

VILNIAUS UNIVERSITETAS
MATEMATIKOS IR INFORMATIKOS FAKULTETAS
PROGRAMŲ SISTEMŲ KATEDRA

Gestų kalbos vienetų atpažinimas iš video srauto

Recognition of Sign language units from a video stream

Bakalauro darbas

Atliko:	Pranciškus Ambrazas	(parašas)
Darbo vadovas:	j. asist. Linas Petkevičius	(parašas)
Darbo recenzentas:	dr. Vytautas Valaitis	(parašas)

Vilnius – 2018

Santrauka

Glaustai aprašomas darbo turinys: pristatoma nagrinėta problema ir padarytos išvados. Santraukos apimtis ne didesnė nei 0,5 puslapio. Santraukų gale nurodomi darbo raktiniai žodžiai.

Raktiniai žodžiai: raktinis žodis 1, raktinis žodis 2, raktinis žodis 3, raktinis žodis 4, raktinis žodis 5

Summary

Santrauka anglų kalba. Santraukos apimtis ne didesnė nei 0,5 puslapio.

Keywords: keyword 1, keyword 2, keyword 3, keyword 4, keyword 5

TURINYS

ĮVADAS	4
Gestų kalba	4
Gestų kalbos specifi­ka	4
Darbo tikslas	5
Darbo uždaviniai	5
Darbo eiga	5
Panaudotos priemonės	6
1. APSIMOKANČIOS SISTEMOS	7
1.1. Prižiūrimas mokymas	7
1.2. Neprižiūrimas mokymas	8
1.3. Praktinis mokymas	9
2. NEURONINIAI TINKLAI	10
2.1. Dirbtiniai neuroniniai tinklai	10
2.2. Daugiasluosknis perceptronas	11
2.3. Konvoliuciniai neuroniniai tinklai	11
2.4. Rekurentiniai neuroniniai tinklai	11
2.5. Apjungiamieji tinklų modeliai	12
3. EKSPERIMENTINĖ DALIS	13
4. MEDŽIAGOS DARBO TEMA DĖSTYMO SKYRIAI	14
REZULTATAI IR IŠVADOS	15
LITERATŪRA	16
SANTRUMPOS	17
PRIEDAI	17
1 priedas. Rankų pirštų numeracija	18

Įvadas

Pasaulyje yra virš 7 milijardų žmonių, kurie kasdien tarpusavyje komunikuoja. Netgi 5% visos žmonijos populiacijos sudaro žmonės, turintys klausos problemų. Vien 34 milijonai iš jų yra vaikai, iš kurių net 60% praradusių klausą vaikystėje galėjo būti girdintys dabar, jei būtų imtasi atitinkamų prevencinių priemonių. Paskaičiuota, kad iki 2050 metų žmonių, turinčių šias problemas, skaičius išaugs netgi iki 900 milijonų, o vien šiuo metu 1,1 milijardo jaunų žmonių nuo 11 iki 35 metų amžiaus yra ant klausos praradimo ribos dėl per didelio triukšmo [Org18].

Gestų kalba

Gestų kalba – tai geriausias būdas klausos negalią turintiems žmonėms bendrauti tarpusavyje. Ja pasaulyje bendrauja didžioji dalis klausos sutrikimus turinčiųjų, o amerikiečių gestų kalba (*angl. American Sign Language (ASL)*) yra trečia pagal populiarumą Amerikoje po anglų ir ispanų kalbų, kuria kalba virš 500 tūkstančių žmonių. Kiekviena šalis turi savo valstybinę kalbą - lietuvių, anglų, ispanų, rusų. Lygiai taip pat kiekviena šalis turi ir savo gestų kalbą. Tai yra tiek jau minėta amerikiečių gestų kalba (ASL), lietuvių, argentiniečių ir kitos gestų kalbos. Netgi tam tikri šalių regionai turi specifinius tos pačios kalbos dialektus, kaip, tarkime, vien Lietuvoje yra aukštaičių, žemaičių, suvalkiečių ar dzūkų tarmės.

Kiekviena gestų kalba turi savo atskirą gramatiką ir sintaksę. Skirtingos gestų kalbos skiriasi tiek abėcėlėmis, tiek pačiais gestais, dėl to skiriasi netgi ta pati gramatika. Taip yra dėl to, kad nėra bendrinės gestų kalbos - vien Amerikoje yra virš 35 skirtingų gestų kalbų.

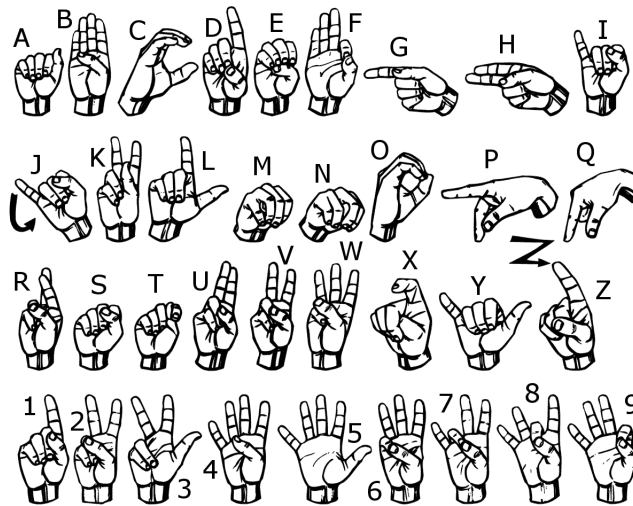
Vienas gestas turi turėti kelias prasmes. Kaip ir lietuvių kalboje žodis „kasa“ turi tris skirtingas reikšmes, taip ir gestų kalboje vienas gestas gali turėti keletą reikšmių. Tačiau iš kitos pusės gestas, parodytas truputėlį kitaip gali turėti visiškai priešingą reikšmę. Tarkime, ASL gestai „geras“ ir „blogas“ skiriasi tik puse į kurią atsuktas deltas, tačiau daugiau neturi jokių skirtumų.

Gestų kalbos specifika

Kiekviena gestų kalba susideda iš **trių** pagrindinių dalių:

1. **Statinė gestų kalba** - dar kitaip vadinama *pirštų kalba* (*angl. fingerspelling*). Tai įvairūs gestai rodomi vienos (ASL, LGK) ar net ir dviejų (britų ar vokiečių gestų kalba) rankų pagalba. Tai nejudantys gestai, rodantys vieną raidę (*žr. 1 pav.*) ar net vieną žodį, kaip, pavyzdžiui, ASL „*I love you*“¹ gestas. Yra galimybė žodžius išreikšti ir abėcėliškai. Lygiai taip pat žmonės kasdieninėje kalboje turi galimybę pasakyti paraidžiui. Tačiau yra įprasta jungti raides į žodžius. O žodžius galiausiai į sakinius. Vienas iš variantų, kuomet naudojama gestų kalba paraidžiui tai vardų pasakyme. Tačiau gestakalbiai prisistatydami parodo gestą, kuris priklauso tik jiems. Tai tarsi parašas tam, kad neberekėtų kreipiantis ar apibūdinant žmogų jo vardo sakyti paraidžiui.

¹liet. Aš tave myliu



1 pav. Amerikiečių gestų kalbos abėcėlė

2. **Dinaminė gestų kalba** - tai žodžių lygio gestų kalba. Nesunku pastebėti, kad 1 paveikslėlyje yra „J“ ar „Z“ raidės, kurios priskiriama dinaminių judesių klasei. Kaip ir yra žodžių, kurie priskiriami statinei gestų kalbai dėl savo kilmės, taip ir yra raidžių, kurios priskiriamos dinaminei gestų kalbai. Dinaminiais judesiais yra išreiškiami įvairūs gestų kalbos žodžiai tokie, kaip, pavyzdžiui, ASL yra „labas“, „gerai“ ar „blogai“.
3. **Kitos ypatybės** - emocijos veide, liežuvis, burna ir kūno laikysena. Tai taip pat labai svarbios gestų kalbos ypatybės. Pavyzdžiui, klausiant gestų kalba klausimo, jei bus pakelti antakiai, tai reikš, kad laukiamas ataskymas „taip“ arba „ne“. Tačiau, jei antakiai bus suraukti, tai reikš, kad klausiama su paaiškinimu „kas“, „kur“, „kaip“, „ką“.

Darbo tikslas

Išanalizuoti gestų kalbos vienetų atpažinimo galimybes ir video srauto.

Darbo uždaviniai

- Gestų kalbos video srautų paieška ir mokomosios medžiagos neuroniniams tinklams surinkimas
- Susipažinimas su rekurentiniais neuroniniais tinklais
- Gestų kalbos vienetų atpažinimas iš video srauto pasinaudojant rekurentiniais neuroniniais tinklais.

Darbo eiga

- Panašių ir jau įgyvendintų projektų paieška
- Esamos sistemos patobulinimai

- Rezultatų palyginimai

Panaudotos priemonės

- Python – programavimo kalba
- TensorFlow – skirta darbui su apsimokančiomis sistemomis²

Įvade nurodomas darbo tikslas ir uždaviniai, kuriais bus įgyvendinamas tikslas, aprašomas temos aktualumas, apibrėžiamas tiriamasis objektas akcentuojant neapibrėžtumą, kuris bus išspręstas darbe, aptiriamos teorinės darbo prielaidos bei metodika, apibūdinami su tema susiję literatūros ar kitokie šaltiniai, temos analizės tvarka, darbo atlikimo aplinkybės, pateikiama žinių apie naudojamus instrumentus (programas ir kt., jei darbe yra eksperimentinė dalis). Darbo įvadas neturi būti dėstyimo santrauka. Įvado apimtis 2–4 puslapiai.

²angl. Machine learning

1. Apsimokančios sistemos

Apsimokančios sistemos (*angl. machine learning*)

1.1. Prižiūrimas mokymas

Prižiūrimas mokymas (*angl. supervised learning*) - tai apsimokančių sistemų apmokymo būdas, kuomet duomenys mokymui yra paruošiami taip, kad kiekvienas duomuo turėtų ir atitinkamą rezultatą. Kitaip tariant, jei yra duomuo a , tai yra ir jį atitinkantis rezultatas, arba dar vadinama etiketė b . Tai būdas, kuris veikia medžio principu.

1 lentelė. Pavyzdinis prižiūrimo mokymo apmokymui paruoštų duomenų rinkinys

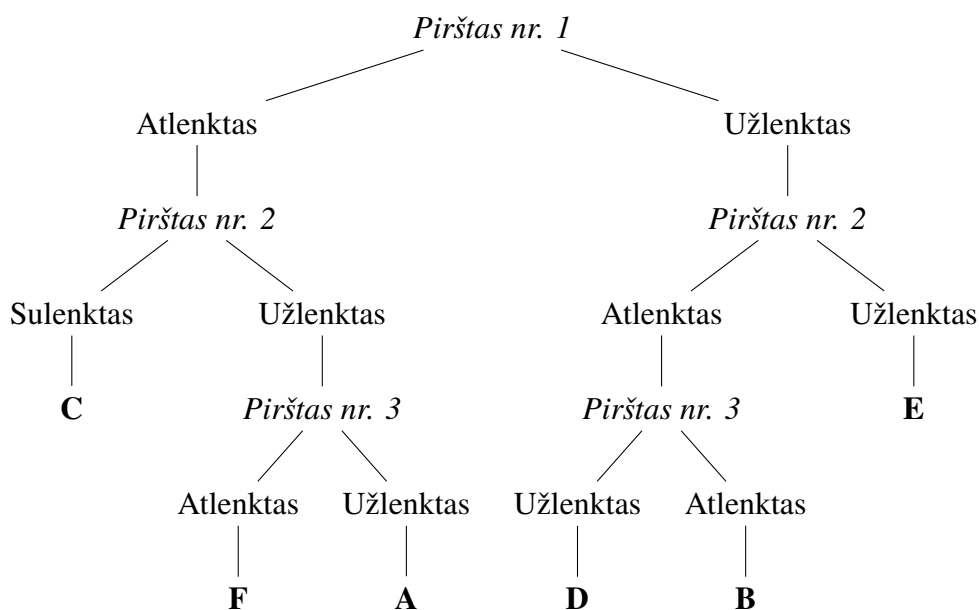
Nr.	Pirštas nr. 1	Pirštas nr. 2	Pirštas nr. 3	Pirštas nr. 4	Pirštas nr. 5	Raidė
1.	Atlenktas	Užlenktas	Užlenktas	Užlenktas	Užlenktas	A
2.	Užlenktas	Atlenktas	Atlenktas	Atlenktas	Atlenktas	B
3.	Atlenktas	Sulenktas	Sulenktas	Sulenktas	Sulenktas	C
4.	Užlenktas	Atlenktas	Užlenktas	Užlenktas	Užlenktas	D
5.	Užlenktas	Užlenktas	Užlenktas	Užlenktas	Užlenktas	E
6.	Atlenktas	Užlenktas	Atlenktas	Atlenktas	Atlenktas	F

1 lentelėje pateikiamas pavyzdys su amerikiečių gestų kalbos abėcėle. Lentelėje pateikiamos pirštų padėties, o pirštai numeruojami pagal 1 priede pateikiamą pirštų numeraciją. Kiekvieno piršto padėtis šiame pavyzdyje gali būti: *atlenktas*, *sulenktas*, *užlenktas*. Ir kiekvienai padėčiai esant pateikiamas rezultatas, arba kitaip - etiketė, kokią raidę abėcėlėje atitinka pavaizduotos pirštų padėties.

2 lentelė. Pavyzdinė praktinė užduotis

Nr.	Pirštas nr. 1	Pirštas nr. 2	Pirštas nr. 3	Pirštas nr. 4	Pirštas nr. 5	Raidė
1.	Atlenktas	Sulenktas	Sulenktas	Sulenktas	Sulenktas	?

2 lentelėje pateikiamas uždavinys, kuriame nurodoma ta pati informacija, kuri buvo pateikta 1 lentelėje. Tačiau rezultatas nėra pateiktas, o jis randamas medžio principu.



2 pav. Galimybių medis

Vien iš šio medžio galimybių medžio galima matyti, kad pilnai užtenka sprendimui nusakyti 3 pirštų, kadangi rezultatų nėra daug. Jei būtų imama visa abėcėlės aibė, tuomet rezultato nustatymui būtų naudojama galimai visų pirštų padėties. Tačiau net ir šį medį optimizavus galima būtų, tarkime, C raidė atsakymą gauti tik iš vieno piršto padėties, kadangi tik ši vienintelė raidė turi sulenktą piršto padėtį. Galiausiai iš šio medžio galima pastebėti, kad 1 lentelėje pateikto pavyzdžio atsakymas yra raidė C.

1.2. Neprižiūrimas mokymas

Neprižiūrimas mokymas (*angl. unsupervised learning*) - mokymas, kuomet duomenims nėra priskiriamos teisingos etiketės ar teisingi rezultatai. Pavyzdžiui, tai galėtų atitikti naujos kalbos mokymąsi be mokytojo ir bet kokio žodyno. Kuomet pastoviai matomas vis tas pats tekstas, žodžiai tampa atpažįstami, tačiau išversti jų neišsina. Tačiau tai nesukelia jokių nepatogumų, jei į tekstą reikia įrašyti tinkamą žodį, kuomet dėl daugybės duomenų yra aišku koks žodis su kokia galūne turėtų būti įrašytas.

3 lentelė. Pavyzdinis neprižiūrimo mokymo apmokymui paruoštų duomenų rinkinys

Nr.	Pirštas nr. 1	Pirštas nr. 2	Pirštas nr. 3	Pirštas nr. 4	Pirštas nr. 5
1.	Atlenktas	Užlenktas	Užlenktas	Užlenktas	Užlenktas
2.	Užlenktas	Atlenktas	Atlenktas	Atlenktas	Atlenktas
3.	Atlenktas	Sulenktas	Sulenktas	Sulenktas	Sulenktas
4.	Užlenktas	Atlenktas	Užlenktas	Užlenktas	Užlenktas
5.	Užlenktas	Užlenktas	Užlenktas	Užlenktas	Užlenktas
6.	Atlenktas	Užlenktas	Atlenktas	Atlenktas	Atlenktas

3 lentelėje pateikiamas pavyzdinis neprižiūrimam mokymui apmokyti paruoštų duomenų rinkinys. Duomenys tokie patys, kaip ir 1 lentelėje, tačiau nėra teisingo atsakymo sulpelio „**Raidė**“.

Apmokius tokią sistemą būtent tokiais duomenimis vienas iš tikėtinų scenarijų, kur galima būtų panaudoti tokią sistemą, tai nuspėti, kokios raidės yra labiausiai tikėtinos ar tiesiog numatyti, kokia labiausiai tikėtina raidžių seka bus rodoma.

1.3. Praktinis mokymas

Praktinis mokymas (*angl. reinforcement learning*) - labiausiai dirbtinį intelektą atitinkančių apsimokančių sistemų apmokymo modelis. Šis mokymas pagrįstas praktiniais bandymais. Kiekvienas teisingai gautas rezultatas yra būdas, kuriuo reikėtų sekti, ir kiekvienas blogai gautas rezultatas, yra būdas, kurio vertėtų atsisakyti. Dažniausiai šis apmokymo būdas naudojamas sistemą apmokant žaisti žaidimus. Vienas iš labiausiai žinomų būtent šiuo apmokymo būdu apmokytų modelių yra *AlphaZero*, kuris sugeba laimėti prieš pasaulio šachmatų čempionus. Tai puikus pavyzdys to, kaip kompiuteris iš laimėjimų, už kuriuos gauna taškus, ir pralaimėjimų, už kuriuos jam taškai atimami, sugeba rasti laimėjimo strategijas kiekviename žingsnyje ir taip, nuolatos tobulėdamas, laimėti dvikovas ar apskritai spręsti uždavinius, kuriuose reikalingas pastabumas ir strategijų kūrimas.

2. Neuroniniai tinklai

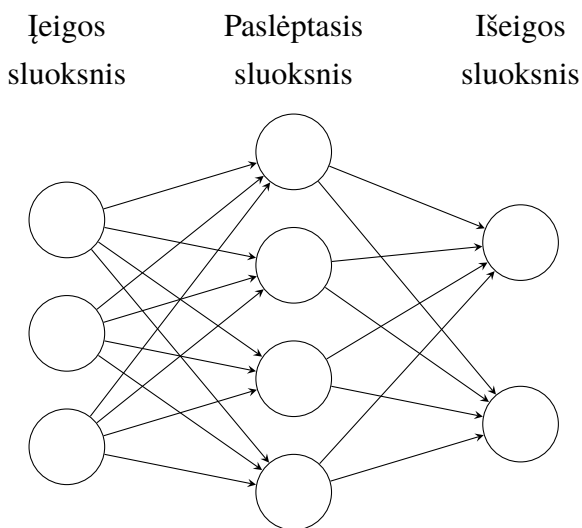
Žmogaus smegenys yra labai sudėtingas, nelijinis ir paralelinis kompiuteris [Hay09]. Kiekvieno kūnas yra sudarytos iš milijardų nervinių ląstelių vadinamų neuronais. Jie sukuria ir/arba perduoda elektrocheminius impulsus. Neuronai tarpusavyje yra sujungti dendritais, ant kurių yra sinapsės.

Kiekvienas sužadintas neuronas dėl pasikeitusios temperatūros, spaudimo, skausmo ar kitų veiksnių, perduoda informaciją į smegenis dėl sprendimo, ką daryti, priėmimo. Tai, kaip ir buvo paminėta, yra siųsti signalą iš vieno neurono į kitą, kol galiausiai signalas pasiekia smegenis. Svarbu ir tai, kad kiekvienas neuronas yra nepriklausomas nuo kito. Tai tik grandis, kuri yra atsakinga už signalo priėmimą ir perdavimą. Smegenims gavus signalą, jį apdorojus ir priėmus sprendimą, signalas tuo pačiu keliu siunčiamas atgal, kol pirmąjį sužadinimą gavęs neuronas sulaukia atsakymo.

2.1. Dirbtiniai neuroniniai tinklai

Dirbtiniai neuroniniai tinklai (*angl. artificial neural networks*) – struktūra, sukurta remiantis žmogaus nervinės sistemos darbu. Dirbtiniai neuroniniai tinklai gali būti išmokinti atlikti klasifikavimo, spėjimo, sprendimų priėmimo ir kitas užduotis.

Dirbtiniai neuroniniai tinklai susideda iš dirbtinių neuronų. Tai trijų sluoksnių tinklas, kuris susideda iš įėjos (*angl. input*), paslėptąjo (*angl. hidden*), kuris gali būti sudarytas iš vieno ar daugiau sluoksnių, ir išeigos (*angl. output*) sluoksnių.



3 pav. Dirbtinio neuroninio tinklo pavyzdys

Kiekvienas neuronas dirbtiniuose neuroniniuose tinkluose gali gauti vieną ar kelias įejas. Visų šių įeigų svorių suma yra apdorojama. Ji apdorojama aktyvacijos funkcija. Yra keletas skirtingų aktyvacijos funkcijų, kaip, pavyzdžiui, tiesinė, *tanh*, *ReLU* ir nelinijinės regresijos funkcijos. Toliau, 1 formulėje pateikiamas *ReLU* aktyvacijos funkcijos principas.

$$f(x) = \begin{cases} 0 & \text{jei } x < 0 \\ x & \text{jei } x \geq 0 \end{cases} \quad (1)$$

2.2. Daugiasluosknis perceptronas

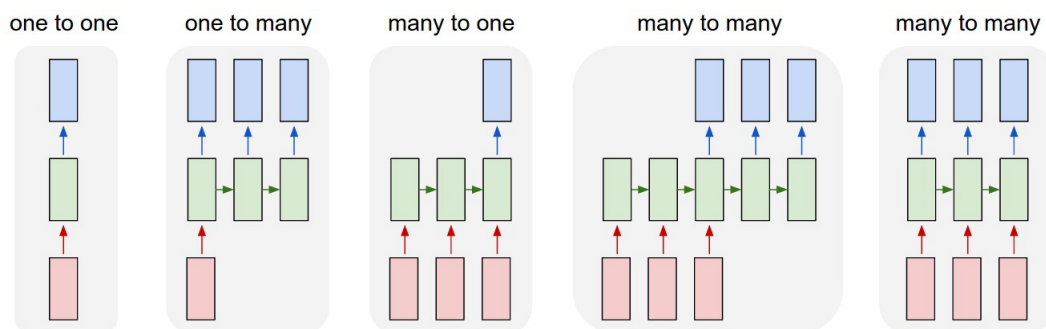
Daugiasluosknis perceptronas (*angl. multilayer perceptron*) - dirbtinių neuroninių tinklų struktūra besivadovaujantis tinklas. Dažniausiai turi tris ar daugiau sluoksnių, kurie dažniausiai kaip aktyvacijos funkciją naudoja nelineines aktyvacijos funkcijas. Dažniausiai tai būna *tanh* ar loginės regresijos funkcijos. Kiekvienas sluoksnio elementas yra sujungtas su kito sluoksnio elementu, todėl tai sudaro pilnai apjungtą (*angl. fully connected*) tinklą. Yra pavyzdžių, kur daugiasluoksniai perceptronai naudojami atpažinti žodinę kalbą ar versti tekstus.

2.3. Konvoliuciniai neuroniniai tinklai

Konvoliuciniai neuroniniai tinklai (*angl. convolitional neural networks*) -

2.4. Rekurentiniai neuroniniai tinklai

Rekurentiniai neuroniniai tinklai (*angl. recurrent neural networks*)



4 pav. tipai

4 paveikslėlyje pavaizduoti keturi skirtingi būdai, kuriais naudojantis neuroniniai tinklai veikia. Rausvos spalvos kvadratai reiškia į eigą (*angl. input*), žalsvas - paslėptuosius sluoksnius (*angl. hidden layers*), o melsvas - iš eigą (*angl. output*). Pateikiami šie būdai:

- **Vienas su vienu** (*angl. one to one*) – būdas, kuriame yra viena į eigą, paslėptasis sluoksnis ir iš eigą. Šis būdas dažniausiai taikomas konstruojant konvoliucinius neuroninius tinklus. Kaip pavyzdį galima pateikti paveikslėlio atpažinimą. Tai galėtų būti statinės gestų kalbos atpažinimas;
- **Vienas su daug** (*angl. one to many*) – būdas, kuriame yra viena į eigą, bet kelios iš eigos. Vienas iš panaudojimo būdų galėtų būti sakinio suformavimas iš paveikslėlio. Toks tinklas ne tik atpažįsta pagrindinį objektą kadre, bet ir apibūdina esančią aplinką, daro kitus sprendimus;

- **Daug su vienu** (*angl. many to one*) – būdas, kuriame yra daug įeigų, bet tik viena išeiga. Tokio būdo pavyzdys galėtų būti vieno žodžio, tarkime, „labas“ atpažinimas iš video sraudo.
- **Daug su daug** (*angl. many to many*) – būdas, kuriame yra daug įeigų ir daug išeigų. Šis būdas gali būti skaidomas į dvi dalis:
 - **Priklausomas** - įeigų skaičius sutampa su išeigų skaičiumi. Kiekviena įeiga turi savo išeigos atitikmenį. Tai būtų dalinai galima gretitinti su *vienas su vienu* būdu. Pavyzdys šios atšakos galėtų būti video srauto klasifikacija pagal kiekvieną kadrą - nuolatinis atnaujinimas, to kas galėjo būti pasakyta, pavyzdžiui, gestų kalboje.
 - **Nepriklausomas** - įeigos skaičius galimai nesutampa su išeigų skaičiumi. Kiekviena įeiga yra nepriklausoma ir išeigos dėliojamos pagal tam tikrus aspektus. Tokio būdo pavyzdys galėtų būti neuroniniai tinklai, kurie atlieka vertėjo funkcijas, pavyzdžiui, iš anglų į lietuvių kalbas, nes skiriasi tiek gramatika, tiek sakinių stilistika.

$$h_t = f_w(h_{t-1}, x_t) \quad (2)$$

ht - hidden state fw - funkcija su parametrais w (svoriais) ht-1 - old state xt - input vector at some time step

[Lip15] LSTM - long-short term memory <https://arxiv.org/pdf/1506.00019.pdf>
<https://medium.com/@erikhallstrm/using-the-dynamicrnn-api-in-tensorflow-7237aba7f7ea>

2.5. Apjungiamieji tinklų modeliai

3. Eksperimentinė dalis

4. Medžiagos darbo tema dėstymo skyriai

Medžiagos darbo tema dėstymo skyriuose išsamiai pateikiamos nagrinėjamos temos detalės: pradiniai duomenys, jų analizės ir apdorojimo metodai, sprendimų įgyvendinimas, gautų rezultatų apibendrinimas.

Medžiaga turi būti dėstoma aiškiai, pateikiant argumentus. Tekste dėstomas trečiuoju asmeniu, t.y. rašoma ne „aš manau“, bet „autorius mano“, „atoriaus nuomone“. Reikėtų vengti informacijos nesuteikiančių frazių, pvz., „...kaip jau buvo minėta...“, „...kaip visiems žinoma...“ ir pan., vengti grožinės literatūros ar publicistinio stiliaus, gausių metaforų ar panašių meninės išraiškos priemonių.

Skyriai gali turėti poskyrius ir smulkesnes sudėtines dalis, kaip punktus ir papunkčius.

Rezultatai ir išvados

Rezultatų ir išvadų dalyje išdėstomi pagrindiniai darbo rezultatai (kažkas išanalizuota, kažkas sukurta, kažkas įdiegta), toliau pateikiamos išvados (daromi nagrinėtų problemų sprendimo metodų palyginimai, siūlomos rekomendacijos, akcentuojamos naujovės). Rezultatai ir išvados pateikiami sunumeruotų (gali būti hierarchiniai) sąrašų pavidalu. Darbo rezultatai turi atitikti darbo tikslą.

Literatūra

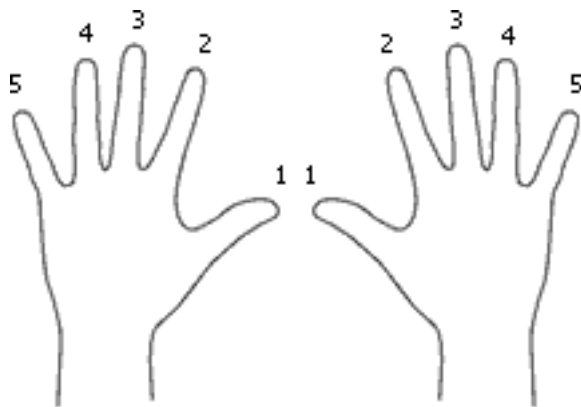
- [Hay09] Simon Haykin. *Neural networks and learning machines*. Pearson Education inc., Upper Saddle River, New Jersey 07458, 2009. 1 psl.
- [Lip15] Zachary Chase Lipton. A critical review of recurrent neural networks for sequence learning. *Corr*, abs/1506.00019, 2015. arXiv: 1506.00019. URL: <http://arxiv.org/abs/1506.00019>.
- [Org18] World Health Organization. Deafness and hearing loss. <http://www.who.int/en/news-room/fact-sheets/detail/deafness-and-hearing-loss>. 2018.

Santrumpos

Sąvokų apibrėžimai ir santrumpų sąrašas sudaromas tada, kai darbo tekste vartojami specialūs paaiškinimo reikalaujantys terminai ir rečiau sutinkamos santrumpos.

Priedas 1

Rankų pirštų numeracija



5 pav. Kairės ir dešinės rankų pirštų numeracija