ROMÂNIA

(19) OFICIUL DE STAT
PENTRU INVENŢII ŞI MĂRCI
Bucureşti

(11) **RO 136008 A1**

(51) Int.Cl.
G10L 15/08 (2006.01);
G10L 15/26 (2006.01);
G06F 40/20 (2020.01)

(12) # CERERE DE BREVET DE INVENŢIE

(21) Nr. cerere: **a 2021 00135**

(22) Data de depozit: **29/03/2021**

(41) Data publicării cererii:
30/09/2022 BOPI nr. **9/2022**

(71) Solicitant:
• **UNIVERSITATEA POLITEHNICA DIN BUCUREŞTI**, *SPLAIUL INDEPENDENŢEI NR.313, SECTOR 6, BUCUREŞTI, B, RO*

(72) Inventatori:
• **CUCU HORIA**, *STR. GENERAL DAVID PRAPORGESCU, NR.98, CHIAJNA, IF, RO;*
• **GEORGESCU ALEXANDRU-LUCIAN**, *STR.NOVACI, NR.10, BL.P60, SC.3, ET.5, AP.79, SECTOR 5, BUCUREŞTI, B, RO;*
• **ONEAŢĂ DAN-THEODOR**, *STR. MARIA CUNTAN, NR.4, BL.S42, SC.B, AP.28, SECTOR 5, BUCUREŞTI, B, RO;*
• **BURILEANU DRAGOŞ**, *STR. JOHANNES KEPLER NR. 2, BL. 2, SC. 2, AP. 61, SECTOR 2, BUCUREŞTI, B, RO;*
• **BURILEANU CORNELIU**, *BL.LACUL TEI, NR.75, BL.16, SC. A, AP.3, SECTOR 2, BUCUREŞTI, B, RO*

Data publicării raportului de documentare:
30.09.2022

(54) # METODĂ DE ESTIMARE A ÎNCREDERII LA NIVEL DE CUVÂNT PENTRU RECUNOAŞTEREA AUTOMATĂ A VORBIRII ŞI METODĂ DE GENERARE A TRANSCRIERII PRECISE PENTRU MATERIALE AUDIO CE CONŢIN VORBIRE CUPRINZÂND ACEASTĂ METODĂ DE ESTIMARE

(57) **Rezumat:**

Invenţia se referă la o metodă de estimare a încrederii la nivel de cuvânt pentru recunoaşterea automată a vorbirii şi la o metodă de generare a unei transcrieri precise a unor materiale audio ce conţin vorbire, cuprinzând această metodă de estimare. Metoda de estimare a încrederii la nivel de cuvânt pentru recunoaşterea automată a vorbirii foloseşte scoruri de tip log-proba (logaritmul probabilităţii) şi neg-entropie (entropie negativă), estimate al nivel de token, pe care le agregă folosind funcţiile sumă, medie sau minim pentru a obţine scoruri de încredere la nivel de cuvânt şi utilizează tehnica de scalare a temperaturii şi/sau tehnica dropout pentru îmbunătăţirea probabilităţilor la nivel de token. Metoda de generare de transcrieri ale unor materiale audio care conţin vorbire foloseşte un sistem de recunoaştere automată a vorbirii pentru generarea unei transcrieri brute şi metoda de estimare a încrederii pentru a genera scoruri de încredere la nivel de cuvânt pentru cuvintele din transcrierea brută şi selectează şi etichetează ca fiind transcrise precis numai acele cuvinte din transcrierea brută care au un scor de încredere mai mare decât un prag configurat manual.

Revendicări: 5
Figuri: 2

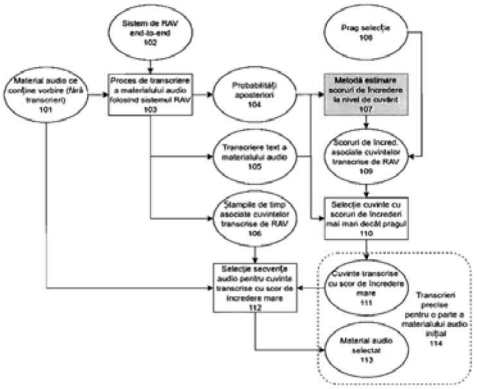

Fig. 1

RO 136008 A1

OFICIUL DE STAT PENTRU INVENŢII ŞI MĂRCI
Cerere de brevet de invenţie
Nr. a 2021 00135
Data depozit 29-03-2021

Metodă de estimare a încrederii la nivel de cuvânt pentru recunoaşterea automată a vorbirii şi metodă de generare a transcrierii precise pentru materiale audio ce conţin vorbire cuprinzând această metodă de estimare

Invenţia aparţine domeniului sistemelor de recunoaştere automată a vorbirii (RAV) care includ subsisteme pentru estimarea încrederii transcrierilor generate.

Antrenarea sistemelor de recunoaştere automată a vorbirii necesită o cantitate foarte mare de date audio ce conţin vorbire asociate cu transcrierile corespunzătoare. Deşi datele audio care conţin vorbire există şi/sau pot fi obţinute relativ uşor indiferent de limbă, transcrierile aferente sunt foarte costisitor de obţinut. Invenţia se referă la o metodă de estimare a încrederii pentru recunoaşterea automată a vorbirii utilizată în cadrul metodei de generare de transcrieri precise pentru materiale audio ce conţin vorbire.

Stadiul cunoscut al tehnicii

În contextul recunoaşterii automate a vorbirii (RAV) şi al adnotării automate a datelor audio, estimarea scorurilor de încredere este de o importanţă crucială pentru că semnalează corectitudinea datelor transcrise automat şi permite o filtrare pe baza acestor scoruri. În plus, metodele pentru estimarea scorurilor de încredere pentru RAV au aplicaţii multiple: îmbunătăţirea robusteţii sistemelor în sarcini critice de siguranţă, evitarea erorilor în sistemele de dialog om-maşină sau facilitarea corecţiilor manuale în sarcinile de transcriere audio prin semnalarea erorilor. Mai mult, diverse lucrări de specialitate au valorificat estimările scorurilor de încredere pentru o serie de sarcini care depind de RAV: selectarea predicţiilor cu grad ridicat de încredere pentru reantrenarea sistemului de bază [Sperber, 2017], propagarea incertitudinilor în traducerea automată a vorbirii [Vesely, 2013], adnotarea manuală a predicţiilor mai puţin sigure pentru învăţarea activă [Yu, 2010].

Există metode de estimare a scorurilor de încredere pentru RAV bazate pe paradigma HMM-GMM. Aceste metode extrag mai întâi un set de caracteristici din reţeaua de decodare, modelul acustic sau de limbă şi, apoi, antrenează un clasificator pentru a prezice dacă transcrierea este corectă sau nu. Exemple tipice de caracteristici includ probabilitatea realizării acustice, scorul modelului de limbă, durata cuvântului, numărul de alternative din reţeaua de confuzie [Kemp, 1997; Weintraub, 1997; Hazen, 2002]. Mai recent, Swarup et al. [Swarup, 2019] a mărit setul de caracteristici folosind embedding-uri profunde ale semnalului acustic de intrare şi ale textului prezis, în timp ce Errattahi et al. [Errattahi, 2018] a arătat că adaptarea la domeniu a caracteristicilor extrase aduce beneficii de performanţă. Clasificatorii folosiţi de metodele de estimare a scorurilor de încredere variază de la câmpuri aleatorii condiţionate (en. conditional random fields) [Seigel, 2013; Cortina, 2016] şi perceptron cu mai multe straturi [Kalgaonkar, 2015] la reţele neuronale recurente bidirecţionale [Ogawa, 2017; Del-Agua, 2018; Li, 2019].

Cea mai cunoscută metodă de estimare a încrederii în rețelele neuronale implică utilizarea directă a probabilității predicției celei mai probabile [Hendrycks, 2016]. S-a observat că rețelele neuronale tind să fie prea sigure pe estimările lor și că acestea pot fi îmbunătățite prin scalarea temperaturii [Hinton, 2015], ceea ce duce de obicei la o mai bună calibrare [Guo, 2017; Ashukha, 2020]. În acest sens se folosește estimarea Monte Carlo: se utilizează tehnica dropout la momentul inferenței pentru a obține predicții multiple, care sunt apoi mediate [Gal, 2016] sau se face media predicțiilor unui ansamblu de rețele antrenate de obicei cu inițializări diferite [Lakshminarayanan, 2017]. Această ultimă metodă este însă costisitoare din punct de vedere computațional, deoarece implică antrenarea a multiple modele de RAV [Ashukha, 2020]. O abordare diferită a estimării încrederii este de a învăța un clasificator (de obicei o altă rețea neuronală) direct deasupra activărilor rețelei de RAV [Corbière, 2019; Chen, 2019].

Drept cel mai apropiat stadiu al tehnicii, menționăm o metodă [Malinin, 2020] care abordează sarcina de estimare a încrederii pentru sistemele RAV end-to-end. Cu toate acestea, această metodă estimează incertitudinea la *nivel de token și frază*, în timp ce metoda conform prezentei invenții estimează scoruri de încredere la nivel de cuvânt. Acest lucru reprezintă un avantaj extrem de important întrucât utilizatorul final primește informații granular, la nivelul fiecărui cuvânt al transcrierii. Mai mult, metoda cunoscută [Malinin, 2020] folosește ansambluri pentru estimarea încrederii, în timp ce metoda conform invenției se bazează pe scalarea temperaturii și dropout. Deși tehnica de dropout a mai fost folosită pentru obținerea scorurilor de încredere pentru RAV [Vyas, 2019], metoda respectivă este este diferită de metoda conform invenției. Autorii metodei cunoscute [Vyas, 2019] generează mai multe ipoteze prin dropout și apoi atribuie scoruri de încredere cuvintelor pe baza frecvenței aparițiilor lor în ipotezele aliniate. Avantajul metodei propuse prin invenție, față de cea existentă, este că se agregă probabilitățile posterioare și nu ipotezele, ceea ce simplifică procedura, deoarece evită pasul de aliniere.

Prezentarea pe scurt a invenţiei

Prezenta invenție se referă la o metodă de estimare a scorurilor de încredere *la nivel de cuvânt* pentru sisteme de RAV de tip end-to-end care folosește scoruri de tip log-proba și neg-entropie estimate la nivel de token pe care le agregă folosind funcțiile sumă, medie sau minim, pentru a obține scoruri de încredere la nivel de cuvânt și, utilizează tehnica de scalare a temperaturii și/sau tehnica de dropout pentru îmbunătățirea probabilităților la nivel de token; și, la o metodă pentru generare de transcrieri precise pentru materiale audio ce conțin vorbire care folosește un sistem de RAV end-to-end pentru generarea transcrierii brute și metodele de estimare a scorurilor de încredere conform prezentei invenții pentru a genera scoruri de încredere la nivel de cuvânt pentru cuvintele din transcrierea brută și care selectează și etichetează ca fiind transcrise precis numai acele cuvinte din transcrierea brută care au un scor de încredere mai mare decât un prag configurat manual.

Invenția de față prezintă următoarele avantaje:

- **Acuratețe crescută de estimare a scorurilor de încredere.** Metoda se aplică sistemelor RAV end-to-end, sisteme ale căror performanțe depășesc, la acest moment, performanțele sistemelor bazate pe HMM;
- **Estimare a scorurilor de încredere la nivel de cuvânt.** Utilizatorul final al scorurilor de încredere este interesat în primul rând de scorurile la nivel de cuvânt, abia apoi de scoruri la nivel de propoziție și în ultimă instanță de scoruri la nivel de token;
- **Simplitate a implementării.** Modelele de tip end-to-end și modul calcul al scorurilor de încredere la nivel de cuvânt prin agregarea scorurilor de încredere la nivel de token conferă simplitate metodei propuse față de metodele cunoscute.

Se dă în continuare un exemplu de realizare a invenției, în legătură cu figurile 1 și 2 care reprezintă:

- Figura 1: Diagrama bloc funcțională a metodei de generare de transcrieri precise pentru materiale audio ce conțin vorbire, metodă care include etapa inovativă de estimare a scorurilor de încredere la nivel de cuvânt 104.
- Figura 2: Diagrama bloc funcțională a metodei de estimare a scorurilor de încredere la nivel de cuvânt.

Metoda de generare de transcrieri precise pentru o parte a materialului audio ce conține vorbire utilizează ca elemente de intrare:

- Materialul audio ce conține vorbire 101
- Sistemul de RAV end-to-end 102
- Pragul de selecție 108

Prima etapă a acestei metode, etichetată 103 în Fig. 1, constă în transcrierea materialului 101 folosind sistemul de RAV 102. În urma procesului de transcriere 103 rezultă

- Probabilitățile aposteriori de la fiecare moment de timp 104
- Transcrierea text 105 a materialului audio
- Ștampilele de timp 106 asociate cuvintelor ce compun transcrierea

A doua etapă a acestei metode, etichetată 107, reprezintă etapa inovativă care utilizează probabilitățile aposteriori 104 și transcrierea 105 pentru a estima scoruri de încredere la nivel de cuvânt 109 pentru toate cuvintele transcrierii 105.

A treia etapă a metodei propuse, etichetată 110, constă în selecția cuvintelor cu scoruri de încredere mai mari decât pragul de selecție 109. În urma acestei etape rezultă un subset de cuvinte transcrise cu scor de încredere mare 111.

A patra etapă a metodei propuse, etichetată 112, implică selecția secvențelor audio pentru cuvintele 111. Selecția se realizează preluând din materialul audio inițial 101 numai porțiunile dintre ștampilele de timp 106 corespunzătoare cuvintelor 111. În urma acestei etape rezultă materialul audio 113.

Materialul audio 113 împreună cu cuvintele 111 reprezintă transcrierile precise 114 pe care această metodă își propune să le genereze.

În continuare, se prezintă detaliat etapa de estimare a scorurilor de încredere 107 în legătură cu Fig. 2 și cu următoarele notații:

- a - secvența audio de intrarea sistemului de RAV;
- θ - parametri sistemului de RAV;
- V - numărul de tokenuri distincte din vocabularul sistemului de RAV
- $t = (t_1,..., t_T)$ - secvența de tokenuri generate la ieșirea sistemului de RAV;
- t_k - tokenul *k* din secvența de tokenuri de ieșire;
- $\hat{t}_{<k}$ - secvența de tokenuri prezisă de sistemul de RAV până la tokenul *k*, exclusiv;
- $p_k = p(t_k|\hat{t}_{<k}, a; \theta)$ - probabilitățile aposteriori 104 generate de sistemul de RAV pentru tokenul cu indexul *k*;
- $\hat{p}_k$ - probabilitățile aposteriori 104 dacă în sistemul de RAV se aplică dropout
- p'_k - probabilitățile aposteriori îmbunătățite 203
- $s_k^{(t)}$ - scorul de încredere la nivel de token 202 pentru tokenul cu indexul *k*;
- $s_j^{(w)}$ - scorul de încredere la nivel de cuvânt 109 pentru cuvântul cu indexul *j*;
- α, β, τ - parametri utilizați pentru scalarea temperaturii. Acești parametri sunt învățați prin optimizarea unei funcții de pierdere de entropie încrucișată (en., cross-entropy loss) pe un set de validare.

Invenția se referă la o metodă de estimare a încrederii pentru sisteme de RAV end-to-end, spre deosebire de metodele proiectate pentru sistemele RAV de tip hibrid bazate pe modele Markov ascunse și rețele neuronale profunde (en. DNN-HMM).

Invenția se referă la o metodă ce folosește un model de tip secvență-la-secvență care asociază unei secvențe audio de intrare a o secvență de token-uri $t = (t_1,..., t_T)$. Modelul este caracterizat de parametri θ, care sunt învățați prin minimizarea pe setul de antrenare a unei funcții de pierdere, cum ar fi funcția de clasificare temporală conexionistă (en. connectionist temporal classification) sau divergența Kullback-Leibler. La inferență, modelul generează probabilități pentru următorul token t_k într-o manieră autoregresivă $p(t_k|\hat{t}_{<k}, a; \theta)$, folosind token-urile prezise

anterior $\hat{t}_{<k}$. Aceste probabilități sunt utilizate pentru efectuarea decodării prin metoda beam search pentru a obține cea mai probabilă secvență de token-uri. Având în vedere că probabilitatea de ieșire condiționată este o distribuție peste V token-uri din vocabular, o notăm cu un vector V-dimensional p_k.

Primul pas al metodei de estimare a scorurilor de încredere, etichetat 201, constă în îmbunătățirea opțională a probabilităților aposteriori 104 (ceea ce va conduce la obținerea unor scoruri de încredere la nivel de cuvânt mai bune) prin aplicarea individuală sau în lanț a tehnicilor de scalare a temperaturii și dropout:

- Fără îmbunătățire: $p'_k = p_k$;
- Îmbunătățire prin scalarea temperaturii. Scalarea temperaturii [Hinton, 2015; Guo, 2017] constă în împărțirea activărilor de tip logit (valorile de dinainte de stratul softmax) la un scalar τ (cunoscut sub numele de temperatură). Pe baza temperaturii τ se actualizează probabilitățile la nivel de token la fiecare moment de timp, astfel: $p'_k = softmax(log(p_k) / \tau)$. Apoi se extrag scorurile la nivel de token $s^{(t)}$ peste probabilitățile actualizate, se agregă în scorul la nivel de cuvânt $s^{(w)}$ și, în cele din urmă, se clasifică cuvântul drept corect sau incorect: $P(correct) = \sigma(\alpha s^{(w)} + \beta)$.
- Îmbunătățire prin dropout. Dropout [Srivastava, 2014] este o tehnică care maschează părți aleatorii ale activărilor într-o rețea, făcând rețeaua mai puțin predispusă la supra-antrenare (en., overfitting). Folosind această tehnică, metoda propusă calculează probabilitățile la nivel de token obținute prin mai multe rulări folosind dropout: $p'_k = \frac{1}{N}\sum_n \hat{p}_k$, unde $\hat{p}$ specifică predicția dropout-ului. Probabilitățile actualizate sunt apoi utilizate pentru a calcula scorurile de încredere la nivel de token (log-proba sau neg-entropie).
- Îmbunătățire prin scalarea temperaturii urmată de dropout.

Al doilea pas al metodei de estimare a scorurilor de încredere, etichetat 204, reprezintă calculul scorurilor de încredere la nivel de token $s_k^{(t)}$ 205 pentru fiecare tokent_k din secvența $t = (t_1, ..., t_T)$. Metoda propusă prevede două modalități de calcul pentru aceste scoruri la nivel de token:

- Logaritmul probabilității (log-proba) celei mai probabile predicții date de clasificator, adică $s^{(t)} = log\ max\ p$.
- Entropie negativă (neg-entropie) calculată peste vocabularul token-urilor la fiecare moment de timp, adică $s^{(t)} = p^T log\ p$. O entropie mare înseamnă o incertitudine mare sau, invers, o entropie negativă mare implică o predicție încrezătoare.

Al treilea pas al metodei de estimare a scorurilor de încredere, etichetat 206, reprezintă calculul scorurilor de încredere la nivel de cuvânt $s_j^{(w)}$ 109 prin agregarea scorurilor de încredere la nivel de token $s_k^{(t)}$ 205 ale tuturor token-urilor asociate respectivului cuvânt. Aceste scoruri de încredere la nivel de token se agregă folosind oricare dintre următoarele trei funcții de agregare: suma, media, minimul. Astfel, scorurile de încredere la nivel de cuvânt 109 se calculează în oricare din următoarele moduri:

- Suma scorurilor de încredere la nivel de token: $s^{(w)} = \sum_T s^{(t)}$
- Media scorurilor de încredere la nivel de token: $s^{(w)} = \frac{1}{T}\sum_T s^{(t)}$
- Minimul scorurilor de încredere la nivel de token: $s^{(w)} = min(s^{(t)})$

Referinţe bibliografice

[Ashukha, 2020] Arsenii Ashukha, Alexander Lyzhov, Dmitry Molchanov, and Dmitry Vetrov, "Pitfalls of in-domain uncertainty estimation and ensembling in deep learning," in International Conference on Learning Representations, 2020.

[Chen, 2019] Tongfei Chen, Jirí Navrátil, Vijay Iyengar, and Karthikeyan Shanmugam, "Confidence scoring using whitebox meta-models with linear classifier probes," in International Conference on Artificial Intelligence and Statistics, 2019, pp. 1467–1475.

[Corbière, 2019] Charles Corbière, Nicolas Thome, Avner Bar-Hen, Matthieu Cord, and Patrick Pérez, "Addressing failure prediction by learning model confidence," in Advances in Neural Information Processing Systems, 2019, pp. 2902– 2913.

[Cortina, 2016] Isaías Sánchez Cortina, Jesús Andrés-Ferrer, Alberto Sanchis, and Alfons Juan, "Speaker-adapted confidence measures for speech recognition of video lectures," Computer Speech & Language, vol. 37, pp. 11–23, 2016.

[Del-Agua, 2018] M. A. Del-Agua, A. Gimenez, A. Sanchis, J. Civera, and A. Juan, "Speaker-adapted confidence measures for ASR using deep bidirectional recurrent neural networks," Transactions on Audio, Speech, and Language Processing, vol. 26, no. 7, pp. 1198–1206, 2018.

[Errattahi, 2018] Rahhal Errattahi, Salil Deena, Asmaa El Hannani, Hassan Ouahmane, and Thomas Hain, "Improving ASR error detection with RNNLM adaptation," in IEEE Spoken Language Technology Workshop, 2018, pp. 190–196.

[Gal, 2016] Yarin Gal and Zoubin Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," in International Conference on Machine Learning, 2016, pp. 1050–1059.

[Guo, 2017] Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q Wein- berger, "On calibration of modern neural networks," in International Conference on Machine Learning, 2017, pp. 1321–1330.

[Hadian, 2018] Hossein Hadian, Hossein Sameti, Daniel Povey, and Sanjeev Khudanpur, "End-to-end speech recognition using lattice-free MMI," in Interspeech, 2018, pp. 12–16.

[Hazen, 2002] Timothy J Hazen,Stephanie Seneff, and Joseph Polifroni, "Recognition confidence scoring and its use in speech understanding systems," Computer Speech & Language, vol. 16, no. 1, pp. 49–67, 2002.

[Hendrycks, 2016] Dan Hendrycks and Kevin Gimpel, "A baseline for detecting misclassified and out-of-distribution examples in neural networks," in International Conference on Learning Representations, 2016.

[Hinton, 2015] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean, "Distilling the knowledge in a neural network," arXiv preprint arXiv:1503.02531, 2015.

[Kalgaonkar, 2015] Kaustubh Kalgaonkar, Chaojun Liu, Yifan Gong, and Kaisheng Yao, "Estimating confidence scores on ASR results using recurrent neural networks," in IEEE International Conference on Acoustics, Speech and Signal Processing, 2015, pp. 4999–5003.

[Kemp, 1997] Thomas Kemp and Thomas Schaaf, "Estimating confidence using word lattices," in Eurospeech, 1997.

[Lakshminarayanan, 2017] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell, "Simple and scalable predictive uncertainty estimation using deep ensembles," in Advances in Neural Information Processing Systems, 2017, pp. 6402–6413.

[Li, 2019] Qiujia Li, PM Ness, Anton Ragni, and Mark JF Gales, "Bi-directional lattice recurrent neural networks for confidence estimation," in IEEE International Conference on Acoustics, Speech and Signal Processing, 2019, pp. 6755–6759.

[Malinin, 2020] Andrey Malinin and Mark Gales, "Uncertainty in structured prediction," arXiv preprint arXiv:2002.07650, 2020.

[Ogawa, 2017] Atsunori Ogawa and Takaaki Hori, "Error detection and accuracy estimation in automatic speech recognition using deep bidirectional recurrent neural networks," Speech Communication, vol. 89, pp. 70–83, 2017.

[Ovadia, 2019] Yaniv Ovadia, Emily Fertig, Jie Ren, Zachary Nado, David Sculley, Sebastian Nowozin, Joshua Dillon, Balaji Lakshminarayanan, and Jasper Snoek, "Can you trust your model's uncertainty? Evaluating predictive uncertainty under dataset shift," in Advances in Neural Infor- mation Processing Systems, 2019, pp. 13991–14002.

[Seigel, 2013] Mathew Stephen Seigel, Confidence estimation for automatic speech recognition hypotheses, Ph.D. thesis, University of Cambridge, 2013.

[Sperber, 2017] Matthias Sperber, Graham Neubig, Jan Niehues, and Alex Waibel, "Neural lattice-to-sequence models for uncertain inputs," in Empirical Methods in Natural Language Processing, 2017, pp. 1380–1389.

[Srivastava, 2014] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov, "Dropout: A simple way to prevent neural

networks from overfitting," Journal of Machine Learning Research, vol. 15, no. 1, pp. 1929–1958, 1 2014.

[Swarup, 2019] Prakhar Swarup, Roland Maas, Sri Garimella, Sri Harish Mallidi, and Björn Hoffmeister, "Improving ASR confidence scores for Alexa using acoustic and hypothesis embeddings," in Interspeech, 2019, pp. 2175–2179.

[Vesely, 2013] Karel Vesely, Mirko Hannemann, and Lukas Burget, "Semi-supervised training of deep neural networks," in Workshop on Automatic Speech Recognition and Understanding, 2013, pp. 267–272.

[Vyas, 2019] Apoorv Vyas, Pranay Dighe, Sibo Tong, and Hervé Bourlard, "Analyzing uncertainties in speech recognition using dropout," in IEEE International Conference on Acoustics, Speech and Signal Processing, 2019, pp. 6730–6734.

[Weintraub, 1997] Mitch Weintraub, Francoise Beaufays, Zeév Rivlin, Yochai Konig, and Andreas Stolcke, "Neural-network based measures of confidence for word recognition," in IEEE International Conference on Acoustics, Speech and Signal Processing, 1997, vol. 2, pp. 887–890.

[Yu, 2010] Dong Yu, Balakrishnan Varadarajan, Li Deng, and Alex Acero, "Active learning and semi-supervised learning for speech recognition: A unified framework using the global entropy reduction maximization criterion," Computer Speech & Language, vol. 24, no. 3, pp. 433–444, 2010.

REVENDICĂRI

1. Metodă de estimare a scorurilor de încredere *la nivel de cuvânt* pentru sisteme de RAV de tip end-to-end **caracterizată prin aceea că** folosește scoruri de tip log-proba și neg-entropie estimate la nivel de token pe care le agregă folosind funcțiile sumă, medie sau minim pentru a obține scoruri de încredere la nivel de cuvânt.

2. Metodă de estimare a scorurilor de încredere la nivel de cuvânt pentru sisteme de RAV de tip end-to-end conform revendicării 1, **caracterizată prin aceea că** utilizează tehnica de scalare a temperaturii pentru îmbunătățirea probabilităților la nivel de token.

3. Metodă de estimare a scorurilor de încredere la nivel de cuvânt pentru sisteme de RAV de tip end-to-end conform revendicării 1, **caracterizată prin aceea că** utilizează tehnica de dropout pentru îmbunătățirea probabilităților la nivel de token.

4. Metodă de estimare a scorurilor de încredere la nivel de cuvânt pentru sisteme de RAV de tip end-to-end conform revendicării 1, **caracterizată prin aceea că** utilizează simultan atât tehnica de scalare a temperaturii, cât și tehnica de dropout pentru îmbunătățirea probabilităților la nivel de token.

5. Metodă pentru generare de transcrieri precise pentru materiale audio ce conțin vorbire care foloseşte un sistem de RAV end-to-end pentru generarea transcrierii brute **caracterizată prin aceea că** folosește oricare dintre metodele de estimare a scorurilor de încredere conform revendicărilor 1, 2, 3 sau 4 pentru a genera scoruri de încredere la nivel de cuvânt pentru cuvintele din transcrierea brută și selectează și etichetează ca fiind transcrise precis numai acele cuvinte din transcrierea brută care au un scor de încredere mai mare decât un prag configurat manual.

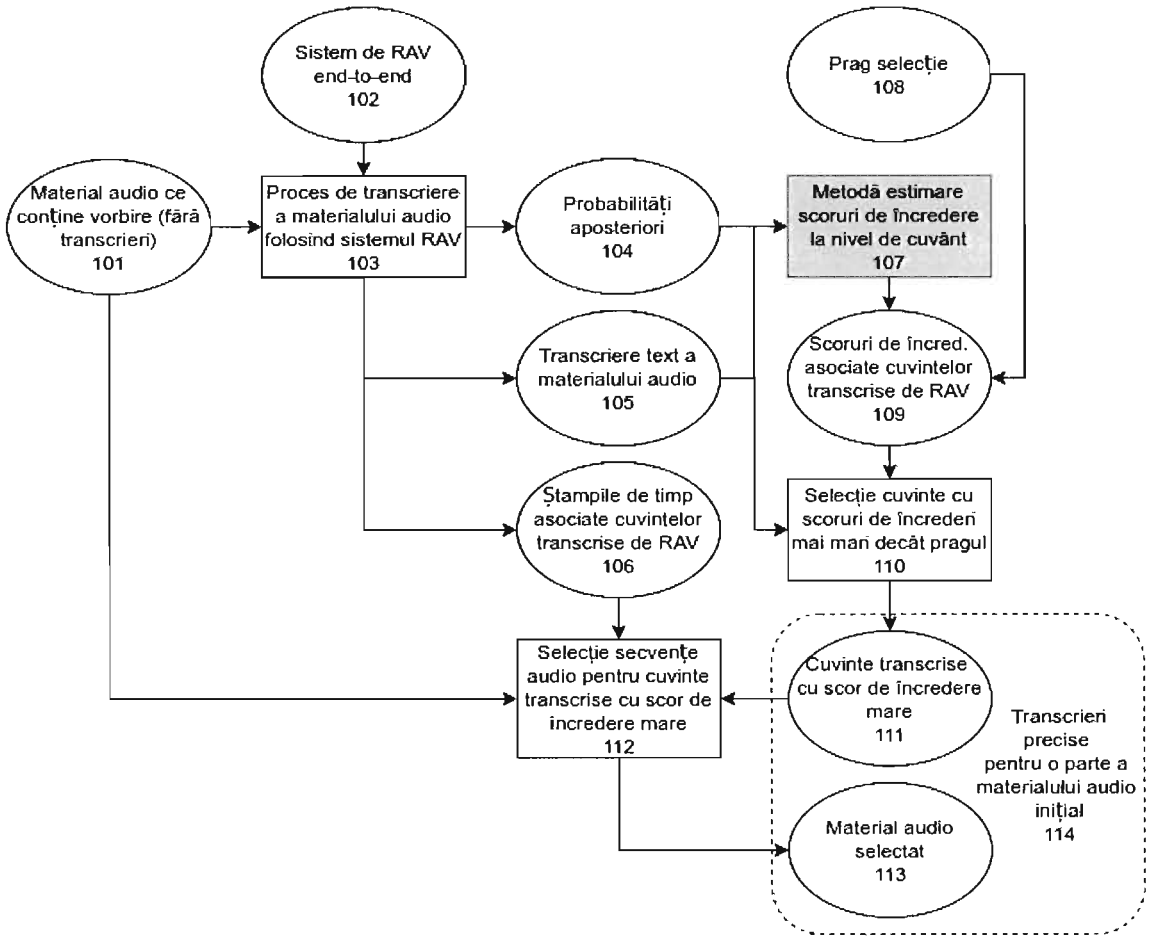

Figura 1 Diagrama bloc funcțională a metodei de generare de transcrieri precise pentru materiale audio ce conțin vorbire, metodă care include etapa de estimare a scorurilor de încredere la nivel de cuvânt 107

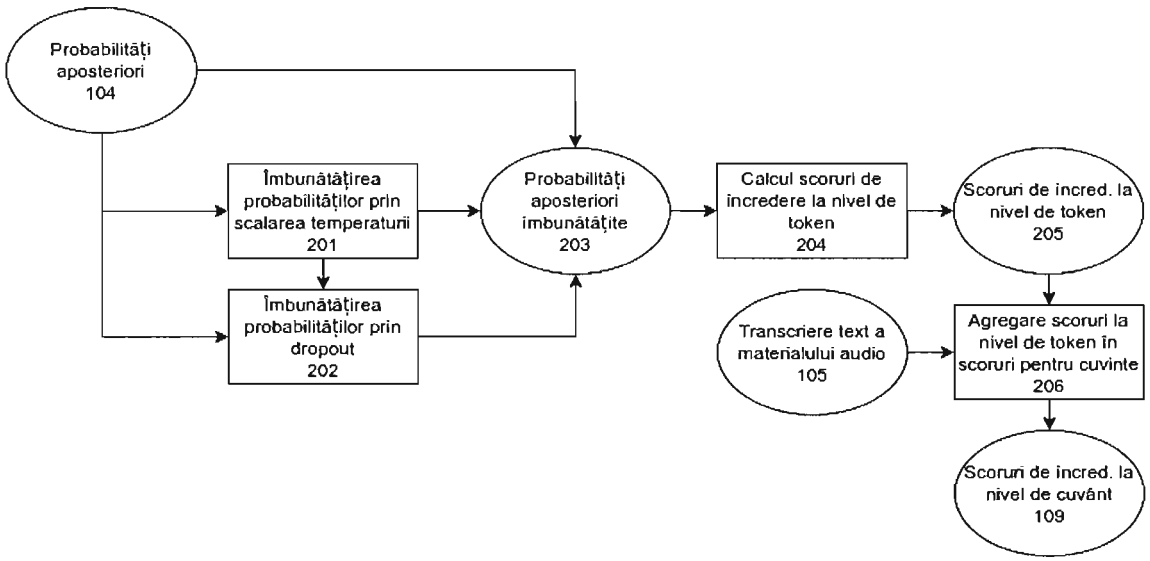

Figura 2 Diagrama bloc funcțională a metodei de estimare a scorurilor de încredere la nivel de cuvânt.

Cont IBAN: RO05 TREZ 7032 0F33 5000 XXXX
Trezoreria Sector 3, Bucureşti
Cod fiscal: 4266081

Serviciul Examinare de Fond: Electricitate-Fizica

RAPORT DE DOCUMENTARE

CBI nr. a 2021 00135	Data de depozit: 29/03/2021	Dată de prioritate
Titlul invenţiei	METODĂ DE ESTIMARE A ÎNCREDERII LA NIVEL DE CUVÂNT PENTRU RECUNOAŞTEREA AUTOMATĂ A VORBIRII ŞI METODĂ DE GENERARE A TRANSCRIERII PRECISE PENTRU MATERIALE AUDIO CE CONŢIN VORBIRE CUPRINZÂND ACEASTĂ METODĂ DE ESTIMARE	
Solicitant	UNIVERSITATEA POLITEHNICA DIN BUCUREŞTI, SPLAIUL INDEPENDENŢEI NR.313, SECTOR 6, BUCUREŞTI, RO	

Clasificarea cererii (Int.Cl.)	***G10L15/08 (2006.01)*** ***G10L15/26 (2013.01)*** ***G06F40/20(2020.01)***
Domenii tehnice cercetate (Int.Cl.)	G01L G06F
Colecţii de documente de brevet cercetate	RO, GB, US, FR, DE, EP, PCT
Baze de date electronice cercetate	RoPatentSearch, Epodoc, Patenw, Google Patents
Literatură non-brevet cercetată	D. Oneaţă, A. Caranica, A. Stan, Horia Cucu, ***"An Evaluation of Word-Level Confidence Estimation for End-to-End Automatic Speech Recognition"***, January 2021, DOI:10.1109/SLT48900.2021.9383570, Conference: 2021 IEEE Spoken Language Technology Workshop (SLT), ianuarie 2021,disponibil la: www.researchgate.net/publication/350394049_An_Evaluation_of_Word-Level_Confidence_Estimation_for_End-to-End_Automatic_Speech_Recognition

Documente considerate a fi relevante

Categoria	Date de identificare a documentelor citate şi, unde este cazul, indicarea pasajelor relevante	Relevant faţă de revendicarea nr.
X	D. Oneaţă, A. Caranica, A. Stan, Horia Cucu, ***"An Evaluation of Word-Level Confidence Estimation for End-to-End Automatic Speech Recognition"***, January 2021, DOI:10.1109/SLT48900.2021.9383570, Conference: 2021 IEEE Spoken Language Technology Workshop (SLT), disponibil la: www.researchgate.net/publication/350394049_An_Evaluation_of_Word-Level_Confidence_Estimation_for_End-to-End_Automatic_Speech_Recognition January 2021 cap. 3	1-5

Documente considerate a fi relevante - continuare		
Categoria	Date de identificare a documentelor şi, unde este cazul, indicarea pasajelor relevante	Relevant faţă de revendicarea nr.
Unitatea invenţiei (art.18)		
Observaţii:		

Data redactării: 25.01.2022

Examinator,
Daniela CRISTUDOR

Litere sau semne, conform ST.14, asociate categoriilor de documente citate	
A - Document care defineşte stadiul general al tehnicii şi care nu este considerat de relevanţă particulară; **D** - Document menţionat deja în descrierea cererii de brevet de invenţie pentru care este efectuată cercetarea documentară; **E** - Document de brevet de invenţie având o dată de depozit sau de prioritate anterioară datei de depozit a cererii în curs de documentare, dar care a fost publicat la sau după data de depozit a acestei cereri, document al cărui conţinut ar constitui un stadiu al tehnicii relevant, **L** - Document care poate pune în discuţie data priorităţii/lor invocată/e sau care este citat pentru stabilirea datei de publicare a altui document citat sau pentru un motiv special (se va indica motivul); **O** - Document care se referă la o dezvăluire orală, utilizare, expunere, etc,	**P** - Document publicat la o dată aflată între data de depozit a cererii şi data de prioritate invocată; **T** - Document publicat ulterior datei de depozit sau datei de prioritate a cererii şi care nu este în contradicţie cu aceasta, citat pentru mai buna înţelegere a principiului sau teoriei care fundamentează invenţia; **X** - document de relevanţă particulară; invenţia revendicată nu poate fi considerată nouă sau nu poate fi considerată ca implicând o activitate inventivă, când documentul este luat în considerare singur; **Y** - document de relevanţă particulară; invenţia revendicată nu poate fi considerată ca implicând o activitate inventivă, când documentul este combinat cu unul sau mai multe alte documente de aceeaşi categorie, o astfel de combinaţie fiind evidentă unei persoane de specialitate. **&** - document care face parte din acelaşi familie de brevete de invenţie.