

SISTEM DE CONTROL ÎN BAZA COMENZILOR VOCALE

Olesea BOROZAN*

*Universitatea Tehnică a Moldovei, Facultatea Calculatoare Informatică și Microelectronică,
Departamentul Informatică și Ingineria Sistemelor,
Școala Doctorală Știința Calculatoarelor, Electronică și Energetică, Chișinău, Republica Moldova*

*Olesea Borozan: olesea.borozan@ia.utm.md

Rezumat: În lucrarea de față sunt prezentate rezultatele proiectării unui sistem de control în baza comenzilor vocale. Funcționalitatea sistemului este bazată pe achiziția comenzilor vocale generate de utilizator, digitalizarea acestor comenzi și recunoașterea acestora în baza modelelor de Rețele Neuronale. Rezultatul recunoașterii comenzilor vocale este transformat în comenzi numerice care prin intermediul porturilor standard sunt transmise la sistemul controlat.

Cuvinte cheie: Rețele Neuronale, comenzi vocale, recunoașterea vorbirii, sistem de comandă.

Introducere

Recunoașterea vorbirii devine un aspect important în viața noastră. Este încorporat în telefoanele noastre, ceasuri inteligente, tablete, dispozitive electronice specializate (ex :Amazon Echo Dot, Apple HomeKit). Care permite să executăm anumite comenzi, traducem un set de cuvinte, să obținem un raport meteo - doar vorbind cu voce tare.

Acest tip de interacțiune este mai ușor și mai rapid, având avantajul că nu necesită cunoștințe suplimentare de operare și control al dispozitivelor, vorbirea fiind un proces natural și automat.

Actualmente se dorește realizarea unor sisteme de recunoaștere automată a vorbirii capabile să învețe rapid noi seturi de eșantioane vocale, independent de vorbitor sau de limbajul vorbit [1-2].

Totuși, sistemele de recunoaștere automată a vorbirii prezintă unele aspecte sensibile: acuratețea sistemului, timpul în care acesta procesează vorbirea, resursele consumate, etc. Fiecare aspect poate fi mai mult sau mai puțin relevant, în funcție de scopul final al aplicației. O acuratețe scăzută va conduce la transcrieri greșite ale semnalului vocal, iar acest fapt va genera interpretări greșite.

Andrew Ng a prezis de mult că pe măsură ce recunoașterea vorbirii va trece de la 95% precizie la 99%, va deveni un mod principal în care oamenii vor interacționa cu dispozitivele inteligente. Datorită învățării profunde pe baza rețelelor neuronale, procentul de recunoaștere a vorbirii va crește considerabil [4-5].

Recunoașterea vorbirii cu ajutorul Rețelelor Neuronale

În ultimii ani, cele mai performante sisteme de inteligență artificială au avut la bază algoritmi din sfera învățării profunde (deep learning). Tehnica reprezintă o abordare a învățării automate (machine learning), folosind rețelele neuronale.

Rețelele neurale sunt nucleul Inteligenței Artificiale și sunt esențiale pentru învățarea mecanică. Acestea sunt rețele interconectate de algoritmi, inspirate de rețelele neurale biologice, existente în creierul uman. Aceste rețele fac schimb de date și pot fi programate pentru a îndeplini anumite sarcini într-un mod complex care încearcă să imite gândirea omenească [3-5].

Aceste sisteme învață să îndeplinească obiective și procese analizând exemplele pe care le primesc din afară, de obicei fără să fie programate să ducă la bun sfârșit o sarcină anume. Ideea de Inteligență Artificială este legată de rețelele neurale, pentru că dezvoltarea acestora este singura metodă actuală prin care se poate dezvolta AI-ul.

Rețelele neuronale sunt sisteme adaptive de tip cutie neagră (black-box) care permit extragerea unui model pornind de la date printr-un proces de învățare [4].

În Figura 1 este reprezentată structura unei Rețele Neuronale cu mai multe nivele, unde:
Input Layer - este nivelul de intrare ce atașează informația de la senzori (sunetul de intrare);
Ouput Layer - este nivelul de ieșire ce generează decizii;
Hidden Layer 1, Layer 2 - nivele ascunse.

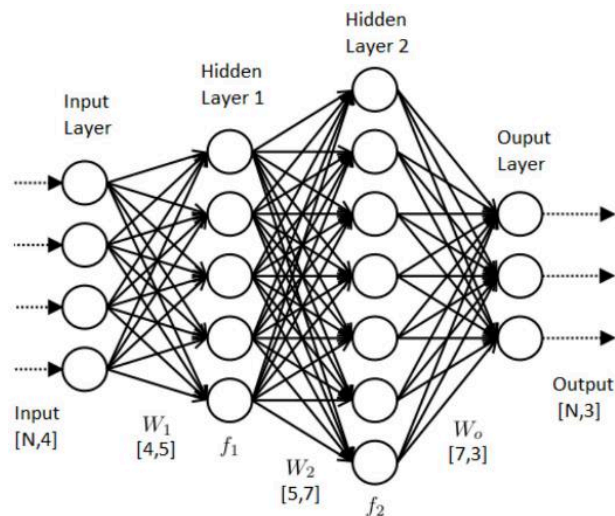


Figura 1. Rețea neuronală

Rețele neuronale recurente sunt foarte eficiente în sarcini ce modelează date cu caracter secvențial, ținând cont de dependențele temporale. Recunoașterea vorbirii sau a scrisului de mână, procesarea limbajului natural, predicția evenimentelor sunt câteva dintre domeniile unde pot fi aplicate [3-5].

Principala caracteristică a acestor rețele este dată de conexiunile de tip ciclu între neuroni. Aceștia învață atât din intrarea curentă, cât și din informațiile învățate pe baza intrărilor din trecut. Din această cauză, rețele neuronale recurente sunt considerate că fiind rețele cu "memorie".

În Figura 2 este prezentată structura unei Rețele Neuronale recurente [3-5].

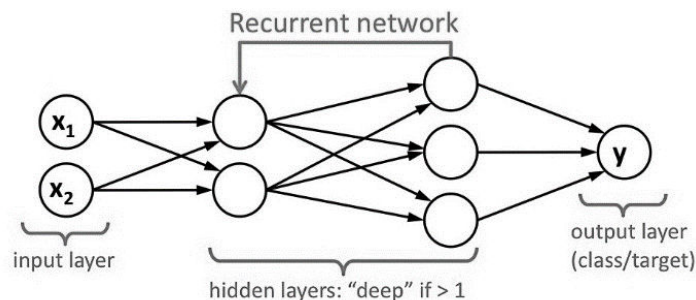


Figura 2. Rețea neuronală recurentă

Procesul de recunoaștere automată a vorbirii poate fi descris ca un proces în patru etape și anume [4-5]:

- a. Înregistrarea eşantioanelor vocale;
- b. Prelucrarea semnalelor;
- c. Extragerea caracteristicilor;
- d. Clasificarea exprimărilor.

Primul pas în recunoașterea vorbirii este introducerea undelor sonore într-un dispozitiv electronic, calculator, smartphone, etc.

Următorul pas este prelucrarea semnalelor și constă în analiza spectrală a semnalului electronic pentru a transforma undele sonore în numere digitale.

Etapa a treia se referă în principal la eliminarea informațiilor nerelevante sau redundante (filtrarea zgomotelor de fond care apar evident în eșantioanele vocale) din prezentarea parametrizată obținută în etapa a doua.

Undele sonore sunt unidimensionale. În fiecare moment de timp, acestea au o singură valoare bazată pe înălțimea undei.

Pentru recunoașterea vorbirii, o rată de eșantionare de 16KHz (16.000 pe secundă) este suficientă pentru a acoperi intervalul de frecvență al vorbirii umane.

Pentru a transforma o undă sonoră în format digital, se înregistrează înălțimea undei în punctele egal distanțate. Ca rezultat vom obține o serie de numere în punctele discrete. În Figura 3 este prezentată o undă de eșantionare [4].

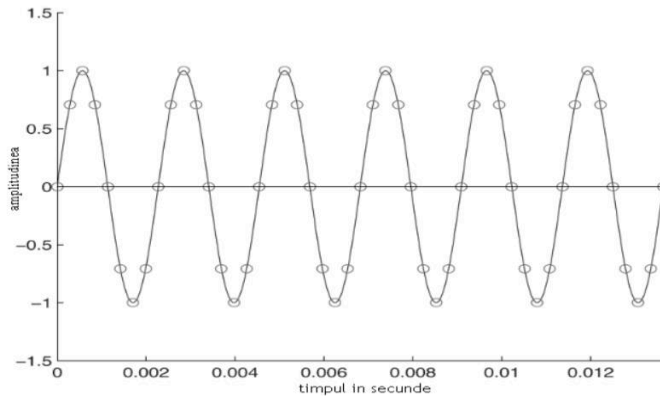


Figura 3. Undă de eșantionare

Odată ce obținem sunetul într-un format ușor de procesat, vom aplica datele originale ale sunetului, împărțite în fragmente de 20 de milisecunde, într-o Rețea Neuronală Recurentă.

Pentru fiecare fragment audio, rețeaua neuronală va încerca să identifice litera care corespunde sunetului care se vorbește în prezent. Rețea neuronală recurentă, rețea care are memorie, care influențează predicțiile viitoare. Acest lucru se datorează faptului că fiecare literă/silabă pe care o prezice, ar trebui să afecteze și probabilitatea silabei pe care o va prezice. De exemplu, dacă am spus „HEL” inițial, este foarte probabil să spunem „LO” la terminarea cuvântului „Hello”. Este mult mai puțin probabil să spunem ceva ne-pronunțabil în continuare, precum „XYZ”. Prin urmare, a avea acea memorie a predicțiilor anterioare ajută rețeaua neuronală să facă predicții mai precise pe viitor [4-5].

Schema funcțională a sistemului de control în baza comenzilor vocale

În Figura 4 este prezentată schema funcțională a unui sistem de control în baza comenzilor vocale.

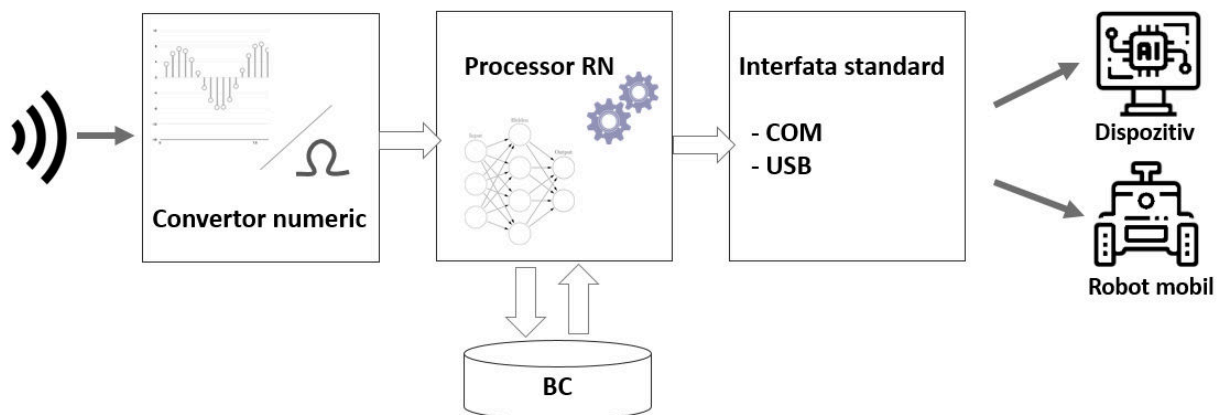


Figura 4. Schema funcțională a sistemului de control în baza comenzilor vocale

În schemă sunt menționate următoarele componente ale sistemului:

1. **Convertor numeric** - transformă undele sonore în secvențe de date numerice.
2. **Procesor RN** - aplicarea modelelor de Rețele Neuronale pentru recunoașterea vorbirii.
3. **BC** - bază de cunoștințe ce include modele și date experimentale acumulate în procesul învățării și recunoașterii vorbirii.
4. **Interfață standard** - pentru comunicarea cu obiectele și actuatorii procesului controlat (Dispozitive, Robot mobil, etc.).

Concluzii

Recunoașterea automată a vorbirii reprezintă un pas important în dezvoltarea sistemelor bazate pe inteligența artificială, în special a sistemelor controlate prin intermediul comenzilor vocale, astfel ușurând accesul persoanelor neinstruite în domeniul informaticii, la resurse controlate de calculator și alte dispozitive inteligente.

În lucrarea dată s-a efectuat o analiză a Rețelelor Neuronale cu mai multe nivele și s-a selectat structura Rețelei Neuronale Recurentă pentru recunoașterea comenzilor vocale în timp real.

Pentru viitor sunt planificate dezvoltarea de noi modele de Rețele Neuronale pentru recunoașterea vorbirii în limba română și dezvoltarea unui set de comenzi digitale pentru comanda unui robot mobil.

Mențiuni

Lucrarea de față a fost elaborată în cadrul tezei de doctorat cu tema "Procesor specializat cu proprietăți cognitive" în cadrul Departamentului Informatică și Ingineria Sistemelor.

Referințe

1. CUCU H., GEORGESCU A.L., Sistem de recunoaștere automată a vorbirii în limba română bazat pe rețele neuronale profunde, LUCRARE DE DISERTATIE, București 2018 , [accesat 10.02.2020] <https://speed.pub.ro/speed3/wp-content/uploads/2018/07/Binder1.pdf>
2. CHIȚĂ M.A, ȘTIRBU C., SISTEM DE RECUNOAȘTERE AUTOMATĂ A VORBIRII BAZAT PE REȚELE NEURONALE FUZZY, Pitești, România, [accesat 15.02.2020] <http://snet.elth.pub.ro/snet2005/cd/papers/s3p12.pdf>
3. LEONARDO ARAUJO DOS SANTOS, Artificial Intelligence, [accesat 20.02.2020] https://leonardoaraujosantos.gitbooks.io/artificial-intelligence/content/recurrent_neural_networks.html
4. MILAN SIGMUND, Automatic Speaker Recognition by Speech Signal, [accesat 20.02.2020] https://www.researchgate.net/publication/221787287_Automatic_Speaker_Recognition_by_Speech_Signal
5. GEITGEY ADAM, How to do Speech Recognition with Deep Learning, [accesat 20.02.2020] <https://medium.com/@ageitgey/machine-learning-is-fun-part-6-how-to-do-speech-recognition-with-deep-learning-28293c162f7a>