

## Corectarea deficiențelor de auz și vedere cu ajutorul serviciilor cognitive

*Iordachescu Anca-Mihaela, grupa 407*

Este bine cunoscut faptul că, aflându-ne într-o societate în care impactul tehnologiei este din ce în ce mai pronunțat, soluțiile software create sunt menite să se adapteze pe nevoile umane și să contribuie la transformarea lumii noastre într-un mediu cât mai inclusiv cu ajutorul accesibilității inteligenței artificiale. Cercetarea și dezvoltarea produselor software care au la bază tehnologia inteligență artificială reprezintă un pas crucial, în special, pentru traiul zilnic al unei persoane care se confruntă cu diferite dizabilități și care are nevoie în permanență de asistență socială pentru a conviețui în societatea actuală.

Serviciile cognitive reprezintă o soluție pentru îmbunătățirea experienței de viață a persoanei cu deficit prin construcția unor modele de învățare automată care îndeplinesc, într-o mare masură, sarcini cognitive asemănătoare comportamental creierului, sarcini cognitive pe care acesta nu le are dezvoltate. Platforma de cloud-computing Azure oferă una din cele mai cunoscute și variate game de servicii cognitive care poate să înlocuiască foarte ușor nevoia unei persoane cu o dizabilitate de a fi asistată social de către o altă persoană în mod constant. Astfel, aceste servicii permit un nivel mai ridicat de accesibilitate, încredere și independența persoanelor în cauză oferindu-le acestora posibilitatea de a descoperi lumea pe cont propriu într-un mediu inclusiv.

O primă dizabilitate a cogniției este reprezentată de insuficiența vizuală. Acestor persoane le este dificil sau, chiar, imposibil să se deplaseze sau să aibă un simț al orientării în spațiu, trebuind să fie asistați sau supervizați întotdeauna. Ca răspuns la această problemă, Azure oferă componenta de Vision în care sunt înglobate diferite soluții de vedere artificială care rezolvă diferite sarcini care includ capacitatea de a vedea. Astfel, cu ajutorul unui dispozitiv care are acces la o cameră video, persoanele în cauză pot să aibă o idee mult mai clară despre ce se întâmplă în jurul lor folosind aplicații care se bazează pe Computer Vision API. *Computer Vision API* permite analizarea input-urilor vizuale cu scopul de a identifica obiecte sau evenimente. De exemplu, cu ajutorul acestui serviciu persoanele cu deficit de vedere pot să realizeze diferite activități din rutina zilnică precum pregătirea unei mese întrucât sistemul este capabil să recunoască ingredientele de care acesta are nevoie pentru a găti, diverse cicluri de mers prin evitarea obstacolelor detectate etc. De asemenea, o altă

nevoie importantă a unei persoane este înțelegerea textului scris, indiferent dacă vorbim despre scrisori, facturi sau marcaje turistice. Azure oferă serviciul de Optical Character Recognition (OCR) care folosit împreună cu serviciul cognitiv specific auzului, numit *Text to Speech*, redă utilizatorului toate informațiile text din jurul acestuia. Este de menționat ca aceste servicii implementate în Azure oferă o gama largă de limbi, având o deschidere spre un număr foarte mare de potențiali beneficiari. Tehnologia din spatele acestui feature se bazează pe inteligență artificială, mai exact pe o rețea adanc neuronală, ce folosește un proces numit *transfer learning* pentru a elimina din timpul de antrenare al acestuia și de a păstra anumite informații învățate de alte rețele deja antrenate pe seturi complexe de date, cu milioane de imagini, precum COCO (Common Objects in Context) [1]. Setul de date pentru antrenare al OCR conține 182 de imagini cu diferite semnături, iar cel de antrenare 100. Modelul folosit pentru crearea acestui serviciu este Faster R-CNN [2], o rețea adanc convolutionala.

Deși este important pentru un nevăzător să identifice conținut în imagini sau video, în ceea ce privește oamenii nu este, în general, de ajuns să identifice oamenii, ci și să îi recunoască. Componenta Vision include Face API care analizează fețe în date vizuale folosind modele de recunoaștere facială state-of-the-art. Mai mult decât atât, acest API poate fi apelat și pentru detecția emoțiilor analizând fața persoanei, persoana cu deficit vizual putând fi astfel informată dacă persoanele din jurul lor sunt triste, fericite, speriate etc.

O altă componentă care, împreună cu componenta Vision, ar facilita traiul persoanelor cu deficit vizual este componenta Decision. *Anomaly Detection API* din cadrul acestei componente identifică evenimente anormale care se pot întâmpla în jurul persoanei în cauza și îl pot atenționa, de exemplu, cu privire la acțiuni periculoase din apropiere (demolarea unui bloc, accident de mașină etc). Mai mulți algoritmi stau în spatele acestei funcționalități, în funcție de secvențele de timp introduse, printre care: transformări Fourier, ESD (Extreme Studentized Deviate), STL Decomposition [3], Dynamic Threshold.

În continuare, tot în cadrul acestei componente, există funcționalitatea *Speaker Recognition*, ce poate recunoaște identitatea unei persoane în baza vocii sale, în cadrul unui apel telefonic sau a unei înregistrări audio. Persoanele nevăzătoare pot folosi aceasta unealtă pentru a-și da seama mai ușor de identitatea persoanei din partea cealaltă a receptorului. Există două tipuri de recunoaștere, *text-dependent* și *text-independent*. Prima dintre acestea

implică ca vorbitorul să utilizeze aceeași frază în timpul procesului de înrolare în baza de date a vorbitorilor și al celui de verificare al vocii, în timp ce în cadrul celei de-a doua categorii, vorbitorul poate utiliza un limbaj obișnuit.

O altă componentă din cadrul Azure Cognitive Services este *Speech*, ce conține diferite module în vederea îmbunătățirii proceselor legate de vorbire. Cea mai relevantă funcționalitate pentru persoanele cu insuficiență auditivă este *Speech to text*, ce permite convertirea discursului unui vorbitor într-un text scris, în timp real, în majoritatea limbilor. Tehnologia din spate folosește modelul de inteligență artificială Universal Language Model, antrenat pe un set de date nedezevăluit, ce aparține Microsoft.

Există și funcționalitatea inversă din cadrul acestei componente, *Text to speech*, ce poate veni în ajutorul nevăzătorilor, permițându-le acestora să acceseze o voce care citește anumite texte. Spre deosebire de tehnologiile tradiționale de realizare a acestei sarcini, Microsoft utilizează rețele neuronale pentru a depăși limitele obișnuite. Astfel, în timp ce în mod obișnuit elementele specifice vorbirii, precum intonația și accentele, și elementele ce țin de convertirea cuvintelor în discursuri ar fi procesate independent, în cazul funcționalității din partea Azure, acestea sunt realizate simultan, rezultatul fiind o voce mai fluidă și mai naturală [4].

Nu în ultimul rand, avem componenta *Language*, ce prezintă diferite funcționalități legate de procesarea textelor și reprezintă o unealtă utilă pentru persoanele cu deficiențe de cogniție. *Named Entity Recognition* oferă utilizatorilor opțiunea de a identifica și a categoriza diferite elemente de interes într-un text, precum persoane, locuri, etc. Pentru această sarcină, există două rețele pre-antrenate ce sunt folosite cel mai des: Stanford NER, considerat ca având cea mai bună performanță, care combină HMM(Hidden Markov Model) și MEMM(Maximum Entropy Markov Model), și Spacy NER, ale cărei detalii tehnice nu sunt publice [5].

O altă funcționalitate din cadrul acestei componente este *Key Phrase Extraction*, utilizat pentru a extrage cele mai importante cuvinte din punct de vedere semantic dintr-un text. Utilizatorii cu deficiențe de vedere pot scădea din timpul petrecut încercând să citească multiple paragrafe, având posibilitatea de a obține un rezumat al acestora. Asemenea modele

funcționează prin utilizarea diferitor ‘sentence embeddings’, metrici ce pot cuantifica importanța cuvintelor și a frazelor în contextul dat, precum *Sent2Vec* și *Doc2Vec*.

În concluzie, suita de programe Azure Cognitive Services reprezintă un pas important în evoluția programelor software menite să ajute persoanele cu deficiențe de cogniție, reușind să aducă mai aproape toți utilizatorii serviciilor acestora de capacitățile complete ale internetului.

## Bibliografie

1. Setul de date COCO: <https://cocodataset.org/>
2. Reteaua convolutională Faster R-CNN:  
<https://towardsdatascience.com/faster-r-cnn-for-object-detection-a-technical-summary-474c5b857b46>
3. ‘STL: A seasonal-trend decomposition procedure based on Loss’, William Cleveland, 1990
4. <https://azure.microsoft.com/es-mx/blog/microsoft-s-new-neural-text-to-speech-service-helps-machines-speak-like-people/>
5. ‘Understanding Named Entity Recognition Pre-Trained Models’, <https://blog.vsoftconsulting.com/blog/understanding-named-entity-recognition-pre-trained-models>