Nyelvimodell-adaptáció ügyfélszolgálati beszélgetések gépi leiratozásához

Tarján Balázs[1], Mihajlik Péter[1,2], Fegyó Tibor[1,3]

[1] Budapesti Műszaki és Gazdaságtudományi Egyetem,
Távközlési és Médiainformatikai Tanszék
{tarjanb, mihajlik, fegyo}@tmit.bme.hu
[2] THINKTech Kutatási Központ Nonprofit Kft.
[3] AITIA International Zrt.

Kivonat: A folyamatos nagyszótáras gépi beszédfelismerés kritikus eleme a statisztikai nyelvi modell, melynek betanításához feladatspecifikus (in-domain) tanítóadatra van szükség. Ilyen tanítóadat azonban a gyakorlatban csak korlátozott mennyiségben áll rendelkezésre, mely felveti a feladattól független vagy ellenőrizetlen (out-of-domain) tanítószövegek felhasználását is. Formálisan nyelvi modell adaptáció révén építhető be az addicionális tanítószövegben tárolt tudás a feladatspecifikus nyelvi modellekbe. Cikkünkben azt vizsgáltuk, hogy telefonos ügyfélszolgálati hanganyagok felismerési pontossága javítható-e a különféle nyelvimodell-adaptációs technikákkal. Kísérleteink szerint mind felügyelt, mind felügyelet nélküli nyelvimodell-adaptációval szignifikánsan növelhető a valós beszélgetéseket leiratozó rendszerek pontossága.

1 Bevezetés

A jelenleg elterjedt nagyszótáras beszédfelismerők statisztikai úton tanított **nyelvi modell**t használnak, így a modell pontosságát döntően befolyásolja, hogy milyen mennyiségű és minőségű tanítószöveg áll rendelkezésünkre. Jó minőségű tanítószöveg általában a felismerési feladathoz illeszkedő hanganyagok kézi leirataiból állítható elő (**in-domain tanítószöveg**). A gyakorlatban azonban a begyűjthető hanganyagok mennyisége és a kézi leiratozás költségei határt szabnak az ilyen úton nyerhető tanítószöveg méretének. Éppen ezért a tudományos közösséget régóta foglalkoztatja, hogyan lehet az akusztikus modellek adaptációjához hasonlóan egy feladattól független (**out-of-domain**), de robosztus nyelvi modellt egy in-domain, de elégtelen mennyiségű adaton tanított modellhez adaptálni.

Cikkünkben különböző méretű és feladatunkhoz különböző mértékben illeszkedő tanítószövegek alapján készült nyelvi modelleket kísérelünk meg adaptálni ügyfélszolgálati beszélgetések felismerésre készített rendszerünkhöz. Megmutatjuk, hogy milyen módon célszerű eljárni, ha kisméretű, de a feladathoz jól illeszkedő kiegészítő szöveghez jutunk, illetve ha egy több tízmillió szót tartalmazó webkorpuszt szeretnénk felhasználni az in-domain modell javítására. **Felügyelt** adaptáció mellett **felügyelet nélküli** adaptációs kísérleteket is végzünk, azaz megvizsgáljuk, hogyan

használhatóak fel a felismerés korábbi kimenetei a nyelvi modell további pontosítására.

A nyelvimodell-adaptációs technikáknak alapvetően két nagy ágát kell megkülönböztetnünk [2]. Az első módszer az ún. maximum a posteriori (MAP) becslésen alapszik [4], és a célja, hogy úgy változtassa meg az out-of-domain modell paramétereit, hogy azok az in-domain modell paramétereinek eloszlását kövessék. A másik adaptációs megközelítésnél az objektív cél az, hogy az out-of-domain nyelvi modell minél kevesebb felismerési hibát vétsen egy kijelölt in-domain tesztanyagon. Itt a paraméterek hangolása diszkriminatív tanítás útján történik. A két megközelítés közül a MAP-adaptáció sok esetben jobban teljesít [2], mint a diszkriminatív tanítás, emellett a megvalósítása is egyszerűbb, így kísérleteinkben ezt módszert alkalmaztuk. A felügyelet nélküli adaptáció hatékonyabbá tehető, ha konfidenciaadatok alapján súlyozzuk vagy szűrjük a felismerési kimeneteket [5], azonban a rendelkezésünkre álló felismerési leiratok nem tartalmaztak megbízhatósági mértéket, így a felügyelet nélküli adaptáció esetén is csakúgy, mint a felügyelt esetben egy más típusú válogatási eljárást alkalmaztuk, melyet a cikkünk későbbi részében ismertetünk.

A következőkben először a kísérletekhez használt tanító és tesztadatbázisokat ismertetjük, majd kitérünk a modellek tanításnál és adaptálásánál alkalmazott módszerekre. A felismerési feladat és módszertan bemutatása után ismertetjük a különböző adaptációs megközelítésekkel kapott eredményeket, míg végül összefoglalását adjuk kísérleteink legfontosabb következményeinek.

2 Tanító és tesztadatbázisok

2.1 Tanító adatbázisok

Két ügyfélszolgálati rendszer in-domain nyelvi modelljének javítását tűztük ki kísérleteink céljaként, melyekre a továbbiakban **MTUBA** (Magyar Telefonos Ügyfélszolgálati Beszédadatbázis) I., illetve II. néven fogunk hivatkozni. Az **MTUBA I.** rendszernél az in-domain modell tanításához egy összesen 380 ezer szavas, kézi leiratokat tartalmazó tanítószöveg állt rendelkezésünkre. Az **MTUBA II.** feladatnál valamivel kisebb, összesen 280 ezer szavas kézi leiratot használhattunk. A felügyelet nélküli adaptációs kísérletekhez további két korpuszt gyűjtöttünk, melyek az egyes rendszerek felismerési kimeneteit tartalmazzák.

Az adaptációs kísérletekhez szükségünk volt egy a feladatokhoz semmilyen módon nem kötődő, out-of-domain korpuszra is. Ideális választásnak tűnt erre a célra a **Magyar Webkorpusz** [6]. Óriási mérete miatt csak a webkorpusz egy tizedét használtuk, mely önmagában 100 millió szót jelent, így elegendően nagynak bizonyult vizsgálatainkhoz. Az eredmények könnyebb értelmezhetősége érdekében egy mind méretében, mind illeszkedésében az in-domain és az out-of-domain korpuszok között elhelyezkedő kiegészítő tanítószöveget is szerettünk volna találni. Erre a megoldást egy ügyfélszolgálati levelezéseket tartalmazó, összesen 1,8 millió szavas korpusz jelentette. Ez az **e-mail korpusz** az in-domain szövegekhez hasonlóan ügyfélszolgálati témájú, így a webkorpusznál jobban illeszkedik a feladathoz, azonban szigorúan véve nem tekinthető in-domain tanítóanyagnak sem, ugyanis a

valódi beszélgetések leiratai sokkal több spontán elemet tartalmaznak, mint az elektronikus levelezés.

1. táblázat: A szöveges tanító adatbázisok méretei

	In-domain		**Felismerési kimenet**		**Kiegészítő korpusz**	
	MTUBA I.	MTUBA II.	MTUBA I.	MTUBA II.	E-mail korpusz	Web-korpusz
Méret [millió szó]	0,38	0,28	32	5,3	1,8	100

2.2 Tesztadatbázisok

A változatos nyelvimodell-konfigurációk kiértékeléséhez minden esetben a tanítóanyagoktól független tesztfelvételeket használtunk. Az MTUBA II. adatbázison több mint 5 órányi felvételt tudtunk tesztelési célokra elkülöníteni, mely megbízható kiértékelést tesz lehetővé, így tesztjeink többségét ezen végeztük. Annak érdekében, hogy minden esetben garantáljuk a független tanítást és tesztelést, egy másik, összesen 2 órás tesztanyagot is definiálnunk kellett az MTUBA II. adatbázison, melynek részletes okaira az 4.2.1 fejezetben térünk ki. Az MTUBA I. adatbázison egy kb. 1 órás tesztanyagot jelöltünk ki, melyen felügyelet nélküli adaptációval kapcsolatos kísérletet végeztünk.

2. táblázat: A teszt adatbázisok jellemzői

	Hossz [min]	Szavak száma [ezer szó]
MTUBA I.	56	5,7
MTUBA II.-5h	300	35
MTUBA II.-2h	120	14

3 Módszertan

3.1 Nyelvimodell-adaptáció

Kísérleteinkben a MAP becslésen alapuló nyelvimodell-adaptáció egy-egy speciális esetét jelentő **korpuszegyesítéses** (count merging) és **nyelvimodell-interpolációs** eljárásokat alkalmaztuk [1]. Két szöveges tudásforrás egyesítésének legegyszerűbb módja, ha n-gram statisztikájukat egyesítjük, és ez alapján készítjük el az n-gram nyelvi modellt. Gyakorlatban ez a két tanítószöveg összemásolásával vitelezhető ki a legegyszerűbben. Ez az eljárás jól működhet, ha hasonló mértékben illeszkedő tanítószövegeket egyesítünk. Abban az esetben azonban, ha egy out-of-domain tanítószöveget szeretnénk egy in-domain tanítószöveghez adaptálni, a korpuszegyesítéssel aránytalanul nagy súllyal kerülhetnek az egyesített modellbe a feladathoz rosszul illeszkedő tanítószöveg n-gram becslései [11]. Ilyenkor

jelenthetnek megoldást az interpolációs eljárások, melyekkel különböző nyelvi modellek n-gram becslései egyesíthetőek tetszőlegesen megválasztott súlyozó tényezővel. Mi az ún. lineáris interpolációt alkalmaztuk [7].

3.2 Perplexitásalapú előválogatás

Nyelvimodell-interpolációval hatékonyan orvosolhatóak az adaptáció során a modellek illeszkedési különbségeiből fakadó problémák. Önmagában használva az adaptáció azonban nem feltétlenül elegendően hatékony. Egy nagyméretű kiegészítő korpusz egyszerre tartalmaz olyan szövegrészeket, melyek a feladatunk szempontjából hasznos n-gramokat hordoznak és olyanokat is, melyek nyugodtan elhagyhatóak lennének. Ha valóban el tudjuk hagyni az adaptáció előtt az adaptálandó nyelvi modellből azokat az n-gramokat, melyek nem illeszkednek a feladatunkhoz, két ponton is nyerhetünk. Egyrészt csökkenthető a nyelvi modell mérete, másrészt a szükségtelen tanítóadatok elhagyásával a modell pontossága is nőhet.

A kiegészítő tanítószövegek sorainak előválogatására egy perplexitásalapú eljárást alkalmazunk. Ennek az egyszerű, de hatékony eljárásnak a lényege abban áll, hogy az in-domain nyelvi modell segítségével kiszámítjuk a kiegészítő korpusz minden sorához az illeszkedési mértéket (**perplexitást**). Ezek után kijelölünk egy küszöböt, amely alatti perplexitással rendelkező sorokat megtartjuk, míg a többit eldobjuk. Tehát az eljárás lényegében arra a feltétezésre épít, hogy azok a sorok, melyeket nagy pontossággal képes megjósolni az in-domain modell, potenciálisan tovább erősítik a modellt, míg azon sorok, melyek rosszul jósolhatóak, nem tartoznak szorosan a felismerési témához, így elhagyhatóak a modellből.

A perplexitást kétféle módon szokás számolni. A hagyományos eljárás szerint, az (**1**)-es képletben w_0-al jelölt mondatkezdő szimbólumot és a w_{K+1} mondatzáró szimbólumot is figyelembe vesszük a *P(s)* mondatvalószínűségek számításakor. Az ez alapján számított perplexitást szokás **PPL**-el jelölni.

$$P(s)_{PPL} = \prod_{i=0}^{K+1} P\left(w_i \mid w_{K-1},..., w_{K-(N-1)}\right) \quad (\mathbf{1})$$

Ezzel szemben a **PPL1**-gyel jelölt metrika a mondatvalószínűségek kiszámításakor nem veszi számításba mondatkezdő és mondatzáró karaktereket (**2**). Vizsgálataink során mindkét mérőszámot kipróbáltuk a gyakorlatban. Az erre vonatkozó eredményeket az 4.1.1 fejezet foglalja össze.

$$P(s)_{PPL1} = \prod_{i=1}^{K} P\left(w_i \mid w_{K-1},..., w_{K-(N-1)}\right) \quad (\mathbf{2})$$

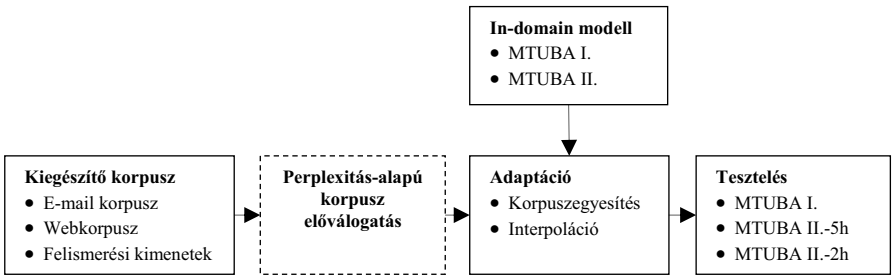
3.3 Tanítás és dekódolás

A vizsgált nyelvi modellek módosított Kneser-Ney simítás [3] használatával készültek az SRI Language Modeling Toolkit (**SRILM**) [10] segítségével. A létrehozott 3-gram, szóalapú modellekben entrópiaalapú metszést egyetlen esetben sem

alkalmaztuk. Interpolált nyelvi modellek készítéséhez és optimalizálásához az SRILM beépített lineáris interpolációs és perplexitásszámító eljárásait használtuk.

Az MTUBA I. feladathoz tartozó akusztikus modell tanításához az erre a célra elkülönített 27 óra, míg az MTUBA II. akusztikus modellhez 38 óra hanganyagot használtuk fel. Az annotált felvételek felhasználásával háromállapotú, balról-jobbra struktúrájú, környezetfüggő rejtett Markov-modelleket tanítottunk a Hidden Markov Model Toolkit [13] eszközeinek segítségével. A létrejött akusztikus modell 4048 egyenként 13 Gauss-függvényből álló állapotot tartalmaz az MTUBA I. modell esetén és 3535 egyenként 16 Gauss-függvényből álló állapotot az MTUBA II. modell esetén. Minden kísérletben a felismerési feladathoz illeszkedő akusztikus modellt használtuk.

A 8 kHz-en mintavételezett, telefonos tesztfelvételek lényegkiemeléséhez 39 dimenziós, delta és delta-delta értékkel kiegészített mel-frekvenciás kepsztrális komponenseken alapuló jellemzővektorokat hoztunk létre, és ún. vak csatornakiegyenlítő eljárást [8] is alkalmaztunk. A súlyozott véges állapotú átalakítókra (**WFST** – Weighted Finite State Transducer) [9] épülő felismerő hálózatok generálását és optimalizálását az Mtool keretrendszer programjaival végeztük, míg a tesztelés során alkalmazott egyutas mintaillesztéshez a VOXerver [12] nevű WFST dekódert használtuk. A felismerő rendszerek teljesítményének értékeléséhez szóhibaarányt (**WER** – Word Error Rate) és karakterhiba-arányt (**LER** – Letter Error Rate) számoltunk, utóbbi gyakran pontosabb képet ad egy felismerő rendszer megbízhatóságáról morfémákban gazdag nyelvek esetén.



1. ábra. Kísérleteink általános módszertani lépései
(a szaggatott vonal opcionális lépést jelöl).

4 Kísérleti eredmények

Ebben a fejezetben a már bemutatott tanító- és tesztadatok felhasználásával, az előző fejezetben ismertetett módszerekkel elért eredményeinket mutatjuk be. Vizsgálataink első felében az MTUBA II. feladat nyelvi modelljéhez kísérelljük meg adaptálni a külső tudásforrásokat, majd a fejezet második felében a felismerési kimenetekkel visszacsatolt felügyelet nélküli adaptációban rejlő lehetőségeket mutatjuk be. Kísérleteink általános módszertani lépéseit az **1. ábra** foglalja össze.

4.1 Felügyelt adaptáció az MTUBA II. nyelvi modellhez

A fejezet során három tudásforrást próbálunk meg adaptálni az MTUBA II. in-domain nyelvi modellhez: nagyméretű, általános tematikájú webkorpuszt, a kisebb méretű, jobban illeszkedő e-mail szövegadatbázist és az MTUBA I. feladat tanítószövegét.

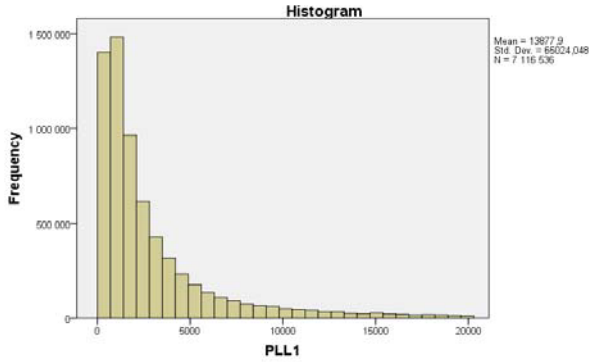
4.1.1 PPL és PPL1 metrika összehasonlítása

Annak eldöntésére, hogy a tanítószövegek sorainak előválogatásához melyik perplexitás-mérőszámot érdemes alkalmazni, terveztünk egy kísérletsorozatot. Első lépésként kerestünk olyan PPL és PPL1 értékpárokat, melyeknél a webkorpuszon végrehajtva a válogatást egyforma méretű tanítószöveget kapunk. A kérdés ezek után úgy módosult, hogy melyik ilyen módon kapott előválogatott tanítószöveggel érhetünk el nagyobb pontosságnövekedést az MTUBA II. felismerési feladaton. Ennek meghatározásához egyesítettük az előválogatott webkorpuszokat az MTUBA II. tanítószövegével, majd az egyesített tanítószövegeken tanítottunk új nyelvi modelleket. Ezután az új nyelvi modellekkel perplexitás- és szótáron kívüli szóarány (OOV – Out of Vocabulary) méréseket hajtottunk végre az MTUBA II.-5h tesztanyagon. A kísérletsorozat eredményeit a **3. táblázatban** foglaltuk össze.

3. táblázat: MTUBA II. in-domain modell és a PPL, valamint PPL1 alapján előválogatott webkorpusz korpuszegyesítéses adaptációjával kapott eredmények az MTUBA II.-5h teszthalmazon kiértékelve.

Válogatási módszer / határ	MTUBA II. tanítószöveg [millió szó]	Kiegészítő webkorpusz [+millió szó]	OOV arány (MTUBA II.-5h) [%]	PPL (MTUBA II.-5h) [-]
PLL-400	0,28	22	1,7	580
PLL1-750			**1,7**	**550**
PPL-200	0,28	7,5	2,1	501
PLL1-400			**2,1**	**454**
PPL-100	0,28	3	2,5	423
PPL1-260			**2,6**	**373**
PPL-50	0,28	1,5	2,9	357
PPL1-200			**2,9**	**320**

A 3. táblázat alapján azt mondhatjuk, hogy azonos kiegészítő korpusz méret mellett a PPL1 metrika segítségével előválogatott webkorpusz nagyobb mértékben járul hozzá az in-domain modell pontosításához. Ez abból olvasható ki, hogy az MTUBA II.-5h tesztanyagon mindkét megközelítés páronként nagyjából megegyező OOV-arány ért el, azonban a PPL1 válogatással kapható perplexitások minden korpuszméret mellett alacsonyabbak. Ennek oka az lehet, hogy a rövid, sok szótáron kívüli szót tartalmazó soroknál a PPL1 metrika reálisabb képet fest az illeszkedés mértékéről. A továbbiakban minden esetben PPL1 alapján végezzük a kiegészítő korpuszok sorainak előválogatását.

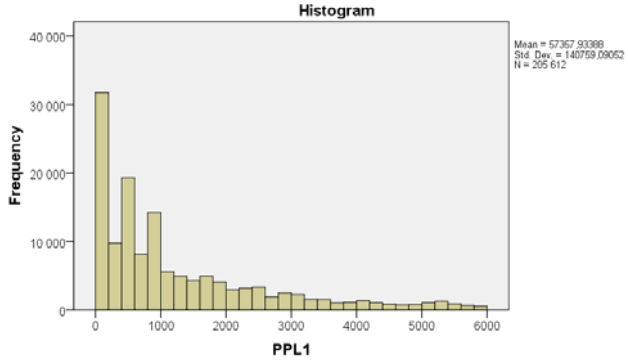


2 ábra. A webkorpusz sorainak PPL1 eloszlása az MTUBA II. in-domain modell alapján, [0;20000] tartományon ábrázolva.

4.1.2 Adaptációs paraméterek

Annak érdekében, hogy megfelelő válogatási küszöböt tudjunk beállítani a **webkorpuszon**, ismerni kell a sorainak PPL1 eloszlását (**2. ábra**). Az adaptációs kísérletekhez a már előző pontban is vizsgált „PPL1-400" illetve „PPL1-260" előválogatási határokat választottunk. 400-nál nagyobb határt megengedve, nagyon megnőtt volna az adaptált modell memóriaigénye, míg 260-nál kisebb határt beállítva már túl sok értékes sort veszítettünk volna. Az interpolációs súly optimalizálásakor mindkét korpuszméret mellett a webkorpuszok 0,1-es súlyozású figyelembevételével kaptuk a legalacsonyabb perplexitásokat az MTUBA II.-5h tesztanyagon.

Az **e-mail korpuszon** a webkorpusz esetében már bemutatott eljárást követtük. Először megvizsgáltuk a korpusz sorainak MTUBA II. in-domain modellel számított PPL1 eloszlását (**3. ábra**), majd ez alapján válogatási küszöbértékeket határoztunk meg. A két kiválasztott küszöbérték az eloszlás első csúcsának határához (1000), illetve a még számottevő mintával rendelkező tartomány határához (6000) illeszkedik. Az e-mail korpusz azonban a webkorpusznál két nagyságrenddel kevesebb szót tartalmaz, ezért a korpusz előválogatás mellett a válogatás nélkül kapható



3. ábra. Az e-mail korpusz sorainak PPL1 eloszlása az MTUBA II. in-domain modell alapján, [0;6000] tartományon ábrázolva.

eredményekre is kíváncsiak voltunk. A perplexitás minimalizálását célzó kísérleteink eredményeként a webkorpuszhoz hasonlóan itt is a 0,1-es kiegészítő modell súly adódott optimálisnak minden esetben.

A kísérletsorozat utolsó állomásaként az **MTUBA I.** modellt adaptáltuk az MTUBA II. modellhez. Mivel a két ügyfélszolgálati feladat szóhasználatában és fordulataiban nagyon hasonlít egymáshoz, az MTUBA I. közel in-domain tanítószövegnek tekinthető, így itt a korpuszegyesítéses eljárást is kiértékeltünk. Az MTUBA I. korpusz kis mérete miatt korpusz-előválogatást nem alkalmaztunk. Az interpoláció során az ideális kiegészítő modell súly 0,2-nek adódott.

4.1.3 Felügyelt adaptációs felismerési eredmények

A MTUBA II.-5h felismerési feladaton kiértékelt felügyelt nyelvimodell-adaptációs eredményeket a **4. táblázatban** foglaltuk össze.

4. táblázat: MTUBA II.-5h tesztanyagon mért felismerési eredmények felügyelten adaptált nyelvi modellek használatával.

Nyelvi modell	Szótár-méret [ezer szó]	OOV arány [%]	PPL [-]	WER [%]	LER [%]
MTUBA II. in-domain	21	4,3	167	46,4	25,0
+0,1 Webkorp. PPL1-400	386	2,1	208	**45,2**	24,6
+0,1 Webkorp. PPL1-260	228	2,6	201	45,5	24,7
+0,1 E-mail korpusz	70	3,3	181	45,4	24,6
+0,1 E-mail korpusz PPL1-6000	55	3,4	178	45,3	24,6
+0,1 E-mail korpusz PPL1-1000	40	3,7	176	45,6	24,7
+MTUBA I. (korpuszegyesítés)	37	3,1	189	45,4	24,6
+0,2 MTUBA I. (interpoláció)	37	3,1	176	**45,2**	**24,5**

A felismerési eredmények alapján látható, hogy a felügyelt adaptációval készült modellek használatával szignifikánsan alacsonyabb felismerési hibát érhetünk el, mint az in-domain MTUBA II. modellel. Bár a kisméretű in-domain nyelvi modellel mérhető a legkisebb perplexitás MTUBA II.-5h tesztanyagon, az adaptált nyelvi modellek ellensúlyozni tudják ezt nagyobb szótárméretükkel, melynek segítségével le tudják szorítani a tesztanyagon mérhető OOV arányukat.

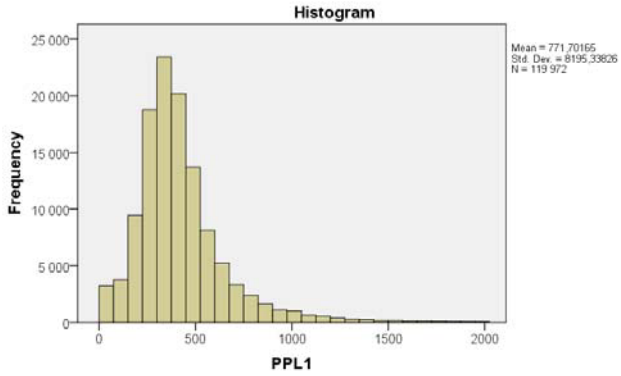
A legalacsonyabb felismerési hibát mind LER mind WER értelemben az MTUBA I. adaptációjával értük el, ráadásul az adaptált modellek közül ehhez tartozott a legkisebb szótárméret is. Igaz tehát, hogy a feladathoz jól illeszkedő tanítóanyagok a legnehezebben hozzáférhetőek és esetenként a legköltségesebbek is, azonban ezekkel lehet a leghatékonyabban végrehajtani az adaptációt. Megfigyelhető továbbá, hogy hasonló mértékben illeszkedő tanítószövegek esetén is eredményesebb eljárás a modell-interpoláció, mint a korpuszegyesítés.

Az MTUBA I.-től nagyon kicsit elmaradva, meglepően jól teljesített a webkorpuszos adaptáció. Igaz, hogy ugyanakkora WER eléréséhez itt tízszer akkora szótárra volt szükség, azonban az MTUBA I.-el ellentétben a webkorpuszt hatékonyan lehet adaptálni más felismerési feladathoz is, így egyfajta univerzális kiegészítő modellnek tekinthető. Az e-mail korpusszal mért eredmények is csak kis

mértékben maradnak el a két korábbi csoport eredményeitől. Itt a valódi érdekességet az adja, hogy összevethetőek a teljes és válogatott kiegészítő korpusszal kapott eredmények. Ez alapján azt mondhatjuk, hogy a túlzott metszés ronthatja az adaptáció hatásfokát (PPL1-1000), azonban az sem igaz, hogy a teljes out-of-domain korpusz alkalmazása jó megoldás. Optimális eredmény akkor született, amikor bár szűrtük a korpuszt, de nem túlzottan nagy mértékben. Mindez arra is utalhat, hogy akár pontosabb felismerési eredmény is elérhető lenne a webkorpusz használatával, ha az adaptáció előtt nagyobb előválogatási küszöböt alkalmaznánk, azonban ilyen nagy szótárméretű felismerő hálózatot szóalapon nem tudunk létrehozni a hálózatépítés nagy memóriaigénye miatt.

4.2 Felügyelet nélküli adaptáció

Felügyelet nélküli adaptációs kísérleteket az MTUBA I. és MTUBA II. feladaton is végeztünk. Vizsgálataink központi kérdése az volt, hogy a felismerő rendszer nyelvi modellje vajon milyen mértékben képes profitálni abból, ha az általa generált korábbi kimenetekkel adaptálunk.



4. ábra. Az MTUBA I. felismerési kimeneteit tartalmazó korpusz sorainak PPL1 eloszlása az MTUBA I. in-domain nyelvi modell alapján, [0;2000] tartományon

4.2.1 Adaptációs paraméterek

Felügyelet nélküli adaptáció esetén egyből adódik a kérdés, hogy vajon szükség van-e perplexitásalapú korpusz előválogatásra. A kérdés megválaszolásához felvettük a 32 millió szavas MTUBA I. felismerési kimenet korpusz PPL1 eloszlását **MTUBA I.** in-domain modell alapján (**4.ábra**). Míg a webkorpusz esetén egy nagyon vegyes szöveggel álltunk szemben, ezért jól különválaszthatóak voltak a jól és kevésbé jól illeszkedő sorok, addig a felismerési kimeneteket tartalmazó korpusznál sokkal egyenletesebb az eloszlás, és az illeszkedés mértéke is átlagosan nagyobb. Ez alapján az feltételezhető, hogy nagymértékű méretcsökkentés csak jól illeszkedő sorok elhagyásának árán valósítható meg. Éppen ezért az eredeti, válogatás nélküli korpusszal is végzünk adaptációt. Az ideális kiegészítő modellsúly 0,9-nek adódott az előválogatott és az eredeti korpusz használatakor egyaránt.

Az MTUBA I. mellett az **MTUBA II.** feladaton is szerettünk volna felügyelet nélküli adaptációs kísérleteket végezni. Ehhez azonban nem használhattuk az MTUBA II.-5h tesztanyagot, ugyanis az MTUBA II. rendszerrel előálló felismerési kimenetek a felismerő egy olyan konfigurációjából származtak, ahol az in-domain nyelvi modell az 5 órás tesztanyag leiratait is tartalmazta. Ez további 2 óra MTUBA II. hanganyag kézi átírását tette szükségessé, melyből megszületett a tanítástól már független MTUBA II.-2h tesztanyag. MTUBA II. esetén csak a teljes, válogatás nélküli kiegészítő korpusszal végeztünk kísérletet. A kiegészítő modellsúly értékét 0,8-nál mértük optimálisnak.

4.2.2 Felügyelet nélküli adaptációs eredmények

A felügyelet nélküli adaptációval készült felismerési eredményeket az **5. táblázatban** foglaltuk össze.

5. táblázat: Felügyelet nélküli adaptációs eredmények az MTUBA I. és MTUBA II.-2h teszthalmazon.

Nyelvi modell	OOV arány [%]	PPL [-]	WER [%]	LER [%]
MTUBA I. in-domain	5,7	310	48,0	25,9
+ 0,9 MTUBA I. felism. PPL1-300	5,7	207	47,5	25,5
+ 0,9 MTUBA I. felism.	5,7	192	**46,8**	**25,1**
MTUBA II. in-domain	5,6	255	50,9	27,5
+ 0,8 MTUBA II. felism.	5,6	173	**49,7**	**26,9**

Megfigyelhető, hogy felügyelet nélküli adaptációval az OOV arányt nem lehet csökkenteni, ami nem meglepő, hiszen ennél az eljárásnál az in-domain nyelvi modell által szolgáltatott felismerési kimeneteket integráljuk, azaz a rendszer szótára elvileg sem bővülhet. Érdekes eredmény azonban, hogy a korábbi kimenetek figyelembevételével jelentősen sikerült csökkenteni a perplexitást és így a szó-, illetve karakter-hibaarányt is. Azaz egy működő rendszerben érdemes lehet a felismerési eredményeket időről-időre adaptálni a nyelvi modellhez, ugyanis ezzel további költségek nélkül pontosabbá tehető a felismerés. A kiegészítő korpusz méretét itt azonban nem érdemes csökkenteni, mert mint az már a perplexitáseloszlás alapján is sejthető volt (**4. ábra**), nehéz olyan vágási határt találni, mely még jelentősen csökkenti a modellméretet, viszont nincs jelentős hatással a felismerési hibára.

5 Összefoglalás

Cikkünkben azt vizsgáltuk, hogy milyen módszerekkel és milyen mértékben lehet felügyelt és felügyelet nélküli adaptációs technikákkal telefonos ügyfélszolgálati hanganyagok felismerésére készített rendszerek in-domain nyelvi modelljeinek pontosságát javítani. Eredményeink alapján azt a következtetést vonhatjuk le, hogy amennyiben a nyelvi modell méretének az alacsonyan tartását tűzzük ki célul, akkor a legjobb eredményt a felismerési feladathoz jól illeszkedő nyelvi modellek

felhasználásával érhetjük el. Ilyen tanítóadatok azonban nem minden esetben állnak rendelkezésre korlátlan mennyiségben, illetve előállításuk a költségek miatt esetenként már nem gazdaságos. Ebben az esetben további pontosságnövekedés érhető el out-of-domain tanítókorpusz felhasználásával is, ha a cikkünkben ismertetett módon kinyerjük a feladathoz jól illeszkedő részeket a korpuszból. El kell azonban fogadni, hogy a nem feladatspecifikus tanítóadatok felhasználása óhatatlanul a modell méretének növekedésével jár.

Különösen értékes és a gyakorlatban jól hasznosítható eredmény továbbá, hogy két már működő ügyfélszolgálati felismerő rendszerben átlagosan 2,4%-os relatív WER-csökkenést sikerült elérni a felismerési kimenetek felügyelet nélküli adaptálásával. Felügyelet nélküli adaptációnál az OOV arány nem csökken, hiszen felismerő rendszer szótára nem bővül, így a javulás egyedül a nyelvi modell jobb előrejelző képességre vezethető vissza, mely a nagy mennyiségű in-domain hanganyag gépi leiratában rejlő tudás felhasználásának köszönhető.

Köszönetnyilvánítás

Kutatásunkat a TÁMOP-4.2.1/B-09/1/KMR-2010-0002-es, a KMOP-1.1.1-07/1-2008-0034-es, a GOP-1.1.1-09/1-2009-0068-as, a KMOP-1.1.3-08/A-2009-0006-os és a NAP-1-2005-0010-es projektek keretében az NFÜ és az NIH támogatta.

Bibliográfia

1. Bacchiani, M., Roark, B.: Unsupervised language model adaptation. In: Proc. of Acoustics, Speech, and Signal Processing (ICASSP ‘03) (2003) 224–227
2. Bacchiani, M., Roark, B., Saraclar,M.: Language model adaptation with MAP estimation and the perceptron algorithm. In: Proc. of HLT-NAACL 2004 (2004) 21–24
3. Chen, S. F., Goodman, J.: An Empirical Study of Smooting Techniques for Language Modeling. Technical Report TR-10-98, Computer Science Group, Harvard University (1998)
4. Gauvain, J.-L., Lee, C.-H.: Maximum a posteriori estimation for multivariate Gaussian mixture observations of Markov chains. In: IEEE Transactions on Speech and Audio Processing Vol.2, No.2 (1994) 291–298
5. Gretter, R., Riccardi, G.: On-line learning of language models with word error probability distributions. In: Proc. of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP ’01) (2001) 557–560
6. Halácsy, P., Kornai, A., Németh, L., Rung, A., Szakadát, I., Trón, V.: Creating open language resources for Hungarian. In: Proc. of the 4th international conference on Language Resources and Evaluation (LREC2004) (2004)
7. Jelinek, F., Mercer, R. L.: Interpolated estimation of Markov source parameters from sparse data. In: Proc.Workshop on Pattern Recognition in Practice (1980)
8. Mauuary, L.: Blind Equalization in the Cepstral Domain for robust Telephone based Speech Recognition. In: Proc. of EUSPICO’98, Vol.1 (1998) 359–363
9. Mohri, M., Pereira, F., Riley, M.: Weighted Finite-State Transducers in Speech Recognition. Computer Speech and Language Vol.16, No.1 (2002) 69–88

10. Stolcke, A.: SRILM – an extensible language modeling toolkit. In: Proc. Intl. Conf. on Spoken Language Processing. Denver (2002) 901–904
11. Tarján B., Mihajlik P.: Magyar nyelvű nagyszótáras beszédfelismerési feladatok adatelégtelenségi problémáinak csökkentése nyelvi modell interpoláció alkalmazásával. In: VII. Magyar Számítógépes Nyelvészeti Konferencia. Szeged, Magyarország (2010). 216–223
12. Tarján, B., Mihajlik, P., Balog, A., Fegyó, T.: Evaluation of Lexical Models for Hungarian Broadcast Speech Transcription and Spoken Term Detection. In: CogInfoCom 2011: 2nd International Conference on Cognitive Infocommunications. Budapest, Hungary (2011) 1–5
13. Young, S., Ollason, D., Valtchev, V., Woodland, P.: The HTK book. (for HTK version 3.2.) (2002)