

Raport științific și tehnic

Denumire proiect (EN)	CULTiVA: Curriculum learning in text mining and visual question answering
Denumire proiect (RO)	CULTiVA: Învățarea pe bază de curriculum în text mining și visual question answering
Acronim	CULTiVA
Cod proiect	PN-III-P1-1.1-TE-2019-0235
Număr contract	72/2020
Contractor	UNIVERSITATEA BUCUREȘTI
Tip proiect	Proiecte de cercetare pentru stimularea tinerelor echipe independente (TE)
Autoritatea contractantă	Unitatea Executivă pentru Finanțarea Învățământului Superior, a Cercetării, Dezvoltării și Inovării
Perioada de raportare	01.01.2021 - 31.12.2021
Etapa de execuție	2/2021
Director de proiect	Radu Tudor Ionescu
Adresă web	https://cultiva-proj.herokuapp.com

1. Rezumat:

În Etapa 2 – „Dezvoltarea și evaluarea unor modele neuronale deep cu rezultate state-of-the-art” am efectuat activitățile 2.1, 2.2 și 2.3. Activitatea 2.1. are ca scop îndeplinirea celui de-al doilea obiectiv, acela de a antrena rețele neuronale state-of-the-art pentru clasificarea textelor și visual question answering, utilizând seturi de date publice. Activitatea 2.2. are ca scop îndeplinirea parțială a celui de-al treilea obiectiv, anume dezvoltarea și evaluarea unui model neuronal deep pentru clasificarea textelor, folosind paradigma de învățare bazată pe curriculum. În mod asemănător, activitatea 2.3 are ca scop îndeplinirea parțială a celui de-al patrulea obiectiv, anume dezvoltarea și evaluarea unui model neuronal deep pentru visual question answering, folosind paradigma de învățare bazată pe curriculum.

2. Procentaj îndeplinire obiective:

Obiectivul 1 – 100%
Obiectivul 2 – 100%
Obiectivul 3 – 50%
Obiectivul 4 – 40%
Obiectivul 5 – 66%

3. Descriere științifică și tehnică:

În conformitate cu activitățile prevăzute în Etapa 2 de raportare din Anexa II a contractului TE72/2020, am efectuat următoarele:

- **Activitatea 2.1. Task 2 – Dezvoltarea și evaluarea unor modele neuronale deep cu rezultate state-of-the-art.**

În vederea îndeplinirii celui de-al doilea obiectiv, am decis să folosim modele neuronale deep de tip transformer, cum ar fi BERT [Devlin et al., NAACL19], pentru clasificarea textelor,

respectiv arhitecturi convoluționale, cum ar fi ResNet [He et al., CVPR16], pentru procesarea imaginilor. În zona procesării limbajului natural, am comparat modelul BERT cu o serie de abordări alternative în diverse probleme de clasificare a textelor, anume detecția polarității opiniei [Tache et al., EACL21], geo-localizarea textelor [Găman et al., VarDial21], detecția satirei [Ionescu et al., IJCNN21; Rogoz et al., ACL21] și recunoașterea sub-dialectelor limbii române [Găman et al., IJIS21]. Totodată, în zona procesării de imagini (naturale sau rezultate în urma aplicării unor transformări pe semnale audio), am comparat mai multe arhitecturi convoluționale de tip ResNet [He et al., CVPR16] pe diverse probleme de clasificare sau detecție, anume recunoașterea și detecția obiectelor [Duță et al., ICCVW21], detecția speciilor de păsări sau a emoției din vorbire [Ristea et al., INTERSPEECH21]. Totodată, în problema de visual question answering, am reușit să reproducem rezultatele din lucrarea [Tan et al., EMNLP19]. Pentru a ajunge la rezultate de top în problemele studiate, am luat în calcul și crearea de ansamble de modele [Găman et al., IJIS21; Ristea et al., INTERSPEECH21]. Atingând performanțe state-of-the-art, putem concluzia că obiectivul 2 din propunerea de proiect a fost îndeplinit.

- **Activitatea 2.2. Task 3 – Dezvoltarea și evaluarea unui model neuronal deep pentru clasificarea textelor, bazat pe paradigma de învățare bazată pe curriculum.**

Această activitate corespunde cu obiectivul 3 din cadrul propunerii de proiect. În vederea îndeplinirii acestui obiectiv, am efectuat experimente pe o problemă de clasificare a textelor, anume detecția polarității opiniei, folosind un model de tip BERT. Pentru a introduce antrenarea graduală, bazată pe curriculum, am utilizat modelul de predicție a complexității dezvoltat în cadrul activității 1.1 din 2020. Pe lângă această abordare, am încercat să obținem o antrenare graduală utilizând dimensiunea vocabularului de cuvinte. Mai precis, vocabularul crește treptat, pe măsură ce modelul se optimizează. Rezultatele preliminare arată că învățarea bazată pe curriculum produce performanțe asemănătoare cu antrenarea standard. În etapa următoare a proiectului, vom încerca să variem hiperparametrii implicați în învățarea bazată pe curriculum, cu scopul de a crește performanța.

- **Activitatea 2.3. Task 4 – Dezvoltarea și evaluarea unui model neuronal deep pentru visual question answering, bazat pe paradigma de învățare bazată pe curriculum.**

Am studiat posibilitatea aplicării învățării pe bază de curriculum pentru a antrena modele neuronale pentru diverse probleme din domeniul imaginilor: detectarea obiectelor, recunoașterea expresiilor faciale, estimarea vârstei persoanelor, clasificarea imaginilor obținute din semnale audio. Pentru problema detecției obiectelor, rezultatele obținute demonstrează că învățarea pe bază de curriculum conduce către performanțe mai bune. Metoda propusă și rezultatele aferente au fost sintetizate în lucrarea [Soviany et al., CVIU21]. De menționat că în vederea dezvoltării unui model pentru visual question answering, una din etapele importante este tocmai detectarea obiectelor în imagini. Astfel, studiul efectuat este relevant în îndeplinirea obiectivului 4. Pentru problemele de recunoașterea expresiilor faciale și estimarea vârstei persoanelor, am aplicat învățarea bazată pe curriculum folosind rețele neuronale de tip profesor-student. Metoda propusă și rezultatele aferente au fost sintetizate în lucrarea [Georgescu et al., MVA21]. Pentru clasificarea imaginilor obținute din semnale audio, am utilizat un algoritm de învățare care stabilește un curriculum pe baza etichetelor prezise de mai multe rețele neuronale convoluționale ce formează împreună un ansamblu. Metoda propusă și rezultatele aferente au fost sintetizate în lucrarea [Ristea et al., INTERSPEECH21].

Metodele de învățare bazată pe curriculum propuse în aceste lucrări [Georgescu et al., MVA21; Ristea et al., INTERSPEECH21; Soviany et al., CVIU21] au aplicabilitate imediată și în visual question answering. Astfel, în etapa următoare a proiectului, vom aplica abordările studiate pentru problema de visual question answering.

4. Diseminarea rezultatelor în articole științifice:

În urma activităților de cercetare fundamentală efectuate în 2021, au rezultat 9 articole publicate în volume ale unor jurnale sau conferințe internaționale, dintre care 2 articole în jurnale din zona roșie (Q1), 1 articol într-un jurnal din zona galbenă (Q2), 1 articol într-o conferință de categoria A* (ACL 2021), 2 articole în conferințe de categoria A (INTERSPEECH 2021, EACL 2021), 1 articol într-o conferință de categoria B (IJCNN 2021) și două articole în workshop-uri (VarDial 2021, NeurArch 2021). Acestea se adaugă celor 2 articole publicate în 2020. În concluzie, au fost îndeplinite cerințele minimale de diseminare a rezultatelor din proiect, care prevedeau publicarea a cel puțin 2 articole de categoria A sau A* (noi publicând 4 astfel de articole), cel puțin 2 articole de categoria B sau workshop-uri (noi publicând 4 astfel de articole) și cel puțin 1 articol de jurnal în zona roșie (noi publicând 2 astfel de articole). Articolele finanțate prin proiectului de cercetare sunt listate în continuare:

- **2020:**

1. A. Bărbălău, A. Cosma, R.T. Ionescu, M. Popescu. Black-Box Ripper: Copying black-box models using generative evolutionary algorithms. In Proceedings of NeurIPS, 2020. **(Rank A* Conference)**
2. M. Găman, R.T. Ionescu. Combining Deep Learning and String Kernels for the Localization of Swiss German Tweets. In Proceedings of VarDial Workshop of COLING, 2020. **(Rank A/2 Workshop)**

- **2021:**

3. P. Soviany, R.T. Ionescu, P. Rota, N. Sebe. Curriculum Self-Paced Learning for Cross-Domain Object Detection. Computer Vision and Image Understanding, 204: 103166, 2021. **(Q1 journal)**
4. A. Tache, M. Găman, R.T. Ionescu. Clustering Word Embeddings with Self-Organizing Maps. Application on LaRoSeDa - A Large Romanian Sentiment Data Set. In Proceedings of EACL, pp. 949–956, 2021. **(Rank A Conference)**
5. M. Găman, S. Cojocariu, R.T. Ionescu. UnibucKernel: Geolocating Swiss German Jodels Using Ensemble Learning. In Proceedings of VarDial (EACL Workshop), pp. 84–95, 2021. **(Rank A/2 Workshop)**
6. R.T. Ionescu, A.G. Chifu. FreSaDa: A French Satire Data Set for Cross-Domain Satire Detection. In Proceedings of IJCNN, pp. 1–8, 2021. **(Rank B Conference)**
7. A.C. Rogoz, M. Găman, R.T. Ionescu. SaRoCo: Detecting Satire in a Novel Romanian Corpus of News Articles. In Proceedings of ACL, pp. 1073–1078, 2021. **(Rank A* Conference)**
8. N.C. Ristea, R.T. Ionescu. Self-paced ensemble learning for speech and audio classification. In Proceedings of INTERSPEECH, pp. 2836–2840, 2021. **(Rank A Conference)**

9. I.C. Duță, M.I. Georgescu, R.T. Ionescu. Contextual Convolutional Neural Networks. In Proceedings of NeurArch (ICCV Workshop), pp. 403–412, 2021. **(Rank A*/2 Workshop)**
10. M. Găman, R.T. Ionescu. The Unreasonable Effectiveness of Machine Learning in Moldavian versus Romanian Dialect Identification. International Journal of Intelligent Systems, 2021. **(Q1 Journal)**
11. M.I. Georgescu, G. Duță, R.T. Ionescu. Teacher-Student Training and Triplet Loss to Reduce the Effect of Drastic Face Occlusion. Machine Vision and Applications, 2021. **(Q2 Journal)**

5. Referințe bibliografice:

- [Devlin et al., NAACL19] J. Devlin, M.W. Chang, K. Lee, K. Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of NAACL, pp. 4171–4186, 2019.
- [Duță et al., ICCVW21] I.C. Duță, M.I. Georgescu, R.T. Ionescu. Contextual Convolutional Neural Networks. In Proceedings of NeurArch (ICCV Workshop), pp. 403–412, 2021.
- [Găman et al., VarDial21] M. Găman, S. Cojocariu, R.T. Ionescu. UnibucKernel: Geolocating Swiss German Jodels Using Ensemble Learning. In Proceedings of VarDial (EACL Workshop), pp. 84–95, 2021.
- [Găman et al., IJIS21] M. Găman, R.T. Ionescu. The Unreasonable Effectiveness of Machine Learning in Moldavian versus Romanian Dialect Identification. International Journal of Intelligent Systems, 2021.
- [Georgescu et al., MVA21] M.I. Georgescu, G. Duță, R.T. Ionescu. Teacher-Student Training and Triplet Loss to Reduce the Effect of Drastic Face Occlusion. Machine Vision and Applications, 2021.
- [He et al., CVPR16] K. He, X. Zhang, S. Ren, J. Sun. Deep Residual Learning for Image Recognition. In Proceedings of CVPR, pp. 770–778, 2016.
- [Ionescu et al., IJCNN21] R.T. Ionescu, A.G. Chifu. FreSaDa: A French Satire Data Set for Cross-Domain Satire Detection. In Proceedings of IJCNN, pp. 1–8, 2021.
- [Ristea et al., INTERSPEECH21] N.C. Ristea, R.T. Ionescu. Self-paced ensemble learning for speech and audio classification. In Proceedings of INTERSPEECH, pp. 2836–2840, 2021.
- [Rogoz et al., ACL21] A.C. Rogoz, M. Găman, R.T. Ionescu. SaRoCo: Detecting Satire in a Novel Romanian Corpus of News Article.s. In Proceedings of ACL, pp. 1073–1078, 2021.
- [Soviany et al., CVIU21] P. Soviany, R.T. Ionescu, P. Rota, N. Sebe. Curriculum Self-Paced Learning for Cross-Domain Object Detection. Computer Vision and Image Understanding, 2020, 204: 103166, 2021.

[Tache et al., EACL21] A. Tache, M Găman, R.T. Ionescu. Clustering Word Embeddings with Self-Organizing Maps. Application on LaRoSeDa - A Large Romanian Sentiment Data Set. In Proceedings of EACL, pp. 949–956, 2021.

[Tan et al., EMNLP19] H. Tan, M. Bansal. LXMERT: Learning Cross-Modality Encoder Representations from Transformers. In Proceedings of EMNLP, pp. 5100–5111, 2019.

Data,
23.11.2021

Director proiect,
Radu Tudor Ionescu