A 2018 őszi egyetemi hallgatóknak és érdeklődő Morgan Stanley kollégáknak szóló NLU kurzus egy innovatív szövegértési modell közös megismerésére, tesztelésére, továbbfejlesztésére irányul. Az elmúlt 2 év során implementált modellből kiindulva remélhetőleg izgalmas közös munkának nézünk elébe.

**A concept háló modell háttere egyéb irányzatokhoz képest**

Mostanában szokatlannak ható megközelítésben modellezzük és implementáljuk a racionális emberi gondolkodást. Szokatlan annyiban, hogy az elmúlt 20 évben inkább az egyes részfeladatokat igen magas szinten megoldó specializált AI megoldásokról lehet hallani. Ezzel szemben igen kevés erőfeszítés történik az „AI szent grál”-jának is nevezhető, a teljes emberi intelligenciát megcélzó kutatásra (lásd még wikipédia: AI-complete). Ennek hátteréről alapos összefoglaló a wikipédián az *AI winter* cikkben.

A concept háló modell közvetlen célja egy implementált NLU (Natural Language Understanding) rendszer, amely az 5 éves gyermek szintjének megfelelő természetes nyelvi outputot képes produkálni angolul.

A szoftver inputja: természetes nyelvű angol szöveg, és modell paraméterek, szabályok, írásban megadva. Outputja: természetes nyelvű angol szöveg, amely az inputban megfogalmazott kérdésekre, felszólításokra adott válasz.

A tesztelés folyamata: Szövegértés teszttel mérjük, hogy a szoftver képes-e az 5 éves gyerek szintjén gondolkodni. Ehhez 5-8 éves szintű szövegeket adunk inputnak, és kérdéseket, amelyekre 5 éves szinten várjuk a választ. A rendszer válaszol természetes nyelven, ezt összehasonlítjuk valódi 5 éves gyerekek teljesítményével.

Az 5 éves szintű teljesítmény mellett a rendszer tervezésénél fontos szempont, hogy felnőtt szintig skálázható, továbbfejleszthető legyen, illetve ilyennek tűnjön (mivel ezt értelemszerűen teljes körűen nem tudjuk még tesztelni).

Van még egy követelmény amit szem előtt tartunk, mégpedig, hogy a modell részleteiben és részfunkcióiban is minél több analógiát, részmodellt tartalmazzon az ember ismert kognitív jelenségeire nézve. Ez egy meglehetős újdonság a megközelítésben, ugyanis közvetlenül felhasználjuk a kísérleti kognitív pszichológia eredményeit, ami egy implementált AI megoldás esetén nem gyakori eljárás.

**Hol tartunk most**

Van egy elég részletesen kidolgozott modellünk, amely a modell kidolgozottsági szintjéig implementálva is van, és a fent leírt tesztet részben teljesíti. De csak kisebb részben, tehát igen sok munka van hátra, és most következik a munka izgalmasabb (nemtriviális) része.

**A concept háló modell röviden**

A modell (és az implementáció) 3 modulja és azok kapcsolódása:

1. Mentalese fordító. Input: angol természetes nyelvű szöveg. Output ugyanaz a szöveg „mentalese” nyelven. (A mentalese a gondolkodás belső nyelve, ahogy Pinker leírja a *The language instinct* c könyvében, ez legálisan letölthető. Lásd továbbá *Language of thought hypothesis* a wikipédián.) Ez nem egy gondolkodó modul, inkább a konkrét nyelv (angol) szintaktikáját dolgozza fel. A szóalakok a mentalese szövegben nagyjából ugyanúgy szerepelnek, mint az angolban.
2. Központi modul. Input: az első modul outputja, mentalese-ben írt szöveg és kérdések. Output: mentalese nyelvű válaszok a kérdésekre. Ez a gondolkodó modul.
3. Visszafordító mentalese-ről angolra. Input a 2. modul outputja, output annak angolra fordított változata. Ez az első modul inverze.

A továbbiakban csak a központi modullal foglalkozom.

**Adatszerkezet**

Egy igen egyszerű reprezentációt alkalmazunk a gondolkodás és a nyelv modellezésére. Reprezentációnkban csupán két elemi adattípus van, az egyik a *szóalak* (az input szöveg szóalakjai a mentalese fordításban is változatlanul megjelennek), a másik a *concept,* amely az összes adattárolási és feldolgozási feladatot ellátja. (Innen kapta a modell a *concept háló* nevet.)

**Szóalak .** A szóalak objektum adattartalma igen egyszerű:

1. A szó, mint karakterlánc. A mentalese fordítóban az input szavak minimális előfeldolgozások esnek át. Ezért tárolnunk tipikusan csak a szótöveket kell. Azonban az angol toldalékok egy része külön szóként jelenik meg.
2. A szó jelentéseire mutató mutatók. Ha egy szónak több jelentése van, akkor több mutató tartozik hozzá.

**Concept, gondolat** (= fogalom, állítás, szójelentés, tudás, koncepció, mondat, bekezdés, történet – ez mind, egyetlen reprezentációval lefedve). A modell alapvető építőköve. Neve a modellben: *concept.* Azért ez a neve, mert a kognitív pszichológiai irodalom concept szavának megfelel. A concept képességei:

1. Egy szó egy jelentése egy concept.
2. Több concept egymáshoz kapcsolódhat, és ezzel új concept jön létre. A modellben minden érdekes információ ezekben a kapcsolatokban (vagyis a különböző szinteken összetett conceptekben) van. Ezzel egyre összetettebb, hierarchikus concept mintázatokat lehet építeni. Ezeken a mintázatokon hajtja végre a modell valamennyi működését.
3. A conceptet mentalese nyelven lehet megadni. Ha ember írja le a conceptet, akkor ez egy mentalese nyelvű string, például C(a,b). A rendszer ezt objektum instance-ként tárolja.
4. A conceptek összekapcsolódásának vannak fajtái. Kb 20 féle kapcsolat van. Például: azonosság, jele D. Ez a mentalese: D(x,y) azt fejezi ki, hogy az x és az y concept azonos (és közben maga D(x,y) is egy concept). Egy másik fontos fajta az osztályba tartozás (C). C(elefánt1, állat1) azt fejezi ki, hogy az elefánt1 concept (ami lehet az elefánt szó egyik jelentése) az állat1 osztályba tartozik. Végül még egy példa: az „A” típusú conceptet a cselekvő és a cselekvés összekapcsolására használjuk: A(he1,run1) a mentalese megfelelője a „*he runs”* mondatnak.
5. A kapcsolat fajtán túl néhány további attribútuma is van a conceptnek. Pl: mennyire igaz, mennyire releváns, mennyire ismert, mennyire mond ellent más concepteknek, mennyire egyedi és mennyire általános. Ezek nem bináris, hanem többfokozatú attribútumok.
6. A concept akármilyen nagyra nőhet. Egy egész könyv is lehet egyetlen concept.
7. A kisebb conceptek, a mondat szintjéig, tipikusan úgy állnak elő, hogy a nyelv szintaktikája eléggé behatárolja (de nem teljesen egyértelműen) hogy mi, milyen kapcsolatban van egymással a mondatban. (pl *An elephant is an animal.*)

**Rövid távú memória.** (working memory.) Jele WM. Az első helyszíne a feldolgozásnak. Ide tároljuk az inputban frissen kapott concepteket (és majd más eredetű concepteket is). Technikailag ez egy concept lista, amiből időnként törlünk. A további feldolgozás a teljes WM tartalmat figyelembe veheti, vagy annak egy részét.

**Hosszú távú memória.** Jele KB. Ez egy másik concept lista, amelyből nem törlünk – ez a hosszú távú emlékezet, a tanulás helye. Ide akkor írunk, mielőtt valamit a WM-ből törlünk. Azt írjuk ide, amit hosszú távra meg akarunk tanulni. A feldolgozásban nem vesz részt a teljes KB tartalom, csak a jelenlegi WM tartalom által aktivált része.

**Szóalakok listája.** A szóalakokat tároljuk egy listában. Ha az inputban megkapjuk a következő szót, kikeressük ebben a listában, és hozzá kapcsolva megtaláljuk az összes jelentését, amik már conceptek. Ezzel indul a feldolgozás.

**A feldolgozás fő mechanizmusai**

A gondolkodás folyamatábrája a következő. Ezeket a lépéseket fogjuk a kurzus során megismerni.

Step 1. Mentalese expression

For all concepts: decrease all pre-activation in WM and KB.

If new paragraph started: erase all activation and pre-activation in KB.

If new input is ambigous (for example a word with multiple meanings, or mapping required): initiate branching of WM to evaluate all options

Step 2.

Store the new input in WM on all living branches (we call this current concept).

Search for the current concept in WM and KB. Update probability and other counters. Update consistency counter of current concept.

For words not found, store them in KB and assign a meaning concept.

Generic concepts (reasoning blueprints) found: mark them activated

Specific concepts found: mark them activated

Search for most relevant concepts among children of activated concepts, mark them activated

All children of concepts activated in this round: increase pre-activation level. Those exceeding threshold: mark them activated.

Step 3. Consider next activated generic concept (reasoning blueprint).

If the current concept now fully matches the reasoning pattern, generate implication concept. This becomes new input, go back to step 2.

Consider next activated concept.

If that one matches the reasoning blueprint, go to completness check.

No more reasoning available.

Update the consistency of all branches in WM. For worst branches, discontinue the branch.

Current concept is not a question.

Current concept is a question.

Step 4. Delete a portion of WM if necessary, store a selection of deleted concepts in KB.

Step 5. Search again for the concept pattern defined by the question. If found, deliver the answer.

Back to step 1.

**Aktiválás.** A rövid távú memóriába bekerülő új conceptet megkeressük a KB-ban, és a hozzá kapcsolódó, valamennyire releváns concepteken az aktiválás szintjét emeljük (majd a valamennyire aktivált többi KB-beli concept aktiválás szintjét csökkentjük). Bizonyos küszöb fölé aktivált KB tartalom vesz részt a jelenlegi feldolgozásban. Ha az input kérdés, akkor a WM-ben és a KB-ban szélesebb körű aktiválást hajtunk végre.

**Szó jelentés meghatározása (word sense disambiguation, WSD).** Ez hagyományosan az NLP irodalom kedves területe, főleg mert sok kész tanító és tesztanyag van hozzá. A modellünkben az 1. modul nem végzi el a WSD-t, hanem a szóalakot a mentalese szövegben átadja a 2. modulnak. A 2. modul tud a szó ismert jelentéseiről (valamint időnként észreveszi, hogy egy eddig ismeretlen új jelentésről van szó), és meg kell határoznia, hogy ezek közül melyik használandó most. Itt történik a WSD az alább részletezett bizonytalanság feldolgozás és hipotézis értékelés segítségével.

**Bizonytalanság, többértelműség (ambiguity).** A természetes nyelvben, és a gondolkodásban, nemcsak a szó jelentése bizonytalan és többértelmű, hanem minden más is. Nem egyértelmű a szintaktika, az igazság, és semmi más sem. A bizonytalanság egy részét a conceptek attribútumaival próbáljuk megragadni (például az igazság nem bináris hanem többértékű), ez sokszor inkább fokozatosságot jelent. Másfajta bizonytalanság a szó többértelműség vagy a szintaxis többértelműsége, mert ezeknek legtöbbször egyetlen jó megoldása van (néha nem, ez a humor, amit ezáltal meg tud ragadni a rendszer, bár ez csak melléktermék). Ez a többértelműség az alábbi hipotézis értékeléssel kezelhető.

**Hipotézis értékelés (branching).** A többértelmű szavak példáján mutatom be. A szóalakhoz kapcsolódóan listázzuk az összes értelmét, és a szöveg eddigi feldolgozását annyi hipotézisre ágaztatjuk el, ahány értelme van. Minden ágon folytatjuk a feldolgozást, különösen az alább ismertetett reasoning-et, melynek során a teljes korábbi input, a KB, és a jelenlegi hipotézis alapján következtetéseket vonunk le. Minden következtetést (ezek conceptek) összehasonlítunk a teljes WM-mel és a KB releváns részével arra nézve, hogy mennyire mond ellent nekik (lásd lentebb igazság), és ennek alapján az egyes ágak között különbségeket tudunk megállapítani. Hamarosan a rosszabb ágakat elvetjük, és egy vagy pár jobb ágon folytatjuk csak a feldolgozást. Ez a folyamat produkálja a kognitív pszichológusok által megfigyelt text comprehension (szövegértés) jelenségeket.

**Következtetés (reasoning).** Az a feldolgozás, amelynek során néhány input concept és egy általános conceptként rögzített szabály (reasoning blueprint) alapján új conceptet hozunk létre, és WM-ben tároljuk. Nézzünk egy példát a reasoning működésére: Legyen KB-beli concept A(animal1,live1) (=az állatok élnek). Jöjjön az inputban hogy C(elephant1, animal1) (=az elefánt állat). Van egy olyan általános conceptünk, ami szerint C(%1,%2) és A(%2,%3) fennállásából következik A(%1,%3). A két input az általános szabály mintát kielégíti. Ezért a reasoning eredménye az A(elephant1, live1) concept. A reasoning egy igen sok részletkérdést rejtő bonyolult témakör.

**Válaszadás mentalese nyelven (question answering).** A régi eredetű KB tartalom, a jelenlegi input, és a jelenlegi reasoning alapján, ha a legutolsó input egy kérdés, akkor mentalese-ben válaszolunk. Ennek során megkeressük a választ WM-ben és KB-ben, és rámásoljuk az outputra.

**Háttértudás (world knowledge, common sense).** A hasonló próbálkozások egyik Achilles-sarka. Olyasmiket kell tudjon a rendszer, hogy ha valaki bement egy ajtón, akkor az vagy nyitva volt, vagy előbb kinyitotta. A modellben erre egy első ránézésre fapados megoldás van: az 5 éves szinten szükséges common sense természetes nyelvi inputként tanítandó meg. Vagyis egyrészt le kell írni az 5 éves gyerek által kapott nyelvi inputot. Másrészt sok olyan inputot kell írni, amit egy 5 éves igazi gyerek nem kap, mint nyelvi inputot. (Például: *If a person moves toward a solid wall, he will stop at the wall and he is not able to move forward. A wall can have a large gap which means it is not solid. A large gap is taller and wider than a person. If a wall has a large gap and a person moves toward the gap, he will not stop at the wall but move forward.)* Látható a modell egy fontos limitációja, mely szerint az ember nem-nyelvi inputjait (például a vizuális inputot és a vizuális képzeletet) is mentalese-be kell hogy konvertálja. Ez egyfelől nehézség, másfelől egy létező kognitív jelenség analógiája (mely szerint egy történet-mesélés feladatban csekély a különbség a humán teljesítményben attól függően, hogy a történetet felolvasták, vagy némafilmen vetítették).

A háttértudás tehát nem más, mint alkalmas módon, emberek által angolul írt input a KB-ben tárolva.

**Indukció, analógia, intuíció, asszociáció, kreatív gondolkodás.** A reasoning deduktív jellegű következtetés. Az indukció más. Egyik esete, ha egy állítás egy speciális esetről található az inputban, de az általánosra következtetünk. Pl. Az input szerint az elefántnak 4 lába van, és ebből arra is következtetünk, hogy az állatnak 4 lába van. Vagy egy asszociáció példa: régen tudjuk, hogy egy konkrét vegyipari termék 10 kg-os zsákokban tárolva 10 évig biztosan eltartható, 100 kg-os kiszerelésben viszont pár éven belül néha felrobban. Most megtudjuk, hogy a gabonasilók néha felrobbannak, és kérdés, ez hogyan előzhető meg. Az asszociáció a *felrobban* concept környezetének aktiválásával indul el, és oda jut(hat), hogy a rendszer a válaszának részeként javasolja kipróbálni a gabonasilóban lévő gabona kisebb rekeszekben történő tárolását.

**Igazság.** A rendszer igazságfilozófiája a következő: egyre inkább igaz az a concept, ami egyre többször, és főleg határozottabban igazként állítva fordul elő az egyre megbízhatóbb forrásnak tekinthető inputban. Az eredmény a modellben az, hogy a következő mondat: *All dogs have four legs*  egy megbízható inputból egyetlen előfordulással is egy igen magas, nehezen megingatható igazságú conceptet eredményez (ezzel szemben a *He said all dogs have four legs* már kevésbé, mert ez a fajta forrás legtöbbször csak közepes lehet). Ez teljesen analóg a gyerekek tanulásával.

A megingathatóság abban az értelemben nem nehéz, hogy újabb egyetlen állítás tudja ezt módosítani, ha elég határozott és jó forrásból való. Ezen a ponton az irodalomban lévő modellek kezdenek nehézségbe ütközni, hogyan kezeljék a háromlábú, a lábatlan döglött kutyákat, azt a kutyaképet amire a lábat nem rajzolták rá, és a kivételek további végtelen sorát. Ezt a modellünk megfelelő reasoning-funkciókkal, valamint a lejjebb részletezett általánosság (generality) attribútummal kezeli. Továbbá a modellben az igazság nem bináris, hanem fokozatos, ennek is van némi szerepe, de nem ez a legfontosabb.

Itt elvágtam a doksit, sok további részletről a kurzuson lesz szó.