Dr. Csapó Tamás Gábor

A mesterséges intelligencia és alkalmazása a beszédtechnológiában

Az első számítógépek megjelenésével, 1952-ben alkották meg a dámajátszó programot. A dáma az első olyan játékok egyike, amelynek fejlesztése során mesterséges intelligenciát (MI) alkalmaztak [4]. A mesterséges intelligencián belül a gépi tanulás olyan rendszerekkel foglalkozik, melyek tanulni képesek, azaz tapasztalatokból tudást generálnak. A gyakorlatban ez azt jelenti, hogy a rendszer példa adatok, minták alapján képes önállóan vagy emberi segítséggel szabályszerűségeket vagy szabályokat felismerni és meghatározni. A gépi tanulás egyik ága mesterséges neurális hálózatokat alkalmaz, melyek eredetileg az élő idegsejt közötti analógiára készültek: az idegsejtben található dendritek, a sejttest és az axonok is szerepelnek a hálózat matematikai modelljében [2]. Bár tagadhatatlan a két elemi egység közötti párhuzam, fontos kiemelni, hogy az élő idegrendszer működése már mai tudásunk alapján is sokkal összetettebb ennél. A mesterséges neurális hálózatok fogalma az 1980-as/1990-es években már nem először járta be a világot, kutatók ezrei és az ipar is óriási reményeket fűzött hozzá. Idővel azonban az akkor kidolgozott eljárás elérte korlátait, és jelentősebb eredmények hiányában a gépi tanulás szakértői más, abban az időben hatékonyabb módszerek felé fordultak. A korábbi nagy reményeket a kiábrándultság, a „neuronhálók tele" váltotta fel. A 2000-es évek közepétől azonban új tudományos eredmények születtek, emellett napjainkban nagyságrendekkel nagyobb számítási teljesítmény és több adat áll rendelkezésre. Ma a kutatók a mély tanuláson alapuló neurális hálózatok segítségével a legtöbb tudományterületen minden korábbinál jobb, pontosabb eredményt érnek el. Jelenleg rengeteg MI kutatás zajlik a világban, számos ország és magánvállalat is nagyon komoly összegeket fektet ebbe a területbe.

Dr. Csapó Tamás Gábor

Dr. Csapó Tamás Gábor

A mesterséges intelligencia és alkalmazása a beszéd-technológiában

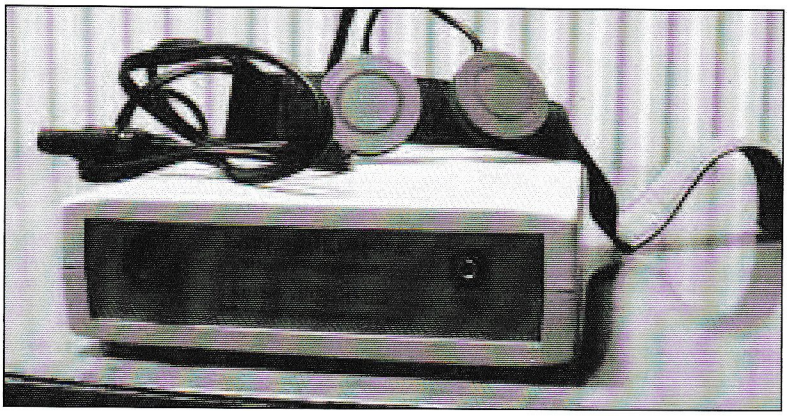
2016-ban a Google által kifejlesztett mesterséges intelligencia első ízben vert meg egy profi játékost a Go kínai játékban, amely a legösszetettebb táblajáték a világon. Ugyanakkor a mesterséges intelligencia-fejlesztésnek nincs egy világosan kijelölt végterméke, ehelyett különféle technológiák, eszközök és módszerek folyamatos fejlesztéséről van szó, melyeket meglehetősen lazán köt össze az MI gyűjtőfogalom. És mire képesek napjaink legfejlettebb mély tanuló rendszerei? Kiválóan tudják modellezni az emberi látást, hallást, és beszédet: például az automatikus képfelismerés és beszédfelismerés területén, valamint a művészet egyes ágaiban néhány éve az embert is meghaladó pontosságot értek el a mesterséges intelligenciát alkalmazó rendszerek. Fontos azonban kiemelni, hogy az MI nem jövőt látó varázsgömb vagy egy fekete doboz, ami minden problémát megold. De a tudomány egy-egy specifikus területén sikerült már segítségével kimagasló eredményeket elérni.

A következőkben azt mutatom be, hogy a gépi beszédkeltés során, mely a beszédtechnológia tudományterületének egyik ága, hogyan alkalmazható a mesterséges intelligencia. Ehhez először vizsgáljuk meg a természetes beszédláncot. Az emberi kommunikációnak számos alapvető feltétele van, például a két partnernek a világról alkotott modellje nagymértékben meg kell, hogy egyezzen. Ez a modell hosszú időszak tanulási folyamata révén alakul ki. A modellhez kapcsolódóan fogalmazódik meg az agyban a beszélő személy kommunikációs szándéka, ami a beszédszerveken keresztül alakul fizikai jelekké. Ezek a fizikai jelek egy átviteli csatornán (természetes közegben a levegőn, gépi megoldásnál hang- vagy videotelefonon) keresztül jutnak el a hallgatóhoz. A hallgató személy érzékszervei adják tovább a megfelelő jelfeldolgozás után az észlelés számára az információt. Beszédtechnológiának a természetes beszédlánc egy vagy több elemének gépi megvalósítását tekintjük [3]. Ennek egyik iránya a gépi beszédkeltés (más néven gépi beszéd-előállítás, szövegfelolvasás vagy beszédszintézis). A probléma klasszikus megoldásai mind szabály alapúak (pl. Kempelen Farkas beszélőgépe 1792-ből, a HungaroVox beszédszintetizátor az 1980-as években, illetve a ProfiVox szövegfelolvasó az 1990-es években), azaz ember által írt szabályok alapján készítik a bemeneti karaktersorozatból a beszéd kimenetet. Napjainkban a gépi beszédkeltés modern megoldásai mesterséges intelligenciát alkalmaznak: a szabályokat automatikusan tanulják meg az adatok alapján. A gépi tanuló rendszer bemenete tehát egy,

esetenként 10-100 órányi beszédet tartalmazó felvételsorozat, melyből az MI képes megtalálni a betűk és a keletkezett beszédjel közti összefüggést [5]. A mai okos-telefonokon már magyar nyelvre is elérhetőek ezek a gépi beszédkeltő rendszerek (pl. Android, beállítások, szövegfelolvasó).

■ A beszédtechnológián belül az ún. „némabeszéd-interfésznek" nevezett terület célja olyan módszerek kidolgozása, melyek a néma artikulációból gépi beszéd előállítására képesek [1]. Az artikuláció (a beszédképző szervek koordinált mozgása) és az akusztikum (a keletkező beszédjel) kapcsolata az 1700-as évek óta foglalkoztatja a beszédkutatókat. Ahhoz, hogy a beszédképző szervek (pl. hangszalagok, nyelv, ajkak) mozgását vizsgálni tudjuk, speciális eszközökre van szükségünk, mivel a legtöbb ilyen szerv nem látható folyamatosan beszéd közben. Lotz János az 1960-as években, Szende Tamás az 1970-es években, majd Bolla Kálmán az 1980-as években röntgenfilm (ún. röntgenogram/kinoröntgenografikus vizsgálat) technológiával vizsgálta a magyar beszéd arti-

Elektroglottográf



kulációját. Az 1980-as évek röntgenes kísérletei után hosszú ideig nem történtek magyar nyelvű artikulációs kutatások. 2016-ban alakult az MTA-ELTE Lendület Lingvális Artikuláció Kutatócsoport, melynek célja a magyar nyelv artikulációjának vizsgálata modern eszközökkel, többek között ultrahanggal. Az ultrahang előnye, hogy egyszerűen használható, elérhető árú, valamint nagy felbontású (akár 800×600 pixel) és nagy sebességű (akár 100 képkocka/másodperc) felvétel készíthető vele. A jó térbeli felbontás azért fontos, hogy a nyelv alakjáról minél pontosabb képet kapjunk, míg a jó időbeli felbontás ahhoz szükséges, hogy a beszédhangok képzésének gyors változását is vizsgálni tudjuk. Az artikuláció és az akusztikai kimenet kapcsolatát az elmúlt években gépi tanulás alapú eszközökkel is vizsgáltuk. Az artikuláció-akusztikum konverzió eredmé-

nyei a szakirodalomban elsősorban a „némabeszéd-interfész" rendszerek fejlesztéséhez járulnak hozzá. Ezeknek lényege, hogy az artikulációs szervek hangtalan mozgását felvéve (pl. nyelvultrahang segítségével) a gépi rendszer ebből beszédet szintetizál, miközben az eszköz használója valójában nem ad ki hangot. Ez egyrészt a beszédsérült embereknek (pl. gégeeltávolítás után) lehet hasznos, másrészt potenciálisan alkalmazható zajos környezetben történő beszédhang kiadására, kiabálás nélkül.

■ Az MI alapú beszédtechnológia, ezen belül a gépi beszédkeltés a beszédkutatás alapkérdéseinek megválaszolása mellett hasznos lehet a vakok és gyengén látók számára (pl. képernyő felolvasó alkalmazás okos-telefonon, beszélő ATM) nyelvoktatásban illetve beszéd-rehabilitációban. A mindennapi ember a gépi beszédkeltéssel leggyakrabban a hangos utas-tájékoztatás keretében találkozhat pl. MÁV vasútállomásokon.

■ Összefoglalásként azt mondhatjuk, hogy a mesterséges intelligencia egy gyűjtőfogalom, amely különféle technológiákat, eszközöket és módszereket alkalmaz. A lehetséges jövőképek közé tartozik, hogy eljöhet az eddigi legnagyobb „MI-tél" (kiábrándultság, mint az 1980-as években), vagy épp az ellenkezője is előfordulhat, vagyis hogy néhány évtizeden belül a számítógép olyan okos lesz, mint az ember.

■ Források

[1] Csapó Tamás Gábor, Gosztolya Gábor, Grósz Tamás, Tóth László, Markó Alexandra: Némabeszéd-interfész nyelvultrahanggal (Beszédgenerálás artikulációs mozgás alapján),= https://www.researchgate.net/project/DNN-based-Silent-Speech-Interfaces-using-Ultrasound-Tongue-Imaging-and-Electromagnetic-Articulography

[2] Gyires-Tóth Bálint: Beszélő számítógépek mély gondolatokkal,= http://www.eletestudomany.hu/neuralis_halozatok

[3] Németh Géza: Kempelentől a WaveNet-ig: a gépi beszédkeltés tudományának fejlődése= http://real.mtak.hu/88670/

[4] Ürögi Réka: Mi az a gépi tanulás és hogyan működik?= https://www.webonic.hu/blog/2018/11/27/mi-az-a-gepi-tanulas-es-hogyan-mukodik/

[5] BME Beszédtechnológia és Intelligens Interakciók Laboratórium=http://smartlab.tmit.bme.hu