

Species Entity Recognition Using a Neural Inhibitory Mechanism

Andrei-Marius Avram, Maria Mitrofan and Vasile Păiș

Research Institute for Artificial Intelligence "Mihai Drăgănescu", Romanian Academy, Bucharest, Romania

Abstract

This paper presents our system that participated in the LivingNER task organized at IberLEF. The system detects species mentioned in Spanish clinical case report documents, as provided by the task organizers. The implementation relies on the XLM-RoBERTa model followed by an additional inhibitory layer based on the lateral inhibition biological process, allowing the model to better focus on the content while potentially removing noise. We compared this approach with other commonly used mechanisms, such as a simple transformer-based NER system and text augmentation utilizing back-translation.

Keywords

Named entity recognition, Contextual embeddings, Species mention entity recognition, Lateral inhibition

1. Introduction

The use of clinical notes written by healthcare providers has long been recognized as a source of valuable information for clinical practice and medical research, helping physicians determine the potential causes of disease, making diagnoses and monitoring clinical outcomes. Clinical natural language processing (NLP) is a valuable field that automatically allows us to extract and analyze this knowledge from written notes. However, although clinical NLP is universally accepted as beneficial, the progress of the field has been slow, having limited impact in the real world.

The LivingNER task [1], part of the IberLEF 2022 shared evaluation campaign for Natural Language Processing (NLP) systems in Spanish and other Iberian languages, focused on a challenge on NER of species mentions and entity linking providing an exhaustively annotated large corpus of Spanish clinical case reports. One of the main aims of the task was to provide the context for developing systems that are ready to be used to solve clinical questions. It offered three sub-tasks: a) LivingNER-Species NER task, in which the participants were provided with a collection of documents containing plain text clinical case reports that they used to identify the exact character offsets of all species mentions, both human and non-human; b) LivingNER-Species Norm task, in which the participant's systems were asked to return all species mentions together with their corresponding NCBI taxonomy concept identifiers; c) LivingNER-Clinical


IberLEF 2022, September 2022, A Coruña, Spain.

✉ andrei.avram@racai.ro (A. Avram); maria@racai.ro (M. Mitrofan); vasile@racai.ro (V. Păiș)

ORCID 0000-0001-7045-038X (A. Avram); 0000-0001-7466-2013 (M. Mitrofan); 0000-0002-0019-7574 (V. Păiș)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

IMPACT task, in which the systems had to perform both a binary classification and to retrieve the list of NCBI Tax IDs that support the binary classification.

Our contribution to LivingNER consists of experimenting with several NER systems for the LivingNER-Species NER task. For all our experiments, we used the power of transfer learning and fine-tuned the multilingual XLM-RoBERTa [2] model on this task employing the standard sequence tagging procedure described in [3]. In addition, we experimented with the recently introduced lateral inhibition layer [4] which improves the neural network focus on more challenging cases that might be found in the training corpus. The best results on the evaluation set of the first task were obtained by a model that employs this layer, with an F1-score of 95.03. The other variants do not make use of this mechanism by at least 0.5%

The rest of the paper is structured as follows. Section 2 presents the related work. Section 3 outlines several statistics about the competition corpus, together with our pre-processing pipeline, while Section 4 describes the lateral inhibition mechanism and how it is used in our overall system. Section 5 provides the evaluation results of our experiments. Section 6 concludes our work and outlines possible future research directions.

2. Related work

In the past few years, several other relevant tasks in developing NER systems have introduced important advances in developing technologies for the biomedical domain. Named Entity Recognition, Concept Normalization and Clinical Coding, Cantemist shared task at IberLEF 2020[5], was structured into three independent sub-tasks, each taking into account a significant use case scenario: Cantemist-NER track that required finding tumor morphology mentioned automatically in the text; Cantemist-NORM track, that required to return all tumor morphology entity mentions together with their corresponding eCIE-O codes, i.e. finding and normalizing tumor morphology mentions; Cantemist-CODING track, that required returning for each of the documents a ranked list of correct eCIE-O code assignments. The best F1-scores were 0.86 for NER, 0.82 for NORM, and 0.83 for CODING.

MEDDOPROF[6] shared task at IberLEF 2021 on automatic recognition, classification and normalization of professions and occupations from medical texts proposed three challenges: NER (recognition of professions, employment statuses and activities in text), CLASS (classifying each occupation mentioned to its holder, i.e. patient or family member) and NORM (normalizing mentions to their identifier in ESCO or SNOMED CT). The best-performing systems were based on deep-learning approaches and achieved 0.81 F-score in occupation detection (NER), 0.79 in classifying occupations to their referent (CLASS) and 0.61 in normalization (NORM).

The ProfNER shared task[7] part of the SMM4H workshop and the shared task[8] organized at NAACL 2021, focused on the identification of professions and occupations from health-relevant Twitter messages written in Spanish. It offered two sub-tasks: a) a binary classification task, deciding if a particular tweet contains a mention of occupation, given the context, and b) extracting the named entities by specifying the entity type, start and end offset, as well as the actual text span. Best-performing participants built systems based on recent NLP technologies (e.g. transformers) and achieved 0.93 F-score in Text Classification and 0.839 in Named Entity Recognition.

[9] stated, once the NEs are identified and classified, they can be used in workflows to perform different functionalities such as document summarization or question answering. A survey of NER using deep learning models[10] emphasized the importance of pre-trained word embedding representations. With the introduction of contextual word representation models, like BERT [3], ELMo[11], ROBERTA [12], XLNet[13], NER systems have been adapted to make use of these new models. Contextualized embeddings were applied with success also in domain-specific settings. Considering the ProfNER shared task[7], the best performing system made use of the BETO model[14] trained on a big Spanish corpus.

Performing a training task with a contextualized embedding model needs heavy processing resources; therefore, not all flavours are available for all languages. Multilingual models such as mBERT and XLM-RoBERTa that use training data in multiple languages can be used in such situations performing especially well on low-resourced languages. For training NER systems for the Romanian language, the XML-ROBERTa model was successfully used in different sub-domains, such as legal [15] and biomedical [?].

3. Dataset and pre-processing

The dataset for the first task of the LivingNER competition contains 1000 files for training, 500 files for validation and 13,467 files for testing. Each file contains text in the Spanish language from the medical domain, each presenting a clinical case from various specialities. The training and the validation subsets are also annotated with two entities that a model must learn to recognize: HUMAN and SPECIES. The HUMAN entity simply denotes the appearance of a human being in text (e.g. "paciente", "hombre"), while the SPECIES entity marks a biological being that is not a human (e.g. "COVID-19", "bacterianas"). There are 23,203 entities in the dataset, quite evenly balanced, 44.4% being a HUMAN entity, while 55.6% being a SPECIES entity.

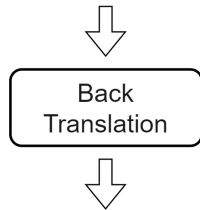
The entity annotations are given as a start/end character offset in a text file. This format does not fit with our internal NER tool, and we had to convert it to a CoNLL-like format where on each row, we have a token annotated with the Inside-Outside-Beginning (IOB) style. To tokenize the text, we used the `spacy es_core_news_sm` Spanish text processing pipeline. In addition, because our employed pre-trained model supports a maximum length of 512 tokens for the input sequence, we further split each text into sentences representing our training samples.

To further experiment with possible improvements on this task, we used the back-translation text augmentation mechanism [16] to generate new sentences that retain mostly the same meaning. However, directly employing this mechanism could potentially affect the annotated entities, and instead, we employ it to augment the text found between entities, thus maintaining the meanings of entities in the given paragraphs. Figure 1 depicts our text augmentation approach through back-translation. We used the OPUS-MT toolkit [17] for both Spanish-English and English-Spanish translation.

4. System architecture

The primary system for participating in the LivingNER task is based on contextualized embeddings extracted from the XLM-RoBERTa model. These are then passed through an additional

[Por lo que respecta a la] madre [de la paciente, hemos sabido que se detectó ARN vírico en muestras de heces de la] madre [el 4 de febrero de 2020 (día 8 de enfermedad)]



[Por lo que se refiere a la] madre [del paciente, hemos sabido que se detectó ARN viral en muestras de heces de la] madre [el 4 de febrero de 2020 (día de la enfermedad 8)]

Figure 1: Illustration of the back-translation mechanism employed in our text augmentation technique. The square brackets are not from the original text and they denote the text fragments that are sent to be back-translated.

layer, recently developed in our institute, inspired by the biological process of *lateral inhibition* [18]. In the human brain, exciting neurons can reduce the activity of other neighbouring neurons. This neural dynamic accounts for increased perception under challenging scenarios, such as low-lighting conditions. Intuitively, based on the biological observations, we consider that the new layer should be able to focus better on challenging cases in the corpus, thus increasing the overall performance. Following the inhibitory layer, a standard contextual-based NER approach is used, consisting of a linear layer taking for input the embeddings values and feeding the output into a classification head, indicating each entity class. The architecture diagram is given in Figure 4.

The lateral inhibition layer, introduced by Păiș [4], was previously used successfully for two other NER tasks: complex named entity recognition [4], and Romanian biomedical NER [?]. Corresponding to the biological process, an embedding value is either allowed to pass unchanged to the next layer or set to zero, depending on the other adjacent values. The implementation is based on two matrices: weights (W) and bias (B). However, different from other layers, the diagonal values of the weights matrix are always set to zero to allow computation based on the interaction with adjacent neurons. The Heaviside function (see Equation 1) is then applied to determine which values pass through the layer or become zero. The complete equation for the forward pass is given in Equation 2, where X is the layer’s input vector (a token embedding representation), $Diag$ represents a matrix having the diagonal set to the vector presented as a parameter, $ZeroDiag$ is the matrix with the value zero on the diagonal, W is the weights matrix, and B corresponds to the bias values.

$$\Theta(x) = \begin{cases} 1, & x > 0 \\ 0, & x \leq 0 \end{cases} \quad (1)$$

$$F(X) = X * Diag(\Theta(X * ZeroDiag(W) + B)) \quad (2)$$

In the backward pass, the Heaviside function was approximated with the sigmoid function with a scaling parameter [19] (Equation 3, where k is the scaling parameter). This approximation

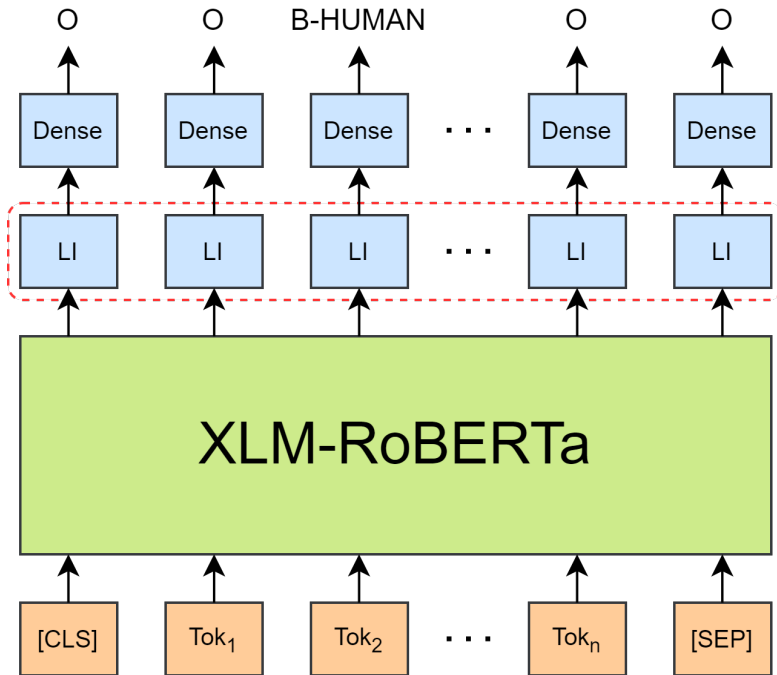


Figure 2: System architecture with the lateral inhibitory (LI) layer included.

accounts for the Heaviside function, which does not have a derivative. The approximation technique is known in the literature as surrogate gradient learning [20]. It allows using a non-differentiable function (in our case, the Heaviside function) in the forward pass while replacing it with another function, with a known derivative, in the backward pass (in our case, the parameterized sigmoid function). The derivative of the sigmoid function is then given in Equation 4, where $\sigma(x)$ is the same as in Equation 3. Figure 4 shows plots for the sigmoid and derivative functions. For higher k values, the function approximates better the Heaviside function. For the experiments reported in this paper, we used a value of $k = 10$. It is foreseeable that different k values will impact the results. Nevertheless, we restricted the experiments to a single value due to the limited time and computation available.

$$\sigma(x) = \frac{1}{1 + e^{-kx}} \quad (3)$$

$$\sigma'(x) = k\sigma(x)\sigma(-x) \quad (4)$$

In addition to the system described in Figure 4 (i.e. XLM-RoBERTa with a lateral inhibitory layer between the contextualized embeddings and the dense layer), we also tried several other approaches, such as plain XLM-RoBERTa without the inhibitory layer, data augmentation through back-translation (described in the previous section), matching in SNOMED-CT, and an ensemble system that uses a voting system to select the entities.

For the SNOMED-CT approach, we extracted the Organism category from the Spanish version

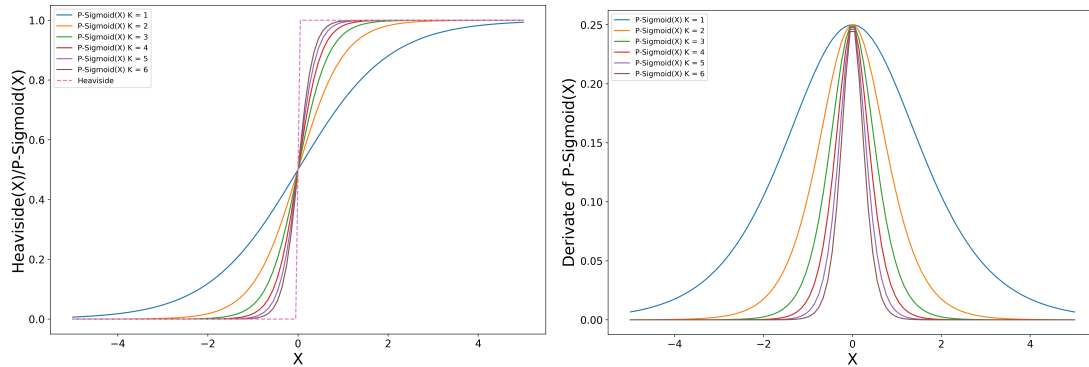


Figure 3: Illustration of the parameterized sigmoid (left) and its derivative for different K values (right).

Model	SPECIES			HUMAN			TOTAL		
	P	R	F1	P	R	F1	P	R	F1
XLM-RoBTa [2]	93.85	94.05	93.95	97.95	97.71	97.83	95.77	95.78	95.78
XLM-RoBTa + LI	93.58	94.19	93.88	97.40	98.05	97.72	95.58	96.00	95.59
XLM-RoBTa + Aug	94.61	93.41	94.00	97.86	97.47	97.66	96.14	95.32	95.73
XLM-RoBTa + LI + Aug	92.79	94.97	93.87	97.31	97.96	97.63	94.89	96.38	95.63
Ensemble	95.92	90.45	92.81	97.89	97.83	97.85	96.11	95.62	95.84

Table 1

Results obtained by our models on the validation set for the first task of the LivingNER competition.

of SNOMED-CT, which includes Bacteria and Virus Species. We then implemented a lookup script that attempted to match a token or a group of tokens with SNOMED-CT data. However, given that this data is missing other living species and the precision of our NER model, we did not find any real benefit in using SNOMED-CT.

5. Results

The generated models were evaluated first on the validation set provided by the organizers. Results are given in Table 1. Considering the overall F1 scores, it seems the ensemble system achieves the highest value. However, the standard deviation is $\sigma = 0.09$, while the variance is $\sigma^2 = 0.008$. Furthermore, looking at individual classes and other indicators, such as precision and recall, different systems achieve the highest results. Therefore there is no clear choice for the best performing system, and thus we submitted all the generated data.

The organizers provided the final evaluation for all the test submissions. The results are presented in Table 2. In this case, the standard deviation is $\sigma = 0.23$, while the variance is $\sigma^2 = 0.056$. The lateral inhibition layer system achieves the highest score of 95.03% F1. It also achieves the highest F1 scores for individual classes.

The data augmentation technique did not improve the overall F1 score. However, the augmented dataset improves the recall for the species class. An interesting result is that in both

Model	SPECIES			HUMAN			TOTAL		
	P	R	F1	P	R	F1	P	R	F1
XLM-RoBTa [2]	93.91	90.76	92.31	97.77	97.40	97.58	95.53	93.58	94.55
XLM-RoBTa + LI	93.91	91.48	92.68	98.10	98.32	98.21	95.69	94.39	95.03
XLM-RoBTa + Aug	92.01	91.51	91.76	97.72	98.06	97.89	94.40	94.30	94.35
XLM-RoBTa + LI + Aug	94.32	91.01	92.64	98.08	97.40	97.74	95.92	93.73	94.81
Ensemble	94.84	90.38	92.56	98.09	97.62	97.85	96.22	93.46	94.82

Table 2

Results obtained by our models on the evaluation set for the first task of the LivingNER competition

not-augmented and augmented scenarios, the introduction of the lateral inhibition layer helps to improve the final F1 metric. The ensemble system improves the precision in both species and overall results. It is also the second-best, after the simple embeddings with the lateral inhibition method.

6. Conclusion

The lateral inhibition layer proved to be beneficial to the overall system performance. This outcome confirms previous results indicating a gain from using this system in difficult cases associated with complex named entities or low-resource settings. As mentioned in Section 4, further experiments with different values for the parameter k are needed to better understand its contribution. Furthermore, given the general nature of the proposed inhibitory layer, it is conceivable that it may be used on other natural language processing tasks for both text and speech. Our future efforts currently focus on integrating this layer into a speech recognition component for the Romanian language and also using it in developing NER systems for Romanian micro-blogging texts [21].

References

- [1] A. Miranda-Escalada, E. Farré-Maduell, S. Lima-López, D. Estrada, L. Gascó, M. Krallinger, Mention detection, normalization classification of species, pathogens, humans and food in clinical documents: Overview of LivingNER shared task and resources, 2022.
- [2] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, É. Grave, M. Ott, L. Zettlemoyer, V. Stoyanov, Unsupervised Cross-lingual Representation Learning at Scale, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020, pp. 8440–8451.
- [3] J. D. Kenton, M.-W. Chang, L. K. Toutanova, BERT: Pre-training of Deep Bidirectional transformers for Language Understanding, in: Proceedings of NAACL-HLT, 2019, pp. 4171–4186.
- [4] V. Păiș, RACAI at SemEval-2022 Task 11: Complex named entity recognition using a lateral inhibition mechanism, in: Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval), 2022.

- [5] A. Miranda-Escalada, E. Farré, M. Krallinger, Named entity recognition, concept normalization and clinical coding: Overview of the CANTEMIST track for cancer text mining in Spanish, corpus, guidelines, methods and results, in: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2020), CEUR Workshop Proceedings, 2020.
- [6] S. Lima-López, E. Farré-Maduell, A. Miranda-Escalada, V. Brivá-Iglesias, M. Krallinger, NLP applied to occupational health: MEDDOPROF shared task at IberLEF 2021 on automatic recognition, classification and normalization of professions and occupations from medical texts, *Procesamiento del Lenguaje Natural* 67 (2021) 243–256. URL: <http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/6393>.
- [7] A. Miranda-Escalada, E. Farré-Maduell, S. Lima-López, L. Gasco Sanchez, V. Briva-Iglesias, M. Agüero-Torales, M. Krallinger, The ProfNER shared task on automatic recognition of occupation mentions in social media: systems, evaluation, guidelines, embeddings and corpora, 2021. doi:10.5281/zenodo.4309356.
- [8] A. Magge, A. Klein, A. Miranda-Escalada, M. Ali Al-Garadi, I. Alimova, Z. Miftahutdinov, E. Farre, S. Lima López, I. Flores, K. O’Connor, D. Weissenbacher, E. Tutubalina, A. Sarker, J. Banda, M. Krallinger, G. Gonzalez-Hernandez, Overview of the Sixth Social Media Mining for Health Applications (#SMM4H) shared tasks at NAACL 2021, in: Proceedings of the Sixth Social Media Mining for Health (#SMM4H) Workshop and Shared Task, Association for Computational Linguistics, Mexico City, Mexico, 2021, pp. 21–32. URL: <https://aclanthology.org/2021.smm4h-1.4>. doi:10.18653/v1/2021.smm4h-1.4.
- [9] Q. Zhong, Y. Tang, An attention-based BILSTM-CRF for Chinese named entity recognition, in: 2020 IEEE 5th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA), IEEE, 2020, pp. 550–555.
- [10] V. Yadav, S. Bethard, A survey on recent advances in named entity recognition from deep learning models, in: Proceedings of the 27th International Conference on Computational Linguistics, Association for Computational Linguistics, Santa Fe, New Mexico, USA, 2018, pp. 2145–2158. URL: <https://aclanthology.org/C18-1182>.
- [11] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, L. Zettlemoyer, Deep contextualized word representations, in: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), Association for Computational Linguistics, New Orleans, Louisiana, 2018, pp. 2227–2237. URL: <https://aclanthology.org/N18-1202>. doi:10.18653/v1/N18-1202.
- [12] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, Roberta: A robustly optimized BERT pretraining approach, *CoRR* abs/1907.11692 (2019). URL: <http://arxiv.org/abs/1907.11692>. arXiv:1907.11692.
- [13] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. R. Salakhutdinov, Q. V. Le, Xlnet: Generalized autoregressive pretraining for language understanding, in: H. Wallach, H. Larochelle, A. Beygelzimer, F. d’Alché-Buc, E. Fox, R. Garnett (Eds.), *Advances in Neural Information Processing Systems*, volume 32, Curran Associates, Inc., 2019. URL: <https://proceedings.neurips.cc/paper/2019/file/dc6a7e655d7e5840e66733e9ee67cc69-Paper.pdf>.
- [14] J. Cañete, G. Chaperon, R. Fuentes, J.-H. Ho, H. Kang, J. Pérez, Spanish pre-trained BERT model and evaluation data, in: PML4DC at ICLR 2020, 2020.
- [15] V. Păiș, M. Mitrofan, C. L. Gasan, V. Coneschi, A. Ianov, Named entity recognition in the Ro-

- manian legal domain, in: Proceedings of the Natural Legal Language Processing Workshop 2021, Association for Computational Linguistics, Punta Cana, Dominican Republic, 2021, pp. 9–18. URL: <https://aclanthology.org/2021.nllp-1.2>. doi:10.18653/v1/2021.nllp-1.2.
- [16] R. Sennrich, B. Haddow, A. Birch, Improving neural machine translation models with monolingual data, in: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2016, pp. 86–96.
- [17] J. Tiedemann, S. Thottingal, OPUS-MT - building open translation services for the world, in: Proceedings of the 22nd Annual Conference of the European Association for Machine Translation, European Association for Machine Translation, 2020.
- [18] R. A. Cohen, Lateral Inhibition, Springer New York, New York, NY, 2011, pp. 1436–1437. URL: https://doi.org/10.1007/978-0-387-79948-3_1379. doi:10.1007/978-0-387-79948-3_1379.
- [19] T. C. Wunderlich, C. Pehle, Event-based backpropagation can compute exact gradients for spiking neural networks, Scientific Reports 11 (2021) 12829. URL: <https://doi.org/10.1038/s41598-021-91786-z>. doi:10.1038/s41598-021-91786-z.
- [20] E. O. Neftci, H. Mostafa, F. Zenke, Surrogate gradient learning in spiking neural networks: Bringing the power of gradient-based optimization to spiking neural networks, IEEE Signal Processing Magazine 36 (2019) 51–63. doi:10.1109/MSP.2019.2931595.
- [21] V. Păiș, M. Mitrofan, V. Barbu Mititelu, E. Irimia, R. Micu, C. L. Gasan, Challenges in creating a representative corpus of Romanian micro-blogging text, in: Proceedings of the Workshop on Challenges in the Management of Large Corpora (CMLC-10), European Language Resources Association, Marseille, France, 2022, pp. 1–7. URL: <https://aclanthology.org/2022.cmlc10-1.1>.