HADTUDOMÁNYI SZEMLE

2017. X. évfolyam 2. szám

NÉGYESI IMRE[1]

A mesterséges intelligencia és a hadsereg II. (Beszédfelismerő rendszerek I.)

The artificial intelligence and the army I. (Discussion Systems I.)

Absztrakt

A mesterséges intelligencia számos terület vizsgálatával foglalkozik, mint a szakértői rendszerek, látás és képfeldolgozás, természetes nyelvek feldolgozása, beszédfelismerés stb. Ennek a cikknek a témája a beszédfelismerés, amelynek területe nem maradhat ki a hadseregekben sem a kutatási irányokból. Ebben a részben az elméleti alapok kerülnek ismertetésre, a második rész a gyakorlati megvalósítás kérdéseit vizsgálja.

Kulcsszavak: informatika, informatikai rendszer, mesterséges intelligencia, beszédfelismerés

Abstract

Artificial intelligence deals with a range of areas such as expert systems, visual and image processing, processing of natural languages, speech recognition, and so on. The subject of this article is speech recognition, the field of which can not be left out of the military either from the research directions. In this section the theoretical basics are described, the second part discusses the issues of practical implementation.

Key words: Informatics, IT System, Artificial Intelligence, Voice Recognition

BEVEZETÉS

Számtalan lehetőség vizsgálata folyt és folyik jelenleg is annak érdekében, hogy az információk széleskörű kezelésével növelhetővé váljon egy-egy tevékenység hatékonysága. Egy alapvetően új, a különböző típusú munkák hatékonyságát növelő lehetőség lehet a

[1] Nemzeti Közszolgálati Egyetem - National University of Public Service, E-mail: negyesi.imre@uni-nke.hu, ORCID: 0000-0003-1144-1912

beszédfelismerő rendszerek alkalmazása a beérkező információk kezeléséhez. A beszédfelismerő rendszer alkalmazása éppen azokon a területeken jelenthet előrelépést, melyeken belül a legutóbbi felmérések jelentős lemaradásokat mutattak. A beszédfelismerő rendszer gyakorlati alkalmazása még nagyon rövid időre tekint vissza és még viszonylag szűk azoknak a szakmáknak a száma, melyekben az alkalmazás feltételei megteremtődtek. A mesterséges intelligencia alkalmazása a beszédfelismerésben a katonai feladatok végrehajtásánál is jelentős távlatokat jelenthet.

A BESZÉDFELISMERŐ RENDSZEREK ALKALMAZÁSÁNAK ELMÉLETI ALAPJAI

A beszédfelismerő rendszer az által, hogy a normál beszéd-sebességgel elmondott szöveget igen nagy pontossággal, szinte az elhangzással egy időben, szerkeszthető, tovább feldolgozható, írott szöveggé alakítja át rendkívül nagy hatékonyság növelést eredményez. A beszédfelismeréső rendszer lényege, hogy egy megfelelő számítógépes program segítségével a leggyorsabb emberi beszédet is közvetlen gépelhető szövegformává alakíthatjuk. Természetesen az információ áramlása a feladatok végrehajtása során több irányú (utasítások, pontosító kérdések), ezért a beszédfelismerő szoftvernek is képesnek kell lenni a többirányú szöveges átalakításra is.

A beszédfelismerés rendszeres alkalmazása, az eddig már használt beszédfelismerő szoftverek esetében, az elemzések alapján megállapíthatóan növelte a munka hatékonyságát, amelyet 30-40%-ban határozták meg. A hatékonyság növelésével elérhető egy megtakarított idő, amely felhasználható lesz (pl.: az elemzések még pontosabb előkészítésére) a munka színvonalának emelésére.

Az alkalmazás érdekében az alábbi előzetes feladatokat kell megoldani:

- be kell szerezni a beszédfelismerő szoftvert;
- el kell készíteni a szakterülethez tartozó szókincs szótárát;
- a beszédfelismerés rendszerét fel kell telepíteni a számítógépekre;
- betanítani a beszédfelismerést alkalmazni kívánó személyeket a rendszer alkalmazására;
- kialakítani és kiadni a szabályozási rendszert a beszédfelismerő rendszer alkalmazásának rendjéről, amelynek illeszkedni kell a teljes munkafolyamathoz.

Melyek lehetnek a beszédfelismerés fontosabb előnyei:

- gyorsítja a reagálási képességet;
- javíthatja az alkalmazók pontos szóbeli kommunikációját;
- lehetővé teszi a feleslegessé vált idők hasznosítását;
- javítja a gépelt anyagok minőségét.

A beszédfelismerés különösen jelentős azokon a területeken, ahol nagyon nagy az írásigény, és sajnos a hadseregek még ebbe a kategóriába tartoznak. Egy adott terület esetében lehet, hogy nem az írásigény nagysága lesz a meghatározó, hanem a beérkező és kimenő adatok pontossága. A beszédfelismerő rendszernek az adott területhez tartozó beszédkészletének kidolgozása azonban biztos, hogy elengedhetetlen lesz, mert a munka

az adott területeken általában olyan speciális szókincskészletet használ. A beszédfelismerő rendszer alapjaiban növeli meg a munka hatékonyságát, azáltal, hogy az igen hosszú, írott formát igénylő anyagokat a hagyományos gépelésnél lényegesen gyorsabban lehet elkészíteni és az archiválás után a hagyományos számítógépes eszközökkel visszakeresni.

A hagyományos beszédtechnológia az alábbi négy fő technológiai területet foglalja magában:

- Az automatikus beszédfelismerés határozza meg, hogy milyen szavakat mondott ki a felhasználó.
- A szintaktikai elemzés és a szemantikai interpretáció segítségével elemezhető a felhasználó közlésének szintaktikai szerkezete, valamint leképezhető annak szemantikai interpretációja az adott rendszer céljainak megfelelően.
- A dialógusvezérlés az input nyelvi jellemzői, az adott felhasználó és feladat egyéni beállításai alapján valósítja meg a rendszer megfelelő lépését, az adatbázislekérdezést.
- A beszédszintézis technológiáját alkalmazzák arra, hogy a gép előállítsa a megfelelő beszédkimenetet.

Azonban, ha azt kérdezzük, hogy megoldott-e a beszédszintézis, más szóval szöveg-beszéd átalakítás, akkor nem kapunk egyértelmű választ. Ennek több oka is lehet, például nem szeretik a felhasználók, nem használják szívesen, mert nem hozott komoly, kimutatható üzleti eredményt senkinek, vagy csak ösztönösen távolságot tartanak a technikai újításoktól.

A beszédszintézis alapelemeiként, illetve fő megoldandó kérdéseiként a következőket tekinthetjük:

- Általános alapeszköz, amely lehet a PC, de csak nagy operatív memóriával, háttértárral, hangkártyával, vagy akár okos telefon, hasonló jó adottságokkal.
- A természetes beszéd alapelemeinek tárolási, módosítási, összefűzési szabályrendszere.
- A tárolt alapelemek meghatározása (pl.: teljes közlendő, mondatok, szavak, szótagok, hangok).
- Hogyan lehet olyan elemeket kialakítani, amelyek jól összefűzhetők és a prozódiai elemek[2] is ráépíthetők?
- Mi az, amit át kell „fogalmazni”, előre le kell „fordítani” írásból beszédre felolvastatás előtt? A fő cél, hogy gépileg „érteni” lehessen a szöveget.

A beszédfelismerés már nagyon régen foglalkoztatta a kutatókat, de a gyakorlati életben is használható eredmények csak a nagyteljesítményű számítógépek elterjedésével születtek meg. A beszéd az emberi kommunikáció alapvető és egyben leggyorsabb formája, így több előny is származik abból, ha az emberi beszédet a számítógép megérti, és ennek eredmé-

[2] A szupraszegmentális tényezők, más néven prozódiai elemek a hangzás színesítésére és a mondanivalónk értelmezésére szolgálnak.

HADTUDOMÁNYI SZEMLE

2017. X. évfolyam 2. szám

nyeként felgyorsul a számítógép és az ember közötti információcsere. A beszédfelismerő rendszerek alkalmazása új megoldásokat nyújthat számos területen, ahol a dokumentálás folyamata és az információk kiemelkedő jelentőséggel bírnak.

A BESZÉDFELISMERŐ RENDSZEREK OSZTÁLYOZÁSA

A **beszédfelismerő rendszerek**et több szempont szerint osztályozhatjuk. Az egyik osztályozási szempont lehet, hogy az alkalmazott beszédfelismerő rendszerek milyen méretű szövegrészeket ismernek fel. Ez alapján lehetnek egyedülálló **szavakat, kifejezéseket, mondatokat, szövegeket** felismerő rendszerek.

Egy másik osztályozási szempont a beszélőhöz kapcsolódik, amely alapján lehetnek beszélő-függő (személy-függő), beszélő-független (személy-független) rendszerek.

A **beszélő-függő rendszerek** egyetlen ember hangjának felismerésére képes, általában adaptív rendszerek, amelyek egy adott személyhez idomulnak. Mivel az egyik ember hangja lényegesen különbözik a másikétól, egyszerűbbek az egyetlen emberi hangra támaszkodó rendszerek, melyek jóval megbízhatóbbak is, ugyanis a rendszer „megtanulja" a beszélő hangszínét, hangsúlyozását, hangerejét.

A **beszélő-független rendszerek előnye, hogy** bárki használhatja. Nincs szükség az előbb említett tanulásra, gyakorlásra, azonban az ilyen rendszer rendkívül komplex, és kevésbé megbízható. Ezek a rendszerek nagyon sok előzetesen létrehozott mintával dolgoznak és megpróbálják a személyfüggőséget átlagolással áthidalni.

A beszédfelismerés egy következő osztályozása szerint a rendszerek lehetnek:

— izolált szavas;
— kapcsolt szavas;
— folyamatos beszéd alapú rendszerek.

Az izolált szavas rendszerek egymástól hosszú idővel elválasztott szavakat használnak, ezért csak rövid utasítások kiadására használhatóak. A kapcsolt szavas rendszereknél a szavak között szünetek minimálisak, még a folyamatos beszédet a diktáló rendszerek kezelik.

A rendszerek osztályozása történhet egy másik jellemző alapján, amely már a konkrét, megoldandó feladathoz kapcsolódik. Ekkor meghatározó lesz az a **feladattípus, szakterület, munkafolyamat**, amelyre az adott rendszert használni akarják. Ebben az esetben azokat a fontos kérdéseket kell megválaszolni, hogy **mekkora méretű szótárkészlettel és mekkora szókinccsel** dolgozik a szoftver.

Lényeges szempont, hogy **milyen környezetben** kívánják használni (pl. mennyire „zajos", milyen az értelmezendő beszéd sebessége). A jó minőségű beszédből felismerő rendszerek adják a kiindulási alapot, ezeket kell később robosztussá tenni, tehát alkalmassá tenni a nagyobb zajterhelésű környezetekben is. Jelenleg a technológia még nem áll azon a szinten, hogy a teljesen szabad beszédet is elfogadható pontossággal ismerje fel, ahhoz ugyanis túlzottan nagyméretű szótárakra s azokat kezelni képes hardverre lenne szükség.

HADTUDOMÁNYI SZEMLE

2017. X. évfolyam 2. szám

Mindezek után tekintsünk át a beszédfelismeréssel kapcsolatosan néhány további olyan kérdést, amelynek tisztázása elősegíthető az alkalmazandó szoftver kiválasztását, létrehozását.

A hatékony beszédfelismerés kritikus része minden hangutasítással működő rendszernek is. A legfontosabb mérőszáma a hangfelismerésnek a felismerés pontossága, amely mérőszámának meghatározása előzetesen kell, hogy megtörténjen, figyelembe véve a rendelkezésre álló lehetőségeket, de mind jobban megközelítve a 100%-ot!

Folyamatos felismerésnél többféle módszer alkalmazható a beszédfelismerés hatékonyságának, ezáltal használhatóságának a mérésére. A leggyakrabban használt mennyiség a WER (Word Error Rate), azaz szóhiba arány, amelynek nézzük meg a rövid leírását. A felismerési eredményt – ha addig nem olyan formában volt – szósorozattá alakítjuk. A referencia átirathoz a dinamikus programozás módszerével hasonlítjuk, ahol a következő súlyokat rendeljük az egyes lehetőségekhez:

- *C* (helyes, „korrekt” felismerés): 0
- *S* (helyettesítés, „szubsztitúció”): 10
- *D* (törlés, „deletálás”): 7
- *I* (beszúrás, „inzerció”): 7

A kiértékelés alapja a legkisebb összsúlyú összerendelés lesz. A fenti betűjelekkel az adott jelenségek darabszámát jelölve, az alábbi felismerési mérőszámok definiálhatók:

$$\text{Felismerési arány (Correct Rate:”Corr”)}=\frac{N-S-D}{N}X100\%,$$
$$\text{Felismerési pontosság (Accuary:”Acc”)}=\frac{N-S-D-I}{N}X100\%$$

1. sz. ábra: Felismerési mérőszám definiálása (Forrás az irodalomjegyzékben.)

Ahol N az összes felismerési egység (szó) száma a referencia-átiratban. A legtöbb alkalmazásnál hibának számít a referenciában nem szereplő szavak beszúrása is, ez csak a felismerési pontosságban jelenik meg. A felismerési pontosság lehet akár negatív is, ha nagy a beszúrások száma.

A felismerési hiba általánosan elfogadott definíciója a következő:

$$\text{Felismerési hiba (Error Rate: ”ER”)}=100\% - \text{Felismerési pontosság} =\frac{S+D+I}{N}100\%$$

2. sz. ábra: Felismerési hiba definíciója (Forrás az irodalomjegyzékben.)

HADTUDOMÁNYI SZEMLE

2017. X. évfolyam 2. szám

WER: Szó felismerési egységeknél tehát a felismerési hiba a WER. A magyar nyelv esetén azonban a szófelismerési hiba bizonyos esetekben túlzottan pesszimista becslést adhat a felismerés pontosságáról.

LER: Elterjedt a LER (Letter Error Rate), azaz a „betű" felismerési hiba, mint metrika használata. (A szóközt is betű értékűnek definiáljuk, egyébként ugyanúgy számoljuk ki karakter egységenként, mint a szóhiba-arányt.)

A gyakorlatban azonban általában nem a felismerési hiba abszolút értéke a kérdés, hanem leggyakrabban annak megváltozása. Ezen belül is tipikusan a javulás relatív mértéke az érdeklődés tárgya. Ezt az alábbiak szerint definiáljuk mind WER, mind LER esetén.

$$\text{Relatív javulás}(-\Delta ER_{rel}) = \frac{\text{ERreferencia-ERúj}}{\text{ERreferencia}} X00\%$$

3. sz. ábra: A javulás relatív értéke (Forrás az irodalomjegyzékben.)

Végül, gyakorlati szempontból igen lényeges metrika lehet a felismerés időigényének az alakulása is, természetesen adott hardver esetén. Erre az RTF (Real Time Factor) a szokásos mérték. (Tehát az alacsonyabb értékek a jobbak.)

$$RTF = \frac{\text{felismerésre fordított idő}}{\text{felismert beszéd hossza}}$$

4. sz. ábra: Felismerés időigénye (Forrás az irodalomjegyzékben.)

Attól, hogy az egyik felismerési teszt során jobb eredményt kaptunk, mint a másikban, még nem jelenthetjük ki 100% biztonsággal, az utóbbi megközelítés általánosságban véve is jobb, hiszen véges méretű teszthalmazzal dolgozhatunk csak. Hasznos lehet a felismerési hiba relatív csökkenése és a hasonló mérőszámok mellett a szignifikancia-szintet is megadni, ami megmutatja, hogy mekkora a tévedés valószínűsége a tekintetben, hogy az eredmények alapján jobbnak minősítettük az egyik megközelítést a másiknál.

Például, a felismerési arányokon alapuló 2 mintás Z-próbával egy valószínűségi becslést kaphatunk arról, hogy két eltérő, de ugyanolyan körülmények között tesztelt felismerési megközelítés közül az egyiket a másiknál jobbnak ítélve, mennyire lehetünk biztosak abban, hogy jól döntöttünk.

Az eljárás gyengéje, hogy csak akkor ad megalapozott becslést, ha a felismert szavak egymástól függetlenek, illetve a fenti definíció szerinti Z valóban jól közelíti a normális eloszlást. Míg izolált szavas teszteknél a szóhiba-arányra vonatkozóan megfelelően nagy mintaszámok esetén ez teljesül is, folyamatos beszéd felismerésekor a szó- és betűhiba-arányra vonatkoztatva már kevésbé (hiszen a nyelvi és kiejtési modellezésnél pont abból indulunk ki, hogy az egymás utáni egységek nem függetlenek egymástól).

HADTUDOMÁNYI SZEMLE

2017. X. évfolyam 2. szám

E problémák miatt – főként a folyamatos beszéd-felismerési eredmények szignifikancia-vizsgálatához – a fentinél összetettebb módszereket szoktak használni. Ilyen például a NIST (National Institute of Standards and Technology) ajánlásban szereplő nem parametrikus Wilcoxon előjeles rang teszt. A módszer alkalmas arra, hogy azonos tesztadatokon futtatott A és B felismerő rendszer szegmenspárokon összehasonlított eredményei alapján becsülje meg annak biztonságát, hogy B jobb, mint A. Ehhez az adott szegmenseken mért WER (vagy LER) értékeket összehasonlítja, a különbségeket rangsorolja, majd a javulás/romlás szerint előjelzi.

A személyes közvetlen kommunikáció során egyértelműen kijelenthetjük, hogy egyszerűbb megérteni valaki beszédét, ha a hallgató ismeri azt, aki beszél hozzá és már hozzászokott annak beszéd stílusához – különösen akkor, ha a beszélőnek egyedi kiejtése vagy erős akcentusa van. A katonai felhasználás során ez külön nehézséget jelenthet elsősorban a soknemzetiségű gyakorlatok folyamán. Az a megállapítás különösen igaz a számítógép alapú hangfelismerésnél, és a hasonló alkalmazásokhoz hasonlóan ezt a tényt kell (lehet) felhasználni a hangfelismerés pontosságának fokozására. Az adatok továbbítására az előzetesen készített dokumentációkban (utasításokban) már kijelölhető konkrét személy (természetesen tartalékok képzésével), aki feladatul kapja az adatok továbbítását. Az emberek könnyen megért olyan beszédet, ahol az egyes szavak összeérnek, nincs közöttük hallható elválasztás, amely szavakra vagy mondatokra osztaná azt. A beszéd hallgatása közben nemcsak az egyes szavak megértésére vagyunk képesek, hanem el tudjuk az egyes szavakat is határolni egymástól.

A szavakat, amelyeket a beszédfelismerő rendszernek kezelni kell, le kell fordítani és szótárba kell foglalni. Az embereknek nagy a szókincse, ezért több ezer szót vagyunk képesek felismerni. Ebből következően a beszédfelismerő rendszereknek, amelyek számítógépes szöveg bevitelre képesek szintén több ezer szavas szótárnak kell rendelkezésre állnia. Ezeket a rendszereket nagy szótáras rendszereknek nevezik. (A nagy szótáras rendszereket más néven kötetlen szótáras rendszereknek is nevezik.) A nagy szótáras (kötetlen szótáras) rendszerek 20-80 000 szót tartalmaznak, ezért bizonyos nyelveken már gyakorlatilag diktáló rendszernek (STT: Speech to Text) tekinthetőek. A magyar nyelv azonban toldalékozó nyelv, így egy adott szó felvétele a szótárba nem jelentheti a szó összes megjelenési formájának a felvételét. Összehasonlítva az angol nyelvvel elmondhatjuk, hogy egy 25 000 szavas szótárral ellátott angol nyelvű rendszer már elfogadható minőséggel működik, addig ugyanennyi szóval egy magyar nyelvű rendszer nem tud megfelelő hatékonysággal működni. A beszédfelismerő rendszer megoldások másik végletét jelentik azon rendszerek, amelyeket arra terveztek, hogy a felhasználó a feltett kérdésekre igennel vagy nemmel válaszolnak, ezek a kis szótáras rendszerek. A kis szótáras rendszereket más néven kötött szótáras rendszereknek is nevezik, amelyek azonban csak mintegy 100 szóval dolgoznak. A katonai feladatok esetében az angol nyelv használata az elfogadott, ezért az említett 25 000 szó elég lehet, de figyelembe kell venni a szakmaspecifikus kifejezéseket. A kis szótáras rendszerek is alkalmazhatóak lehetnek, hiszen rövid parancsok, utasítások „belefférnek” a 100 szavas keretbe.

HADTUDOMÁNYI SZEMLE

2017. X. évfolyam 2. szám

A BESZÉDFELISMERÉS MÓDSZEREI

A következőkben nézzük meg, hogy a különböző beszéd-felismerési feladatoknak a végrehajtása során, milyen módszereket alkalmazhatunk.

A beszédfelismerésnek alapvetően három komponense van:

- lényegkiemelés (feature extraction) a hanghullám változásaiból olyan elemeket próbálunk kiemelni, melyeknek kicsi az intrindividuális és az interindividuális jellemzője (függetlenül attól, hogy ki mondta, milyen érzelmi állapotban mondta);
- mintaillesztés ugyanazt a szót nem lehet kétszer ugyanabban a ritmusban kimondani, ezért a mintaillesztés feladata a különböző ritmikájú és spektrális karakterű kiejtések közötti különbségek kiküszöbölése;
- utó/előfeldolgozás: az utófeldolgozásnak ma már nem jellemző a használata. Az előfeldolgozás (pl. zajcsökkentés) robusztusabbá teszi a felismerést.

A lényegkiemelés során az időfüggvényt keretekre (ablakokra) bontjuk, amelyek 10-30 ms hosszú ablakok, és az ablakokat 50%-os fedésben helyezzük egymásra. Az ablak alakja kétféle lehet: Rectangular (négyszögletes) vagy Hamming ablak.

Ha a beszéd időfüggvénye f(t) és az ablak időfüggvénye w(t), akkor a kiablakolt függvény a(t)=f(t)·w(t). A spektrális jellemzésre igen jó a Fourier transzformáció: Miután előzetesen általában minták állnak rendelkezésre, ezért DFT-t (diszkrét Fourier transzformációt) alkalmazunk. A DFT a halmozott spektrum mintáit adja (számsorozat DFT-je a számsorozathoz tartozó spektrum elégséges mintáit adja).

$$A(\omega) = F(\omega) * W(\omega) = F(\alpha) W(\omega - \alpha) d\alpha$$

5. sz. ábra: Fourier transzformáció(Forrás az irodalomjegyzékben.)

Ezek után tekintsünk bele a mintaillesztési alapfogalmaiba. Adottak pl. izolált szavak (lényegkiemelt vektorsorozatokkal), prototípusok, illetve fonémák (ezekhez is vektorsorozatok tartoznak), vagyis a felismerés alapjául szolgáló nyelvi egységet reprezentáló vektorsorozatok. Feladat, hogy az ún. tesztkiejtésből meg tudjuk állapítani, hogy melyik referenciához hasonlít a legjobban. Legnagyobb probléma, hogy ugyanazt a szót az emberek különböző ritmusban képesek kiejteni, de ugyanez igaz egy embernél ugyanazon szó kétszeri kiejtésénél. Meg kell tehát találnunk azt a technikát, amivel a megfelelő dolgok lesznek összeillesztve.

Erre három módszer létezik, ebből kettőnek statisztikai megfigyelés az alapja, a harmadik sablon (template) alapú.

- HMM – Hidden Markov Model (statisztikus);
- ANN – Artificial Neural Network (statisztikus);
- DTW – Dynamic Time Warping (sablon alapú).

HADTUDOMÁNYI SZEMLE

2017. X. évfolyam 2. szám

A HMM módszernél a ritmikai változások figyelembevétele történik (pl. izolált szavas beszédfelismerésnél). A modell lépni kényszerül minden 10. ms-ban, de nem kényszerül ellépni onnan. Így ezzel a technikával alapvetően ki lehet küszöbölni az alapvető ritmusbeli különbségeket. Lehetséges ugró él is, ha valamelyik hangot nem ejtjük ki. Miért hívják ezt rejtett Markov modellnek? Azért, mert a véges automatáknál megszokott módtól eltérően itt nem tudjuk, hogy a folyamat milyen állapotban van. Erre a megfigyelésből kell következtetnünk. A modell kiad P_T vektort, miközben az állapotokban eljut az „N” állapotig. A megfigyelési sorozatot O-val jelöljük (mint observation). Aközben az emisszióból (azon vektorok, amelyeket a Markov folyamat emittál) nem tudjuk megállapítani, hogy melyik állapotban vagyunk. Egy állapothoz sokféle vektor tartozhat, ezért inkább valószínűségekkel számolunk. $P(O_t|q_j)$ egy valószínűségre jellemző érték, ahol q_j a j-edik állapot O_t pedig folytonos értékkészletű. Minden egyes ponthoz sűrűségfüggvény-értéket rendelünk. Ezt a sűrűségfüggvényt adatbázisokból kell meghatározni.

A Dinamikus idővetemítés (Dynamic Time Warping) módszere csak egyszerű felismerési feladatokra alkalmas, ezért részletesen nem foglalkozunk a módszerrel, annak ellenére, hogy akkor is jól jöhet, ha egy új szótárelem felvételére csak bemondás útján van lehetőség.

A Mesterséges neurális háló (ANN) izolált szavas beszédfelismerésre alkalmas, illetve nemlineáris osztályozásra feature extraction részeként, ezekre bevált, de nem jelent lényegi többletet a GMM(Gauss Mixture Modell)-hez képest.

A következő lehetőséget akkor kell alkalmaznunk, ha nem csak izolált szavakat akarunk használni. Az alapötlet az, hogy az izolált szavas HMM Markov modelljeit összevonjuk, így a rendszerbe nyelvtani információt viszünk a modell topológián keresztül Ebben az esetben előnyt jelenthet a statisztikai megközelítés, amely elég egyszerű, egyúttal itt is a leghatékonyabb megközelítésnek bizonyult. Azonban számtalan negatív jelenség is megjelenik, mert a természetes nyelvek nem írhatók le determinisztikus nyelvtannal, illetve a szöveg nem ugyanaz, mint a hangsor, tehát kiejtés-modellezés is szükséges, valamint méret probléma is felmerülhet. Ugyanakkor a magyar nyelv esetében előnyt jelenthet, hogy a kiejtésre az írott formából jól következtethetünk.

A beszédfelismeréshez szervesen kapcsolódik a beszélő-felismerés is, amelynek fő kérdése, hogy megállapítható-e az elhangzó beszéd alapján a beszélő személye, ha ismerjük az illetőt, illetve ha nem. (Katonai feladatok végrehajtása során nagyon kicsi a valószínűsége annak, hogy ismerjük a beszélőt!) A vizsgálat során a kiinduló feltételezés az, hogy az agyunkban létrejövő neurális spektrogram tartalmazza a beszélő ismérveit. De vajon ez az információ, ezek a paraméterek olyan mértékben jellemzőek-e a beszélőre, mint pl. az ujjlenyomat?

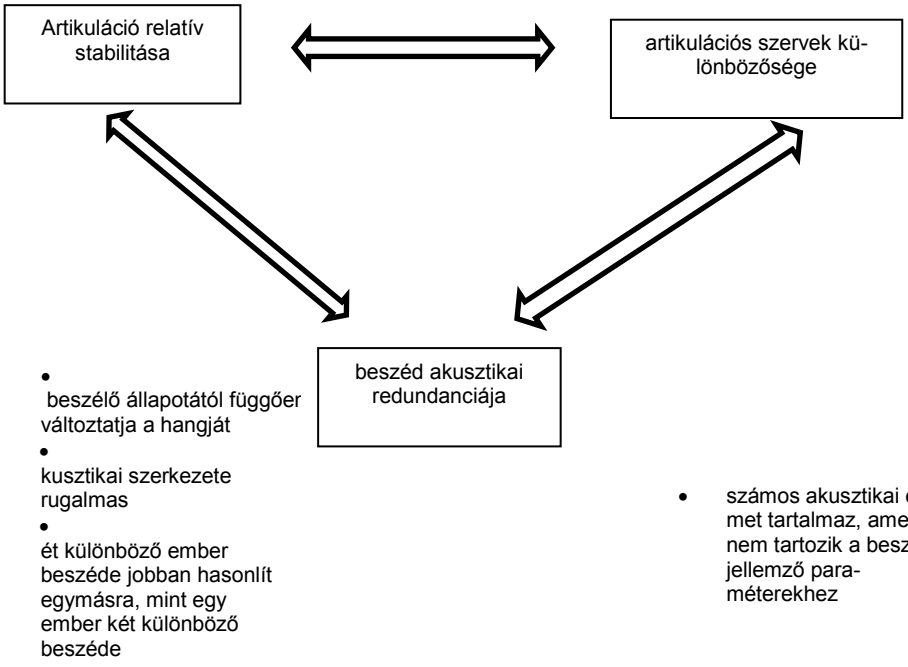
A beszélő-felismerésnek alapvetően két iránya van:

— az n lehetőségből kizárható-e nx személy;
— az n lehetőségből melyik az nx esemény;
— a kettő kombinációja: benne van-e, és ha igen, ki lehet?

A beszélő-felismerés paradoxona a következő ábrával szemléltethető:

HADTUDOMÁNYI SZEMLE

2017. X. évfolyam 2. szám



6. sz. ábra: A beszélő-felismerés problémái (Forrás: saját szerkesztés.)

A beszélő vizsgálata során számtalan egyéb problémára is megoldást kell keresni bármely alkalmazandó szoftvernek. Csak felsorolásszerűen megjelenítve nézzük meg melyek lehetnek ezek a problémák illetve vizsgálandó kérdések és a felsorolások mögött zárójelben megjelenítjük a katonai alkalmazás relevanciáját:

- A beszélő azonosítása a hangszínezet alapján (fonetika) (Nem releváns!)
- Milyen mértékben jellemző az emberre a hangja, beszéde? (Nem releváns!)
- Miképpen határozható meg az egyéni hangszínezet? (Nem releváns!)
- Melyik beszédképzési konfigurációval mutatja a legszorosabb kapcsolatot? (Nem releváns!)
- Miként fejezhető ki a hangszínezet: akusztikai-fonetikai, percepciós-fonetikai vagy mind együtt? (Nem releváns!)
- Milyen szubjektív benyomásokat kell figyelembe venni a hangszínezet tekintetében? (Nem releváns!)
- A beszéd elhangzásának helyszíne (pragmatika). (Releváns!)
- Szemantikai szerkezet, jelentés (milyen kifejezéseket válogatok ki) – stratégiák. (Előzetesen leszabályozott!)

- Szintaktikai szerkezet (milyen sorrendben, hogyan mondom őket) – transzformációs szabályok. (Előzetesen leszabályozott!)
- Fonológiai szabályok, fonetikai szerkezet. (Nem releváns!)
- Artikulációs működések. (Nem releváns!)
- Akusztikai hullámforma. (Nem releváns!)
- Egyetlen beszélő egyetlen hangja laboratóriumi körülmények között és egyéb tényezők közül többet is elhanyagolva is nagyon sokféle lehet. (Nem releváns!)
- Az alaphang magasságának változása: a korral változik (a nőké felnőttkorig kicsit mélyül, de alapvetően nem változik, idősebb korra jobban mélyül, a férfiaké kamaszkorban nagyon mélyül, felnőttkorban egész mély, majd idősebb korban újra magasodik). (Nem releváns!)
- Az érzelmek is befolyásolják (öröm, bánat, stb., de ez a kettő azonosítható a legjobban) a beszédet. (Nem releváns!)
- A prozódia jellemzőbb, mint a szegmentális szerkezet: alaphang-magasság, tempó, intenzitás, szünetstruktúra, ritmus, artikulációs változás. (Nem releváns!)
- A beszéd tempójának változásánál meghatározóak a környezeti feltételek is. (Releváns!
- ha automatizálni szeretnénk a beszélő-felismerést, az időviszonyok változása gondot jelenthet: két beszédminta tempója nem azonos, akkor most ugyanaz a beszélő volt-e vagy sem, továbbá befolyásolhatja a beszéd sebességét az érzelmi állapot és legfőképpen a zajviszonyok. (Előre leszabályozott!)
- Sok leadott jelentés után monoton jellegű lesz a beszéd (moduláció csökken). (Előre leszabályozott!)

ÖSSZEFOGLALÁS, KÖVETKEZTETÉSEK

Folyamatosan folynak a lehetőségek vizsgálatai, hogy az információk széleskörű kezelésével növelhetővé váljon egy-egy tevékenység hatékonysága. A vizsgált mesterséges intelligencia területének egy viszonylag új eleme, a különböző típusú munkák hatékonyságát növelő lehetőség a beszédfelismerő rendszerek alkalmazása lehet a beérkező információk kezeléséhez. A katonai feladatok végrehajtása során a beszédfelismerő rendszer alkalmazása éppen azokon a területeken jelenthet előrelépést, melyeken belül a legutóbbi felmérések jelentős lemaradásokat mutattak. A különböző többnemzetiségű műveletek (gyakorlatok) ideje alatt az egységes alapelveken alapuló információátadás kiemelt kérdésként vetődik fel a NATO hadseregeinek együttműködése során. A beszédfelismerő rendszer gyakorlati alkalmazása még nagyon rövid időre tekint vissza és még viszonylag szűk azoknak a szakmáknak a száma, melyekben az alkalmazás feltételei megteremtődtek. Végkövetkeztetésként kimondhatjuk, hogy a mesterséges intelligencia alkalmazása a beszédfelismerésben a katonai feladatok végrehajtásánál is jelentős távlatokat jelenthet, növelve ezzel a feladatok végrehajtásának hatékonyságát. Ez a cikk ezért vállalkozott az elméleti

alapok ismertetésére és reményeim szerint megteremtette az alapot a gyakorlati megvalósítás lehetőségeinek elemzésére, amely majd a cikk második részében kerül kifejtésre.

FELHASZNÁLT IRODALOM

1. Vicsi Klára: A beszédfelismerés fejlődése, a mai beszédfelismerési módszerek ismertetése, alpha.tmit.bme.hu/speech/docs/education/beszedkomm_felism3.PPT, 2006, Letöltve: 2017.05.20.
2. Wilcoxon, F.: "Individual Comparisons by Ranking Methods." Biometrics 1, 80-83, 1945.
3. L. R. Bahl, P. F. Brown, P. V. de Souza, R. L. Mercer: Maximum mutual information estimation of hidden Markov model parameters for speech recognition. Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, Vol. 1, pp. 49–52, Tokyo, Japan, April 1986.
4. L. E. Baum, J. A. Eagon: An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model of ecology. Amer. Math. Soc. Bull., Vol. 73, pp. 360–362, 1967.
5. M. H. Cohen: Phonological structures for speech recognition. Ph.D. dissertation, University of California, Berkeley, USA, 1989.
6. Creutz, M. and Lagus, K.: “Unsupervised Morpheme Segmentation and Morphology Induction from Text Corpora Using Morfessor 1.0.”, Publications in Computer and Information Science, Report A81, Helsinki University of Technology, March, (2005)
7. Czap László: Audiovizuális beszédfelismerés és szintézis, PhD értekezés, BME, Budapest, 2005.
8. Gordos G., Takács Gy.: Digitális beszédfeldolgozás, Műszaki Könyvkiadó, Budapest, 1983.
9. Németh B., Mihajlik P., Tikk D., Trón V.: Statisztikai és szabály alapú morfológiai elemzők kombinációja beszédfelismerő alkalmazáshoz. MSZNY 2007: V. Magyar Számítógépes Nyelvészeti Konferencia, pp. 95-105, Szeged, 2007.