Nagyszótáras beszédfelismerés morfémaalapú rekurrens nyelvi modell használatával

Grósz Tamás

Aalto University, Finland
tamas.grosz@aalto.fi

Kivonat A klasszikus beszédfelismerő rendszerek számára hatalmas kihívást jelentenek az agglutináló nyelvek, hiszen pontos eredmények eléréséhez hatalmas szótárakra van szükség a ragozás és a szóösszetétel miatt. A probléma főleg a nyelvi modell részét érinti a felismerőnek, tekintve, hogy túl nagy szótárméret esetén a tanulási fázis rendkívül nehéz, ez pedig szuboptimális modellhez vezethet. Ezen problémára megoldást jelenthet, ha szavak helyett azoknál kisebb egységet, morfémákat használunk a nyelvi modellezés során. A cikkben bemutatásra kerül egy morfémaalapú, rekurrens neuronhálós nyelvi modellt alkalmazó beszédfelismerő, amely használatával szignifikánsan jobb eredményeket tudtunk elérni egy magyar nyelvű beszédkorpuszon mint a hagyományos szószintű megközelítéssel.
Kulcsszavak: beszédfelismerés, nyelvi modell, morféma, rekurrens neuronháló

1. Bevezetés

Az elmúlt pár évben elfogadott ténnyé vált, hogy mély neuronhálós akusztikus és nyelvi modellekkel lehet elérni a legjobb beszédfelismerési pontosságot (Hinton és mtsai, 2012). Ezen új beszédfelismerő rendszerek többsége a nyelvi modell építése során szavakat használ építőelemként, ami angol nyelv esetén jól működik, azonban komoly problémát okoz agglutináló nyelvek esetében.

A legnagyobb problémát a szóalaki változatosság okozza, amely egy fontos jellemzője a morfológiailag gazdag nyelveknek. Sok szóalak esetén rendkívül nagy méretű szótárat kell használnunk, hogy elfogadható pontosságot tudjunk elérni, ez pedig megnehezíti a nyelvi modell tanítását, mivel nagy szótár esetén viszonylag kevés tanítóminta áll rendelkezésünkre osztályonként.

Megoldásként módosíthatjuk a nyelvi modellünket, hogy szavak helyett azoknál kisebb egységeket használjon. Egy ilyen lehetséges egység a morféma, amit korábban már sikeresen használtak finn és magyar nyelvű beszédfelismerőkben. Extrém esetben átválthatunk akár karakter szintű nyelvi modellre is, az ún. end-to-end beszédfelismerő rendszerek jelentős része ezt a megoldást használja. Mindkét megközelítés esetén számottevően csökken a szótárméret, ezáltal könnyebbé válik a nyelvi modell tanítása. Munkánkban mi a morfémaalapú megközelítést vizsgáltuk.

Cikkünkben egy általános módszert mutatunk be, amelynek segítségével morfémaalapú beszédfelismerő rendszereket tanítunk magyar nyelvű híradós adatbázison. A felismerőnk akusztikus modellként egy modern mély neuronháló struktúrát alkalmaz, nyelvi modell oldalon pedig a hagyományos n-gram megközelítést hasonlítjuk össze mély rekurrens hálókkal. Eredményeink alapján kijelenthetjük, hogy a morfémaalapú nyelvi modell használatával nem csak a szótár méretét csökkentettük, de a felismerés pontosságot is szignifikánsan javítottuk.

2. Kapcsolódó irodalom

Morfémaalapú rendszer esetén első lépésként szegmentálnunk (a szavakat morfémákra bontani) kell a tanítóadatunkat, ezt többféle módon is megtehetjük. A szegmentáláshoz használhatunk nyelvspecifikus szabályokon és szótáron alapuló módszert, például a HunMorph (Trón és mtsai, 2005) rendszer alkalmazásával.

Alternatívaként használhatunk statisztikai szegmentáló eljárást is, ennek előnye, hogy nem igényel semmilyen külső tudást, a rendelkezésére álló szöveget felhasználva keres egy optimális felbontást. Ezen módszerek közül mi a Morfessor Baseline (Creutz és Lagus, 2002) eljárást használtuk, amely egy Minimum Description Length (MDL) elven működő módszer. Célja, hogy felügyelet nélkül létrehozzon egy optimális lexikont, amely segítségével szegmentálható a tanító szöveg.

Magyar nyelvű beszédfelismerésen belül morfémaalapú nyelvi modell használatával már több mű is foglalkozott (Mihajlik és mtsai, 2007; Németh és mtsai, 2007; Tarján és mtsai, 2009; Tarján és mtsai, 2014), melyek több lehetséges szegmentálási módszert hasonlítanak össze. Eredményeikből megállapítható, hogy a Morfessor Baseline módszer képes hatékonyan szegmentálni magyar nyelvű szövegeket. Az eddigi munkákban közös, hogy nyelvi modellként a hagyományos n-gram módszert alkalmazták, ezzel ellentétben mi mély rekurrens neuronhálókat is alkalmaztunk kísérleteink során.

A közelmúltban megmutatták, hogy más nyelveken (finn és észt) is számottevő javulások érhetőek el automatikusan konstruált morféma szintű nyelvi modell használatával (Smit és mtsai, 2017). A javasolt eljárásukban a Morfessor Baseline-t alkalmazták a szegmentálási lépés során, majd n-gram modelleket hasonlítottak össze rekurrens neuronhálókkal, vizsgálataink során mi is ezt a módszert követtük.

3. Morfémák szegmentálása

Szavak szegmentálása során célunk meghatározni, hogy az egyes szak mely morfémákból épülnek fel. A feladat elvégzésére alkalmazhatunk nyelvspecifikus szabályalapú rendszereket vagy automatikus módszereket, esetleg ezek kombinációját. Fontos megjegyezni, hogy mi az automatikus módszerekre fókuszáltunk, az általuk javasolt egységek azonban nyelvészeti szempontból nem feltétlenül tekinthetőek morfémáknak, de az egyszerűség kedvéért mi morfémaként fogunk ezekre az egységekre hivatkozni.

Az itt alkalmazott Morfessor Baseline algoritmus a felügyelet nélküli módszerek családjába tartozik. Tanítás során egy mohó, lokális keresést hajt végre az optimális morféma lexikon meghatározásához, amely a következő hibafüggvény optimalizálja:

$$L(\Theta, D_w) = -logp(\Theta) - \alpha logp(D_w|\Theta), \tag{1}$$

ahol Θ a modell paraméterei, D_w a tanító adat, α pedig a hibafüggvény paramétere. A prior valószínűség ($p(\Theta)$) kizárólag a lexikontól függ, számítása MDL alapú módszerrel történik (Virpioja és mtsai, 2013). Az adat likelihood valószínűségét a tanító adatbázisban található szavak aktuális analízise ($Y = (y_1 \dots y_N)$) alapján becsülhetjük;

$$p(D_w|\Theta) = \sum_{j=1}^{N} logp(w_b) \sum_{i=1}^{|y_j|} logp(m_{ji}|\Theta), \tag{2}$$

ahol m_{ji} a j-edik szó felbontásának i-edik morfémája, w_b pedig a szavak közötti határoló szimbólum. Az α paraméter segítségével tudjuk kontrollálni a lexikonban található morfémák számát, kicsi érték esetén a prior lesz a meghatározó tag, így az optimalizáló próbál minél kisebb lexikont létrehozni. Nagy α érték esetén a likelihood lesz a domináns, ami miatt a modell hosszú morfémákat preferál, ez pedig nagyobb lexikont eredményez.

A tanítás kezdetén az összes szó, amely előfordul a tanító adatbázisban bekerül a lexikonba, majd az algoritmus kiválaszt ezek közül egyet, amelynek megkeresi az optimális felbontását a 1. képlet alapján. Az algoritmus ez után iteratívan folytatja a felbontások keresését, amíg egy optimális lexikont nem kap.

A tanítási lépés után a dekódolási lépés következik, amikor is szavakat próbálunk morfémákra bontani, a legvalószínűbb felbontás meghatározására a Viterbi algoritmust használhatjuk.

Kísérleteink során a Morfessor-2.0 (Virpioja és mtsai, 2013) szoftvert használtuk a szegmentáló modell létrehozására. Az egyszerűség kedvéért csak a szegmentálás végrehajtása után, a nyelvi modell tanítás során különböztettük meg a prefix, szuffix és közbülső morfémákat. A 1. táblázat egy példa mondat szegmentálását tartalmazza. Megfigyelhető, hogy az α értékének csökkenésével egyre kisebb egységekre bontja a modell a szavakat.

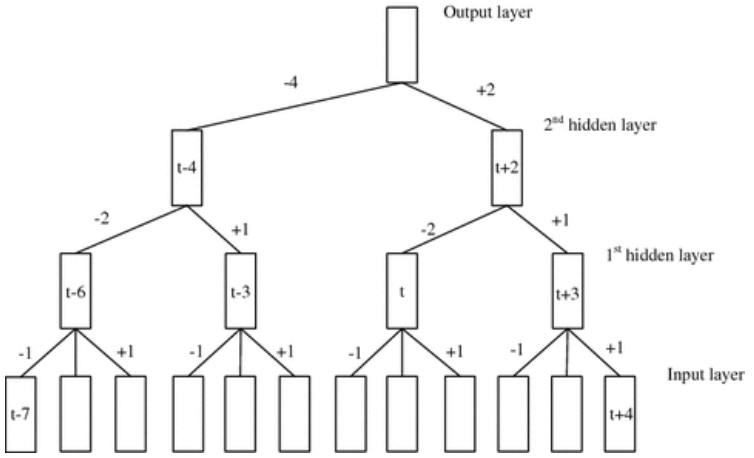
α	szegmentált példamondat
0.1	közösség+ +ét minden oldalról fenyegető veszélyeket
0.01	közösség+ +ét minden oldalról fenyegető veszély+ +eket
0.001	közös+ +ség+ +ét minden oldal+ +ról fenyeget+ +ő veszély+ +eket

1. táblázat. Példa szegmentálásra különböző α paraméterek esetén.

4. Akusztikus modell

Egy standard akusztikus modell feladata, hogy a bementi spektrális jellemzők alapján megbecsülje az egyes fonémák valószínűségét. Tanítás során a kiejtési szótár segítségével határozzuk meg az egyes szakhoz tartozó fonémákat, ez a megközelítés sajnos esetünkben nem alkalmazható, mivel a nyelvi modellünk morféma szinten működik. A problémát az okozza, hogy minden morfémához definiálnunk kellene annak kiejtését a kontextus (környező morfémák) ismerete nélkül. Szerencsére a probléma viszonylag könnyen kezelhető, amennyiben fonémák helyett grafémákat használunk akusztikus egységként, ebben az esetben a kiejtési szótár könnyen generálható.

Kísérleteinkben graféma alapú akusztikus modelleket használtunk, amelyeket a Kaldi (Povey és mtsai, 2011) rendszer segítségével tanítottunk. Végső modellként egy időkésleltett neuronhálót (time-delay neural network, TDNN) (Peddinti és mtsai, 2015) használtunk, amelyet lattice-mentes maximális kölcsönös információ (lattice-free maximum mutual information)(Povey és mtsai, 2016) módszerrel tanítottunk.



1. ábra: Egy három réteges TDNN neuronháló struktúrája.

A TDNN hálók specialitása, hogy rejtett rétegeik időbeli konvolúciót végeznek, az első rejtett réteg csak egy kis időbeli kontextust dolgoz fel, a későbbi rétegek pedig egyre nagyobb időablakot fednek le a korábbi rejtett rétegek segítségével. Működését a 4. ábra szemlélteti. Tanításuk során a Kaldi keretrendszerben elérhető ún. chain receptet követtük. A neuronháló 10 rejtett réteget tartalmazott, amelyek mindegyike 1000 darab relu aktivációs függvényt alkalmazó neuronból állt. Bemenetként standard MFCC jellemzővektorokat használtunk, összesen 13 koefficienst illetve azok Δ-ját és $\Delta\Delta$-ját.

5. Nyelvi modell

Tradicionálisan nyelvi modellezésre az ú.n. n-gram modelleket szokás használni, amelyek az előző n – 1 darab szó alapján becsülik a következő szó valószínűségét. Ezen modellek tanítása során a szükséges statisztikákat a rendelkezésre álló szövegből számítjuk. A pontosabb eredmények elérése érdekében több finomítása is létezik a módszernek, mi ezek közül a Kneser-Ney simítást alkalmaztuk a VariKN (Siivola és mtsai, 2007) rendszer használatával. Kísérleteink során a hagyományos 3-gram modellek mellett számottevően nagyobb n-gram-okat is felhasználtunk, abban bízva, hogy morfémaalapú modellek esetén hasznos lehet a nagyobb kontextus használata.

A hagyományos n-gram megközelítés mellett a manapság nagy népszerűségnek örvendő rekurrens neuronhálókat is kipróbáltuk. Az utóbbi években a rekurrens neuronhálók kiemelkedően jó eredményeket értek el természetes nyelvi feldolgozásban. Beszédfelismerésben a rövid- és hosszú-távú memória cellákat (long short-term memory, LSTM) alkalmazó változatuk terjedt el leginkább(Young és mtsai, 2018). A legfőbb különbség a hagyományos rekurrens neuron és az LSTM cella között, hogy utóbbi nem csak a korábbi kimenetét kapja meg bemenetként, hanem rendelkezik egy belső állapottal is, amely a hosszú-távú emlékezésben segít.

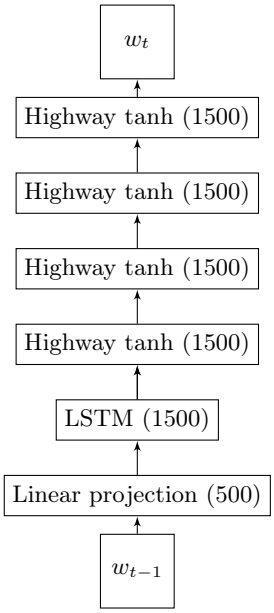
Formálisan, egy bemeneti vektor (x_{t-1}) esetén egy LSTM cella első lépésben a következő számításokat végzi:

$$\begin{aligned} f_t &= \sigma(W_f x_{t-1} + U_f h_{t-1} + b_f) \\ i_t &= \sigma(W_i x_{t-1} + U_i h_{t-1} + b_i) \\ o_t &= \sigma(W_o x_{t-1} + U_o h_{t-1} + b_o), \end{aligned} \tag{3}$$

ahol h_{t-1} az előző kimenet, σ pedig a sigmoid függvény. A kiszámított bemeneti (i_t), kimeneti (o_t) és felejtő (f_t) kapuk értékei alapján pedig a végső kimenet (h_t) illetve a belső memória (c_t) új értéke kerül meghatározásra;

$$\begin{aligned} c_t &= f_t c_{t-1} + i_t tanh(W_c x_{t-1} + U_c h_{t-1} + b_c) \\ h_t &= o_t tanh(c_t) \end{aligned} \tag{4}$$

Munkánkban a nyelvi modellként használt neuronhálók struktúráját a 2 ábra szemlélteti. Első lépésben a bementet egy projekciós réteg dolgozza fel, amely a beágyazást (embedding) végzi, ezt a réteget nem tanítottuk külön, a tanítás elején véletlenszerűen inicializáltuk. A beágyazó réteg után következik az LSTM réteg, ami a belső memória segítségével próbál információt tárolni a korábbi szavakról vagy morfémákról, majd négy highway réteg dolgozza fel ennek kimenetét. A highway rétegek lényege, hogy kimenetük az eredeti bemenet és a rejtett neuronok kimenetének lineáris kombinációja, ez megkönnyíti a gradiens propagálását tanítás során, ami pedig lehetővé teszi, hogy sok rejtett réteget használjuk hatékonyan. A lehetséges következő szavak valószínűségeit egy softmax réteg segítségével becsüljük, a neuronhálók tanításhoz a TheanoLM (Enarvi és Kurimo, 2016) keretrendszert használtuk.



2. ábra: A kísérleteink során használt rekurrens nyelvi modell felépítése.

5.1. Kiértékelés neuronhálós nyelvi modellel

A felismerési folyamat során sajnos nem realisztikus egyből a neronhálós nyelvi modellt használni, hiszen ismert, hogy a dekódolás keresési tere exponenciálisan növekszik a hipotézis hosszával, ez pedig lelassítja a rendszert. További ellenérv, hogy a neuronhálók kiértékelése számottevően több időt igényel mint egy egyszerűbb n-gram használata. Ezen problémára több megoldás is létezik, az egy lehetőség, hogy a neuronháló felhasználásával szöveget generálunk, melyből hagyományos n-gram modellt tanítunk és ezt használjuk a felismerés során (Mittul és mtsai, 2018; Tarján és mtsai, 2019), így ugyan veszítünk némi információt, de lehetőségünk van gyors, akár online dekódolásra is.

Talán a legelterjedtebb megoldás mégis a kétkörös dekódolás (two pass decoding). Ekkor első körben egy egyszerű n-gram nyelvi modell (tipikusan 3-gram) segítségével ún. lattice-t hozunk létre, majd a második körben újrasúlyozzuk (rescore) a felismerési hipotéziseket a lattice-ben a neuronháló kimenetei alapján. Kísérleteinkben mi is ezt a megközelítést alkalmaztuk, hiszen így tisztább képet kaphatunk a neuronhálók pontosságáról.

Alternatívaként használhatunk n-legjobb listákat (n-best list) (Deoras és mtsai, 2011), azonban kezdeti kísérleteink alapján ez a megközelítés rosszabb eredményeket ad mint a kétkörös módszer. Megemlítenénk, hogy közelmúltban megjelentek új módszerek, amelyek képesek a dekódolást csak neuronhálós nyelvi modellel hatékonyan végrehajtani (Jorge és mtsai, 2019), sajnos ezt a megközelítést nem volt időnk tesztelni.

Nyelvi modell egysége	Szótár méret teszt	OOV ráta
Szó	420520	9.9%
Morf. α=0.1	183803	0.5%
Morf. α=0.01	53667	0.3%
Morf. α=0.001	11562	0.2%

2. táblázat. Tanító adatbázis statisztikái.

6. Tanító adatbázisok

Az akusztikus modellek tanítására az Origo korpuszt használtuk, amely összesen 2.7 millió mondatot tartalmaz, a szóalakok száma pedig meghaladja az 50 milliót. A Morfessor modellek tanítása előtt véletlenszerűen kiválasztottunk 10000 mondatot, ezeket validációs halmazként használtuk.

Az akusztikus modell tanításához egy magyar nyelvű híradós adatbázist (Tóth és Grósz, 2013) használtunk, amely megközelítőleg 30 órányi beszédanyagot tartalmazott, ebből 2 órányit használtunk validációs, 4 órányit pedig teszt halmazként.

7. Eredmények

Első lépésben a szószintű és a morféma szegmentálással kapott szótárakat hasonlítottuk össze (2. táblázat). Ezek létrehozása során kizárólag a szöveges adatbázist használtuk (az akusztikus tanítóadat átirata nem lett hozzáadva a tanítóadathoz). A szószintű megközelítés esetén a VariKN rendszert használtuk a szótár létrehozására, a kiválasztott nagyjából 420000 szavas szótár a szöveges tanítóadat leggyakoribb szavaiból lett kiválasztva, ez az akusztikus teszthalmazban található szavak 9.9%-át nem tartalmazza. Természetesen nagyobb szótár esetén ez az arány csökkenthető, ám ekkor a nyelvi modell mérete számottevően megugrik, különösen a nagy n-gram esetén.

Morfémaalapú megközelítések esetén látható, hogy sokkal kisebb szótárral is sokkal jobban le tudjuk fedni a teszt adatot, ezzel lehetővé téve a pontosabb felismerést. Ahogy egyre jobban csökken a lexikon mérete (annak eredményeként, hogy a prior tagra koncentrál a szegmentáló algoritmus), egyre kevesebb szót találunk a teszt halmazban, amit nem tudunk a morfémákkal lefedni (out-of-vocabulary, OOV arány). Természetesen a kisebb szótár azt is jelenti, hogy egyre kisebb egységekre bontjuk az egyes szavakat, ami nem feltétlenül előnyös a nyelvi modell számára.

Vizsgálataink során három különböző nyelvi modellt alkalmaztunk, a felismerés első fázisát mindig a 3-gram modellel végeztük. A második körben pedig egy nagy n-gram modellt illetve a neuronhálós rendszerünket használtuk. Az összehasonlításokhoz a szóhiba-arány (word error rate, WER) metrikát használtuk, a morféma alapú felismerő kiértékelésénél a WER ugyanazt a szószintet jelenti-e, mint a szóalapúnál. A szóalakok rekonstrukciójához a felismerés végén a morfémákat a '+' határoló jelzés esetén összevontuk.

Nyelvi modell egysége	Nyelvi modell típusa	Validációs halmaz	Teszt
Szószintű	VariKN (3-gram)	20.91%	19.73%
	VariKN (16-gram)	20.95%	19.65%
	LSTM	19.30%	17.98%
Morfessor α=0.1	VariKN (3-gram)	17.60%	16.17%
	VariKN (16-gram)	17.48%	16.17%
	LSTM	16.69%	15.29%
Morfessor α=0.01	VariKN (3-gram)	18.92%	17.48%
	VariKN (21-gram)	19.00%	17.49%
	LSTM	15.28%	14.09%
Morfessor α=0.001	VariKN (3-gram)	19.18%	18.00%
	VariKN (24-gram)	18.70%	17.44%
	LSTM	15.69%	14.41%

3. táblázat. Beszédfelismerési eredmények.

A 3. táblázatban láthatóak a különböző megközelítésekkel elért eredményeink. A szószintű rendszereket tekintve megállapítható, hogy nagy méretű (16-gram) modell használata nem javít a felismerés pontosságán, a neuronhálós megoldás viszont szignifikánsan jobb eredményt képes produkálni, mint amit n-gram használatával el tudunk érni. Ez utóbbi megfigyelés a morfémaalapú rendszerek esetén is igaz. Morfémákat alkalmazó felismerők minden esetben jobban teljesítettek mint a hagyományos szószintűek, így megállapíthatjuk, hogy magyar nyelvű beszéd esetén célszerű használatuk.

Érdekességként megfigyelhető, hogy kicsi α esetén, amikor is a szavakat sok kicsi egységre bontjuk, akkor a 23-gram modell már jobban teljesít mint a sima 3-gram. Ennek magyarázata abban keresendő, hogy ekkor már fontos a nagy kontextus használata, hiszen a 3-gram használatával előfordulhat, hogy hosszabb szavakat (amik több mint 3 morfémára lettek bontva) nem tudunk lefedni és így semmi információval nem rendelkezünk a korábbi szavakról.

A legjobb eredményeket neuronhálós nyelvi modellel értük el $\alpha = 0.01$ használatával. Ekkor 3.9% javulást láthatunk a szószintű változathoz hasonlítva, ami közel 22%-os relatív javulást jelent. A magyarázat arra, hogy miért pont ez a szegmentálás bizonyult legjobbnak az lehet, hogy ekkor már kellően lecsökkent a szótár mérete ahhoz, hogy hatékonyan tudjon a neuronháló tanulni és a szavakat nem bontottuk túl sok egységre, így nem jelentet túl nagy kihívást a korábbi morfémákra való "emlékezés" sem.

Megfigyelhető továbbá, hogy egyre kisebb morfémaszótár esetén az n-gramok egyre rosszabb eredményt értek el. Ebből arra lehet következtetni, hogy ezen modellek a nagy méretű morfémákat preferálják, ami nagy szótárat eredményez.

8. Konklúzió

Cikkünkben morfémaalapú rekurrens nyelvi modelleket alkalmazó beszédfelismerők teljesítményét vizsgáltunk egy magyar nyelvű korpuszon. Megállapítható,

hogy a szavak felbontása morfémákra megkönnyíti a nyelvi modell feladatát, így pontosabb felismerő rendszereket taníthatunk. A morfémákat alkalmazó modellek előnye a szószintűekkel szemben két fő tényezőnek köszönhető, egyrészről a lényegesen kisebb felismerési szótárnak, másrészről pedig annak, hogy morfémák segítségével lényegesen több szót tudunk felépíteni így csökkentve az OOV rátát. Fontos azonban megtalálni az egyensúlyt a szótár és a morfémák mérete között, hiszen a túl kicsi egységekre bontás ugyan lényegesen csökkenti a lexikon méretét, de nehezebbé is teszi a pontos modell tanítását.

Eredményeink alapján az is nyilvánvaló, hogy a hagyományos n-gram modelleknél számottevően jobban teljesítenek a neuronhálót alkalmazók, ahogy ezt már több korábbi munka is igazolta. További kutatásaink során a neuronhálós nyelvi modell továbbfejlesztésére tervezünk fókuszálni. Érdekes kérdés például, hogy vajon a szószintű modellek esetén rendkívül jól teljesítő figyelem (attention) mechanizmus (Bahdanau és mtsai, 2015) vajon morfémaalapú rendszer esetén is hasznos-e?

Hivatkozások

Bahdanau, D., Cho, K., Bengio, Y.: Neural machine translation by jointly learning to align and translate. In: 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings (2015), http://arxiv.org/abs/1409.0473

Creutz, M., Lagus, K.: Unsupervised discovery of morphemes. In: Proceedings of the ACL-02 Workshop on Morphological and Phonological Learning - Volume 6. pp. 21–30. MPL '02, Association for Computational Linguistics, Stroudsburg, PA, USA (2002), https://doi.org/10.3115/1118647.1118650

Deoras, A., Mikolov, T., Church, K.: A fast re-scoring strategy to capture long-distance dependencies. In: Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing. pp. 1116–1127. Association for Computational Linguistics, Edinburgh, Scotland, UK. (Jul 2011), https://www.aclweb.org/anthology/D11-1103

Enarvi, S., Kurimo, M.: TheanoLM — An Extensible Toolkit for Neural Network Language Modeling. In: Interspeech 2016. pp. 3052–3056 (2016), http://dx.doi.org/10.21437/Interspeech.2016-618

Hinton, G., Deng, L., Yu, D., Dahl, G.E., Mohamed, A.r., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T.N., és mtsai: Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. IEEE Signal processing magazine 29(6), 82–97 (2012)

Jorge, J., Giménez, A., Iranzo-Sánchez, J., Civera, J., Sanchis, A., Juan, A.: Real-Time One-Pass Decoder for Speech Recognition Using LSTM Language Models. In: Proc. Interspeech 2019. pp. 3820–3824 (2019)

Mihajlik, P., Fegyó, T., Tüske, Z., Ircing, P.: A Morpho-graphemic Approach for the Recognition of Spontaneous Speech in Agglutinative Languages - like Hungarian. In: Interspeech 2007. pp. 1497–1500 (2007)

Mittul, S., Peter, S., Sami, V., Mikko, K.: First-pass decoding with n-gram approximation of RNNLM: The problem of rare words. In: Machine Learning in Speech and Language Processing Workshop (2018)

Németh, B., Mihajlik, P., Tikk, D., Trón, V.: Statisztikai és szabály alapú morfológiai elemzők kombinációja beszédfelismerő alkalmazáshoz. In: Magyar Számítógépes Nyelvészeti Konferencia. pp. 95–105 (2007)

Peddinti, V., Povey, D., Khudanpur, S.: A time delay neural network architecture for efficient modeling of long temporal contexts. In: INTERSPEECH (2015)

Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., Hannemann, M., Motlicek, P., Qian, Y., Schwarz, P., Silovsky, J., Stemmer, G., Vesely, K.: The Kaldi Speech Recognition Toolkit. In: IEEE 2011 Workshop on Automatic Speech Recognition and Understanding. IEEE Signal Processing Society (Dec 2011), iEEE Catalog No.: CFP11SRW-USB

Povey, D., Peddinti, V., Galvez, D., Ghahremani, P., Manohar, V., Na, X., Wang, Y., Khudanpur, S.: Purely Sequence-Trained Neural Networks for ASR Based on Lattice-Free MMI. In: INTERSPEECH (2016)

Siivola, V., Creutz, M., Kurimo, M.: Morfessor and VariKN machine learning tools for speech and language technology. In: INTERSPEECH. pp. 1549–1552. ISCA (2007)

Smit, P., Virpioja, S., Kurimo, M.: Improved subword modeling for wfst-based speech recognition. In: Proc. Interspeech 2017. pp. 2551–2555 (2017)

Tarján, B., Fegyó, T., Mihajlik, P.: A bilingual study on the prediction of morph-based improvement. In: Spoken Language Technologies for Under-Resourced Languages (2014)

Tarján, B., Fegyó, T., Mihajlik, P.: Ügyfélszolgálati beszélgetések nyelvmodellezéserekurrens neurális hálózatokkal. In: Magyar Számítógépes Nyelvészeti Konferencia. pp. 23–33 (2019)

Tarján, B., Mihajlik, P., Tüske, Z.: Nagyszótáras híranyagok felismerési pontosságának növelése morfémaalapú, folyamatos beszédfelismerővel. In: Magyar Számítógépes Nyelvészeti Konferencia. pp. 185–194 (2009)

Tóth, L., Grósz, T.: A comparison of deep neural network training methods for large vocabulary speech recognition. In: Text, Speech, and Dialogue. pp. 36–43. Springer Berlin Heidelberg (2013)

Trón, V., Gyepesi, Gy., Halácsy, P., Kornai, A., Németh, L., Varga, D.: Hunmorph: Open source word analysis. In: Proceedings of Workshop on Software. pp. 77–85. Association for Computational Linguistics, Ann Arbor, Michigan (Jun 2005), https://www.aclweb.org/anthology/W05-1106

Virpioja, S., Smit, P., Grönroos, S.A., Kurimo, M.: Morfessor 2.0: Python Implementation and Extensions for Morfessor Baseline. D4 julkaistu kehittämistai tutkimusraportti tai -selvitys (2013), http://urn.fi/URN:ISBN:978-952-60-5501-5

Young, T., Hazarika, D., Poria, S., Cambria, E.: Recent Trends in Deep Learning Based Natural Language Processing [Review Article]. IEEE Computational Intelligence Magazine 13(3), 55–75 (Aug 2018)