NLP4SM: Natural Language Processing for social media

Gonzalo Medina Medina[1], Jose Camacho Collados[2] and Eugenio Martínez Cámara[1]

[1]Department of Computer Science and Artificial Intelligence, Andalusian Research Institute in Data Science and Computational Intelligence (DaSCI), University of Granada, Spain

[2]School of Computer Science and Informatics, Cardiff University, United Kingdom

Abstract

NLP4SM is a website for the execution, analysis and comparison of tweet classification methods based on language models. Currently, NLP4SM supports the text classification tasks considered in TweetEval, but it aims at integrating additional text clasification tasks and to wider the number of language models available with the goal of becoming to a benchmark platform for assessing text classification methods with real data from social media.

Keywords

Language models, text classification, social media.

1. Introduction

The most likely source of the vertiginous progress of Natural Language Processing (NLP) in the recent years is the proposal of the Word2Vec model [1], which eases the generation of unsupervised linguistic features that are known as word embeddings and they represent the meaning of words in vectors of real numbers. The strong results reached by word embeddings based on Word2Vect enhanced the design of new word embeddings models, such as Glove.[1] These models set an embedding vector to each word regardless of its context, and for this reason the next landmark were starred by the contextual word embeddings models [2]. The transformers models stand out as contextual word embeddings, with BERT [3] as outstanding example. These models are known as language models, and their capacity of representing the meaning of words couple with the possibility of using them as pre-trained models have driven the progress of a broad branch of NLP tasks, especially those mostly linked to the classification of the semantic meaning of text, such as the opinion polarity of a review, the offensive meaning or the underlying emotional meaning of a message.

The potential of language models has made them the baseline of a wide range of NLP tasks, and they can even be used for developing learning models in production environments. On the other hand, the ease of tuning these models to specific NLP tasks has led the development and release of a huge amount of pre-trained language models in a large bunch of NLP task, with HuggingFace and especially its Transformers library [4] standing out. This vast variety of language models makes their comparison and analysis really difficult as a previous step of the particular language model to fine-tune to a specific use case.

The certain use of language in social networks makes to adapt the NLP methods to the specific use of language of each social network, as for instance to Twitter [5]. Language models also needs this fitting to the use of language of social networks, which makes them to be at the top of most NLP shared-tasks.

The great availability of language models has not been coupled with the release of web platforms for comparing and analysing the different language models in specific NLP tasks. Nevertheless, the issue of the great availability of training corpora and the evaluation of learning models begins to be resolved by the publication of leader boards of learning models trained on gold standards, such as SuperGLUE [6] or TweetEval [7].

Following the example of the NLP classification tasks leader boards, we present the web platform NLP4SM,[2,3] whose demonstrative prototype is described in this paper. NLP4SM is a web application for analysing the performance of Twitter language

SEPLN-PD 2022. Annual Conference of the Spanish Association for Natural Language Processing 2022: Projects and Demonstrations, September 21-23, 2022, A Coruña, Spain

gmedina95@correo.ugr.es (G. Medina Medina); camachocolladosj@cardiff.ac.uk (J. Collados); emcamara@decsai.ugr.es (E. Martínez Cámara)
0000-0003-1618-7239 (J. Collados); 0000-0002-5279-8355 (E. Martínez Cámara)

[1] https://nlp.stanford.edu/projects/glove/

[2] Prototype: https://nlp4sm.on.fleek.co/

[3] Production [8]: https://tweetnlp.org/demo/

models fine-tuned to the tasks of (1) sentiment analysis, (2) emotion analysis, (3) offensive language classification, (4) hate speech classification, (5) irony detection and (6) stance classification on abortion, climate change, atheism, feminism and Hillary Clinton. NLP4SM allows on one hand the classification of a free span of text, and on the other hand the classification of the meaning of a bunch of tweets returned by Twitter. Furthermore, the classification results are shown as charts to ease their understanding. NLP4SM can be used by non-NLP experts and NLP scientists that need to compare different language models in one of the mentioned tasks on real data. The design of the system allows the consideration of new language models of the previous NLP tasks, as well as the incorporation of new result visualisation methods.

2. Language Models in NLP4SM

The first version of NLP4SM incorporates learning models that classify the meaning of tweets. The learning models are based on the fine-tuning of Twitter language models to the specific NLP tasks, which we subsequently describe.

2.1. NLP tasks

We select the NLP tasks according to their scientific relevancy, as well as the high social demand to have automatic systems that can identify specific kind of messages. The tasks are also part of TweetEval, and we present them as what follows.

Emotion analysis It identifies the underling emotion of a text. Although it is a multi-label task, we redefined it as a multi-class classification task. The corpus "Affect in Tweets" [9] was used to fit the model to the most frequent emotions of the corpus: joy, optimism, anger and sadness.[4]

Sentiment analysis It classifies the opinion polarity in positive, negative or neutral. The corpus of the subtask A of "Sentiment Analysis in Twitter" task of SemEval17 [10] was used to fit the model.[5]

Hate speech It aims at classifying whether a tweet express hate. The corpus of HateEval from SemEval19 was used to fit the model [11].[6]

Irony detection The goal is to classify whether a tweet is ironic. The corpus of the Irony Detection task from SemEval18 was used to fit the model [12].[7]

Offensive language It identifies whether a span of text has an offensive meaning. The corpus of OffensEval from SemEval19 was used to fit the model [13].[8]

Emoji prediction It aims at predicting the emoji that best represent the meaning of a tweet. The corpus of Emoji Prediction from SemEval18 was used to fit the model [14].[9]

Stance classification It classifies the author stance according to a topic. The corpus of the task Detectin Stance from SemEval16 was used to fit the model. The topics considered are: abortion,[10] atheism,[11] feminism,[12] climate change[13] and Hillary Clinton.[14]

Multinguality Social networks are multilingual, and for this reason NLP4SM also allows to analyse multilingual language models, namely those ones based on XLM-R [15] that is fitted on a large set of tweets written in more than 50 languages. NLP4SM also provides the XLM-T language model fitted to the sentiment analysis task in eight different languages [16].

2.2. Language Models

The language models currently included in NLP4SM match with the ones in TweetEval and they are available in HuggingFace. We have used the RoBERTa-base model [17] pre-trained on English text from social networks [7].

The fine-tuning of RoBERTa-base to each NLP task is based on a output layer with the same output units than the number of classes of each task [17].

[4] https://huggingface.co/cardiffnlp/twitter-roberta-base-emotion
[5] https://huggingface.co/cardiffnlp/twitter-roberta-base-sentiment
[6] https://huggingface.co/cardiffnlp/twitter-roberta-base-hate
[7] https://huggingface.co/cardiffnlp/twitter-roberta-base-irony
[8] https://huggingface.co/cardiffnlp/twitter-roberta-base-offensive
[9] https://huggingface.co/cardiffnlp/twitter-roberta-base-emoji
[10] https://huggingface.co/cardiffnlp/twitter-roberta-base-stance-abortion
[11] https://huggingface.co/cardiffnlp/twitter-roberta-base-stance-atheism
[12] https://huggingface.co/cardiffnlp/twitter-roberta-base-stance-feminist
[13] https://huggingface.co/cardiffnlp/twitter-roberta-base-stance-climate
[14] https://huggingface.co/cardiffnlp/twitter-roberta-base-stance-hillary

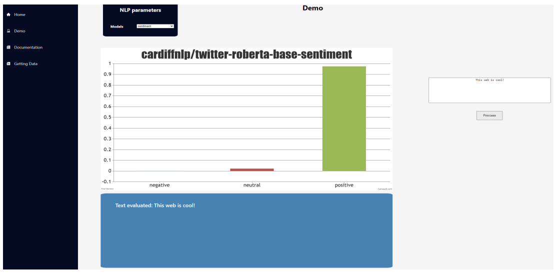

Figure 1: Sentiment analysis, 'text mode' mode.

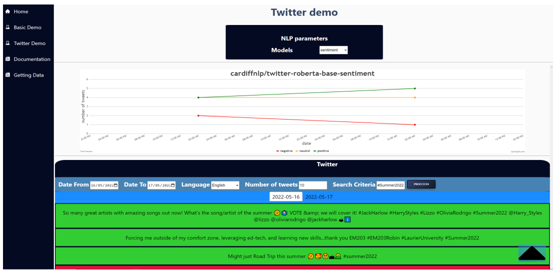

Figure 2: Sentiment analysis, 'Twitter mode'.

The languages models used are described and linked in section 2.1.

3. Description of NLP4SM

We aim at providing an unified and accessible platform for assessing and analysing social network text classification models. Hence, we have developed a web application for the first version of NLP4SM.

NLP4SM is built upon a client-server architecture led by a REST API. Moreover, we have relied on external services for running the language models. NLP4SM uses Huggingface because it is currently the on-cloud service that hosts the language models included in NLP4SM, it is the artificial intelligence service platform most used by the NLP research community and it provides a high quality service.

The server side is developed in Python and it is based on the micro-framework Flask. The server side is responsible of the communication with HuggingFace through using its API. Moreover, the server side queries Twitter according to the user query.

The client side is a web interface based on JavaScript React. It allows two different forms of evaluating the models, namely:

Text mode It evaluates any language model described in section 2 with a span of text written down by the user in a text box. Several charts show the result of the evaluation. Figure 1 depicts and example of the text mode.

Twitter mode It process a set of tweets returned in real-time from Twitter according to the user query. The user can configure his query according to the language, the time and the specific text of the query. NLP4SM retrieves the tweets and shows with different kind of charts the result of running the selected language model. Figure 2 depicts and example of the text mode.

4. Conclusions and future work

In this paper, we presented the prototype demonstration NLP4SM, which aims at easing the access, analysis and comparison of classification models based on language models of different NLP tasks with real data from social networks. NLP4SM allows the evaluation of any span of text, and the evaluation of tweets from a user query.

We plan as future work: (1) to integrate more NLP tasks, (2) to extend the number of language models considered, and (3) to add a greater number of visualisation methods of results.

Acknowledgments

This research work is supported by the R&D&I grant PID2020-116118GA-I00 funded by MCIN/AEI/10.13039/501100011033.

References

[1] T. Mikolov, K. Chen, G. Corrado, J. Dean, Efficient estimation of word representations in vector space, in: Proc. of Workshop at ICLR, 2013.

[2] M. T. Pilehvar, J. Camacho-Collados, Embeddings in natural language processing: theory and advances in vector representations of meaning, Synthesis Lectures on Human Language Technologies 13 (2020) 1–175.

[3] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, in: Proc. of the 2019 Conf. of the NAACL, Vol. 1 (Long and Short Papers), 2019, pp. 4171–4186. URL: https://aclanthology.org/N19-1423. doi:10.18653/v1/N19-1423.

[4] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, J. Davison, S. Shleifer,

P. von Platen, C. Ma, Y. Jernite, J. Plu, C. Xu, T. Le Scao, S. Gugger, M. Drame, Q. Lhoest, A. Rush, Transformers: State-of-the-art natural language processing, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, Association for Computational Linguistics, Online, 2020, pp. 38–45. URL: https://aclanthology.org/2020.emnlp-demos.6. doi:10.18653/v1/2020.emnlp-demos.6.

[5] E. Martínez-Cámara, M. T. Martín-Valdivia, L. A. Ureña-López, A. Montejo-Ráez, Sentiment analysis in twitter, Natural Language Engineering 20 (2014) 1–28. doi:10.1017/S1351324912000332.

[6] A. Wang, Y. Pruksachatkun, N. Nangia, A. Singh, J. Michael, F. Hill, O. Levy, S. Bowman, Superglue: A stickier benchmark for general-purpose language understanding systems, 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/4496bf24afe7fab6f046bf4923da8de6-Paper.pdf.

[7] F. Barbieri, J. Camacho-Collados, L. Espinosa Anke, L. Neves, TweetEval: Unified benchmark and comparative evaluation for tweet classification, in: Findings of the ACL: EMNLP 2020, 2020. URL: https://aclanthology.org/2020.findings-emnlp.148. doi:10.18653/v1/2020.findings-emnlp.148.

[8] J. Camacho-Collados, K. Rezaee, T. Riahi, A. Ushio, D. Loureiro, D. Antypas, J. Boisson, L. Espinosa-Anke, F. Liu, E. Martínez-Cámara, et al., Tweetnlp: Cutting-edge natural language processing for social media, arXiv preprint arXiv:2206.14774 (2022).

[9] S. Mohammad, F. Bravo-Marquez, M. Salameh, S. Kiritchenko, SemEval-2018 task 1: Affect in tweets, in: Proc. of The 12th Int. Workshop on Semantic Evaluation, 2018, pp. 1–17. URL: https://aclanthology.org/S18-1001. doi:10.18653/v1/S18-1001.

[10] S. Rosenthal, N. Farra, P. Nakov, SemEval-2017 task 4: Sentiment analysis in Twitter, in: Proc. of the 11th Int. Workshop on Semantic Evaluation (SemEval-2017), 2017, pp. 502–518. URL: https://aclanthology.org/S17-2088. doi:10.18653/v1/S17-2088.

[11] V. Basile, C. Bosco, E. Fersini, D. Nozza, V. Patti, F. M. Rangel Pardo, P. Rosso, M. Sanguinetti, SemEval-2019 task 5: Multilingual detection of hate speech against immigrants and women in Twitter, in: Proceedings of the 13th International Workshop on Semantic Evaluation, Association for Computational Linguistics, Minneapolis, Minnesota, USA, 2019, pp. 54–63. URL: https://aclanthology.org/S19-2007. doi:10.18653/v1/S19-2007.

[12] C. Van Hee, E. Lefever, V. Hoste, SemEval-2018 Task 3: Irony detection in English tweets, in: Proc. of The 12th Int. Workshop on Semantic Evaluation, 2018. URL: https://aclanthology.org/S18-1005. doi:10.18653/v1/S18-1005.

[13] M. Zampieri, S. Malmasi, P. Nakov, S. Rosenthal, N. Farra, R. Kumar, SemEval-2019 task 6: Identifying and categorizing offensive language in social media (OffensEval), in: Proceedings of the 13th International Workshop on Semantic Evaluation, Association for Computational Linguistics, Minneapolis, Minnesota, USA, 2019, pp. 75–86. URL: https://aclanthology.org/S19-2010. doi:10.18653/v1/S19-2010.

[14] F. Barbieri, J. Camacho-Collados, F. Ronzano, L. Espinosa-Anke, M. Ballesteros, V. Basile, V. Patti, H. Saggion, SemEval 2018 task 2: Multilingual emoji prediction, in: Proceedings of The 12th International Workshop on Semantic Evaluation, Association for Computational Linguistics, New Orleans, Louisiana, 2018, pp. 24–33. URL: https://aclanthology.org/S18-1003. doi:10.18653/v1/S18-1003.

[15] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. 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, Association for Computational Linguistics, Online, 2020, pp. 8440–8451. URL: https://aclanthology.org/2020.acl-main.747. doi:10.18653/v1/2020.acl-main.747.

[16] F. Barbieri, L. Espinosa-Anke, J. Camacho-Collados, A Multilingual Language Model Toolkit for Twitter, in: arXiv preprint arXiv:2104.12250, 2021.

[17] 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, arXiv preprint arXiv:1907.11692 (2019).