Magyar nyelvű Fedőnevek-ágensek létrehozása nyers szövegek alapján épített gráf felhasználásával

Cserháti Réka, Kolláth István, Kicsi András, Berend Gábor

Szegedi Tudományegyetem, Informatikai Intézet
Szeged, Árpád tér 2.
{cserhatir, kollathistvan}@gmail.com, {akicsi, berendg}@inf.u-szeged.hu

Kivonat A Fedőnevek társasjátékot játszó "kémfőnök" ágens feladata olyan szavak keresése, amelyeket egy emberi játékos bizonyos előre megadott szavak közül minél többhöz kapcsolódónak talál, de más megadott szavakhoz nem. Ezt a mai fejlett nyelvtechnológiai módszerekkel is kihívást jelent jól megvalósítani. Cikkünkben kísérletet teszünk ennek a feladatnak a matematikai leírására, és bevezetünk új, elméleti szempontból megalapozott módszereket, amivel egyben létrehozzuk az első magyar nyelven működő Fedőnevek-ágenseket is, melyekkel a véletlenszerű tippelést jelentősen felülmúló eredményeket érünk el valós, emberekkel folytatott játékok keretében.

Kulcsszavak: Fedőnevek, Játék, PMI-gráf, Szóasszociációk

1. Bevezetés

A mesterséges intelligencia kutatásának már nagyon régóta egyik központi témája a különféle játékokat emberi szinten, vagy annál jobban játszó ágensek fejlesztése. A területen a legtöbb tanulmány egymás ellen játszott, matematikailag könnyen formalizálható játékokra fókuszál (ld. pl. Allis és mtsai, 1994). A Fedőnevek című, napjainkban igen népszerűnek számító társasjáték viszont ezektől nagyban különbözik.

Az eredeti játékban két csapat versenyez egymással. Egy 25 szókártyából álló táblán vannak a kék vagy a piros csapathoz tartozó, illetve semleges kártyák, valamint egy azonnali vereséget jelentő kártya (fekete). Egy csapat akkor nyer, ha az ő csapatukhoz tartozó összes kártya felfedésre kerül a másik csapat kártyáinál korábban, vagy az ellenfél felfedi a fekete kártyát. Azt viszont, hogy melyik kártya milyen színű, mindkét csapatból csak 1-1 ember (a kémfőnök) tudja. Ezért a kémfőnökök minden körben egy-egy utalást adnak a csapatnak, mely egy utalószóból és egy számból áll. A csapat többi tagja (ügynökök) egymással egyeztetve sorban felfedik a táblán azokat a kártyákat, amelyeket az utalószóhoz kapcsolódónak gondolnak, amíg rossz kártyára nem tippelnek, vagy el nem érik a kémfőnök által számként megadott limitet.

Tehát a játékhoz kétféle ágens, egy kémfőnök és egy ügynök(csapat) létrehozása is lehetséges. Mindkettő fő feladata az, hogy az emberi játékosokkal jó együttműködésre legyen képes. Ehhez azonban a játékban megfigyelhető emberi viselkedés modellezésére is szükség van, jelen esetben annak kiszámítására, hogy az emberek általánosan mely szavakat tartják egymáshoz kapcsolódónak.

Ez a feladat meglehetősen kapcsolódik a pszicholingvisztikában régóta és sokat kutatott szóasszociáció-modellezéshez (Palermo és Jenkins, 1964; McNeill, 1966), azonban semmiképp sem nevezhető vele ekvivalensnek. A szóasszociációs kísérleteknél az alanyoknak a lehető leggyorsabban kell adott szóhoz egy hozzá kapcsolódó tetszőleges szót megnevezni, de ebben az esetben a kémfőnök feladata egy olyan szó megtalálása, ami minél több szóhoz kapcsolódik a sajátjai közül, a többi szóhoz viszont nem, vagy jelentősen kevésbé szorosan. A feladatra szánható ideje is legfeljebb nagyon lazán (a többi játékos türelmében) van korlátozva, és – személyes tapasztalatok alapján – a kémfőnökök gyakran több perc gondolkodási időt is felhasználnak a megfelelő utalás kigondolására. Emiatt előfordul, hogy az összekapcsolt szavak bonyolult módon, akár csak áttételesen függenek össze. Az ügynökök feladata – a tábla szavai közül az utalószóhoz kapcsolódók megkeresése – ennél jobban hasonlít egyszerű asszociációkhoz, de az idő itt sem lényeges, és az asszociációknak fontos korlátját képezi a táblán szereplő szavak száma. Egy személyes játékban a játékosok közötti viszony és közös ismeretek is szerepet játszhatnak, egy robottal való játék esetén viszont ez nem befolyásol.

2. A játék matematikai leírása

Tételezzük fel, hogy V szótárra létezik egy $D \in \mathbb{R}^{|V|\times|V|}$ távolság- vagy hasonlóságmátrix, amelyben $D_{ij} = d(v_i, v_j)$ a kapcsolat egzakt mérőszáma két tetszőleges v_i, v_j szó között, vagyis minden ember tudatában pontosan ilyen erősek a kapcsolatok. Ekkor az ügynök-ágens megvalósítása egyszerű: a táblán szereplő szavak közül mindig azt kell választani, amelyik a legszorosabban kapcsolódik az utalószóhoz, vagy a legközelebb van hozzá. Így egy mohó kémfőnök-ágens sem túl bonyolult: legyen v_i a szótár i-edik szava, és minden i-re $[w_{i1}, w_{i2}, ..., w_{in}]$ az adott körben a táblán szereplő felfedetlen szavak, sorba rendezve a v_i-hez legszorosabban kapcsolódótól a legkevésbé kapcsolódóig. Ekkor azt az i-t keressük, amelyre a legnagyobb olyan k szám létezik, hogy $w_{i1}, w_{i2}, ..., w_{ik}$ mind az ágens csapatához tartozó szavak.

Valójában azonban ilyen feltételek mellett az ügynökök viselkedése determinisztikus, tehát tulajdonképpen a két kémfőnök játszik egymás ellen. A szótár, vagyis a lehetséges döntéseik száma véges, és a kémfőnökök ismerik minden döntés kimenetelét, azaz egymás lehetséges stratégiáit is. Így a játék már kétszemélyes diszkrét matematikai játékká, azaz véges fával leírhatóvá válik. A mohó döntés nem feltétenül optimális, mert egy kémfőnöknek figyelembe kell vennie, hogy a saját és a másik kémfőnök döntései függvényében milyen lehetőségeik lesznek később, és ez alapján kell optimalizálnia a döntését. Ilyen keretek között az optimális stratégia kialakítása még további kutatás tárgyát képezheti, de nem a nyelvtechnológia területén, így ebben a cikkben ezzel többet nem foglalkozunk.

A fenti feltételek persze a valóságtól nagyon távol állnak, hiszen ilyen távolságfüggvény, ami minden ember mentális reprezentációinak tökéletesen megfelel, biztosan nem létezik. Ezt már csak onnan is láthatjuk, hogy a klasszikus asszociációs teszteken, ahol tulajdonképpen a legközelebbi szomszédok keresése a feladat, a kísérleti alanyok sosem adják mindannyian ugyanazt a választ (Palermo és Jenkins, 1964; Postman és Keppel, 2014). Értelmes probléma viszont az, hogy olyan távolságfüggvényt hozzunk létre, melyben a szavak közelsége átlagosan a lehető legkisebb mértékben tér el az emberek által megítélt kapcsolat-erősségtől. Sőt, ha korlátlan mennyiségű emberi adat állna rendelkezésünkre, azt is megállapíthatnánk, hogy egy véletlenszerűen választott emberi ügynök-játékos bármely utalásra mekkora valószínűséggel választja adott tábla egyes szavait, így egy sztochasztikus, de matematikailag továbbra is megoldható játékot kapnánk.

Korlátlan mennyiségű adat hiányában azonban a célunk egy a priori távolságmátrix kialakítása, ami viszonylag jól közelíti az ideális, az emberek megítéléséhez átlagosan legközelebb álló távolságmátrixot; továbbá ezen egy olyan pontozófüggvény definiálása (szintén a priori), ami a lehetséges utalásokat valósághűen rangsorolja aszerint, hogy arra várhatóan hány jó tippet tud majd adni az emberi ügynök-játékos. Ezzel a távolságmátrixunk és a pontozófüggvényünk együtt már egy (mohó) kémfőnök-ágenst is meghatároz.

Tanulmányunkban kémfőnök-ágenseket hozunk létre kétféle távolságfüggvény és ezeken négy pontozófüggvény alkalmazásával. Ezek tesztelésére egy saját fejlesztésű webes játékban gyűjtjük az emberi ügynökjátékosok döntéseit, így először értékelünk ki Fedőnevek-ágenseket valós játékkörnyezetben. Az eredmények szerint sikerül a véletlenszerű tippelést jelentősen meghaladó, és más korábbi (de más körülmények között tesztelt) Fedőnevek-ágensekével összemérhető eredményeket elérnünk.

3. Kapcsolódó irodalom

3.1. Asszociációk

A szóasszociációkat régóta, számos okból vizsgálják pszichológiai, pszicholingvisztikai és nyelvészeti kutatások: a segítségükkel vizsgáltak bizonyos betegségeket, használták a memória és kognitív folyamatok működésébe való betekintésre, valamint a kognitív lexikonnak és a nyelv bizonyos folyamatainak modellezésére is (Bel-Enguix és mtsai, 2019).

Az asszociációkkal foglalkozó tanulmányok közül számunkra Spence és Owens (1990) eredményei a legfontosabbak. Ők kimutatták, hogy egy korpuszban a szavak együttelőfordulásainak mennyisége jól jelzi a köztük lévő szemantikai kapcsolat szorosságát, és asszociációk erősségének mérésére is alkalmas. Jelen munkánkban ezért a szavak kapcsolódását modellező távolságfüggvényt súlyozott kookkurrenciák alapján valósítjuk meg (részletesen ld. 4.1.).

Habár létezik olyan gráf (Bel-Enguix és mtsai, 2014), és ebből olyan szóbeágyazás-modell is (Bel-Enguix és mtsai, 2019), amelyet speciálisan az asszociációk alapján hoztak létre, ezek megvalósításához nehezen megszerezhető asszo-

ciációs adatokra van szükség, ami nagy erőforrásigénynek számít a nyers korpuszokkal összevetve, és nehézzé teszi az ezt felhasználó módszerek alkalmazását újabb nyelveken.

3.2. Fedőnevek-ágensek

A – legjobb tudomásunk szerint – első Fedőnevek-ágensre hasonlító algoritmusokat kifejezetten azért hozták létre Shen és mtsai (2018), hogy a segítségükkel az emberi asszociációkat modellezhessék. Az ő egyszerűsített játékukban a tábla mindig három főnévből áll, ezek közül előre meghatározott kettőhöz kell megfelelő utalószót találni, ami viszont csak három melléknév közül lehet az egyik. Az alábbi öt különböző távolságszámítás alapján állították elő az utalásokat:

- Bigramok valószínűsége a szavak gyakoriságához viszonyítva,
- Skip-gram (Mikolov és mtsai, 2013) szerinti koszinusz hasonlóság,
- GloVe (Pennington és mtsai, 2014) szerinti koszinusz hasonlóság,
- A ConceptNet5 (Speer és Havasi, 2013) tudásgráf szerinti kapcsolat,
- Topikmodellezéssel számított távolság.

Úgy találták, a bigramok valószínűsége alapján modellezhető az emberi játékosok viselkedése a legjobban, ami összevág Spence és Owens (1990) eredményeivel (bár ők az együttelőfordulásokat sokkal nagyobb ablakmérettel számolták).

Kim és mtsai (2019) célja már közvetlenül a játékot jól játszó ágensek építése volt. Metrikáik hátttereként a

- CBOW, Skip-gram és GloVe szóbeágyazásokat (több konfigurációban),
- valamint a WordNet gráf alapú adatbázist (Miller, 1995)

használták, utóbbit számos különböző távolságfüggvénnyel. Tanulmányukban azonban nem emberi adatokkal értékelik az ágensek teljesítményét, hanem kémfőnök- és ügynökágensek párosításával, amivel viszont csak két ágens működésének hasonlósága deríthető ki, az emberekkel való együttműködési képességüktől függetlenül.

Jaramillo és mtsai (2020) a távolságot a következő reprezentációk alapján számítják:

- Wikipédia-cikkekből és szótári definíciókból számolt TF-IDF hasonlóság,
- a szavak naiv-Bayes alapú osztályozása, illetve
- a GPT2 transzformer nyelvi modell (Radford és mtsai, 2019) első rétegéből kinyert szóbeágyazások.

Ezen módszerek közül ők a GPT2 vektorait találják legalkalmasabbnak az emberekkel való együttműködésre.

A legfrissebb a témában Koyyalagunta és mtsai (2021) cikke, amelyben a korábban is használt Skip-gram és GloVe szóbeágyazások mellett távolságmátrixaik előállítására

- a FastText-et (Bojanowski és mtsai, 2017),

– a BERT modellt (Devlin és mtsai, 2018),
– valamint a BabelNet tudásgráfon (Navigli és Ponzetto, 2010) egy erre a célra kifejlesztett, szavakat speciális szabályok mentén egymáshoz rendelő keretrendszert

is felhasználnak.

A fenti munkák a szavak közti távolság számításán kívül az utalások pontozófüggvényeiben is eltérnek. Koyyalagunta és mtsai (2021) jelöléseihez igazodva, legyen $\tilde{c}$ egy lehetséges utalószó, I_n a célzott szavak halmaza, vagyis a $\tilde{c}$ utalószóhoz legközelebbi n db kék szó, R az összes, nem a csapathoz tartozó (rossz) szó halmaza, és $s(\cdot,\cdot)$ a két szó hasonlóságát számító függvény. Kim és mtsai (2019) pontozófüggvénye ekkor

$$g_{kim}(\tilde{c}, n) = \begin{cases} \min_{b \in I_n} s(\tilde{c}, b), & \text{ha } \min_{b \in I_n} s(\tilde{c}, b) > \max_{r \in R} s(\tilde{c}, r) \\ 0, & \text{különben.} \end{cases} \quad (1)$$

Jaramillo és mtsai (2020) ugyanezt a függvényt veszik át, emellett kiegészítik egy veszélyességi súlyozással a kártyák színe szerint. Koyyalagunta és mtsai (2021) viszont definiálnak egy másikat is:

$$g_{koy}(\tilde{c}, n) = \lambda_B \Big(\sum_{b \in I_n} s(\tilde{c}, b) \Big) - \lambda_R \big(\max_{r \in R} s(\tilde{c}, r) \big), \quad (2)$$

ahol λ_B és λ_R beállítható paraméterek.

Ezenkívül pedig bevezetnek egy másik módszert is, amellyel nem kizárólag a szóhasonlóságok alapján pontozzák az utalásokat, hanem a gyakoriságuk és a szótári definícióikból létrehozott vektorok (Dict2vec, Tissier és mtsai, 2017) hasonlósága alapján is – de ez tulajdonképpen az eredeti távolságmátrix módosításának tekinthető.

Az ő eredményeik szerint a GloVe alapján számolt távolságok teljesítenek legjobban a szótári definíciókkal és gyakorisággal kombinálva, anélkül viszont a FastText vektorok hasonlósága bizonyul a legjobb metrikának.

3.3. Gráfos szóreprezentációk

A szavak reprezentálására a legjobban bevált módszer azok vektorokhoz rendelése, de emellett találkozhatunk gráfos szóreprezentációkkal is. Vektorok esetén minden szót egy-egy vektor reprezentál, melyek között a távolságot leggyakrabban a két vektor bezárt szöge alapján számítják ki. Gráfos szóreprezentációk esetén a szótár összes szava egy nagy gráf csúcsainak feleltethető meg, ezek között a távolságot a gráftól függően sokféleképpen lehet definiálni, egy lehetőség például a két csúcs közti legrövidebb út hossza. Ilyenek a korábbi módszerekben használt tudásgráfok (Miller, 1995; Speer és Havasi, 2013; Navigli és Ponzetto, 2010) is például. Ahogy az előbb láttuk, egy Fedőnevek-ágens távolságmátrixának előállítására mindkét struktúra használható.

Jelen munkánkban egy gráfot (is) használunk, ami valamennyire hasonlít Hope és Keller (2013) együttelőfordulások alapján létrehozott gráfjára. Ezt a

Word Sense Induction (többjelentésű szavak jelentéseinek szétválasztása háttérinformációk nélkül) feladatra használják, később Pelevina és mtsai (2016) egy hasonló módszert alkalmaznak szóbeágyazás-modellek többértelműsítésére.

Másik szempontból hasonlónak tekinthető gráf a GraphGlove (Ryabinin és mtsai, 2020) is, melyben a GloVe célfüggvénye szerint optimalizálják a távolságokat egy gráf csúcsainak megfeleltetett szavak között.

4. Saját ágenseink

Az általunk létrehozott ágensek[1] nemcsak abban különböznek a korábbiaktól, hogy magyar nyelven működnek, hanem javaslunk két új távolságszámítási módszert és több pontozófüggvényt is, amelyeket egymással kombinálva valós, emberekkel játszott játékban tesztelünk.

4.1. Távolságszámítás

A korábbi, asszociációk és együttelőfordulások kapcsolatáról szóló eredmények (Spence és Owens, 1990; Shen és mtsai, 2018) figyelembe vételével a távolságmátrixainkat nem a nyelvtechnológia legfrissebb neurális módszereivel, hanem nyers szövegben számolt együttelőfordulások alapján hozzuk létre.

Legyen $\mathbf{V} = \{v_1, ..., v_n\}$ egy szótárban szereplő összes szó. Az egyszerűbb módszer, hogy két szó távolságának tekintjük a következőt:

$$d_0(v_i, v_j) = \begin{cases} 1/\mathrm{PMI}(v_i, v_j), & \text{ha } \mathrm{PMI}(v_i, v_j) > 1 \\ \infty, & \text{különben,} \end{cases} \tag{3}$$

ahol $\mathrm{PMI}(v_i, v_j) = \ln\left(\frac{p(v_i, v_j)}{p(v_i) \cdot p(v_j)}\right)$, vagyis az együttes előfordulás valószínűsége a gyakoriságokhoz viszonyítva.

A másik javaslatunk távolságszámításra (egyben egy saját, új gráfos szóreprezentáció) ennek kiterjesztése egy $G(V, E, w)$ súlyozott gráffá, ahol a csúcsok a $\mathbf{V}$ szótár szavainak felelnek meg, és az élsúly két csúcs között $w(e(v_1, v_2)) = d_0(v_1, v_2)$. Ekkor két szó távolságát Ryabinin és mtsai (2020)-hoz hasonlóan a G-ben köztük lévő legkisebb összsúlyú út összsúlyaként definiálhatjuk, azaz

$$d_G(v_i, v_j) = \min_{\pi \in \Pi_G(v_i, v_j)} \sum_{e_k \in \pi} w(e_k), \tag{4}$$

ahol $\Pi_G(v_i, v_j)$ a G-ben v_i-ből v_j-be vezető összes út halmaza. Ennek előnye, hogy (4)-ben olyan szópárokra is értelmeződik a távolság, ahol (3) végtelent adott.

Jelen tanulmányunkban az együttelőfordulásokat a Magyar Webkorpusz 2.0 (Nemeskey, 2020) egy részén (a Wiki alkorpuszon és a 2019-es szövegeken, összesen 1,414 milliárd token), lemmatizált alakok között számláljuk, a szavak távolsága szerint is súlyozva, a GloVe szóbeágyazások tanításához használt módszerrel[2], maximum 10-es ablakmérettel. Csak a legalább 700-szor előforduló szavakat vesszük figyelembe, a szótár mérete így 12 267.

[1] `https://github.com/xerevity/CodeNamesAgent`

[2] `https://github.com/stanfordnlp/GloVe/blob/master/src/cooccur.c`

4.2. Pontozófüggvények

Mondjuk azt, hogy az ágens a kék csapatban játszik, azaz a kék szavakhoz kapcsolódó utalásokat szeretnénk generálni, természetesen a fenti távolságfüggvények alapján. Kim és mtsai (2019) függvényei egy lehetséges utalás pontszámát az alapján határozták meg, hogy az adott szó mennyire hasonló a kék szavakhoz. Ennek viszont hátránya, hogy lehetnek az utalószóhoz hasonlító kék (jó) szavak mellett az utalószóhoz csak egy nagyon kicsivel kevésbé hasonlító más színű (rossz) szavak. Feltehetjük, hogy ilyen esetben az ügynökök kisebb valószínűséggel választják a célzott szavakat; vagy általánosan, mennél kisebb a különbség két szó utalástól való távolsága között a mi távolságfüggvényünk szerint, annál nagyobb valószínűséggel lesz az emberi játékos szerint fordított a két szó sorrendje az utaláshoz való hasonlóságban.

A korábbi munkáktól eltérően, a gráffal történő távolságszámítás miatt nekünk nem a szavak hasonlóságáról, hanem a szavak távolságáról vannak adataink, melyek alapján a lehetséges utalásokat rangsorolni szeretnénk, ezért a pontozófüggvényeinknek ebben különbözniük kell a korábbiaktól. Mivel hasonlóságról sokféle módon át lehet térni távolságra (és fordítva), a távolságokon és hasonlóságokon alapuló pontozófüggvények között ekvivalencia csak bizonyos, ezen áttérésre vonatkozó feltételek ismeretében állapítható meg.

Pontozófüggvényeinkben a célszavak távolságának és a rossz szavak távolságának különbségét, illetve arányát vizsgáljuk. Ha előre megadott számú szót kell összekötni (ahogy például Koyyalagunta és mtsai (2021) csak 2 szóra célzott utalásokkal tesztelték az ágenseiket), akkor a mi pontozófüggvényünk a legközelebbi rossz szó és a legtávolabbi célzott szó utalástól való távolságának különbségét számítja ki:

$$g_K(\tilde{c}, n) = \min_{r \in R} d(\tilde{c}, r) - \max_{b \in I_n} d(\tilde{c}, b), \tag{5}$$

a 3.2 részben használt jelölésekkel.

Vagy a legközelebbi rossz szó és a legtávolabbi célzott szó utalástól való távolságának arányát ugyanígy:

$$g_H(\tilde{c}, n) = \min_{r \in R} d(\tilde{c}, r) / \max_{b \in I_n} d(\tilde{c}, b). \tag{6}$$

Egy teljes játékban viszont realisztikusabb, hogy az ágens maga választja ki azt is, hogy az utalással hány szót céloz meg. Ezért egy másik pontozófüggvényünk a következő:

$$g_K(\tilde{c}) = \sum_{b \in I} \left(\min_{r \in R} d(\tilde{c}, r) - d(\tilde{c}, b) \right), \tag{7}$$

ahol I az összes olyan kék szó halmaza, amelyek a lehetséges utaláshoz az összes rossz szónál közelebb vannak. Amennyiben itt is csak az n legközelebbi kék szót számítanánk, ez a függvény megfelelő hasonlóság-távolság áttérés esetén ekvivalens lenne a (2) függvénnyel, ha abban $\lambda_R = n \cdot \lambda_B$.

Ennek a különbség helyett hányadossal számolt változata pedig:

$$g_H(\tilde{c}) = \sum_{b \in I} \left(\min_{r \in R} d(\tilde{c}, r)/d(\tilde{c}, b_i) \right). \tag{8}$$

Az utóbbi pontozófüggvényeknek viszont két fontos hibája van. Egyrészt lehetséges, hogy egyetlen, az utaláshoz nagyon közeli szó miatt adnak egy utalásnak magas pontszámot. Ennek orvoslására a legközelebbi rossz szó távolságától való különbségnek szabunk egy felső határt, g_{K_S} esetében 0.3-at, g_{H_S}-nél 5-öt.

Másrészt pedig lehetnek olyan szavak a célzottak között, melyek utalástól való távolsága csak minimálisan tér el a legközelebbi rossz szóétól, így az ügynökök túl nagy valószínűséggel választják majd azokat. Ennek kiküszöbölésére pedig csak azokat a szavakat tekintjük a pontszámításnál a legközelebbi rossz szónál közelebbinek, amelyeknek az utalástól számított távolsága legalább 0.05 értékkel kisebb, mint a legközelebbi rossz szóé. Ezek önkényesen, intuíció alapján választott értékek, melyek finomhangolása az emberi ügynököktől gyűjtött adatok alapján javíthatja az ágensek teljesítményét.

1. ábra: Egy kiértékeléshez használt tábla, a kékkel keretezett szavak felfedendők.

5. Kiértékelés és eredmények

Az ágensek kiértékelésében nagyrészt Koyyalagunta és mtsai (2021)-hoz igazodunk. Ők a táblákra csak olyan szavakat sorsolnak, amelyek az eredeti Fedőnevek társasjáték szavai, és mi is ezt tesszük. A játékhoz tartozó 400 szóból 245-öt használunk, mert ennyi szerepel a saját adatbázisunkban, vagyis a korpuszban legalább 700-szor (az idézett tanulmányban ez a szám 208). Szintén az ő mintájukra egy táblára 20 szót sorsolunk, melyek közül 10 felfedendő, a másik 10 között

Távolság	Pontozás	Utalószó	Célzott szavak
d_G	$g_K(\cdot, 2)$	tűzijáték	torta, bomba
d_G	$g_K(\cdot, 3)$	meghökkentő	háló, csavar, egyenes
d_G	$g_{K_s}(\cdot)$	kapus	háló, bomba, kesztyű
d_G	$g_H(\cdot, 2)$	kapus	háló, kesztyű
d_G	$g_H(\cdot, 3)$	sarok	háló, csavar, kesztyű
d_G	$g_{H_s}(\cdot)$	szeglet	csavar, bomba, kesztyű, bánya, fűszer, egyenes
d_0	$g_{K,H}(\cdot, 2)$	tripla	torta, csavar
d_0	$g_{K,H}(\cdot, 3)$	hagyományos	torta, bomba, fűszer
d_0	$g_K(\cdot)$	szárít	háló, csavar, kesztyű, fűszer, egyenes
d_0	$g_H(\cdot)$	szárít	fűszer, egyenes, kesztyű, csavar, háló

1. táblázat. Távolság- és pontozófüggvényeink kombinációjával generált utalások az 1. ábrán látható táblához

nem teszünk különbséget (a kiértékelés szempontjából ezek színe nem lényeges). 100 táblát generálunk a kiértékeléshez, közülük az egyik látható az 1. ábrán.

Koyyalagunta és mtsai (2021) csak 2 szóra irányuló utalásokkal értékelik az ágenseket; egy emberi játékostól legalább 2, legfeljebb 4, az utaláshoz kapcsolódó, táblán lévő szót kérve. Mi ezért a játékosoknak legalább n, legfeljebb $n + 2$ szó bejelölését engedjük meg (n értelemszerűen a célzott szavak száma), a sorrendet is figyelembe véve. Az adatokat egy saját fejlesztésű webes alkalmazással[3] gyűjtöttük, részletesebb információk találhatók a 3. táblázatban.

Koyyalagunta és mtsai (2021)-hoz hasonlóan, a kiértékelés mérőszámaiként kiszámítjuk a játékosok által megjelölt szavak pontosságát és fedését a kémfőnök ágensek által célzott szavakon. Legyen A az ágens célzott szavainak halmaza, u_i az ügynökjátékos által i-ediknek megjelölt szó, $U_k = \{u_1, u_2, ..., u_k\}$, U pedig az ügynök által megjelölt összes szó. Ekkor a következőképpen számítjuk ki az értékelés mérőszámait:

– Pontosság:

$$\text{P@n} = \frac{|A \cap U_n|}{|A|}$$

– Fedés:

$$\text{R@n+2} = \frac{|A \cap U|}{|A|}$$

Ezeken kívül bevezetünk egy saját, az eredeti játék szempontjából relevánsabb mérőszámot, mely egy tényleges, ágens és ember által végigjátszott játék szempontjából pontosabban mutatja meg az ágens teljesítményét. A tényleges játékban ugyanis csak addig tippelhetnek az ügynökök, amíg egy rossz kártyát el nem találnak, vagyis a bejelölés sorrendje is fontos. Az előrehaladás mérőszáma azt mutatja meg, ilyen szabályok mellett hány jó kártyát sikerült az ügynököknek a körben felfedni, ha csak a célzott szavakat tekintjük jónak, vagyis azt is

[3] `https://fedonevekagens.herokuapp.com/`

rossznak számítjuk, ha egy nem célzott, de jó szót fed fel véletlenül. Képlettel:

$$H = k \mid \forall u \in U_k : u \in A, u_{k+1} \notin A.$$

Ezek várható értéke véletlenszerű tippek esetén a következő:

– Pontosság:

$$\mathbb{E}(\text{P@n}) = \frac{|A|}{20}$$

– Fedés:

$$\mathbb{E}(\text{R@n+2}) = \frac{|U|}{20}$$

– Előrehaladás[4]:

$$\mathbb{E}(H) = \frac{21}{21 - |A|} - 1.$$

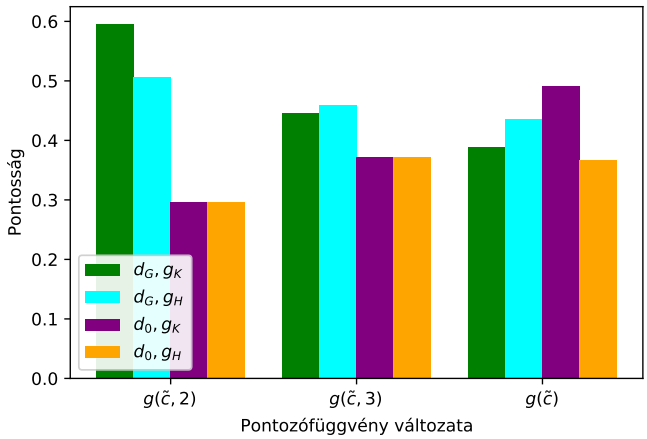
Távolság	Pontozás	Pontosság	Fedés	Előrehaladás
d_G	$g_K(\cdot, 2)$	0.595 (0.1)	0.611 (0.11)	1.111 (0.105)
d_G	$g_K(\cdot, 3)$	0.446 (0.15)	0.457 (0.165)	0.903 (0.167)
d_G	$g_K(\cdot)$	0.389 (0.233)	0.398 (0.243)	0.8 (0.285)
d_G	$g_H(\cdot, 2)$	0.507 (0.1)	0.529 (0.115)	0.865 (0.105)
d_G	$g_H(\cdot, 3)$	0.46 (0.15)	0.503 (0.164)	0.818 (0.167)
d_G	$g_H(\cdot)$	0.435 (0.273)	0.435 (0.276)	1.1 (0.351)
d_0	$g_{K,H}(\cdot, 2)$	0.296 (0.1)	0.304 (0.106)	0.453 (0.105)
d_0	$g_{K,H}(\cdot, 3)$	0.372 (0.15)	0.389 (0.166)	0.728 (0.167)
d_0	$g_K(\cdot)$	0.491 (0.248)	0.502 (0.256)	1.142 (0.309)
d_0	$g_H(\cdot)$	0.367 (0.216)	0.468 (0.227)	0.738 (0.260)

2. táblázat. Távolság- és pontozófüggvényeink eredményei, zárójelben a célzott szavak és a tippek számának átlagából számított várható értékekkel.

A kiértékelés eredményeit a 2. táblázat tartalmazza. A d_0 távolságfüggvény mellett az (5) és (6) pontozófüggvények a 100 felhasznált táblán pontosan ugyanazokat az utalásokat generálták, így ezeket együtt értékeltük ki. A random játékot minden konfiguráció magasan felülmúlja, és bár a tesztelés körülményei között számos különbség van (nyelv, táblákra sorsolható szavak, emberi játékosok), a 2 szóra irányuló, gráffal történő távolságszámítással generált utalásaink jól teljesítenek Koyyalagunta és mtsai (2021) eredményeihez képest is. Az ő ágenseik pontosságai 0.3 és 0.667 közöttiek, a Dict2vec nélküliek között pedig 0.6 a legmagasabb pontosság.

Megfigyelhető ugyanakkor, hogy a gráffal történő távolságszámítás esetén a 2 szóra célzott utalások jelentősen jobban teljesítenek, mint 3 vagy nagyobb számra, pedig véletlenszerűen a nagyobb számhoz tartozó utalások eredményeinek

[4] A levezetést ld. pl.: Ahlgren (2014).



2. ábra: Az ágenseink pontossága

várható értéke magasabb. Sajnos az eddigi irodalomban nincs kiértékelés 2-nél több szóra célzott utalásokra, így az nyitott kérdés marad, hogy ezt a feladat nehézsége vagy a módszereink hiányosságai okozzák.

A meghatározatlan számú célzott szóra generált utalások általános problémája, ahogy az 1. táblázat példáján és a 3. táblázatban is látszik, hogy túl sok szóra célozzák az utalásokat, így azok feltehetőleg kevésbé egyértelműek. Egy ötlet ennek orvoslására a későbbiekben, ha a távolságkülönbségek harmonikus közepét vesszük a számukkal megszorozva, az összeg helyett, ami másképp a számtani közép az elemszámmal megszorozva.

Azt is láthatjuk, hogy a pontozófüggvények nem egyformán működnek a különböző távolságszámítási módszerekkel kombinálva. Sajnos eszerint a későbbiekben a legjobb konfiguráció megtalálásához szükséges lesz erre vonatkozó heurisztikákat bevezetni, vagy minden távolságfüggvényt minden pontozófüggvénnyel tesztelni kell.

6. Összegzés és fejlesztési irányok

Jelen munkánkban kísérletet tettünk a Fedőnevek társasjáték matematikai leírására egy kémfőnök-ágens feladatának két részre bontásával. Az emberekkel való együttműködéshez először egy olyan távolság- vagy hasonlóságmátrix megadása szükséges, amely kellően jól közelíti az emberek megítélése szerinti kapcsolatokat, majd ezen egy olyan pontozófüggvényt kell definiálni, amely a lehetséges utalásokat aszerint rangsorolja, hogy azokra hány jó tippet fog várhatóan egy emberi játékos adni.

Korábbi, asszociációkkal kapcsolatos kutatások alapján a távolságmátrixainkat egy korpusz szavai közti együttelőfordulások alapján hoztuk létre. Pontozó-

Távolság	Pontozás	Játékok	Célzott szavak	Jelölt szavak	Gondolkodási idő
d_G	$g_K(\cdot, 2)$	63	2	2.190	00:34 min
d_G	$g_K(\cdot, 3)$	62	3	3.306	00:36 min
d_G	$g_K(\cdot)$	50	4.66	4.860	01:09 min
d_G	$g_H(\cdot, 2)$	67	2	2.298	00:24 min
d_G	$g_H(\cdot, 3)$	55	3	3.290	00:43 min
d_G	$g_H(\cdot)$	50	5.46	5.520	00:47 min
d_0	$g_{K,H}(\cdot, 2)$	64	2	2.125	00:35 min
d_0	$g_{K,H}(\cdot, 3)$	59	3	3.322	00:34 min
d_0	$g_K(\cdot)$	49	4.96	5.122	00:55 min
d_0	$g_H(\cdot)$	111	4.33	4.558	00:48 min

3. táblázat. A lejátszott játékok száma, a célzott szavak és a tippek számának átlaga, és az egy táblára jutó átlagos gondolkodási idő a kiértékelésben.

függvények tekintetében is vezettünk be újításokat, egyrészt figyelembe vettük a legközelebbi rossz szó, illetve a célzott szavak utalástól való távolságának különbségét, másrészt olyan függvényt is definiáltunk, amely képes a célzott szavak számának eldöntésére is.

A továbbiakban mindenképpen célunk a saját távolság- és pontozófüggvényeink finomítása a kiértékeléssel gyűjtött adatok alapján, és az összehasonlításuk a korábbi megfelelőikkel, azonos kiértékelési körülmények között. Ezenkívül szeretnénk adatot gyűjteni emberi kémfőnök-játékosok döntéseiről is, és ezekhez ügynök-ágenseket létrehozni, ami közvetlenül a távolságfüggvény optimalizálását teszi lehetővé.

Bár számtalan nyelvtechnológiai módszer került már felhasználásra a távolságmátrixok létrehozásához, vannak még olyanok, amiket érdemes lenne kipróbálni. Ilyen például az asszociációk gráfba foglalása (Bel-Enguix és mtsai, 2014) és a GraphGlove (Ryabinin és mtsai, 2020). A távolságmátrix bevezetésének köszönhetően lehetőség nyílik több különböző módon előállított mátrix aggregálására is, létrehozva például egyszerre együttelőfordulásokon, tudásgráfokon és neurális szóreprezentációkon alapuló távolságmátrixokat, ami nagyon érdekes új iránynak ígérkezik. A különböző módszerekkel előállított távolságmátrixok hasonlóságának vizsgálata pedig más területeken, például kognitív modellezésben és neurális szóreprezentációk interpretálhatóságában is fontos tanulságokat tartogathat.

Köszönetnyilvánítás

Köszönjük a játékot tesztelő ismerőseinknek a kiértékeléshez szükséges adatokat, valamint a bírálóknak, hogy hasznos javaslataikkal hozzájárultak a cikk végső verziójához.

A cikkben bemutatott eredmények részben az Innovációs és Technológiai Minisztérium ÚNKP-21-1 kódszámú Új Nemzeti Kiválóság Programjának a Nem-

zeti Kutatási, Fejlesztési és Innovációs Alapból finanszírozott szakmai támogatásával készültek. A dolgozatban szereplő kutatási eredmények létrejöttét az Innovációs és Technológiai Minisztérium és a Nemzeti Kutatási, Fejlesztési és Innovációs Hivatal is támogatta a Mesterséges Intelligencia Nemzeti Laboratórium keretében.

Hivatkozások

Ahlgren, J.: The probability distribution for draws until first success without replacement (2014)

Allis, L.V., és mtsai: Searching for solutions in games and artificial intelligence (1994)

Bel-Enguix, G., Gómez-Adorno, H., Reyes-Magaña, J., Sierra, G.: Wan2vec: Embeddings learned on word association norms. Semantic Web 10(6), 991–1006 (2019)

Bel-Enguix, G., Rapp, R., Zock, M.: A graph-based approach for computing free word associations. In: Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14). pp. 3027–3033 (2014)

Bojanowski, P., Grave, E., Joulin, A., Mikolov, T.: Enriching word vectors with subword information. Transactions of the Association for Computational Linguistics 5, 135–146 (2017)

Dagan, I., Lee, L., Pereira, F.C.: Similarity-based models of word cooccurrence probabilities. Machine learning 34(1), 43–69 (1999)

Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding pp. 4171–4186 (2018), `https://aclweb.org/anthology/papers/N/N19/N19-1423/`

Hope, D., Keller, B.: Uos: A graph-based system for graded word sense induction. In: Second Joint Conference on Lexical and Computational Semantics (* SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013). pp. 689–694 (2013)

Jaramillo, C., Charity, M., Canaan, R., Togelius, J.: Word autobots: Using transformers for word association in the game codenames. In: Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment. vol. 16, pp. 231–237 (2020)

Kim, A., Ruzmaykin, M., Truong, A., Summerville, A.: Cooperation and codenames: Understanding natural language processing via codenames. In: Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment. vol. 15, pp. 160–166 (2019)

Koyyalagunta, D., Sun, A., Draelos, R.L., Rudin, C.: Playing codenames with language graphs and word embeddings. Journal of Artificial Intelligence Research 71, 319–346 (2021)

McNeill, D.: A study of word association. Journal of Verbal Learning and Verbal Behavior 5(6), 548–557 (1966)

Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J.: Distributed representations of words and phrases and their compositionality. In: Advances in neural information processing systems. pp. 3111–3119 (2013)

Miller, G.A.: Wordnet: a lexical database for english. Communications of the ACM 38(11), 39–41 (1995)

Navigli, R., Ponzetto, S.P.: Babelnet: Building a very large multilingual semantic network. In: Proceedings of the 48th annual meeting of the association for computational linguistics. pp. 216–225 (2010)

Nemeskey, D.M.: Natural Language Processing Methods for Language Modeling. Ph.D.-értekezés, Eötvös Loránd University (2020)

Palermo, D.S., Jenkins, J.J.: Word association norms: Grade school through college. (1964)

Pelevina, M., Arefiev, N., Biemann, C., Panchenko, A.: Making sense of word embeddings. In: Proceedings of the 1st Workshop on Representation Learning for NLP. pp. 174–183. Association for Computational Linguistics, Berlin, Germany (Aug 2016), https://aclanthology.org/W16-1620

Pennington, J., Socher, R., Manning, C.D.: Glove: Global vectors for word representation. In: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). pp. 1532–1543 (2014)

Postman, L., Keppel, G.: Norms of word association. Academic Press (2014)

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I., és mtsai: Language models are unsupervised multitask learners. OpenAI blog 1(8), 9 (2019)

Ryabinin, M., Popov, S., Prokhorenkova, L., Voita, E.: Embedding words in non-vector space with unsupervised graph learning. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). pp. 7317–7331 (2020)

Shen, J.H., Hofer, M., Felbo, B., Levy, R.: Comparing models of associative meaning: An empirical investigation of reference in simple language games. In: Proceedings of the 22nd Conference on Computational Natural Language Learning. pp. 292–301 (2018)

Speer, R., Havasi, C.: Conceptnet 5: A large semantic network for relational knowledge. In: The People's Web Meets NLP, pp. 161–176. Springer (2013)

Spence, D.P., Owens, K.C.: Lexical co-occurrence and association strength. Journal of Psycholinguistic Research 19(5), 317–330 (1990)

Tissier, J., Gravier, C., Habrard, A.: Dict2vec: Learning word embeddings using lexical dictionaries. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. pp. 254–263 (2017)