEVO-perusmallin, ja BERT-sisällytystä sisältävä NN-malli on suorituskyvyltään paras.

**Tulos**

Parantavatko yhteistyöhön perustuvat sulautukset tuloksia?

**Esimerkki 5.2089**

LastStateRNN on klassinen RNN-malli, jossa viimeinen tila kulkee MLP:n läpi ja LR-kerros arvioi vastaavan todennäköisyyden. AvgRNN-mallissa sen sijaan tarkastellaan kaikkien soluista lähtevien tilojen keskimääräistä vektoria. AttentionRNN-malli on se, joka on esitetty BIBREF9:ssä. Lisäksi esitellään häirintäkielen havaitsemista varten MultiAttentionRNN-malli, joka sisältää yhden huomion sijasta neljä huomiota, yhden kutakin luokkaa varten.

**Tulos**

Mitä perusmallia käytetään?

**Esimerkki 5.2090**

Siamilaisessa neuroverkossa on kaksi rinnakkaista konvoluutioverkkoa, INLINEFORM0 , joilla on sama painojen joukko, INLINEFORM1 , kuten kuvassa KUVA 5 (a) on esitetty.

**Tulos**

Mikä on siamilaisen neuroverkon arkkitehtuuri?

**Esimerkki 5.2091**

Meidän asetuksessamme lähtökohtana on perusmalli, joka on koulutettu NLI-tiedoilla. Sen sijaan, että käytettäisiin automatisoituja vastapuolimenetelmiä, tässä mallin "vastapuolena" on ihmisen kommentoija. Kun annetaan konteksti (jota kutsutaan NLI:ssä usein myös "premissioksi") ja haluttu kohdemerkintä, pyydämme ihmisen kirjoittajaa esittämään hypoteesin, joka huijaa mallia luokittelemaan merkinnän väärin. Kirjoittajaa voidaan pitää "valkoisen hatun" hakkerina, joka yrittää tunnistaa järjestelmän haavoittuvuuksia. Jokaisen väärin luokitellun ihmisen luoman esimerkin kohdalla pyydämme kirjoittajaa myös ilmoittamaan syyn, miksi hän uskoo, että se luokiteltiin väärin.

**Tulos**

Käyttävätkö he aktiivista oppimista tietokokonaisuuksiensa luomiseen?

**Esimerkki 5.2092**

Tässä julkaisussa keskitytään sellaisten tietojen esittämiseen, jotka ovat arvokkaita näissä tehtävissä, mutta jotka on tähän asti jätetty suurelta osin huomiotta koneoppimisessa, joka keskittyy yksinkertaistettuun kieleen, erityisesti tekstin rakenteeseen (esim. kappaleet, rivit), typografiaan (esim. kirjasintyyppi, kirjasintyyli) ja kuviin (sisältö, sijainti ja mitat). Fyysistä sivun segmentointia (vain PDF-tiedostojen osalta), kappaleiden segmentointia ja rivien segmentointia koskevat tiedot lisättiin tekstin rakennekerroksen textspan-elementin osaksi.

**Tulos**

Mitä tietoja tekstin rakenteesta sisältyy korpukseen?

**Esimerkki 5.2093**

Suoritamme kokeita käyttämällä seuraavia uusimpia malleja: (1) SEQ2SEQ: sekvenssistä sekvenssiin -malli, jossa on huomiomekanismit BIBREF21, (2) CVAE: ehdollinen variaatiomalli, jossa on KL-annealing ja BOW-häviö BIBREF2, (3) Transformer: kooderi-dekooderiarkkitehtuuri, joka perustuu pelkästään huomiomekanismeihin BIBREF22, (4) HRED: yleistetty sekvenssistä sekvenssiin -malli hierarkkisen RNN-kooderin kanssa BIBREF23, (5) DialogWAE: ehdollinen Wassersteinin autokooderi, joka mallintaa datan jakaumaa kouluttamalla GAN:ia latenttien muuttujien tilassa BIBREF6.

**Tulos**

Mitä uusimpia malleja käytettiin kokeissa?

**Esimerkki 5.2094**

Asiakirjaluokittelijana käytämme Kimin kaltaista sanapohjaista CNN:ää, joka koostuu seuraavista kerroksista: $ \texttt {Conv} \xrightarrow{} \texttt {ReLU} \xrightarrow{} \texttt {1-Max-Pool} \xrightarrow{} \texttt {FC} \\ $ Tulevaan työhön kuuluisi LRP:n soveltaminen muihin neuroverkkoarkkitehtuureihin (esim. merkkipohjaisiin tai rekursiivisiin malleihin) uusissa NLP-tehtävissä sekä sen tutkiminen, miten relevanssitieto voitaisiin ottaa huomioon luokittelijan koulutusmenettelyn tai ennustuskyvyn parantamiseksi.

**Tulos**

Tutkitaanko kokeiluissa, miten eri arkkitehtuurit ja kerrokset edistävät tiettyjen päätösten tekemistä?

**Esimerkki 5.2095**

Käytämme kilpailun tietokokonaisuutta, joka sisältää tekstiä twiiteistä, joilla on edellä mainitut kategoriat.

**Tulos**

Mitä tietokokonaisuutta tässä työssä käytetään?

**Esimerkki 5.2096**

BIBREF:ssä5 mainittujen luokkien mukaan ironia voidaan jakaa kolmeen luokkaan: sanallinen ironia polariteettikontrastin avulla eli lauseet, jotka sisältävät ilmaisun, jonka polariteetti on käänteinen tarkoitetun ja kirjaimellisen merkityksen välillä; muunlainen sanallinen ironia eli lauseet, joissa ei ole polariteettikontrastia kirjaimellisen ja tarkoitetun merkityksen välillä, mutta jotka ovat silti ironisia; ja tilannekohtainen ironia eli lauseet, joissa kuvataan tilanteita, jotka eivät täytä joitakin odotuksia. Koska kahteen jälkimmäiseen kategoriaan kuuluva ironia on epäselvää ja vaikeasti ymmärrettävää, päätämme keskittyä tässä työssä vain ensimmäiseen kategoriaan kuuluvaan ironiaan. Koska aiempia ironian tuottamista koskevia töitä ja peruslinjoja ei ole, toteutamme mallin tyylinsiirtoon perustuen.

**Tulos**

Mitä vaikeuksia ironisen kuvion mallintamisessa on?

**Esimerkki 5.2097**

Mallit sisältävät sekä neuroverkkoja (esim. RNN:t, CNN:t) että tavanomaisia koneoppimisvälineitä (esim. Naive Bayes ja Laplace Smoothing, k-klusterointi, SVM lineaarisella ytimellä).

**Tulos**

Mitä koneoppimismalleja käytetään?

**Esimerkki 5.2098**

Ensin työntekijöille annettiin alaan liittyvä kvalitatiivinen suhde q+/-( INLINEFORM0 ), joka ilmaistiin englanniksi (esim. "Jos pinnalla on enemmän kitkaa, esine kulkee hitaammin"), ja heitä pyydettiin syöttämään kaksi vertailtavaa esinettä, ihmistä tai tilannetta. Tämän jälkeen he loivat kysymyksen, jota ohjasivat lukuisat esimerkit, ja heitä rohkaistiin olemaan mielikuvituksellisia ja käyttämään omia sanojaan. Toiseksi, LF:t saatiin esiin käyttämällä uutta tekniikkaa, jossa ne muodostettiin käänteisesti seurantakysymysten joukosta ilman, että työntekijät joutuivat tutustumaan niiden taustalla olevaan formalismiin. Tämä on mahdollista, koska LF:t ovat rajallisia. Viitaten aiemmin esitettyihin LF-malleihin (1) ja (2) (jakso SECREF13 ), nämä kysymykset ovat seuraavat: Näistä tiedoista voimme päätellä kohde-LF:n ( INLINEFORM0 on INLINEFORM1:n komplementti, INLINEFORM2 , asetamme mielivaltaisesti INLINEFORM3 =maailma1, joten kaikki muut muuttujat voidaan päätellä).

**Tulos**

Miten he saavat kysymystensä loogiset muodot aineistostaan?

**Esimerkki 5.2099**

Muistimekanismi on otettu käyttöön, jotta malli voi tarkastella paikallisia piirteitä laajemmin ja saada käyttöönsä koko sarjan.

**Tulos**

Mitä ominaisuuksia he käyttävät?

**Esimerkki 5.2100**

Jaoimme Dryhootchin valitsemat kyselyt 1 200 käyttäjälle (305 käyttäjää on itse ilmoittanut kärsivänsä PTSD:stä ja loput on valittu satunnaisesti edellisistä 2 423 käyttäjästä) ja saimme 210 onnistunutta vastausta. Dryhootch määrittelee PTSD:n voimakkuuden neljään luokkaan kaikkien kolmen kliinisen kyselytyökalun (DOSPERT, BSSS ja VIAS ) viikoittaisten kyselytulosten perusteella.

**Tulos**

Mitä kliinisesti validoituja kyselytyökaluja käytetään?

**Esimerkki 5.2101**

Lähtökohtana käytimme DIP-korpusta BIBREF37 , joka koostuu 49:stä 100 verkkosivun klusterista, jotka käsittelevät koulutusaiheita (esim. kiusaaminen, kotiopetus, huumeet) ja sisältävät lyhyen kuvauksen kustakin aiheesta.

**Tulos**

Mitkä verkkoasiakirjojen kokoelmat sisältyvät korpukseen?

**Esimerkki 5.2102**

Etnisyys/rotuYksi mielenkiintoinen kuvio on, että vauvojen etnistä alkuperää/rotua ei näytä mainittavan, ellei vauva ole musta tai aasialainen. Toisin sanoen: valkoinen näyttää olevan oletusarvo, ja muut näyttävät olevan merkittyjä. Miten voimme sanoa, ovatko tiedot todella puolueellisia vai eivät? Taulukon TABREF22 luvut ovat silmiinpistäviä: etnisen alkuperän merkitsemisessä näyttää olevan todellinen, systemaattinen ero ryhmien välillä. Voimme mennä askeleen pidemmälle ja tarkastella kaikkia 697 kuvaa, joissa on sana "vauva". Jos valkoisia vauvoja on suhteettoman paljon, tämä vahvistaa päätelmää siitä, että aineisto on vääristynyt.

**Tulos**

Mitä vääristymiä aineistosta löytyy?

**Esimerkki 5.2103**

Kukin tietokokonaisuus koostuu tietueiden kokoelmasta, jossa on yksi laadunvarmistusongelma tietuetta kohti. Jokaiseen tietueeseen sisältyy kysymysteksti, kysymyksen kannalta merkityksellinen asiayhteysasiakirja, joukko ratkaisukandidaatteja ja oikea ratkaisu. Kunkin tietueen kontekstidokumentti koostuu luettelosta kysymyksen kannalta merkityksellisistä luokitelluista ja pisteytetyistä pseudodokumenteista. Useat perusmenetelmät perustuvat haettuun kontekstiin vastauksen saamiseksi kysymykseen. Näiden osalta viittaamme hakutarkkuuteen (Search Accuracy) niiden tapausten osuutta, joiden kohdalla oikea vastaus löytyy kontekstista. Luonnollisesti hakutarkkuus kasvaa kontekstin koon kasvaessa, mutta samalla lukusuorituskyky heikkenee, koska vastauksen poimiminen on vaikeampaa pidemmissä asiakirjoissa.

**Tulos**

Mitä hakujärjestelmää käytettiin peruslinjoihin?

**Esimerkki 5.2104**

Olemme luoneet englanninkielisten kotiutusyhteenvetojen ja hoitajien muistiinpanojen tietokokonaisuuden, jossa keskitytään usein takaisin otettaviin potilaisiin, jotka on merkitty 15 kliinisellä potilasfenotyypillä, joiden uskotaan olevan yhteydessä toistuvan teho-osastolle takaisinottamisen riskiin, kuten asiantuntijamme (LAC, PAT, DAG) ja kirjallisuuslähteet ovat osoittaneet. BIBREF10 BIBREF11 BIBREF12

**Tulos**

Kuinka monta eri fenotyyppiä aineistossa on?

**Esimerkki 5.2105**

Kokeissamme RNN ottaa vastaan lausumien spektrogrammeja, jotka kulkevat kahden 2D-konvoluutiokerroksen läpi, minkä jälkeen ne kulkevat seitsemän kaksisuuntaisen rekursiivisen kerroksen ja täysin kytketyn kerroksen läpi, jossa on softmax-aktivointi. Kaikki rekursiiviset kerrokset on normalisoitu eräkohtaisesti. Jokaisella aikaportaalla softmax-aktivoinnit antavat todennäköisyysjakauman hahmoille. CTC-tappio BIBREF8 lasketaan sitten aikavaiheittaisten todennäköisyyksien perusteella.

**Tulos**

Mitä mallia ne käyttävät päästä päähän -puheentunnistukseen?

**Esimerkki 5.2106**

Arvioimme malliamme kahdella julkisella tietokokonaisuudella, jotka ovat Penn Treebank (PTB) BIBREF9 ja End-to-end (E2E) -tekstintuottamiskorpus BIBREF10, joita on käytetty useissa aiemmissa teoksissa tekstintuottamiseen BIBREF0, BIBREF5, BIBREF11 ja BIBREF12. PTB koostuu yli 40 000 lauseesta Wall Street Journalin artikkeleista, kun taas E2E-tietokanta sisältää yli 50 000 lausetta ravintola-arvosteluista. Taulukossa TABREF11 on yhteenveto näiden kahden tietokokonaisuuden tilastoista.

**Tulos**

Mitä tietokokonaisuutta he käyttävät tekstin mallintamiseen?

**Esimerkki 5.2107**

Jotta ihmisen edustaja voisi nopeasti määrittää eskaloitumisen syyn, luomme visualisoinnin käyttäjän käänteistä käyttämällä huomiopainoja korostamaan eskaloitumispäätökseen vaikuttavia käänteitä.

**Tulos**

Selittävätkö ne mallin ennusteita pelkästään huomiopainoilla?

**Esimerkki 5.2108**

Näin ollen marcheggiani2017 esitti neuromallin, jossa syntaksi jätetään syrjään riippuvuuspohjaisen SRL:n osalta ja saadaan suotuisia tuloksia, mikä kumoaa luontaisen uskomuksen, että syntaksi on välttämätön SRL-tehtävässä BIBREF11 .

**Tulos**

Onko olemassa syntaksin suhteen tunnistamattomia SRL-malleja?

**Esimerkki 5.2109**

Kuten Sequicity-kehyksessä, raportoimme entiteettien vastaavuusprosentin, BLEU-pisteet ja menestyksen F1-pisteet.

**Tulos**

Mitä arviointimittareita käytettiin?

**Esimerkki 5.2110**

Syväoppiminen on kiistatta edistänyt tekniikan tasoa monissa luonnollisen kielen käsittelytehtävissä, syntaktisten riippuvuuksien jäsennyksestä BIBREF0 nimettyjen yksiköiden tunnistamiseen BIBREF1 ja konekääntämiseen BIBREF2 . Sama pätee varmasti myös kielen mallintamiseen, jossa viimeaikaiset edistysaskeleet neuraalisissa kielimalleissa (NLM) ovat johtaneet huomattavasti parempiin lähestymistapoihin, kun niitä mitataan standardimittareilla, kuten perpleksisyydellä BIBREF3 , BIBREF4 .

**Tulos**

Mikä on yleisesti käytetty kielimallien arviointimittari?

**Esimerkki 5.2111**

Havaintomme mukaan, jos kansankielinen kappale sisältää enemmän klassisessa kirjallisuudessa käytettyjä runokuvia, sen tuottama runo saa yleensä korkeamman pistemäärän. Havaitsimme myös, että kuvailevista kappaleista luodut runot saavat korkeammat pisteet kuin loogisista tai filosofisista kappaleista luodut runot.

**Tulos**

Mitkä ovat joitakin ohjeita kirjoitettaessa tulo kansankielellä niin malli voi tuottaa

**Esimerkki 5.2112**

Aloitamme harjoitteluprosessin soveltamalla sääntöä INLINEFORM4 luonnollisen kielen kysymyksiin INLINEFORM5 . Tuloksena saatua tietokokonaisuutta pidetään harjoitusaineistona, jolla alustetaan sekä semanttinen jäsentäjä että kysymysgeneraattori. Sen jälkeen molempia malleja parannetaan noudattaen takaisinkäännösprotokollaa, jonka mukaan kohdejaksojen pitäisi noudattaa todellisen datan jakaumaa, mutta lähdejaksot voidaan kuitenkin tuottaa kohinaa sisältäen.

**Tulos**

Miten takaisinkäännösmalli koulutetaan?

**Esimerkki 5.2113**

Lisäksi toteamme, että lähestymistapamme voi edelleen johtaa virheellisiin tosiseikkoihin tai jopa hallusinaatioihin. Mielenkiintoinen näkökulma voisi olla mallin rajoittaminen tietorakenteeseen, jotta vältettäisiin epätarkat tai jopa ristiriitaiset kuvaukset.

**Tulos**

Mitä tulevia mahdollisia parannuksia on lueteltu?

**Esimerkki 5.2114**

Tämän hankkeen tiedot ovat kaksiosainen: ensimmäinen osa on historialliset S&P 500 -osakkeet, jotka on ladattu Yahoo Finance -palvelusta. Käytämme tietoja ajanjaksolta 12/07/2017 - 06/01/2018. Toinen osa on uutisartikkeli talousalalta, joka on kerätty samalla ajanjaksolla kuin varastotiedot. Näin ollen kerätään vain talousalan uutisartikkelit. Tiedot otetaan pääasiassa Webhose-arkistoidusta datasta, joka koostuu 306242 uutisartikkelista JSON-muodossa joulukuusta 2017 kesäkuun 2018 loppuun.

**Tulos**

Mitä tietokokonaisuutta käytetään tutkimuksessa?

**Esimerkki 5.2115**

Pyysimme kustakin lauseesta kahta erillistä parafraasia, koska uskomme, että hyvän lauseen sulautuksen pitäisi asettaa parafraasit lähelle toisiaan vektoriavaruudessa.Useita muutostyyppejä valittiin erityisesti, jotta sulautuksia voitaisiin testata perusteellisesti. Erilaisessa merkityksessä annotoijien tulisi luoda lause, jolla on jokin muu merkitys, käyttäen samoja sanoja kuin alkuperäisessä lauseessa. Muita muunnoksia, joiden pitäisi olla vaikeita upotuksille, ovat minimaalinen muutos, jossa lauseen merkitys pitäisi muuttaa merkittävästi käyttämällä vain hyvin pientä muunnosta, tai hölynpöly, jossa lähdelauseen sanoja pitäisi sekoittaa niin, että se on kieliopillisesti oikein, mutta ilman mitään merkitystä.

**Tulos**

Mitä merkintöjä tietokokonaisuudessa on saatavilla?

**Esimerkki 5.2116**

Kolme selkeää aluetta on havaittavissa $\langle cc \rangle $:n käyttäytymisessä suhteessa $\wp $:hen $t\_f$:n kohdalla, kuten kuvassa FIGREF15 on esitetty (siniset neliöt). Vaihe I: $\langle cc \rangle $ kasvaa tasaisesti, kun $\wp < 0.4$, mikä osoittaa, että tällä alueella sanojen naapurustojen välillä on pieni korrelaatio. Täydet sanastot saavutetaan myös, kun $\wp < 0.4$; Vaihe II: kriittisellä alueella $\wp ^\* \ (0.4,0.6)$ tapahtuu jyrkkä siirtymä, jossa $\langle cc \rangle $ siirtyy jyrkästi kohti 1:tä. Myös $V(t\_f)$:n äkillinen muutos $\wp ^\*$:n suhteen $\wp ^\* $:n suhteen on havaittavissa (kuva FIGREF16); vaihe III: yhden sanan kielet ovat vallitsevia, kun $\wp > 0.6$. $\langle cc \rangle $:n maksimiarvo osoittaa, että sanojen naapurustot ovat täysin korreloituneita.

**Tulos**

Mitkä ovat kielen muodostumisen kolme mahdollista vaihetta?

**Esimerkki 5.2117**

Arvioimme havaintomallejamme kolmella vertailukohteella: FCE-testiaineisto (41 000 merkkiä) ja CoNLL 2014 Shared Task -tietokannan kaksi vaihtoehtoista merkintää (30 000 merkkiä) BIBREF3 .

**Tulos**

Mitä annotoitua korpusta he käyttivät?

**Esimerkki 5.2118**

Kuvassa FIGREF10 esitetään UTCNN-malli. Koska useampi kuin yksi käyttäjä voi olla vuorovaikutuksessa tietyn viestin kanssa, lisäämme ensin maksimipoolauskerroksen käyttäjämatriisin upotuskerroksen ja käyttäjävektorin upotuskerroksen jälkeen muodostaaksemme moderaattorimatriisin INLINEFORM0 ja moderaattorivektorin INLINEFORM1 moderaattorille INLINEFORM2, jossa INLINEFORM3:a käytetään asiakirjan koostamisprosessin semanttiseen muunnokseen, kuten edellisessä jaksossa mainittiin. Termi moderaattori tarkoittaa tässä yhteydessä pseudokäyttäjää, joka tarjoaa kaikkien sitoutuneiden käyttäjien yleisen semantiikan/sentimentin yhtä asiakirjaa varten. Upotus INLINEFORM4 mallintaa moderaattorin kannanottopreferenssiä, toisin sanoen paljastetun käyttäjän kannanoton mallia: haluaako käyttäjä osoittaa mieltymyksensä, haluaako käyttäjä osoittaa puolueettomuutta neutraaleilla lausunnoilla ja järkevillä argumenteilla vai haluaako hän vain osoittaa vahvaa tukea yhdelle kannalle. Ihannetapauksessa käyttäjän latentti kanta mallinnetaan INLINEFORM5 -ohjelmalla kunkin käyttäjän osalta. Vastaavasti aiheeseen liittyvää tietoa varten lisätään aihematriisin upotuskerroksen ja aihevektorin upotuskerroksen jälkeen maksimipoolauskerros, jotta muodostetaan yhteinen aihematriisin upotus INLINEFORM6 ja yhteinen aihevektorin upotus INLINEFORM7 aiheen INLINEFORM8 osalta, jossa INLINEFORM9 mallintaa aiheen INLINEFORM10 semanttista transformaatiota käyttäjissä ja INLINEFORM11 mallintaa aiheen kannan taipumusta. Latentti aiheen asenne mallinnetaan myös INLINEFORM12:lla kullekin aiheelle.Mitä tulee kommentteihin, pidämme niitä lyhyinä dokumentteina, joissa on vain kirjoittajat, mutta ei tykkääjiä eikä heidän omia kommenttejaan. Siksi sovellamme asiakirjojen koostumusta kommentteihin, vaikka tässä käyttäjät ovat kommentoijia (käyttäjät, jotka kommentoivat). Huomataan, että saman sanan INLINEFORM0-sanan upotukset viesteissä ja kommenteissa ovat samat, mutta sen jälkeen kun ne on muunnettu INLINEFORM1:ksi kuvassa KUVA 4 esitetyssä asiakirjojen koostamisprosessissa, ne saattavat muuttua erilaisiksi, koska niissä on eri sitoutuneita käyttäjiä. Lähtökommenttikuvaus yhdessä kommenttivektorin INLINEFORM2 ja aihevektorin INLINEFORM3 upotuksen kanssa ketjutetaan ja lisätään maksimipoolauskerros kommenttien tärkeimpien ominaisuuksien valitsemiseksi. Sen sijaan, että vaadittaisiin, että kommentin kannan on oltava samaa mieltä postauksen kanssa, UTCNN yksinkertaisesti poimii kommentin sisällöstä tärkeimmät piirteet; ne voivat olla hyödyllisiä riippumatta siitä, osoittavatko ne ilmeistä yhteisymmärrystä vai erimielisyyttä. Kun tässä yhdistetään kommenttitietoja, maksimipooling-kerros on siis sopivampi kuin muut pooling- tai yhdistämiskerrokset. Uskomme, että tämä on yksi syy UTCNN:n suorituskyvyn paranemiseen.Lopuksi yhdistetty kommenttiesitys yhdessä käyttäjävektorin INLINEFORM0 upottamisen, aihevektorin INLINEFORM1 upottamisen ja asiakirjan esityksen kanssa syötetään täysin kytkettyyn verkkoon, ja softmaxia sovelletaan lopullisen kannanottotarran ennusteen saamiseksi viestille.

**Tulos**

Kuinka monta kerrosta UTCNN-mallissa on?

**Esimerkki 5.2119**

Viime vuosina tutkijat ovat ehdottaneet erilaisia suodattimiin perustuvia ominaisuuksien valintamenetelmiä asiakirjojen tekstiluokittelun suorituskyvyn parantamiseksi BIBREF19. Lisäksi käytämme suodattimiin perustuvaa ominaisuuksien valintamenetelmää nimeltä Normalized Difference Measure (NDM) BIBREF5, jotta termit voidaan asettaa paremmuusjärjestykseen niiden erottelukyvyn perusteella luokkien välillä. Rehman et al. BIBREF5 ehdottivat, että kaikki piirteet, jotka ovat piirteen vasemmassa yläkulmassa ja oikeassa alakulmassa, ovat erittäin merkittäviä verrattuna piirteisiin, jotka ovat lävistäjien ympärillä. Nykyaikaiset suodattimiin perustuvat piirteiden valinta-algoritmit, kuten ACC2, käsittelevät samalla tavalla kaikkia niitä piirteitä, jotka ovat diagonaalien ympärillä BIBREF5.

**Tulos**

Onko suodattimiin perustuva ominaisuuksien valinta (FSE) eräänlainen regularisointi?

**Esimerkki 5.2120**

Määrittelimme aikomukset ohjaamalla kyselyitä, jotka kerättiin joukkoistamistehtävän avulla, jossa joukkoistajia kehotettiin esittämään aihealueisiin liittyviä kysymyksiä ja komentoja tavalla, jolla he olisivat vuorovaikutuksessa tekoälyavustajan kanssa. Ryhmittelimme käsin scoping-tehtävien tuottamat tiedot intenteiksi.

**Tulos**

Miten tietokokonaisuutta kommentoitiin?

**Esimerkki 5.2121**

Ajoimme UTD:n kaikkiin 104 puhelinkeskusteluun, jotka yhdistävät 11 tuntia ääntä ja espanjankieliset transkriptiot sekä niiden englanninkieliset käännökset. Pöytäkirjat sisältävät 168 195 espanjankielistä sanamerkkiä (10 674 tyyppiä) ja käännökset 159 777 englanninkielistä sanamerkkiä (6 723 tyyppiä).

**Tulos**

Mikä on puhekorpuksen koko?

**Esimerkki 5.2122**

tämä vähentää merkittävästi tarvittavaa koulutusaikaa: SBERT voidaan virittää alle 20 minuutissa, ja samalla saadaan parempia tuloksia kuin vastaavilla lauseiden upottamismenetelmillä.

**Tulos**

Kuinka kauan sen koulutus kestää?

**Esimerkki 5.2123**

Tulevaisuutta ajatellen tässä asiakirjassa tarkasteltujen töiden erinomainen laajennus olisi se, että tutkimuksiin osallistuville useille oppimismenetelmille annettaisiin enemmän riippumattomuutta (esim. vähemmän ihmisen väliintuloa) ja että tuloskuvien kokoa kasvatettaisiin.

**Tulos**

Mitä haasteita on vielä ratkaisematta?

**Esimerkki 5.2124**

Syvästi liikuttuneet lukijat vuodattavat kyyneleitä tai saavat kylmiä väreitä ja kananlihaa jopa laboratorio-olosuhteissa BIBREF4. Tällaisissa tapauksissa tunnereaktio merkitsee itse asiassa esteettistä arviointia: kertomukset, joilla on kyky liikuttaa lukijoita, arvioidaan juuri tästä syystä hyviksi ja voimakkaiksi teksteiksi. Vastaavasti kertomuksissa koettu jännityksen tunne ei ainoastaan reagoi juonen sisällön kulkuun, vaan se myös ennustaa suoraan esteettistä mieltymystä (tai vastenmielisyyttä). Tunteet, joilla on tämä kaksitahoinen kyky, on määritelty "esteettisiksi tunteiksi" BIBREF2.

**Tulos**

Mitä ovat esteettiset tunteet virallistettuina?

**Esimerkki 5.2125**

Bi-LSTM-malli koostuu Bi-LSTM-kerroksesta, jota seuraa lineaarinen kerros syötteen ominaisuuksien poimimiseksi. Bi-LSTM-kerroksessa on 300-ulotteinen piilotila kutakin suuntaa varten. Lopullista luokittelua varten lisätään lineaarinen lisäkerros ennustettujen luokkajakaumien tulostamiseksi. BERT-malli saadaan hienosäätämällä esivalmennettuja BERT-kuvioita NER-tietoihin ylimääräisellä kouluttamattomalla CRF-luokittelijalla. Hienosäädimme kaikkia BERT:n parametreja, myös CRF:n parametreja, päästä päähän.

**Tulos**

Mitä luokittelijoita käytettiin kokeissa?

**Esimerkki 5.2126**

Pointer-Gen+ARL-SEN-mallimme saavuttaa parhaan tuloksen, 60,8 %. Tämä on 18,2 prosentin absoluuttinen parannus Pointer-Gen-perusmalliin verrattuna.

**Tulos**

Mikä on tämän menetelmän parannus parhaaseen nykyaikaiseen menetelmään verrattuna?

**Esimerkki 5.2127**

Täydellinen tietokokonaisuutemme koostuu kaikista Redditissä tammikuusta 2013 joulukuuhun 2014 olleista alaryhmistä, joiden sanastossa on vähintään 500 sanaa, joita käytämme mittareidemme arvioinnissa, vähintään neljän kuukauden ajalta alaryhmän historiasta. Laskemme mittarimme yhteisön käyttäjien kirjoittamien kommenttien perusteella kuukausien aikaikkunoissa jokaiselle riittävän aktiiviselle kuukaudelle ja poistamme manuaalisesti yhteisöt, joissa suurin osa kommenteista on vieraskielisiä.

**Tulos**

Miten vertailun kohteena olevat 300 Reddit-yhteisöä valittiin?

**Esimerkki 5.2128**

Tiedonkeruuasetelmamme käyttää dialogisimulaattoria, jolla luodaan ensin dialogin hahmotelmat ja sitten parafrasoidaan ne luonnollisten lausumien saamiseksi.

**Tulos**

Mistä tietokokonaisuus on peräisin?

**Esimerkki 5.2129**

Kun kyseessä on uutistili Twitterissä, luemme sen twiitit tilin aikajanalta. Sitten lajittelemme twiitit julkaisupäivän mukaan nousevasti ja jaamme ne $N$-kappaleisiin. Kukin palanen koostuu lajitellusta twiittien sarjasta, joka on merkitty vastaavan tilin tunnisteella. Näin ollen tutkimme tapoja havaita epäilyttävät tilit tarkastelemalla niiden twiittejä ryhmissä (chunkeissa). Hypoteesimme on, että epäilyttävillä tileillä on ainutlaatuinen malli twiittien lähettämisessä. Koska heidän tarkoituksenaan on johtaa harhaan, heidän tapansa siirtyä twiittien sarjasta seuraavaan on piilotettu allekirjoitus, joka perustuu heidän aikomuksiinsa. Näin ollen näiden twiittien lukeminen kokonaisuuksina voi parantaa valeuutistilien havaitsemista.

**Tulos**

Miten palaset määritellään?

**Esimerkki 5.2130**

Kun ultraääni- ja äänisegmenttipari on annettu, voimme laskea niiden välisen etäisyyden mallin avulla. Jotta voisimme ennustaa lausuman synkronointivirheen, tarkastelemme diskreettiä joukkoa ehdokkaita, laskemme kunkin etäisyyden keskiarvon lausuman segmenttien välillä ja valitsemme sen, jonka keskimääräinen etäisyys on pienin.

**Tulos**

Ennustaako heidän neuroverkkonsa yksittäisen offsetin tallenteessa?

**Esimerkki 5.2131**

Käytimme Mikolovin 2013 julkaisemaa julkista word2vec-työkalua sanojen upotusten saamiseksi. Heidän neuroverkkomenetelmänsä on samanlainen kuin feed-forward-neuroverkot BIBREF5 , BIBREF6 . Käyttämässämme Skip-gram-malliarkkitehtuurissa valitsimme saatujen sanavektoreiden ulottuvuudeksi 200.

**Tulos**

Minkä tyyppisiä ja kokoisia sanasulkeumia käytettiin?

**Esimerkki 5.2132**

Tietokokonaisuutemme on muodostettu Twitterin etävalvonnan avulla.

**Tulos**

Mistä he saivat tiedot tätä hanketta varten?

**Esimerkki 5.2133**

Tässä jaksossa arvioimme menetelmäämme ja vertaamme sen suorituskykyä kilpaileviin lähestymistapoihin. Käytämme INLINEFORM0 -kertaista arviointiprotokollaa INLINEFORM1:n kanssa satunnaisella tietokokonaisuuden jaolla. Mittaamme suorituskykyä käyttämällä vakiotarkkuusmittaria, jonka määrittelemme testidatajoukon oikein luokiteltujen tietonäytteiden ja kaikkien testinäytteiden väliseksi suhteeksi.

**Tulos**

Mitä arviointimittareita käytetään?

**Esimerkki 5.2134**

Neuraalisen g2p-järjestelmän kouluttamiseksi tarvitaan suuri määrä ääntämistietoja. G2p:n vakiotietoaineisto on Carnegie Mellon Pronouncing Dictionary BIBREF12 . Se on kuitenkin yksikielinen englanninkielinen resurssi, joten se ei sovellu monikieliseen tehtävään. Sen sijaan käytämme kaikissa kokeissa deri2016grapheme-ohjelmalla kerättyä monikielistä ääntämiskorpusta. Tämä korpus koostuu Wiktionarystä poimituista oikeinkirjoitus-äännevastaavuuspareista. Se on jo jaettu harjoitus- ja testijoukkoihin. Korpuksen tilastot on esitetty taulukossa TABREF10 Wiktionarystä poimittujen raakojen IPA-äännevastaavien lisäksi korpus tarjoaa automaattisesti puhdistetun version äännevastaavista. Puhdistus on välttämätön vaihe, koska verkosta poimitut tiedot ovat usein kohinaisia ja niiden transkriptiot saattavat olla epäjohdonmukaisen yksityiskohtaisia. Tässä käytetyllä tietojen puhdistuksella pyritään saamaan transkriptiot yhdenmukaisiksi Phoible BIBREFissä4 käytettyjen foneemisten luetteloiden kanssa.

**Tulos**

mitä tietokokonaisuuksia he käyttivät?

**Esimerkki 5.2135**

Tietokanta sisältää noin 30 miljoonaa lausumaa fiktiivisistä hahmoista.

**Tulos**

Kuinka suuri tietokokonaisuus on?

**Esimerkki 5.2136**

Hyödynnämme äskettäin julkaistua YK:n vuotuisen yleiskeskustelun aikana pidettyjen valtiollisten puheiden korpusta, joka on ensimmäinen valtioiden tekstimuotoista tuotosta sisältävä tietokokonaisuus, joka on tallennettu säännöllisin aikasarjavälein ja joka sisältää otoksen kaikista puheita pitävistä maista BIBREF11 .

**Tulos**

Mitä tietokokonaisuutta he käyttävät?

**Esimerkki 5.2137**

Korpuksen upottaminenEmme ole kiinnostuneita akustisen mallin optimoinnista paljon suuremmalla harjoituskorpusjoukolla INLINEFORM1 , jossa INLINEFORM2 on korpusten lukumäärä ja INLINEFORM3 . Jokainen korpus INLINEFORM4 on kokoelma INLINEFORM5-pareja, joissa INLINEFORM6 on tulo-ominaisuudet ja INLINEFORM7 on sen kohde. tarkoituksemme tässä on laskea sulautus INLINEFORM0 jokaiselle korpukselle INLINEFORM1, jossa INLINEFORM2 odotetaan koodaavan tietoa sen korpuksesta INLINEFORM3 . Nämä upotukset voidaan kouluttaa yhdessä monikielisen standardimallin BIBREF4 kanssa. Ensin alustetaan kaikkien korpusten upotusmatriisi INLINEFORM4 , jonka INLINEFORM5 -rivi INLINEFORM6 vastaa korpuksen INLINEFORM7 upotusta INLINEFORM8 . Seuraavaksi harjoitusvaiheessa INLINEFORM9 voidaan käyttää tulo-ominaisuuden INLINEFORM10 vinouttamiseen seuraavasti. DISPLAYFORM0 jossa INLINEFORM0 on lausuma, joka on poimittu satunnaisesti INLINEFORM1:stä, INLINEFORM2 on sen piilotetut piirteet, INLINEFORM3 on akustisen mallin parametri ja Encoder on pinottu kaksisuuntainen LSTM kuvan mukaisesti. KUVA5 . Seuraavaksi sovellamme kielikohtaista softmaxia logittien INLINEFORM4 laskemiseen ja optimoimme ne CTC-tavoitteella BIBREF29 . Upotusmatriisi INLINEFORM5 voidaan optimoida yhdessä mallin kanssa koulutusprosessin aikana.

**Tulos**

Miten ne laskevat korpustason sulautumat?

**Esimerkki 5.2138**

Tarkastelimme myös tapauksia, joissa kaikki kommentoijat olivat täysin yksimielisiä siitä, että kolmion poiminta oli virheellinen. Näitä kolmioita oli yhteensä 138 kappaletta, jotka olivat peräisin 76 eri lauseesta.

**Tulos**

Mikä on yleisin virhetyyppi?

**Esimerkki 5.2139**

Mathur et al. ehdottivat artikkelissaan loukkaavien twiittien havaitsemiseksi Ternary Trans-CNN -mallia, jossa he kouluttavat malliarkkitehtuurin, joka koostuu kolmesta konvoluution 1D-kerroksesta, joiden suodatinkoot ovat 15, 12 ja 10 ja ytimen koko 3, ja jota seuraa kaksi tiheää täysin kytkettyä kerrosta, joiden koko on 64 ja 3. Ensimmäisessä tiheässä FC-kerroksessa on ReLU-aktivointi ja viimeisessä tiheässä kerroksessa Softmax-aktivointi. He pystyivät kouluttamaan tämän verkon Davidsonin et al. toimittamalla rinnakkaisella englanninkielisellä tietokokonaisuudella. Kirjoittajat pystyivät saavuttamaan 83,9 %:n tarkkuuden, 80,2 %:n tarkkuuden ja 69,8 %:n palautuksen.Lähestymistapa näytti lupaavalta, kun otetaan huomioon, että tietokokonaisuus koostui vain 3189 lauseesta, jotka oli jaettu kolmeen luokkaan, ja näin ollen toistimme kokeen, mutta emme onnistuneet toistamaan tuloksia. Tulokset olivat heikompia kuin alkuperäisten tekijöiden saavuttamat tulokset. Suurin osa mallin hyperparametrien valinnoista oli kuitenkin peräisin tästä työstä.

**Tulos**

Mikä on edellisen työn malli?

**Esimerkki 5.2140**

Tässä käytän Informapin algoritmia BIBREF12. Vertailun vuoksi käytän tässä kokeessa myös CCM:ää ja SCA:ta etäisyyden mittaamiseen. UPGMA-algoritmia käytettäisiin vastaavasti näissä kahdessa tapauksessa.

**Tulos**

Verrataanko ehdotettua menetelmää aiempiin menetelmiin?

**Esimerkki 5.2141**

WikiTableQuestions sisältää 22 033 kysymystä, ja se on kertaluokkaa suurempi kuin aiemmat huipputason tietokokonaisuudet. Kysymyksiä ei ole suunniteltu ennalta määriteltyjen mallien mukaan, vaan käyttäjät ovat laatineet ne käsin, mikä osoittaa suurta kielellistä vaihtelua.

**Tulos**

Miten he keräävät tietoja kyselyn selitysongelmaa varten?

**Esimerkki 5.2142**

Tätä tarkoitusta varten valitsimme monipuolisen joukon kahdeksan eri kieliperheisiin kuuluvaa lähdekieltä - baskia, ranskaa, saksaa, unkaria, italiaa, navajoa, turkkia ja quechua - ja kolme kohdekieltä - englantia, espanjaa ja zulua.

**Tulos**

Mitä puun kohdekieliä tutkielmassa tarkastellaan?

**Esimerkki 5.2143**

Tässä asiakirjassa ehdotamme kolmea erityyppistä CRU-mallia: matala fuusio, syvä fuusio ja syväparannettu fuusio, perustavimmasta perustavimpaan ja ilmaisuvoimaisimpaan. Yksinkertaisin malli on soveltaa CNN-kerrosta suoraan upotuskerroksen jälkeen, jotta saadaan sekoitettuja kontekstuaalisia esityksiä. Sen jälkeen käytetään GRU-kerrosta.

**Tulos**

Miten CNN syötetään uusiin yksiköihin?

**Esimerkki 5.2144**

Takaisinottoriskiluokittimen syötteeksi poimittiin 45 kliinisesti tulkittavissa olevaa piirrettä ottoa kohti. Nämä piirteet voidaan ryhmitellä kolmeen ryhmään (täydellinen luettelo piirteistä on taulukossa TABREF5): Sosiodemografiset tiedot: sukupuoli, ikä, siviilisääty jne. Aikaisempi sairaushistoria: aiempien sairaalahoitojen määrä, itsetuhoisuus, keskimääräinen hoitojakson pituus (kyseiseen sairaalahoitoon asti) jne. Nykyisen sairaalahoidon tiedot: hoitojakson pituus, itsetuhoisuusriski, muistiinpanojen määrä ja pituus, kotiuttamisajankohta, arviointipisteet jne. Nämä piirteet voidaan jakaa edelleen kahteen ryhmään: "strukturoidut" kliiniset piirteet ja "strukturoimattomat" kliiniset piirteet. feature extraction ::: Strukturoidut piirteetSrukturoidut piirteet ovat piirteitä, jotka tunnistettiin EHR:stä säännöllisen lausekkeen täsmäytyksen avulla ja jotka sisältävät luokituspisteitä, joiden on raportoitu psykiatrisessa kirjallisuudessa korreloivan lisääntyneen takaisinottoriskin kanssa, kuten Global Assessment of Functioning, Insight and Compliance:Global Assessment of Functioning (GAF): Potilaan psykososiaalinen toimintakyky, joka vaihtelee 100:sta (erittäin hyvä toimintakyky) 1:een (vakavasti heikentynyt) BIBREF13.Insight: Se, missä määrin potilas tunnistaa ja hyväksyy sairautensa (joko hyvä, tyydyttävä tai huono): Potilaan kyky noudattaa lääkitystä ja noudattaa lääkärin ohjeita (joko Kyllä, Osittain tai Ei lainkaan).Näitä ominaisuuksia käytetään laajasti kliinisessä käytännössä, ja niillä arvioidaan potilaan yleistä tilaa ja ennustetta potilaan arvioinnin aikana.Feature Extraction ::: Strukturoimattomat piirteetSrukturoitujen piirteiden tarkoituksena on tallentaa potilaan tila suhteessa seitsemään riskitekijäalueeseen (ulkonäkö, ajatusprosessi, ajatusten sisältö, ihmissuhteet, päihteiden käyttö, ammatti ja mieliala) sähköisen potilastietojärjestelmän vapaan tekstin kertomuksista. Näiden seitsemän osa-alueen on todettu olevan yhteydessä takaisinottoriskiin aiemmassa BIBREF-työssä14 : 1) niiden lausekkeiden suhteellinen määrä vastaanottokertomuksissa, jotka liittyvät kuhunkin riskitekijäalueeseen (vastaanottokertomuksen lausekkeiden kokonaismäärästä) ja 2) kliiniset tunnepisteet jokaiselle näistä riskitekijäalueista, eli tunnepisteet, joilla arvioidaan potilaan psykososiaalisen toiminnan tasoa (positiivinen, negatiivinen tai neutraali) kunkin riskitekijäalueen osalta.

**Tulos**

Mitä ominaisuuksia käytetään?

**Esimerkki 5.2145**

Uutuuksiamme ovat: Itsepelattavan oppimisen käyttäminen neuraalisessa vastausluokittelijassa (kuvattu yksityiskohtaisesti alla).Neuraalisten mallien optimointi tietyille metriikoille (esim. monimuotoisuus, johdonmukaisuus) ensemble-asetuksessamme.Erillisen dialogimallin kouluttaminen jokaiselle käyttäjälle, mikä personoi sosiaalista robottiamme ja tekee siitä johdonmukaisemman.Vastausluokittelun ennustajan ja vastausluokittelijan käyttäminen vastausten näkökohtien, kuten tunnelman, aiheen, loukkaavuuden, monimuotoisuuden jne., ennustamiseen ja hallintaan.Mallin ennustajan käyttäminen parhaan vastausmallin ennustamiseen ennen vastausehdokkaiden luomista, mikä vähentää laskentakustannuksia. entropiapohjaisen suodatustekniikan käyttäminen kaikkien dialogidatajoukkojen suodattamiseen, jolloin saadaan laadukkaampaa harjoitusdataa BIBREF3. suurten, valmiiksi koulutettujen hierarkkisten BERT- ja GPT-dialogimallien rakentaminen BIBREF6, BIBREF7, BIBREF8. käyttäjän syötteen jatkuva seuranta automaattisten mittareidemme avulla, mikä varmistaa, että käyttäjä pysyy mukana.

**Tulos**

Mikä on uutta kirjailijan lähestymistavassa?

**Esimerkki 5.2146**

Näiden seitsemän kielen merkistöissä on vain vähän päällekkäisyyksiä, paitsi että i) ne sisältävät kaikki yhteiset latinalaiset perusaakkoset ja ii) sekä hindi että marathi käyttävät devanagari-kirjoitusta. Otimme 7 merkkijoukon liiton monikieliseksi grafeemijoukoksi (SECREF2), joka sisälsi 432 merkkiä.

**Tulos**

Kuinka suuri osa ASR-grafeemien joukosta on jaettu eri kielten välillä?

**Esimerkki 5.2147**

Hashtagien ennustamista sosiaalisessa mediassa on käsitelty aiemmin, esimerkiksi BIBREF15 , BIBREF16 . BIBREF15 käyttää myös neuraalista arkkitehtuuria, mutta muodostaa tekstin upotukset sanojen hakutaulukosta.

**Tulos**

Onko tämä hashtagien ennustamistehtävä vakiintunut tehtävä vai jotain uutta?

**Esimerkki 5.2148**

Suurin osa tunneanalyysin sanastoresursseista on englanninkielisiä. Jotta näitä lähteitä voitaisiin silti hyödyntää, AffectiveTweets-paketin leksikoni käännettiin espanjaksi käyttäen konekäännösalustaa Apertium BIBREF5.

**Tulos**

Miten harjoitusaineisto käännettiin?

**Esimerkki 5.2149**

TextCat on tunnetuin Perl-toteutus out-of-place-menetelmästä, ja se sisältää mallit 76 kielelle valmiissa kokoonpanossaan; ohjelmaa ei ylläpidetä aktiivisesti. TextCat on tunnetuin Perl-toteutus out-of-place-menetelmästä, se luettelee mallit 76 kielelle valmiissa kokoonpanossaan; ohjelmaa ei ylläpidetä aktiivisesti. TextCat ei ole ainoa esimerkki out-of-place-menetelmän valmiista toteutuksesta: muita toteutuksia ovat muun muassa libtextcat, jossa on 76 kielimallia, JTCL, jossa on 15 kieltä, ja mguesser, jossa on 104 mallia eri kielikoodipareille. Tärkein ongelma, johon myöhemmät toteutukset puuttuvat, on luokittelun nopeus: TextCat on toteutettu Perl-kielellä eikä sitä ole optimoitu nopeuden kannalta, kun taas libtextcatin ja mguesserin kaltaiset toteutukset on kirjoitettu nimenomaan nopeiksi ja tehokkaiksi. whatlang-rs käyttää algoritmia, joka perustuu merkkitrigrammaan, ja se viittaa käyttäjälle BIBREF7-artikkeliin. Se on valmiiksi koulutettu 83 kielen kanssa. on Google Chrome -selaimeen upotettu kielitunniste. Se käyttää NB-luokittelijaa ja käsikirjoituskohtaisia luokittelustrategioita. olettaa, että kaikki syötteet ovat UTF-8-muodossa, ja antaa koodauksen tunnistamisen ja transkoodauksen käyttäjän tehtäväksi. käyttää Unicode-tietoa syötteen käsikirjoituksen määrittämiseen. toteuttaa myös useita esikäsittelyn heuristiikkoja, jotka auttavat parantamaan suorituskykyä kohdealueella (verkkosivut), kuten poistamaan merkkijaksot, kuten .jpg. Vakiototeutus tukee 83 kieltä, ja saatavilla on myös laajennettu malli, joka tukee 160 kieltä. on Java-kirjasto, joka toteuttaa kielitunnisteen, joka perustuu NB-luokittimeen, joka on koulutettu merkkien . Ohjelmiston mukana toimitetaan valmiiksi koulutetut mallit 53 kielelle käyttäen Wikipediasta saatuja tietoja. käyttää erilaisia normalisointiheuristiikkoja parantaakseen suorituskykyä tietyillä kielillä, mukaan lukien: (1) kiinalaisten/japanilaisten/korealaisten merkkien klusterointi harvojen merkkien vähentämiseksi; (2) "kielestä riippumattomien" merkkien poistaminen ja muu tekstin normalisointi; ja (3) arabialaisten merkkien normalisointi.on Python-toteutus BIBREF150 :n kuvaamasta menetelmästä, jossa hyödynnetään saman kielen harjoitteluaineistoa useista eri tekstilähteistä sellaisten merkkisarjojen tunnistamiseksi, jotka ennustavat vahvasti tiettyä kieltä tekstin lähteestä riippumatta. Tämä ominaisuusjoukko yhdistetään NB-luokittimeen, ja se jaetaan valmiiksi koulutetun mallin kanssa 97 kielelle, joka on laadittu käyttäen dataa viidestä eri tekstilähteestä. BIBREF151 vertailee empiirisesti mallia ja , ja toteaa, että se on hyvä sekä tarkkuuden että luokittelunopeuden suhteen. BIBREF153 käyttää vektoriavaruusmallia, jossa merkkisekvenssejä painotetaan ominaisuuksittain. Yksi sen erityispiirre on se, että se käyttää piirteiden valinnassa syrjivää koulutusta, eli se käyttää erityisesti sellaisia piirteitä, jotka ovat vahvoja todisteita tiettyä kieltä vastaan, mitä NB-mallit eivät yleensä huomioi. Toinen ominaisuus on se, että se käyttää merkkijonojen välistä tasoitusta hyödyntääkseen lausetason paikallisuutta kieliennusteiden tekemisessä olettaen, että vierekkäiset lauseet ovat todennäköisesti samaa kieltä. BIBREF153 raportoi, että tämä parantaa huomattavasti tunnisteen tarkkuutta. Whatthelang on Python-kielellä kirjoitettu tuore kielitunnistin, joka käyttää FastTextin NN-pohjaista tekstinluokittelualgoritmia. Se tukee 176 kieltä.implementoi valmiiksi koulutetun luokittelijan, joka on koulutettu Wikipedian datan avulla ja joka kattaa 122 kieltä. Vaikka sitä ei ole kuvattu sellaisenaan, varsinainen käytetty luokittelualgoritmi on lineaarinen malli ja siten läheistä sukua sekä NB:lle että kosinipohjaiselle vektoriavaruusmallille. edellä mainittujen yleiskäyttöisten kielitunnistimien lisäksi on pyritty tuottamaan valmiiksi koulutettuja kielitunnistimia, jotka on suunnattu erityisesti Twitter-viesteille. on Twitter-kohtainen työkalu, jossa on sisäänrakennetut mallit 19:lle kielelle. Se käyttää asiakirjan esitystapaa, joka perustuu tries BIBREF401 . Algoritmi on LR-luokittelija, joka käyttää datan kaikkia mahdollisia osajonoja, mikä on tärkeää, jotta suhteellisen lyhyistä Twitter-viesteistä saadaan mahdollisimman paljon tietoa.BIBREF152 vertailee kahdeksaa valmista kielitunnistinta, joita sovelletaan ilman uudelleenkoulutusta Twitter-viesteihin. Heidän mukaansa yksi ongelma on se, että valmiiden järjestelmien tarkkuuden vertailu on vaikeaa, koska kunkin järjestelmän tukema kielten osajoukko on erilainen, eikä se välttämättä myöskään kata kaikkia kohdedatassa esiintyviä kieliä. Kirjoittajat päättävät verrata tarkkuutta koko kielivalikoiman osalta, koska tämä kuvastaa parhaiten todennäköistä käyttötapaa, jossa valmisjärjestelmä sovelletaan uuteen dataan. He havaitsevat, että parhaat yksittäiset järjestelmät ovat , ja , mutta hieman korkeampaan tarkkuuteen päästään yksinkertaisella äänestykseen perustuvalla yhdistelmäluokittimella, jossa on mukana nämä kolme järjestelmää.Tämän lisäksi on olemassa kaupallisia tai muita suljetun lähdekoodin kielitunnistimia ja kielitunnistuspalveluja, joista mainitsemme muutamia. Polyglot 3000 ja Lextek Language Identifier ovat itsenäisiä kielitunnistimia Windowsille. Open Xerox Language Identifier on verkkopalvelu, jossa on saatavilla REST- ja SOAP API:t.

**Tulos**

Mitkä ovat asiakirjassa käsitellyt valmiit järjestelmät?

**Esimerkki 5.2150**

JESSI koulutetaan käyttämällä ainoastaan jaetussa tehtävässä annettuja tietokokonaisuuksia, eikä siinä käytetä mitään muita ulkoisia tietoja.

**Tulos**

Mitä tietokokonaisuuksia käytettiin?

**Esimerkki 5.2151**

Arvioimme TN-teemamallia kvantitatiivisesti tavanomaisilla teemamallien mittareilla, kuten testijoukon perpleksisyydellä, todennäköisyyden konvergenssilla ja klusterointimittareilla. Laadullisesti arvioimme mallia visualisoimalla aiheyhteenvetoja, kirjoittajien aihejakaumia ja suorittamalla automaattisen merkintätehtävän.

**Tulos**

Mitkä ovat tässä asiakirjassa käytetyt "suorituskyvyn" mittarit?

**Esimerkki 5.2152**

Tässä asiakirjassa ehdotetaan, että lähdetietokannan merkityksellisimpiä näytteitä käytetään kohdetietokannan harjoitteluun. Yksi tapa löytää samankaltaisimmat näytteet on löytää kohdetietokannan ja lähdetietokannan kehitysjoukon kaikkien näytteiden välinen pareittainen etäisyys. ehdotamme, että kehitysjoukkoon käytetään klusterointialgoritmia. Tässä käytetty klusterointialgoritmi on hierarkkinen klusterointialgoritmi. Kosinuksen samankaltaisuutta käytetään kriteerinä kunkin kysymyksen ja vastauksen klusterointiin. Näin ollen nämä klusterit edustavat kohdetietokannan kehitysjoukkoa, ja kunkin klusterin vastaava keskus edustaa kaikkia kyseiseen klusteriin kuuluvia näytteitä. Seuraavassa vaiheessa kunkin keskuksen etäisyyttä käytetään kosinin samankaltaisuuden laskemiseen. Lopuksi lähdetietokannan näytteet, jotka ovat kaukana näistä keskuksista, jätetään huomiotta. Toisin sanoen poikkeavat näytteet eivät osallistu siirto-oppimiseen.

**Tulos**

Miten malli siirretään?

**Esimerkki 5.2153**

Arvioimme ehdotettua lähestymistapaa kiinalaisen sosiaalisen median tekstin tiivistämistehtävässä, joka perustuu sekvenssistä sekvenssiin -malliin. Suuren mittakaavan kiinalaisen lyhyen tekstin tiivistämistietokanta (Large-Scale Chinese Short Text Summarization Dataset, LCSTS) on muodostettu BIBREF1 -tietokannan avulla. Tietokanta koostuu yhteensä yli 2,4 miljoonasta teksti-tiivistelmäparista, jotka on muodostettu kuuluisasta kiinalaisesta sosiaalisen median mikroblogipalvelusta Weibosta.

**Tulos**

Ilmoitetaanko tulokset vain englanninkielisistä tiedoista?

**Esimerkki 5.2154**

Kuten kohdassa "sec:coalitionpolicy" todettiin, Euroopan yhdentymisen kannattajat ja vastustajat jakautuvat voimakkaasti kahteen lohkoon, mikä näkyy vielä selvemmin kuviossa FIGREF42 B. Unionin valtio ja kehitys -alueella havaitaan jälleen voimakas jakautuminen kahteen lohkoon (ks. kuvio FIGREF42 E). Tämä eroaa kuitenkin talous- ja rahajärjestelmästä, jossa havaitaan äärivasemmiston ja äärioikeiston yhteistyötä, jossa jako on perinteisen vasemmisto-oikeistoakselin suuntainen.Twitterissä muodostuvien koalitioiden mallit muistuttavat läheisesti Euroopan parlamentin malleja.

**Tulos**

Havaitaanko analyysissä, että koalitiot muodostuvat samalla tavalla eri politiikan aloilla?

**Esimerkki 5.2155**

Yleisemmin ottaen tiedämme vain vähän siitä, millaisia kehityskysymyksiä eri maat pitävät ensisijaisina tai ovatko maakohtaiset tekijät, kuten vauraus tai demokratia, syynä siihen, että maat ajavat todennäköisemmin tiettyjen kehityskysymysten ottamista maailmanpoliittiselle asialistalle. Huomasimme, että maakohtaiset tekijät, kuten vauraus, väestö, demokratia, julkisen kehitysavun määrä ja konfliktit, eivät vaikuta merkittävästi aihealueesta 2 käytävään keskusteluun (joskin alueellisia vaikutuksia on havaittavissa).

**Tulos**

Mitkä ovat kansainvälisen kehitysretoriikan maakohtaiset ajurit?

**Esimerkki 5.2156**

Ternaarinen ja hienojakoinen tunteiden luokittelu olivat osa SemEval-2016 "Sentiment Analysis in Twitter" -tehtävää BIBREF16 . Käytämme haasteen järjestäjien julkaisemia korkealaatuisia tietokokonaisuuksia.

**Tulos**

Mitä tietokokonaisuutta he käyttivät?

**Esimerkki 5.2157**

Tavoitteenamme on antaa NMT-järjestelmälle tietoa ensimmäisen persoonan lauseiden puhujasta ja keskustelukumppanista, jotta se voi tuottaa halutun kohdepuolen morfologian, kun tietoa ei ole saatavilla lähdelauseesta. Lähestymistapamme tässä työssä on mustan laatikon injektio, jossa yritämme syöttää tietoa syötteeseen vaikuttaaksemme koulutetun NMT-järjestelmän tulosteeseen ilman, että meillä on pääsyä sen sisäisiin osiin tai sen koulutusmenettelyyn, kuten vanmassenhove-hardmeier-way:2018:EMNLP:ssä ehdotetaan. Tämän todentamiseksi kokeilemme lauseiden kääntämistä seuraavilla variaatioilla: Ei etuliitettä - GMT-järjestelmän palauttama peruskäännös. "He said:"-merkki miespuolisesta puhujasta. Odotamme järjestelmän vinoutuvan entisestään kohti maskuliinisia muotoja. "She said:"-merkki naispuolisesta puhujasta ja tuntemattomasta yleisöstä. Koska tämä vastaa puhujan todellista sukupuolta, odotamme, että ensimmäisen persoonan pronominien ja verbien, joiden subjektina on ensimmäisen persoonan pronomini, käännös paranee. "I said to them:"-merkki tuntemattomasta puhujasta ja monikossa olevasta yleisöstä. "He said to them:"-Maskuliininen puhuja ja monikossa oleva yleisö. "She said to them:"-Nainen puhuja ja monikossa oleva yleisö - täydellinen, oikea tila. Odotamme parasta käännöstarkkuutta tässä asetelmassa. "He/she said to him/her"- Tässä asetamme (virheellisen) yksikön sukupuoleen merkityn yleisön tutkiaksemme kykyämme hallita yleisön morfologiaa.

**Tulos**

Mitkä ovat black-box-kontekstin injektiojärjestelmän osat?

**Esimerkki 5.2158**

Tässä artikkelissa ehdotamme kahta neuraalista upotusmallia, joiden avulla voidaan oppia jatkuvia käsitevektoreita, jotka perustuvat BIBREF11 -hyppygrammimalliin. Ensimmäinen mallimme on Concept Raw Context -malli (CRC), joka hyödyntää käsitteiden mainintoja laajamittaisessa tietokannassa oppiakseen yhdessä sekä sanojen että käsitteiden upotukset. Toinen mallimme on Concept-Concept Context -malli (3C), joka oppii käsitteiden upotukset niiden käsitteellisistä konteksteista (eli konteksteista, jotka sisältävät vain ympäröiviä käsitteitä).

**Tulos**

Mitkä ovat kaksi neuraalista upotusmallia?

**Esimerkki 5.2159**

Käytämme kahta erilaista valvomatonta lähestymistapaa sanojen merkityksen erotteluun. Toinen, nimeltään "tiheä malli", edustaa synsetit ja kontekstit tiheässä, matalaulotteisessa avaruudessa keskiarvoistamalla sanojen upotukset. Synsettien sulautusmallia käytettäessä noudatamme SenseGram BIBREF14 -mallia ja sovellamme sitä synonyymien graafista indusoituihin synsetteihin. Havaitsemme, että SenseGram-pohjainen lähestymistapa sanojen merkitysten erotteluun tuottaa huomattavasti parempia tuloksia kaikissa tapauksissa (taulukko TABREF25 ). Pääsyy tähän on samankaltaisten sanojen implisiittinen käsittely, joka johtuu semanttisesti toisiinsa liittyvien sanojen tiheiden sanavektoreiden keskiarvoistamisesta.

**Tulos**

Esittävätkö kirjoittajat hypoteeseja siitä, miksi tiheä tila oli parempi kuin harva tila?

**Esimerkki 5.2160**

Tässä tutkimuksessa tutkimme harhojen vähentämisen tehokkuutta harjoittelun aikana ottamalla käyttöön uuden tappiofunktion, joka kannustaa kielimallia tasoittamaan sukupuolittuneiden sanaparien, kuten hän ja nainen, ennustamisen todennäköisyyksiä.

**Tulos**

millaisia mies- ja naissanoja tarkastellaan?

**Esimerkki 5.2161**

Käytämme MedWeb-tietokokonaisuutta BIBREF4 ("Medical Natural Language Processing for Web Document"), joka toimitettiin osana NTCIR-13-konferenssin BIBREF5 osatehtävää. Tietojen yhteenveto on taulukossa TABREF1. Pseudotwiittejä on yhteensä 2560 kappaletta kolmella eri kielellä: Japani (ja), englanti (en) ja kiina (zh).

**Tulos**

Kuinka suuri on tietokokonaisuus, jota käytetään mallin hienosäätöön punaisen lipun lääketieteellisten oireiden havaitsemiseksi yksittäisissä lausunnoissa?

**Esimerkki 5.2162**

Opetussuunnitelman yleismaailmallisuuden ja yleisen sovellettavuuden varmistamiseksi suoritamme perusteellisen tutkimuksen kolmella julkisesti saatavilla olevalla keskustelukorpuksella, jotka ovat PersonaChat BIBREF12, DailyDialog BIBREF13 ja OpenSubtitles BIBREF7 ja jotka koostuvat 140 248, 66 594 ja 358 668 todellisesta keskustelunäytteestä.

**Tulos**

Mitä kolmea julkisesti saatavilla olevaa coroporaa käytetään?

**Esimerkki 5.2163**

DNN-pohjaiset mallit yhdistettynä siirto-oppimiseen päihittävät parhaat tunnetut tulokset kaikissa kolmessa tietokokonaisuudessa. Aiemmat parhaat F1-tulokset Wikipedia BIBREF4- ja Twitter BIBREF8 -tietokannoille olivat 0,68 ja 0,93. Saavutamme molemmille tietokokonaisuuksille F1-pisteet 0,94 käyttämällä BLSTM-malleja, joissa on huomio- ja ominaisuustason siirto-oppiminen (taulukko TABREF25 ). Formspring-tietokannan tekijät eivät ole ilmoittaneet F1-tulosta. Heidän menetelmänsä tarkkuus on 78,5 % BIBREF2 . Meidän F1-tuloksemme on 0,95 ja tarkkuus 98 % samasta tietokokonaisuudesta.

**Tulos**

Millaisia tuloksia he saavuttivat?

**Esimerkki 5.2164**

Poistamme käyttäjän iän soveltamalla säännöllisten lausekkeiden malleja profiilin kuvauksiin (kuten "17-vuotias, itsensä vahingoittaminen, ahdistus, masennus") BIBREF41 . Kootaan "ikäprefiksit" ja "ikäsuffiksit" ja käytetään kolmea iän poimintasääntöä: 1. Olen X vuotta vanha 2. Syntynyt vuonna X 3. X vuotta vanha, jossa X on "päivämäärä" tai ikä (esim. 1994). Valitsimme INLINEFORM1:stä 1464 käyttäjän INLINEFORM0 osajoukon, jotka paljastavat sukupuolensa profiilikuvauksessaan.

**Tulos**

Mistä yksilötason demografisia tietoja saadaan?

**Esimerkki 5.2165**

Teemme kokeilumme 8,7 miljoonan annotoidun ja anonymisoidun käyttäjän lausumilla. Ne on annotoitu ja johdettu 23 toimialueen pyynnöistä.

**Tulos**

Minkä tietokokonaisuuksien/korporaatioiden perusteella tätä työtä arvioidaan?

**Esimerkki 5.2166**

Profiilien todentamisprosessissa havaitsimme, että useimmat jengiläisten profiilit kuvaavat jengikulttuuria edustavaa kontekstia. Esimerkkejä tällaisista profiilikuvista on esitetty kuvassa KUVA32 , jossa käyttäjällä on kädessään tai osoittamassa aseita, hänet nähdään jengikulttuuria ilmentävässä ryhmäkuvassa tai hän esittelee graffiteja, käsimerkkejä, tatuointeja ja suuria määriä käteistä rahaa.

**Tulos**

Mitä eroja kuvien käytössä on jengiläisten ja muun Twitter-väestön välillä?

**Esimerkki 5.2167**

Pääkirjat: 9) Klassiset teostyypit ("sinfonia", "ouvertuuri" jne.); 9) Klassiset teostyypit ("sinfonia", "ouvertuuri" jne.).); 10)Soittimet; 11)Opus-muodot ("op", "opus"); 12)Teoksen numeromuodot ("no", "number"); 13)Teoksen avaimet ("C", "D", "E", "F", "G", "A", "B", "tasa", "terävä"); 14)Teoksen moodit ("duuri", "molli", "m").

**Tulos**

Mitä tehtäväkohtaisia ominaisuuksia käytetään?

**Esimerkki 5.2168**

Hienosäätöön käytämme seuraavaa GLUE-tehtävien osajoukkoa BIBREF4: MRPC: Microsoft Research Paraphrase Corpus BIBREF13STS-B: Semantic Textual Similarity Benchmark BIBREF14SST-2: Stanford Sentiment Treebank, kaksisuuntainen luokittelu BIBREF15QQP: Quora Question Pairs datasetRTE: Recognizing Textual Entailment datasetsQNLI: Stanford Question Answering Dataset -tietokantaan perustuva Question-answering NLI BIBREF3MNLI: Multi-Genre Natural Language Inference Corpus, täsmäytetty osa BIBREF16

**Tulos**

Mitä GLUE-tehtävien osajoukkoa käytetään?

**Esimerkki 5.2169**

GE-FL vähentää instanssien annotaation aiheuttamaa raskasta kuormitusta ja toimii hyvin, kun annamme ennakkotietoa ilman ennakkoasenteita. Kokeissamme havaitsimme, että kullekin luokalle on toimitettava vertailukelpoinen määrä merkittyjä piirteitä. Valitsemme satunnaisesti $t \ in [1, 20]$ piirteitä piirrejoukosta yhdelle luokalle ja vain yhden piirteen toiselle luokalle. Menetelmiämme arvioidaan myös tietokokonaisuuksilla, joissa on erilaisia epätasapainoisia luokkajakaumia. Rakennamme manuaalisesti useita elokuvatietoaineistoja, joiden luokkajakaumat ovat 1:2, 1:3 ja 1:4, poistamalla satunnaisesti 50 %, 67 % ja 75 % positiivisia asiakirjoja. KL-divergenssin sisällyttäminen on riittävän vankka keino hallita epätasapainoa sekä tietokannassa että merkityissä ominaisuuksissa, kun taas kolme muuta menetelmää eivät ole yhtä kilpailukykyisiä.

**Tulos**

Miten ne määrittelevät mallin kestävyyden?

**Esimerkki 5.2170**

Tulokset ::: Tässä asetelmassa vertaamme ympäristössä suoritettujen toimintojen määrää (kehykset) ja agentin saavuttamaa pistemäärää (eli +1 palkkio, jos kolikko on kerätty). Go-Exploressa laskemme myös toiminnot, joita käytettiin ympäristön palauttamiseen valittuun soluun, eli agentin saattamiseen valitun solun edustamaan tilaan. Tämä mahdollistaa Go-Exploren ja algoritmien, jotka käyttävät laskentaan perustuvaa palkitsemista tekstipohjaisissa peleissä, välisen tutkimustehokkuuden yksiselitteisen vertailun. Tärkeää on, että BIBREF8 osoitti, että DQN ja DRQN eivät ilman tällaisia laskentapalkkioita koskaan pysty löytämään onnistunutta rataa vaikeissa peleissä, kuten kokeissamme käytetyissä peleissä. Kuvassa FIGREF17 esitetään vuorovaikutusten määrä ympäristön kanssa (kehykset) suhteessa saavutettuun maksimipistemäärään keskiarvona 10:stä saman vaikeusasteen pelistä. Kuten BIBREF8 osoittaa, DRQN++ löytää radan, jolla saavutetaan maksimipisteet, nopeammin kuin DQN++. Toisaalta Go-Exploren vaihe 1 löytää optimaalisen radan noin puolet harvemmalla vuorovaikutuksella ympäristön kanssa. Lisäksi Go-Exploren löytämän liikeradan pituus on aina optimaalinen (30 askelta), kun taas DQN++:n keskimääräinen pituus on 38 ja DRQN++:n 42 askelta. CookingWorld-pelin CoinCollector tapauksessa vertailimme malleja kolmessa aiemmin mainitussa asetelmassa, jotka olivat yksittäinen, yhteinen ja nollalaukaus. Kaikissa kokeissa mittasimme kaikkien pelien loppupisteiden summan ja niiden liikeradan pituuden (askelten määrä). Taulukossa TABREF26 on yhteenveto tuloksista näissä kolmessa asetelmassa. Go-Exploren vaiheessa 1 yksittäisillä peleillä saavutetaan kokonaispisteet 19 530 (kaikkien pelien summa), joka on hyvin lähellä suurinta mahdollista pistemäärää (eli 19 882 pistettä), 47 562 askeleella. Voittorata löytyi 4 279 pelissä 4 440 pelistä. Tämä tulos vahvistaa jälleen kerran, että Go-Exploren tutkimusstrategia on tehokas tekstipohjaisissa peleissä. Seuraavaksi arvioimme Go-Exploren vaiheessa 1 poimittujen liikeratojen avulla koulutetun yksinkertaisen jäljittely-oppimiskäytännön tehokkuutta ja yleistettävyyttä edellä mainituissa kolmessa ympäristössä. Tässä asetelmassa kukin malli koulutetaan tyhjästä jokaisessa 4440 pelissä Go-Exploren vaiheessa 1 (edellinen vaihe) löydetyn liikeradan perusteella. Kuten taulukosta TABREF26 käy ilmi, LSTM-DQN BIBREF7, BIBREF8 -lähestymistapa, jossa ei käytetä sallittuja toimintoja, toimii huonosti. Yksi selitys tälle voi olla se, että tämän mallin on vaikea tutkia sekä kieltä että pelistrategiaa samanaikaisesti; mallin on vaikea löytää palkintosignaalia ennen kuin se on oppinut mallintamaan kieltä, koska lähes yksikään sen toimista ei ole sallittu, ja juuri nämä palkintosignaalit ovat välttämättömiä kielimallin oppimiseksi. Kuten taulukosta TABREF26 nähdään, LSTM-DQN:n saavuttama pistemäärä kasvaa kuitenkin dramaattisesti käyttämällä sallittuja toimintoja $\epsilon $-greedy-vaiheessa (+ADM-rivi taulukossa TABREF26). DRRN BIBREF10 saavuttaa erittäin korkean pistemäärän, koska se oppii nimenomaisesti asettamaan hyväksyttävät toiminnot paremmuusjärjestykseen (eli paljon yksinkertaisempi tehtävä kuin tekstin tuottaminen). Lopuksi, lähestymistapamme, jossa käytetään Seq2Seq-mallia, joka on koulutettu Go-Exploren vaiheen 1 antamalla yksittäisellä liikeradalla, saa kaikista menetelmistä korkeimman pistemäärän, vaikka emme käytä sallittuja toimintoja tässä vaiheessa. Tässä kokeessa Seq2Seq-malli ei kuitenkaan pysty täydellisesti jäljittelemään annettua liikerataa, ja sen saavuttama kokonaispistemäärä on itse asiassa 9,4 prosenttia pienempi kuin Go-Exploren vaiheessa 1 saavutettu kokonaispistemäärä. Kuvassa FIGREF61 (liitteessä SECREF60) esitetään pisteiden jakautuminen kunkin tason ja mallin osalta, ja siinä näkyy, että mallimme ja muiden menetelmien välinen ero kasvaa, kun pelit vaikeutuvat tarvittavien taitojen osalta. Tässä asetelmassa 4440 peliä on jaettu harjoitus-, validointi- ja testipeleihin. Jako tehdään satunnaisesti, mutta siten, että eri vaikeustasot (reseptit 1, 2 ja 3) ovat edustettuina yhtä suurella osuudella kaikissa kolmessa jaossa eli vaikeusasteen mukaan jaoteltuina. Kuten taulukosta TABREF26 käy ilmi, RL-perusratkaisujen suorituskyky nollapisteissä on heikko, mikä voi johtua samoista syistä, joiden vuoksi RL-perusratkaisut alisuoriutuvat Joint-tapauksessa. Erityisen mielenkiintoista on se, että DRRN-mallin suorituskyky on huomattavasti heikompi kuin Go-Explore Seq2Seq -mallin, vaikka DRRN-mallilla on pääsy sallittuihin toimintoihin testiajankohtana, kun taas Seq2Seq-malli (samoin kuin LSTM-DQN-malli) joutuu rakentamaan toimintoja merkki kerrallaan koko 20 000 merkin sanastosta. Toisaalta Go-Explore Seq2Seq osoittaa lupaavia tuloksia, sillä se ratkaisee melkein puolet näkemättömistä peleistä. Kuva FIGREF62 (liitteessä SECREF60) osoittaa, että suurin osa hävityistä peleistä kuuluu vaikeimpaan joukkoon, jossa pelin voittaminen edellyttää hyvin pitkää toimintasarjaa. Nämä tulokset osoittavat sekä Seq2Seq-mallin kouluttamisen suhteellisen tehokkuuden Go-Explore-pelien liikeradoilla että sen, että tarvitaan lisäponnistuksia sellaisten vahvistusoppimisalgoritmien suunnittelemiseksi, jotka yleistyvät tehokkaasti näkymättömiin peleihin.

**Tulos**

Kuinka paremmin uusi lähestymistapa toimii kuin nykyiset ratkaisut?

**Esimerkki 5.2171**

Ehdotetussa mallissa yhdistetään kaksi solmua (sanaa) toisiinsa esiintymisreunojen lisäksi, jos vastaava sanan upotuskuvaus on samanlainen.

**Tulos**

Käyttävätkö ne pelkkiä sanojen upotuksia vai korvaavatko ne jotkin mallin aiemmat piirteet sanojen upotuksilla?

**Esimerkki 5.2172**

Otokseen valittiin kaikki julkaisut, jotka julkaistiin seuraavissa tietotekniikan alaluokissa: Artificial Intelligence (cs.AI), Machine Learning (cs.LG), Social and Information Networks (cs.SI), Computational Linguistics (cs.CL), Computers and Society (cs.CY), Information Retrieval (cs.IR) ja Computer Vision (CS.CV), Computer Science (stat.ML) ja Social Physics (physics.soc-ph). Suodatimme artikkelit, joiden otsikossa tai tiivistelmässä oli vähintään yksi sanoista "machine learning", "classif\*" tai "supervi\*" (isoja ja pieniä kirjaimia ei huomioida). Tämän jälkeen suodatimme paperit, joiden otsikossa tai tiivistelmässä oli vähintään sana "twitter" tai "tweet" (isoja ja pieniä kirjaimia ei huomioida), jolloin tulokseksi saatiin 494 paperia. Käytimme samaa kyselyä Elsevierin vertaisarvioituja artikkeleita sisältävään Scopus-tietokantaan ja valitsimme 30 satunnaisotannalla poimittua artikkelia, jotka oli valittu enimmäkseen konferenssijulkaisuista. Yksi Scopus-otoksen artikkeli oli vioittunut, joten tarkasteltiin vain 29 artikkelia.

**Tulos**

Miten ArXivin koneoppimisjulkaisut poimittiin?

**Esimerkki 5.2173**

Käytämme seuraavia halpoja tapoja luoda pseudolähdetekstejä: copy: tässä asetelmassa lähdepuoli on pelkkä kopio kohdepuolen tiedoista. copy-marked: toinen tapa integroida kopioita ilman OOV:ien käsittelyä on täydentää lähdesanastoa kohdesanaston kopiolla. copy-dummies: varsinaisten kopioiden sijasta korvaamme jokaisen sanan "dummy"-tavuilla.

**Tulos**

mitä tietojen simulointitekniikoita otettiin käyttöön?

**Esimerkki 5.2174**

Useita kierroksia: Keskimääräinen lausumien määrä vuoropuhelua kohden on noin 23, mikä takaa kontekstirikkaan kielenkäyttäytymisen.

**Tulos**

Kuinka monta kierrosta keskimäärin tehdään vuoropuhelua kohden?

**Esimerkki 5.2175**

On kaksi näkökohtaa, jotka siirretään tulevaan työhön. Ensinnäkin tässä suunnitelluissa järjestelmissä oletettiin, että syötteet ovat kelvollisia väitelauseita. Jotta tällaisia järjestelmiä voitaisiin käyttää, on kehitettävä mekanismeja, joilla voidaan tunnistaa kelvollisia argumenttirakenteita. Lisäksi jätämme huomiotta luotettavuuteen ja uskottavuuteen liittyvät kysymykset, jotka ovat tärkeitä tutkimuskysymyksiä ja joita käsitellään muissa töissä.

**Tulos**

Mitä haasteita korostetaan?

**Esimerkki 5.2176**

MT-tehtävässä käytämme WMT 2014 En $\leftrightarrow $ Fr -rinnakkaiskorpusta. Tietokokonaisuus sisältää 36 miljoonaa En $\rightarrow $ Fr-lauseparia. Vaihdoimme lähde- ja kohdelauseet saadaksemme rinnakkaisdataa Fr $\rightarrow $ En -käännöstehtävää varten. Käytämme näitä kahta tietokokonaisuutta (72 miljoonaa lauseparia) kouluttaaksemme yhden monikielisen NMT-mallin oppimaan molemmat käännössuunnat samanaikaisesti.

**Tulos**

Mitä tietoja käytettiin monikielisen koodaimen kouluttamiseen?

**Esimerkki 5.2177**

BIBREF6 esittelee KG-A2C:n, joka käyttää tietämysgraafiin perustuvaa tilojen esittämistä apuna toimintojen osion määrittämisessä kombinatorisesti mitoitetussa toiminta-avaruudessa - erityisesti he käyttävät tietämysgraafia rajoittaakseen, millaisia entiteettejä voidaan täyttää tyhjät kohdat mallitoiminta-avaruudessa. He testaavat lähestymistapaansa Zork1:llä ja osoittavat, että tietämysgraafin ja mallitoimintojen valinnan yhdistelmä parantaa nykyisiä menetelmiä. He huomauttavat, että heidän lähestymistapansa saavuttaa 40 pisteen tuloksen, joka vastaa Zork1:n pullonkaulaa, jossa "grue" syö pelaajan (mikä johtaa negatiiviseen palkintoon), jos pelaaja ei ole ensin sytyttänyt lamppua.

**Tulos**

Mitkä ovat perustasot?

**Esimerkki 5.2178**

Tässä työssä analysoidaan, miten esivalmennetut kielimallit ymmärtävät kieltäviin lausumiin tallennettua fakta- ja maalaisjärjen tietoa. Tätä varten esitellään negaatioita sisältävä LAMA-tietokanta. Rakennamme sen yksinkertaisesti lisäämällä negaatioelementtejä (esim. "ei") LAMA-lausekkeisiin (esim. "Suhteellisuusteoriaa ei kehittänyt [MASK].").

**Tulos**

Miten he laajensivat LAMA-arviointikehystä keskittymällä negaatioon?

**Esimerkki 5.2179**

Siksi analyysissämme keskitymme ketjun piirteisiin, jotka liittyvät läpivalaisun ja selittämisen ilmiöihin.

**Tulos**

Mitä coreference-ilmiöitä analysoidaan?

**Esimerkki 5.2180**

Kielimalli (LM): Koulutamme kaksikerroksisen rekursiivisen neuraalisen kielimallin, jonka GRU-solujen piilokoko on 512. Sequence-to-Sequence Attention Model (S2S): Koulutetaan kaksikerroksinen neurologinen sekvenssistä sekvenssiin -malli, joka on varustettu bilineaarisella huomiofunktiolla ja jossa on piilotetun koon 512 GRU-soluja. Lineaarinen dynaaminen järjestelmä (LDS): Koulutamme myös jaksossa SECREF1 käsitellyn lineaarisen dynaamisen järjestelmän, joka on yksi vertailulinjoistamme oikeudenmukaista vertailua varten. Puolivalvottu SLDS (SLDS-X%): Koulutamme puolivalvonnan käytettävyyden arvioimiseksi myös puolivalvottuja SLDS-malleja, joissa on vaihteleva määrä tunnetunnisteita, toisin kuin alkuperäisessä mallissa, joka käyttää 100-prosenttisesti tunnisteilla varustettua dataa.

**Tulos**

Mitä perustasoja käytetään?

**Esimerkki 5.2181**

Tämän ongelman ratkaisemiseksi BIBREF1 esitteli semanttisesti ehdollistetun sukupolvimallin, jossa käytetään hierarkkista epäyhtenäistä itsehuomiota (HDSA) .

**Tulos**

mihin semanttisesti ehdollistettuihin malleihin niitä verrattiin?

**Esimerkki 5.2182**

Kompensoidaksemme aukkokohtaisten parametrien poissulkemisen otamme käyttöön paremman ominaisuuksien esityksen käyttäjän lausumasta ja dialogin tiloista käyttämällä syntaktista tietoa ja konvoluutio-neuraaliverkkoja (CNN, convolutional neural networks).

**Tulos**

Mitä verkkoarkkitehtuuria he käyttävät SIM-kortissa?

**Esimerkki 5.2183**

Vertaamme HR-VAE-malliamme kolmeen vahvaan perusmalliin, jotka käyttävät VAE:tä tekstin mallintamiseen: VAE-LSTM-pohja: Variationaalinen automaattinen koodausmalli, joka käyttää LSTM:ää sekä koodauksessa että dekoodauksessa. KL-hehkutusta käytetään latenttien muuttujien romahdusongelman ratkaisemiseen BIBREF0;VAE-CNN: Variationaalinen autokooderimalli, jossa on LSTM-kooderi ja laajennettu CNN-dekooderi BIBREF7;vMF-VAE: Variationaalinen autokooderimalli, jossa käytetään LSTM:ää sekä kooderissa että dekooderissa ja jossa ennakkojakauma on von Mises-Fisher (vMF) -jakauma Gaussin jakauman sijasta BIBREF5.

**Tulos**

Verrataanko niitä uusimpaan tekstin tuottamiseen?

**Esimerkki 5.2184**

Täydellinen aineisto koostuu sekä edellä mainitusta otoksesta että alustavan työn 819 parista, eli yhteensä 2677 parista. Jaoimme vSTS-tietokannan harjoittelu-, validointi- ja testiosioihin satunnaisotannalla ja säilytimme kokonaispistejakaumat. Käytämme yhteensä 1338 paria harjoitteluun, 669 validointiin ja loput 670 paria lopulliseen testaukseen.

**Tulos**

Kuinka suuri vSTS-koulutusdata on?

**Esimerkki 5.2185**

Testaamme myös olemassa olevia debiasing-menetelmiä, CDA:ta ja REG:ää, mutta koska BIBREF5 raportoi, että tulokset vaihtelevat huomattavasti REG:n eri regularisointikertoimilla, suoritamme hyperparametrin virityksen ja ilmoitamme parhaat tulokset taulukossa TABREF12 .

**Tulos**

mitä nykyisiä strategioita verrataan?

**Esimerkki 5.2186**

Health Level Seven Fast Healthcare Interoperability Resources (HL7 FHIR)FHIR BIBREF5 on uusi avoin standardi terveydenhuollon tiedoille, jonka on kehittänyt sama yritys, joka kehitti HL7v2:n. Resource Description Framework (RDF)RDF on semanttisen webin selkärankaBIBREF8.

**Tulos**

Mitä FHIR ja RDF tarkoittavat?

**Esimerkki 5.2187**

Jotta suurta datakokonaisuutta $\mathcal {A}$ ja $\mathcal {M}$ voidaan hyödyntää riittävästi, malli on esiharjoitteluvaiheessa esikoulutettu CTC-pohjaisella ASR-tehtävällä ja MT-tehtävällä.

**Tulos**

Mihin huomiomoduuli on esivalmennettu?

**Esimerkki 5.2188**

Vaikka suunnitelman luomisvaihe on taatusti uskollinen syötteelle, käännösprosessi suunnitelmista tekstiksi perustuu neuraaliseen seq2seq-malliin, ja se voi kärsiä tällaisten mallien tunnetuista ongelmista: se voi harhailla tosiasioita, joita ei ole syötteessä, toistaa tosiasioita tai jättää pois tosiasioita. Vaikka suunnitelmien ja tekstin selkeä yhdistäminen auttaa vähentämään näitä ongelmia huomattavasti, BIBREF0-järjestelmässä on edelleen 2 % tällaisia virheitä. Viimeaikaiset työt neuraalisen tekstin tuottamisen ja tiivistämisen alalla pyrkivät puuttumaan näihin ongelmiin pyrkimällä kartoittamaan tekstin tuotokset takaisin strukturoituihin predikaatteihin ja vertaamalla näitä predikaatteja syöttötietoihin.

**Tulos**

Mikä on suunnitelman tehokkuus?

**Esimerkki 5.2189**

Huomaa, että noudatamme mallintamisessa säästäväistä suunnitteluperiaatetta: sekä Centroid että Naïve Bayes ovat parametrittomia malleja, $k$NN riippuu vain $k$:n valinnasta, ja KDE käyttää yhtä kaistanleveysparametria $h$. Centroid-malli tiivistää jokaisen siemensanojen joukon sen odotetun vektorin mukaan upotusavaruudessa ja luokittelee käsitteet luokkaan, jonka odotettu upotus on lähimpänä euklidista etäisyyttä softmax-säännön mukaisesti; Naïve Bayes -malli ottaa huomioon sekä keskiarvon että varianssin olettaen, että upotusulottuvuudet ovat riippumattomia, sovittamalla normaalijakauma, jolla on keskivektori ja diagonaalinen kovarianssimatriisi, kunkin luokan siemensanojen joukkoon;

**Tulos**

Miten parametriton malli toimii?

**Esimerkki 5.2190**

Lataamme 210 käyttäjän kaikki Twitter-viestit, jotka ovat sotaveteraaneja ja kliinisesti diagnosoituja PTSD-oireyhtymästä kärsiviä henkilöitä, ja saimme yhteensä 12 385 twiittiä.

**Tulos**

Kuinka monta Twitterin käyttäjää tutkitaan kliinisesti validoidulla kyselyllä?

**Esimerkki 5.2191**

Perusmalli anglikaanien automaattista uuttamista varten luotiin käyttämällä äsken esittelemäämme annotoitua korpusta harjoitusmateriaalina. Kuten luvussa 3 mainittiin, anglikaanien havaitsemistehtävää voidaan lähestyä sekvenssien merkintäongelmana, jossa vain tietyt tekstikatkelmat merkitään anglikaanisiksi (samaan tapaan kuin NER-tehtävässä). Malliksi valittiin ehdollinen satunnaiskenttämalli (CRF), joka oli myös suosituin malli molemmissa Shared Tasks on Language Identification for Code-Switched Data -tehtävissä BIBREF23 ja BIBREF24.

**Tulos**

Perustellaanko asiakirjassa CRF:n käyttöä perusmallina?

**Esimerkki 5.2192**

Lähtötilanteen tulokset ::: Aiemmassa työssämme BIBREF7, jossa rakensimme puhejärjestelmiä 700 kielestä löydettyyn aineistoon, käsittelimme yhdenmukaistamisongelmia (kun ääntä ei ole segmentoitu vuoron/lauseen kokoisiin kappaleisiin) ja oikeellisuusongelmia (kun ääni ei vastaa transkriptiota). Käytimme tässä samoja tekniikoita kuin edellä on kuvattu.Parhaanlaatuista puhesynteesiä varten tarvitsemme muutaman tunnin foneettisesti tasapainotetun, yhden puhujan luetun puheen. Ensimmäisenä askeleenamme käytimme dialogien kunkin vuoron alku- ja loppupisteitä ja valitsimme useimmin esiintyvän puhujan, nmlch, alku- ja loppupisteet. Näin saimme noin 18250 segmenttiä. Lisäksi poistimme automaattisesti liiallisen hiljaisuuden näiden vuorojen alusta, keskeltä ja lopusta (F0:n esiintymisen perusteella). Näin saimme 13 tuntia ja 48 minuuttia puhetta.Tasasimme tämän datan foneettisesti ja rakensimme puheklusteringen tilastollisen puhesyntetisaattorin BIBREF9 kaikesta tästä datasta. Puheentunnistukseen (ASR) käytimme Kaldi BIBREF11 -ohjelmistoa. Rakensimme neuraalisia päästä päähän -koneellisia käännösjärjestelmiä Mapudungunin ja espanjan välille molempiin suuntiin käyttäen uusinta Transformer-arkkitehtuuria BIBREF14 ja BIBREF15-työkalupakettia.

**Tulos**

Mitä malleja käytetään kolmen NLP-tehtävän perustasolla?

**Esimerkki 5.2193**

Akustinen sanan upotus on funktio, joka ottaa syötteenä sanaa vastaavan puhesegmentin, INLINEFORM0 , jossa jokainen INLINEFORM1 on kehystason akustisten piirteiden vektori, ja antaa tuloksena segmenttiä kuvaavan kiinteäulotteisen vektorin, INLINEFORM2 .

**Tulos**

Miten he esittävät mallinsa syöttöominaisuudet upotusten kouluttamiseksi?

**Esimerkki 5.2194**

W2V:n kouluttamiseen tarvittava tietokokonaisuus saatiin käyttämällä tietoja, jotka oli poimittu italialaisen Wikipedian dumppausdatasta (päivätty 2019.04.01), italialaisen Google Newsin pääluokista (MAAILMA, KANSAT, LIIKETOIMINTA, TEKNOLOGIA, VIIHDE, URHEILU, TIETO, TERVEYS) ja joistakin anonymisoiduista keskusteluista, jotka käytiin käyttäjien ja asiakaspalvelukeskustelurobotin (Laila) välillä.

**Tulos**

Mitä tietokokonaisuutta käytetään Word2Vecin harjoitteluun italiankielellä?

**Esimerkki 5.2195**

Kooderi on konvoluutiohermoverkko (Convolutional Neural Network, CNN) ja dekooderi on LSTM-verkko (Long Short-Term Memory, LSTM), kuten kuvassa 2 on esitetty. Kuva kulkee koodaajan läpi ja tuottaa kuvan esityksen, jota dekooderi käyttää kuvan sisällön tuntemiseen ja kuvauksen tuottamiseen sana sanalta.

**Tulos**

Mitä mallia käytetään kuvien koodaamiseen?

**Esimerkki 5.2196**

Sovellamme adaptiivisesti harvoja Transformers-ohjelmia neljään konekäännöstehtävään. IWSLT 2017 saksa $\rightarrow $ englanti BIBREF27: 200K lauseparia.KFTT japani $\rightarrow $ englanti BIBREF28: 300K lauseparia.WMT 2016 romania $\rightarrow $ englanti BIBREF29: 600K lauseparia.WMT 2014 englanti $\rightarrow $ englanti BIBREF30: 4,5M lauseparia.

**Tulos**

Mitä tehtäviä käytetään arvioinnissa?

**Esimerkki 5.2197**

EmotionLines BIBREF6 on dialogitietoaineisto, joka koostuu kahdesta osajoukosta, Friends ja EmotionPush, dialogien lähteen mukaan.

**Tulos**

mitä tietokokonaisuuksia käytettiin?

**Esimerkki 5.2198**

Korjaamattomat merkintöjen ristiriitaisuudet Jotkin muistamisen puutteet johtuvat korjaamattomista merkintöjen ristiriitaisuuksista. Suurelta osin meitä haittaavat erot annotoitavien attribuuttien valinnassa. Kun yksi resurssi merkitsee sukupuolen ja toinen tavun, emme voi päätellä sanan sukupuolta pelkästään sen pintamuodosta. Itse resursseja pitäisi päivittää, jotta ne koodaisivat asiaankuuluvan morfosyntaktisen tiedon. Joissakin kielissä päällekkäisiä muotoja oli hyvin vähän, eikä niiden välillä ollut yhtään tunnistetapausta tai lähes yhtäläisyyttä: Arabia, hindi, liettuan kieli, persia ja venäjä. Täydellinen luettelo havaituista, korjaamattomista ristiriidoista esitetään koodipohjan ohella. Pystyimme löytämään tiettyjä tapauksia, joissa UniMorph-merkintöjä oli sovellettu väärin, sekä erityisiä tapauksia, joissa molemmissa resursseissa oli kieltenvälisiä epäjohdonmukaisuuksia.

**Tulos**

Mitkä ovat kartoituksessa esiintyvien muistivirheiden pääasialliset lähteet?

**Esimerkki 5.2199**

Kysymyksiä: BIBREF-korpuksen31 7 787 luonnontieteiden tenttikysymystä Aristo Reasoning Challenge (ARC) -korpuksesta, joka sisältää standardoituja 3.-9. luokan luonnontieteiden tenttikysymyksiä 12 yhdysvaltalaisesta osavaltiosta viime vuosikymmeneltä.

**Tulos**

Miten aineisto kerättiin?