

AZ AUTOMATIKUS BESZÉDFELISMERÉS ÉS A GÉPI FORDÍTÁS

Kiefer Ferenc

A gépi fordítás kettős értelemben komplex tudomány: egyrészt megvalósításához több határtudomány eredményeire van szükség, másrészt eredményei több határtudomány számára értékesek. A gépi fordítás megvalósításához nyelvészeti és matematikai alap kutatásokra, továbbá műszaki jellegű alkalmazott kutatásokra van szükség. Ezek a kutatások azonban nemcsak a gépi fordítás számára hasznosíthatók, hanem szinte mindenütt, ahol gép és ember együttműködésére van szükség. Így ezek a munkák az automatizálás fontos láncszemét alkotják.

A gépi fordítás hasznosításáról lehetnek viták, hiszen viszonylag még nagyon keveset tudunk a nyelvről és a fordításról. De a gépi fordítás előkészítésével kapcsolatos eredményeknek felhasználásáról más területen máris tudunk. Így például mind a nyelvészeti, mind a matematikai alap kutatások eredményei felhasználhatók a gépi adatfeldolgozásban, az információ visszanyerésben, szótárak összeállításában, fordításelméletben, beszédfelismerésben.

E cikk keretén belül csak az automatikus beszédfelismerés kérdésével kívánunk foglalkozni. Ez a kérdés nemcsak a fent vázolt módon kapcsolódik a gépi fordításhoz, hanem közvetlenül is. Tudniillik már évek óta kísérleteznek ugynevezett tolmácssgéppel, olyan géppel tehát, amely a beszédet automatikusan felismeri, regisztrálja és lefordítja, majd outputként az idegen megfelelőt adja. Persze ennek megvalósítása távolról sem csak pénzkérdés, mint ahogyan azt egyik napilapunk állította, hiszen nagyon sok alapvető problémát még eddig egyáltalán nem sikerült megoldani. Még hosszú évek munkájára van szükség ahhoz, hogy a gyakorlati megvalósítással komolyan foglalkozni lehessen.

Mi is az automatikus beszédfelismerés? A beszédfelismerés folyamata lényegében abban áll, hogy a gép a hang-hullámokat átalakítja jelekké, amelyek nyelvi egységeknek felelnek meg. A felismerés pontossága a hibák számával mérhető. Más szóval, a géptől azt követeljük, hogy ahányszor ugyanaz a hanghullám jelentkezik, mindannyiszor ugyanaz a jel szerepeljen outputként. Ha ez nem következik be, a gép hibát követ el.

Olyan gép konstruálása, amely tetszőleges nyelven beszélőnek a beszédét felismerne, mégha lehetséges lenne is, nagyon komplikált és hatalmas méretű lenne, még akkor is, ha az elektronikus számológépek mai szintjén gondolkozunk. A feladatot tehát csökkenteni kell. Mégpedig a géptől csak azt szokták megkövetelni, hogy csak egy-két beszélőnek a beszédét ismerje fel, sőt még a beszédre /vagyis arra, hogy a beszélő mit mondjon/ is szoktak különböző megszorítást tenni.

A beszéd felismerésének két módját ismerjük. A kezdeti kísérleteknél kizárólag az úgynevezett akusztikai felismerést alkalmazták. Ez azt jelenti, hogy a gép tulajdonképpen csak akusztikai mutatókat hasonlít össze. Az összehasonlításához csak a bejövő akusztikai jelet és a tárolt akusztikai jeleket veheti figyelembe. A Bell Társaság /Bell Telephone Laboratories/ kutatói /DAVIS, BIDDULPH, BALASHEK és DUDLEY/ számokkal végeztek kísérletet, a számokat egy beszélő ejtette ki és az elért pontosság /pusztán akusztikai felismerésnél/ 97 %-os volt. A második kísérletet OLSON és BELAR végezte, mégpedig 10 egyszótagu angol szóval, melyeket szintén egy beszélő ejtett ki. Az elért pontosság 98 %-os volt. Bármilyen biztatók voltak is ezek az eredmények, hamarosan kiderült, hogy a szótár megkészszerzésével a nehézségek exponenciálisan nőnek. Viszont már ezek is vezettek gyakorlatilag felhasználható eredményhez: telefonhívás elvileg most már számok beolvasásával is lehetségessé vált.

A további kutatások az akusztikai felismerést nyelvi felismeréssel egészítették ki. Szavak helyett fonémák felismerésére törekedtek, hiszen ha a gép képes egy nyelv bármely fonémáját tetszőleges környezetben felismerni /természetesen kielégítő pontossággal/, akkor a szavak felismerése már nem fog lényeges többlet-munkát jelenteni. FRY a londoni University College-ben új módszert dolgozott ki a beszéd felismerésre, amely a fenti megfontoláson alapszik. Megjegyzendő, hogy FRY segítőtársa matematikai és műszaki oldalról a magyar származású P.DÉNES volt.

A gép 13 angol fonéma felismerésére volt képes, ugyanakkor azonban minden olyan angol szó felismerésére is, amelynek fonémái a 13 adott fonéma között vannak. A fonéma-felismerés pontossága 60 %-os volt. Látszólag ez a pontosság sokkal kisebb, mint a fentemlitett szó-felismerés pontossága. Ha azonban tekintetbe vesszük, hogy a 13 fonéma az angol fonémarendszer 1/3-át alkotja, míg viszont a 10 szó az angol szókincsnek alig 0,001 %-át, akkor a fenti pontosság egyáltalán nem megvetendő.

Ez a pontosság tovább növelhető két irányú módszerfejlesztéssel. Egyrészt az akusztikai felismerést kell javítani. WIREN és STUBBS javaslata szerint nem elég egy-egy akusztikai tulajdonságra csak "igen-nem"-mel válaszolni. A bináris módszer nem vezethet kielégítő eredményhez. Az akusztikai tulajdonságokat /éppúgy mint a fonémák tulajdonságait/ kétdimenziós mátrixban kell megadni. Más szóval ez azt jelenti, hogy egy-egy akusztikai jel több akusztikai alaptulajdonságnak is "eleget tehet" egyidőben. Aszerint, hogy hány alaptulajdonságnak tesz eleget, más és más valószínűséggel reprezentál egy adott fonémát. Ehhez az elvhez járul még a jelek időtartamának és intenzitásának figyelembevétele.

A másik kiegészítés nyelvészeti természetű. Ebben FRY és munkatársai a beszéd emberi megértésének folyamatát modellálták. Ez azt jelenti, hogy a beszédet halló ember a hangokat fonémaként ismeri fel /azaz osztályozza, más szóval, felismeri, hogy az elhangzott "hang" anyanyelvének fonémája/. A fonémából megalkotja azután a morfémákat, a morfémákból a szavakat és a szavakból a mondatokat. Azaz minden "szinten" bizonyos megszorítások alkalmazhatók a fonémákra. Ezek a megszorítások segítik az embert is a hibák kiküszöbölésére. Ezért van az, hogy sokszor a hibásan beszélőt is megértjük. Sőt a tapasztalat azt mutatja, hogy a legtöbb hiba kiküszöbölése nyelvi "meggondolások" alapján megy végbe, nem pedig akusztikai alapon. Kísérleti tény, hogy ha egy angol anyanyelvű egyén a nyelvi közlés 50%-át ismeri csak fel a fonémák szempontjából pontosan, akkor is megérti a közlést. Az automatikus beszédfelismerésben is ezt a folyamatot kell, amennyire lehet, figyelembe venni.

FRY és munkatársai az akusztikai felismerést először csak fonéma diagrama-statisztikával egészítették ki és máris jobb eredményhez jutottak. Akusztikai úton szavakat csak 24 %-os pontossággal tudtak felismerni, a fenti kiegészítéssel a pontosság 44 %-os volt.

A fonéma-pár, fonéma-hármas stb. statisztikai vizsgálata már csak azért is fontos, mert bizonyos kombinációkat teljesen kizár, másokat pedig valószínűségük szerint rangsorolja. Így például magyarban nem fordulhat elő három mássalhangzó egymásután, vagy nem fordulhat elő /ha a szóösszetételt figyelmen kívül hagyjuk/ az scs kettős s.i.t.

A modern számítógépek lehetővé teszik azt, hogy hatalmas mennyiségű információt tároljunk. Így például a szavak fonéma-sorokként tárolhatók a gépben, a szófelismerés tehát ily módon egyszerűen, visszavezethető fonémfelismerésre. A szófelismeréstől azonban még hosszú ut vezet a mondatfelismerésig. Ha a gép szavak egymásutánját ismeri fel, még nem ismeri fel magát a mondatot. /Tehát nem is tudja lefordítani./ Ezért az eddigieken kívül még grammatikai információkat is kell a gépnek tárolnia. Mégpedig valószínűleg predikciók formájában. Más szóval, a gépnek magának kell a szó-sor és a szavakra vonatkozó információk /tehát például, hogy ez és ez a szó főnév, hímnemű stb./, továbbá a grammatikai predikciók /azaz azoknak a szabályoknak az összessége, amely megmondja, hogy bizonyos szavak milyen grammatikailag megengedhető nagyobb egységet alkothatnak/ felhasználásával a mondatot rekonstruálnia.

Tulajdonképpen itt már pontosan átvehető az a módszer, amelyeket a gépi fordítás kidolgozott. Az akusztikai felismerésen túl mindaz, amit felhasználunk, már gépi analízis, úgy ahogyan azt a gépi fordításnál alkalmazzák. És itt kanyarodhatunk vissza eredeti gondolatmenetünkhöz: a gépi fordítás eredményei számos más területen használhatók. Még a di-, tritetragramm stb. statisztikákra is szükség lehet a gépi fordításnál, illetőleg annak előkészítésénél. Különösen áll ez olyan nyelvekre mint a magyar, amely tehát végződésekben igen gazdag, tehát nem mindig egyszerű eldönteni, hogy hogyan építsünk fel a magyar nyelvre egy-egy részalgoritmust. Ehhez sokszor tudnunk kell valamit azokról a kisebb egységekről is, amelyekből a magyar szókincs felépül.

...□...

BIBLIOGRÁFIA

- DAVIS, K.H., BIDDULPH, R., and BALASHEK, S., Automatic recognition of spoken digits, in Communication Theory /W.Jackson, ed./ pp. 433-441, Academic Press, New York, 1956.
- DENES, P., MATHEWS, M.V., Spoken digit recognition using timefrequency pattern matching, J. Acoust. Soc. Amer., 1960. 11.sz. 1450-1455.p.
- DUDLEY, H., BALASHEK, S., Automatic recognition of phonetic patterns in speech, J. Acoust. Soc. Amer., 1958. 6.sz. 721-732.p.
- FRY, D.B., DENES, P., On presenting the output of a mechanical speech recognizer, J. Acoust. Soc. Amer. 1957. 3.sz. 364-367.p.
- FRY, D.B., DENES, P., The solution of some fundamental problems in mechanical speech recognition, Language and Speech, vol 1/ part 1-1958. jan-march, 35-58.p.
- FRY, D.B., DENES, P., An analogue of the speech recognition process, Mechanization of Thought Processes: National Physical Laboratory Symposium No.10, 275-283.p. London, H.M. Stationary Office, 1959.
- OLSON, H.F., BELAR, H., Phonetic typewriter, J. Acoust. Soc. Amer., 1956. 6.sz. 1072-1081.p.
- STEVENS, K.N., Towards a model for speech recognition, J. Acoust. Soc. Amer., 1960. 1.sz. 47-55.p.
- WIREN, J., STUBBS, H.L., Electronic binary selection system for phoneme classification, S. Acoust. Soc. Amer., 1956. 6.sz. 1082-1091.p.

ooo°ooo

KIEFER, F.: Automatic Speech Recognition and Machine Translation

The present paper deals with different aspects of automatic speech recognition. First, automatic speech recognition was carried out solely on acoustic basis. However, this approach turned out to be unsatisfactory. It seemed reasonable to complement the method acoustic by a linguistic one. The linguistic information used in automatic speech recognition consists of different phoneme statistics and on the other hand, of grammatical information of various kinds. The grammatical information allowed for in automatic speech recognition links up with some results obtained in the mechanical analysis of texts well known in machine translation. In this way, automatic speech recognition may profitably make use of the different methods developed within the investigations on mechanical analysis.

←+0+←