Profiling Irony and Stereotype Spreaders on Twitter Using Multi-View Learning

Mohammad Karami Sheykhlan and Saleh Kheiri Abdoljabbar

June 19, 2022

Profiling Irony and Stereotype Spreaders on Twitter using multi-view learning

Mohammad Karami Sheykhlan[1], Saleh Kheiri Abdoljabbar[2]

[1]Independent Researcher
[2]University of Tabriz

Abstract
With the rise of social media, millions of people are using it every day. They may publish content about everything. In addition to maintaining freedom of speech, social media executives must restrict the spread of harassing speech. For this purpose, the Sheykhlan team developed a system with multi-view learning in combination with an SVM, to identify malicious users. The proposed approach achieved 94.63% accuracy on 5-fold cross-validation on the English dataset.

Keywords
TF-IDF, SVM, Irony and Stereotype Spreaders, Twitter, RoBERTa, Deep Learning, Machine Learning

1. Introduction

Twitter is one of the social networks where a large number of tweets are published daily. While users interact with many people and express personal views, users who post annoying tweets about the LGBT community, women, color, etc. should be restricted.

In PAN @ CLEF 2021, the Profiling Hate Speech Spreaders on Twitter task[1] was in both English and Spanish, and the best overall accuracy was 73% obtained by [2]. However, Dukic et al.[3] has achieved 75% accuracy on English datasets, which is the best accuracy in English dataset.

This paper details the proposed model for the PAN@CLEF 2022 Author Profiling Shared Task, Profiling Irony and Stereotype Spreaders on Twitter (IROSTEREO).—-

In Section 2 we present our method and preprocessing strategies that we have tested. and finally, In Section 3 we show our results.

2. Methodology

This section presents the dataset and the models utilized in the experimentation. For this we have used python and toolkits of NLTK [4], emoji[1], keras[5], sklearn[6], and transformers[7]. We then experimented with several feature representations, and finally, the combination of TF-IDF with n-grams weight and support vector machine classifier[8] achieved the best accuracy.

CLEF 2022: Conference and Labs of the Evaluation Forum, September 5–8, 2022, Bologna, Italy
mohammadkaramisheykhlan@gmail.com (M. K. Sheykhlan); salehkheiri@gmail.com (S. K. Abdoljabbar)

[1]https://pypi.org/project/emoji/

2.1. Corpus

The corpus of this task is composed of author tweets in English. The training corpus contains 420 XML files and each file includes 200 tweets from an author. Unlike in previous years, the test set was provided to participants without ground truth with 180 XML files, and we can upload test set results to the TIRA Integrated Research Architecture[9].

2.2. Preprocessing

Firstly, like [10][11] each author's tweets are concatenated with \n in a single string per author. Then author tweets converted to lowercase, removing punctuations, stopwords, emojis, and numbers. After That, performed lemmatization and replaced #USER#, #MENTION#, and #URL# with usr, mntn, and url, respectively.

2.3. SVM Classifier

First of all, We split the original training dataset into two subsets to test our models. based on the train_test_split method in sklearn, We randomly assign 80%-20% from the training dataset to training-valid splits. Then we expremented the TF-IDF with n-gram weights, pre-trained word2vec, and RoBERTa representation.

For TF-IDF, A feature extraction module as shown in Figure 1 is used to extract char (1, 9) and word (1, 6) n-grams. Also we convert per word of tweets to sequence of Part Of Speech tagging (POS) Then used POS word ngram (1, 3).
For example:

"Still no use. Still no banks ." = "RB DT NN . RB DT NNS ."

The best parameters is shown in Table 1.

The RoBERTa models are built using the Transformers module and roberta-base pretrained model. The models consist of 12 hidden layers, 12 attention heads, a single dense layer classifier, and uses Adam optimizer.We tuned maximum length token equal 128. Like dukic et al.[3] we tested following strategy:

1. Only embedding from the last hidden state (768)
2. Sum of embedding vectors from the 12 hidden states (768)
3. Average over the embedding vectors from the 12 hidden states (768)
4. Sum of embedding vector from the last 4 hidden states (768)
5. Average over the embedding vectors from the last 4 hidden states (768)

For pre-trained word2vec, we used Google News Negative 300 Dimention[2]. For each author, we sum the embedding vector per word that exists in the pre-trained dictionary. So, the author tweets embedding vector dimension is 300.

[2]https://github.com/mmihaltz/word2vec-GoogleNews-vectors

Table 1
Hyperparameters found via Grid Search

Analyzer	n-gram range	max_df	min_df	sublinear_tf	SVM C parameter	kernel
word	(1, 6)	0.9	0.02	True	0.85	linear
char	(1, 9)	0.9	0.02	True	0.85	linear
POS	(1, 3)	1.0	1.0	True	1000	linear

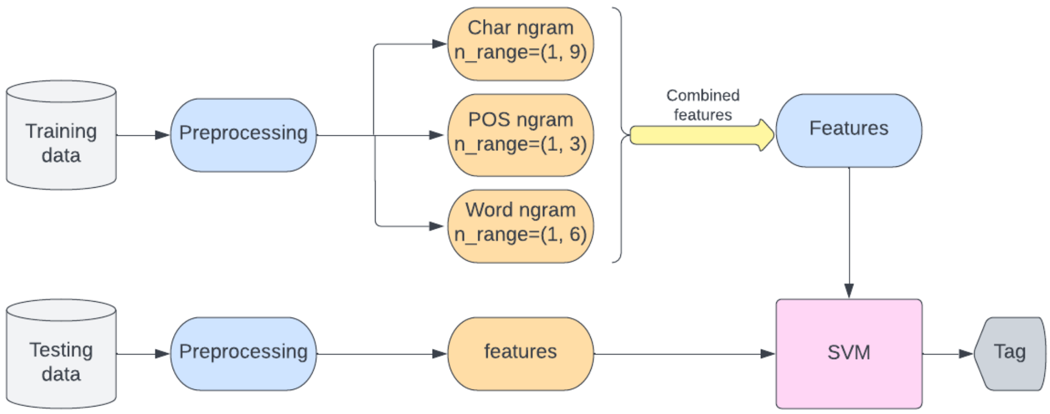

Figure 1: The final proposed approach

3. Result

The evaluation metric used by the task organizer in this task is accuracy. To analyze the effectiveness of the proposed approach, we applied 5-fold cross-validation only on the training set. based on Table 2 the best accuracy achieved by combination of char, word, and POS ngram.

References

[1] M. Kestemont, E. Stamatatos, E. Manjavacas, J. Bevendorff, M. Potthast, B. Stein, Overview of the Authorship Verification Task at PAN 2021, in: A. J. M. M. F. P. Guglielmo Faggioli, Nicola Ferro (Ed.), CLEF 2021 Labs and Workshops, Notebook Papers, CEUR-WS.org, 2021.

[2] M. Siino, E. Di Nuovo, I. Tinnirello, M. La Cascia, Detection of hate speech spreaders using convolutional neural networks—Notebook for PAN at CLEF 2021, in: G. Faggioli, N. Ferro, A. Joly, M. Maistro, F. Piroi (Eds.), CLEF 2021 Labs and Workshops, Notebook Papers, CEUR-WS.org, 2021. URL: http://ceur-ws.org/Vol-2936/paper-189.pdf.

[3] D. Dukić, A. S. Kržić, Detection of Hate Speech Spreaders with BERT—Notebook for PAN at CLEF 2021, in: G. Faggioli, N. Ferro, A. Joly, M. Maistro, F. Piroi (Eds.), CLEF 2021 Labs and Workshops, Notebook Papers, CEUR-WS.org, 2021. URL: http://ceur-ws.org/Vol-2936/paper-164.pdf.

model	feature	C	kernel	accuracy
	Word ngram	.85	linear	94.0%
	Char ngram	.85	linear	93.7%
	POS ngram	1000	linear	80.34%
	POS+Word ngram	0.1	poly	94.33%
	POS+char ngram	1.1	poly	93.14%
	Char+word ngram	0.9	linear	94.33%
SVM	Char+word+POS ngram	0.9	linear	**94.63%**
	Word2vec	13	rbf	88.95%
	Last hidden state embedding	10	linear	73.81%
	Sum of 12 hidden states embedding vec.	0.09	linear	76.19%
	Avg. the 12 hidden states embedding vec.	1e-06	linear	46.42%
	Sum of the last 4 hidden states embedding vec.	1e-06	linear	46.42%
	Avg. the las 4 hidden states embedding vec.	1e-06	linear	46.42%

Table 2
Best accuracy obtained performing a 5-fold cross validation on SVM

[4] S. Bird, Nltk: The natural language toolkit, in: Proceedings of the COLING/ACL on Interactive Presentation Sessions, COLING-ACL '06, Association for Computational Linguistics, USA, 2006, p. 69–72. URL: https://doi.org/10.3115/1225403.1225421. doi:10.3115/1225403.1225421.

[5] F. Chollet, et al., Keras, 2015. URL: https://github.com/fchollet/keras.

[6] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in Python, Journal of Machine Learning Research 12 (2011) 2825–2830.

[7] 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. L. Scao, S. Gugger, M. Drame, Q. Lhoest, A. M. 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://www.aclweb.org/anthology/2020.emnlp-demos.6.

[8] C. Cortes, V. Vapnik, Support-vector networks, Machine learning 20 (1995) 273–297.

[9] M. Potthast, T. Gollub, M. Wiegmann, B. Stein, Tira integrated research architecture, in: Information Retrieval Evaluation in a Changing World, 2019.

[10] M. Rahgouy, H. Giglou, T. Rahgooy, M. Sheykhlan, E. Mohammadzadeh, Cross-domain Authorship Attribution: Author Identification using a Multi-Aspect Ensemble Approach, in: L. Cappellato, N. Ferro, D. Losada, H. Müller (Eds.), CLEF 2019 Labs and Workshops, Notebook Papers, CEUR-WS.org, 2019. URL: http://ceur-ws.org/Vol-2380/.

[11] H. B. Giglou, M. Rahgouy, T. Rahgooy, M. K. Sheykhlan, E. Mohammadzadeh, Author profiling: Bot and gender prediction using a multi-aspect ensemble approach, in: L. Cappellato, N. Ferro, D. E. Losada, H. Müller (Eds.), Working Notes of CLEF 2019 - Conference and Labs of the Evaluation Forum, Lugano, Switzerland, September 9-12, 2019, volume 2380 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2019. URL:

http://ceur-ws.org/Vol-2380/paper_231.pdf.