

Az enyhe kognitív zavar automatikus azonosítása beszédfelismerési technikák használatával

Tóth László¹, Gosztolya Gábor¹, Vincze Veronika¹, Hoffmann Ildikó^{2,3},
Szatlóczi Gréta⁴, Biró Edit⁴, Zsura Fruzsina⁴,
Pákási Magdolna⁴, Kálmán János⁴

¹MTA-SZTE Mesterséges Intelligencia Kutatócsoport,
{ tothl, ggabor, vinczev } @inf.u-szeged.hu
6720 Szeged, Tisza Lajos krt. 103.

²Szegedi Tudományegyetem, Magyar Nyelvészeti Tanszék,
6722 Szeged, Egyetem u. 2.

³MTA Nyelvtudományi Intézet,
1068 Budapest, Benczúr u. 33.

⁴Szegedi Tudományegyetem, Pszichiátriai Klinika,
6725 Szeged, Kálvária sgt. 57.

Összefoglaló: Az enyhe kognitív zavar (EKZ) olyan tünetegyüttes, melynek fontos szerepe van néhány demenciátípus, például az esetlegesen később kialakuló Alzheimer-kór korai megjelölésében, és így a kezelés minél korábbi elkezdésében. Korábbi kutatásainkban megmutattuk, hogy az EKZ jó eséllyel detektálható a páciens (spontán) beszéde alapján, a megfelelő beszédjellemzők kinyerése révén. Ebben a cikkben egy beszédfelismerésen alapuló megoldást mutatunk a jellemzők kinyerésének automatizálására, míg a betegség fennállására vonatkozó döntést gépi tanulási módszerekkel hozzuk meg. Ezzel a megoldással a teljes feldolgozási folyamat automatizálható, ami megteremti az alapjait egy későbbi automatizált betegszűrő teszt kidolgozásának.

Bevezető

Az enyhe kognitív zavar (EKZ) olyan tünetegyüttes, melynek fontos szerepe van néhány demenciátípus, többek között az esetlegesen később kialakuló Alzheimer-kór korai megjelölésében [1]. Sok esetben a páciensek nyelvhasználatának szupraszegmentális jellemzői alapján már a demencia tényleges klinikai megjelenése előtti fázisban megállapíthatók az enyhe kognitív zavar jelei.

Az EKZ detektálására korábban bemutattunk egy módszert, amely spontán beszédből számolt akusztikus jellemzőkre épül [2]. Kísérletileg igazoltuk, hogy a javasolt akusztikus jellemzők (pl. beszédtempó, artikulációs tempó, néma és kitöltött szünetek száma és hossza) valóban olyan információt hordoznak a spontán beszédben, melyek szignifikánsan eltérnek az EKZ-s páciensek és a kontrollcsoport tagjai esetében. Abban a dolgozatban a hangfelvételek szöveges átírását és annotálását kézilleg

végeztük, a Praat szoftvercsomag [3] felhasználásával. Mivel ez nagyon munkaigényes, a gyakorlati használhatósághoz a jellemzők automatikus kinyerésére és az alapján a hipotézis automatikus meghatározására lenne szükség. Jelen dolgozatban ennek a két lépésnek az automatizálására javasolunk egy módszert, gépi beszédfelismerési technikákra és statisztikai alapú gépi tanulási módszerekre alapozva.

A korábbi dolgozatunkban ([2]) rámutattunk, hogy a néma szünetek mellett a kitöltött szüneteknek („ööö”, „hmm” stb.) is fontos szerepük van, és míg a néma szünetek felismerése egyszerű jelfeldolgozási eszközzel is lehetséges, a kitöltött szünetek detektálása nem triviális. Az irodalomban számos munka foglalkozik ugyan az EKZ automatikus felismerésével (pl. [4,5,6]), de ezek általában beszédnek tekintik a kitöltött szüneteket, ami meghamisíthatja a későbbi jellemzőkinyerési lépés (pl. beszédtempó-számítás) eredményeit. Jelen cikkben egy olyan eljárást mutatunk, amely egy beszédfelismerőre alapozva nyeri ki a szükséges akusztikus paramétereket, majd gépi tanulási módszerekkel jelzi az EKZ fennállásának gyanúját. A kísérletek alapján a javasolt gépi megoldás csupán kicsivel ad rosszabb eredményt, mint a kézi feldolgozás, viszont lehetővé teszi a folyamat teljes automatizálását, ami alapot adhat egy későbbi automatizált betegszűrési metodika kidolgozásához.

EKZ detektálására használt beszédjellemzők

Az enyhe kognitív zavar kimutathatóan befolyásolja a páciens beszédét (ld. pl. [2,5,7,8]). Jelen kísérletünkben a korábbi munkánkban ([2]) bemutatott módon készítjük pácienseinket spontán beszédre, az alábbi forgatókönyv szerint. Miután megnéznak egy kimondottan erre a célra tervezett, egyperces animációs filmet, tesztalanyainkat megkérjük, hogy meséljék el a filmben látottakat (*azonnali felidézés*). Ezt követi egy másik, hasonló film levetítése, majd az alanyokat megkérjük, hogy meséljék el a tegnapi napjukat (*spontán beszéd*). Végül az alanyoknak el kell mesélniük a második filmben látottakat (*késleltetett felidézés*).

Az alanyonként három hangfelvételtől a következő akusztikus paramétereket nyerjük ki: artikulációs tempó (1. jellemző, a hezitációk nélkül számított másodpercenkénti beszédhang-szám), beszédtempó (2. jellemző, a hezitációkkal együtt számított másodpercenkénti beszédhang-szám), a felvétel hossza (3. jellemző), kitöltött, illetve kitöltetlen szünetek összhossza (4-5. jellemzők), kitöltött, illetve kitöltetlen szünetek száma (6-7. jellemzők), hezitációs ráta (8. jellemző, a kitöltött és kitöltetlen szünetek

összhosszának és a felvétel hosszának aránya). Hezitációnak a beszéd legalább 30 ms hosszú hiányát tekintjük.

A jellemzők automatikus kinyerése beszédfelismerési technikákkal

A főnti jellemzők manuális kiszámítása meglehetősen munkaigényes, ráadásul képzett személyt igényel. Ezért kívánatos a jellemzők kinyerésének automatizálása, melyet mi automatikus beszédfelismerési technikák használatával oldottunk meg. A jellemzők egy része ugyan egyszerű jelfeldolgozási eszközökkel is meghatározható (pl. beszéd/csend részek elkülönítése), azonban a beszédtempó és egyéb más, a beszédhangok hosszán alapuló jellemzők kinyerésére ezek alkalmatlanok.

A főntiek miatt egy beszédfelismerő rendszert tanítottunk be egy spontán beszédet tartalmazó adatbázisra, ami esetünkben a BEA Spontánbeszéd-adatbázis volt [9]. Mivel fontos volt, hogy a kitöltött szüneteket képesek legyünk azonosítani, az elérhető annotációt módosítanunk kellett, hogy tartalmazzon bizonyos, a spontán beszédben előforduló elemeket (pl. kitöltött szünetek, be- és kilégzések, nevetés, köhögés). Vegyük észre, hogy a korábban ismertetett jellemzők kiszámításához általában nem szükséges az egyes beszédhangok megkülönböztetése, csupán azok *megszámolása*, illetve hosszának megmérése. Mivel a beszédhangok téves azonosítása nem okoz gondot, ezért egy egyszerűsített beszédfelismerő modellt használtunk, amely nagy pontossággal képes a beszédjelben szereplő beszédhangokat megtalálni és azonosítani [10,11]. A kitöltött szünetekhez önálló modelleket rendeltünk, így a beszédfelismerő rendszer kimenete minden felvételre egy időzített beszédhang-sorozat. Ennek alapján a korábban javasolt jellemzőket már automatikusan is ki tudjuk számítani.

A felhasznált EKZ hangadatbázis

Az EKZ felismerésére vonatkozó kísérleteinket saját hangadatbázison végeztük, melynek rögzítését folyamatosan végezzük. Jelenleg már több, mint száz tesztalany beszédét vettük fel, mely az irodalomban már a nagyobb adatbázisok közé számít. Különböző okokból (rossz hangminőségű felvételek, ellentmondó diagnózis stb.) jelen cikkünkben 51 tesztalany felvételeivel foglalkozunk, melyből 32 páciens szenved EKZ-ban és 19 tartozik a kontrollcsoportba.

Kétlépéses EKZ-felismerés

Korábbi munkánkban megmutattuk a (háromszor) nyolc akusztikus jellemző szignifikanciáját. Egy automatikus betegségrő rendszerhez azonban azt is meg kell oldani, hogy a jellemzők alapján a gép eldöntse, hogy az alany beteg-e vagy sem. Rendszerünkben az EKZ felismerése két lépésre bomlik: az elsőben egy beszédfelismerő rendszerre támaszkodva kiszámoljuk a korábban ismertetett jellemzőket, a másodikban pedig egy modellt építünk, amely a kinyert jellemzők alapján megkísérli elkülöníteni az EKZ-ban szenvedőket ill. a kontrollcsoport tagjait. Utóbbi célra statisztikai alapú gépi tanulási módszereket alkalmaztunk: lineáris szupportvektor gépet (SVM), illetve véletlen erdőt. Kísérleteinkben ezen algoritmusok Weka programcsomagban [12] található megvalósításait használtuk. Tekintve, hogy – gépi tanulási értelemben – ez egy nem túl nagy adatbázis, külön tanító- és tesztalalmaz helyett mindig kihagytuk egy-egy alany felvételeit, és a maradék 50-en tanítottuk a modellünket, melyet a kihagyott felvételen teszteltünk. A hagyományos pontosságmérteken (*accuracy*) kívül számoltunk pontosságot (*precision*), fedést (*recall*), illetve ezek harmonikus közepét, az F_1 értéket (*F-measure*) is.

Kibővített jellemzőkészlet

A manuális jellemzőkészlet hátránya, hogy minden újabb jellemző hozzávétele a meglévő listához nagyon lelassítaná az egyes felvételek feldolgozását. Az automatikus beszédfelismerésen alapuló megközelítésünk esetén ugyanakkor a beszédfelismerési lépés időigénye az, ami jelentős; ennek eredményéből (az időzített fonémasorozatból) további jellemzők kinyerése már könnyen és gyorsan megtehető. Mivel előfordulhat, hogy ezek a további jellemzők lényegesen javítanak a második lépésben használt gépi tanuló módszer pontosságát, ezért bevezettünk egy kibővített jellemzőkészletet is, mely az eddig vizsgált "beszédhangok" (beleértve a kitöltött és néma szüneteket is) száma és összhossza mellett az előfordulási hosszok átlagát és szórását is tartalmazza. Megvizsgálva a beszédfelismerő rendszerünk kimenetét, azt is észrevettük, hogy a felismerő a kitöltött szüneteket időnként összetéveszti bizonyos beszédhangokkal (pl. „ö”, „m”). Emiatt a fenti négy statisztikai értéket kiszámítottuk néhány ilyen beszédhangra is; így felvételenként 27 jellemzőt kaptunk. Egy-egy tesztalanyra így összesen 81 jellemzőt számítottunk ki, melyeket kiegészítettünk az alany nemével és életkorával. Így összesen 83 jellemzőt kaptunk minden tesztalanyra.

1. sz. táblázat

Módszer	Jellemzőkészlet	Prec.	Felid.	F ₁	Pont.
SVM	Manuális	82,4%	87,5%	86,2%	82,4%
	Automatikus	83,9%	81,3%	82,5%	78,4%
	Kibővített	80,6%	90,6%	85,3%	80,4%
Véletlen erdő	Manuális	76,5%	81,3%	78,8%	72,5%
	Automatikus	81,8%	84,4%	83,1%	78,4%
	Kibővített	76,3%	90,6%	82,9%	76,5%

Eredmények

Az elért pontosságértékek az 1. számú táblázatban találhatóak. Látható, hogy az SVM jobban teljesített, mint a véletlen erdő; ez alól a fedésértékek jelentenek csak kivételt. A legjobb pontosságértékeket a manuális jellemzőkészlet használatával kaptuk. Mikor ugyanezeket a jellemzőket automatikusan nyertük ki a hangfelvételekből, némileg rosszabb értékeket kaptunk, ami valószínűleg a beszédfelismerő rendszer kimenetében található pontatlanságoknak tudható be. Ugyanakkor a kibővített jellemzőkészlet használatával a gépi tanulók pontossága nőtt, olyannyira, hogy majdnem elérték a manuális jellemzőkészlet használata mellett mért értékeket: a 85,3%-os F₁-érték csak kis mértékben marad el a kézi jellemzőkinyerés mellett kapott 86,2%-tól.

Eredményeink nehezen összevethetők más szerzők eredményeivel, hiszen eltér a használt adatbázis. Ugyanakkor az általunk ismert dolgozatokban is 75-90% közti eredmények szerepelnek, így az általunk elért pontosságértékek versenyképesnek számítanak. Természetesen azt jelen pillanatban még nem tudjuk megmondani, hogy milyen mértékű pontosságra lesz majd szükség a gyakorlati alkalmazhatósághoz.

Köszönetnyilvánítás

Jelen kutatást a Telemedicina fókuszú kutatások orvosi, matematikai és informatikai tudományterületeken című, TÁMOP-4.2.2.A-11/1/KONV-2012-0073 számú projekt támogatta, valamint a Bolyai János Kutatói Ösztöndíj. A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Hivatkozások

- [1] S. Negash, L.E. Petersen, Y.E. Geda, D.S. Knopman, B.F. Boeve, G.E. Smith, R.J. Ivnik, D.V. Howard, J.H. Howard Jr, and R.C. Petersen. "Effects of ApoE genotype and Mild Cognitive Impairment on implicit learning," *Neurobiology of Aging*, vol. 28, no. 6, pp. 885-893, 2007

- [2] I. Hoffmann, D. Németh, C. Dye, M. Pákási, T. Irinyi, and J. Kálmán. "Temporal parameters of spontaneous speech in Alzheimer's disease," *International Journal of Speech-Language Pathology*, vol. 12, no. 1, pp. 29-34, 2010
- [3] P. Boersma. "Praat, a system for doing phonetics by computer," *Glott international*, vol. 5, no. 9/10, pp. 341-345, 2002
- [4] K.L. de Ipina, J.B. Alonso, J. Solé-Casals, N. Barroso, P. Henriquez, M. Faundez-Zanuy, C.M. Travieso, M. Ecaz-Torres, P. Martínez-Lage, and H. Eguiraun. "On automatic diagnosis of Alzheimer's disease based on spontaneous speech analysis and emotional temperature," *Cognitive Computation*, vol. 7, no. 1, pp. 44-55, 2015
- [5] A. Satt, R. Hoory, A. König, P. Aalten, and P.H. Robert. "Speech-based automatic and robust detection of very early dementia," *Interspeech*, pp. 2538-2542, 2014
- [6] M. Lehr, E. Prud'hommeaux, I. Shafran, and B. Roark. "Fully automated neuropsychological assessment for detecting Mild Cognitive Impairment," *Interspeech*, 2012
- [7] B. Roark, M. Mitchell, J.-P. Hosom, K. Hollingshead, and J. Kaye. "Spoken language derived measures for detecting mild cognitive impairment," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 7, pp. 2081-2090, 2011
- [8] V. Baldas, C. Lampiris, C. Capsalis, and D. Koutsouris. "Early diagnosis of Alzheimer's type dementia using continuous speech recognition," *MobiHealth*, Ayia Napa, Cyprus, pp. 105-110, 2011
- [9] M. Gósy. "BEA: A multifunctional Hungarian spoken language database," *The Phonetican*, vol. 105-106, pp. 50-61, 2012
- [10] T. Grósz, Gy. Kovács, and L. Tóth. "Új eredmények a mély neuronháló magyar nyelvű beszéd felismerésben," *MSZNY*, pp. 3-13, 2014
- [11] L. Tóth. "Phone recognition with hierarchical Convolutional Deep Maxout Networks," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2015, no. 25, pp. 1-13, 2015
- [12] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. Witten. "The WEKA data mining software: an update," *ACM SIGKDD explorations newsletter*, vol. 11, no. 1, pp. 10-18, 2009