Reshape or Update? Metric Learning and Fine-tuning for Low-Resource Influencer Profiling

Profiling Cryptocurrency Influencers with Few-shot Learning

Roberto Labadie Tamayo
rlabtam@posgrado.upv.es

Areg Mikael Sarvazyan
arsar1@inf.upv.es

Contents

Introduction

- Profilin crypto influencers
  - Understand impact of users in crypto community
  - Predict or understand market trends
  - **Study the behavior of crypto communities**
- Influential users significantly influence sentiment and adoption of cryptocurrencies through social media activities [1]

[1] M. Ortu, S. Vacca, G. Destefanis, C. Conversano, Cryptocurrency ecosystems and social media environments: An empirical analysis through hawkes' models and natural language processing, Machine Learning with Applications 7 (2022). doi:https://doi.org/10.1016/j.mlwa.2021.100229

Related work

- General profiling tasks use ML or DL approaches:
  - Bag-of-word or TfIdf features with SVM, [2] Logistic regression [3], etc.
  - LSTM [4], CNN [5] or Transformers [6]
- Low resource learning:
  - In-context learning of LLMs [7]
  - Bi-encoders / dual-encoder & tri-encoders / triplet networks [8]
  - Entailment based models [9]

[2] Basile, Angelo, et al. "N-GrAM: New Groningen Author-profiling Model." Conference and Labs of the Evaluation Forum (CLEF 2017): Information Access Evaluation meets Multilinguality, Multimodality, and Visualization. 2017.
[3] Pizarro, Juan. "Using N-grams to detect Fake News Spreaders on Twitter." CLEF (working notes). 2020
[4] Labadie, R., Castro-Castro, D., & Bueno, R. O. (2020, September). Fusing Stylistic Features with Deep-learning Methods for Profiling Fake News Spreader. In CLEF (Working Notes).
[5] Siino, M., Di Nuovo, E., Tinnirello, I., & La Cascia, M. (2021). Detection of hate speech spreaders using convolutional neural networks. In CLEF (Working Notes) (pp. 2126-2136).
[6] Labadie Tamayo, R., Castro Castro, D., & Ortega-Bueno, R. (2021, September). Deep Modeling of Latent Representations for Twitter Profiles on Hate Speech Spreaders Identification Task. In Proceedings of the Working Notes of CLEF 2021, Conference and Labs of the Evaluation Forum, Bucharest, Romania, September 21st to 24th, 2021 (pp. 2035-2046). CEUR.
[7] Dong, Q., Li, L., Dai, D., Zheng, C., Wu, Z., Chang, B., ... & Sui, Z. (2022). A survey for in-context learning. arXiv preprint arXiv:2301.00234.
[8] Hoffer, E., & Ailon, N. (2015). Deep metric learning using triplet network. In Similarity-Based Pattern Recognition: Third International Workshop, SIMBAD 2015, Copenhagen, Denmark, October 12-14, 2015. Proceedings 3 (pp. 84-92). Springer International Publishing.
[9] Chinea-Rios, M., Müller, T., De la Peña Sarracén, G.L., Rangel, F., Franco-Salvador, M. (2022). Zero and Few-Shot Learning for Author Profiling. In: Rosso, P., Basile, V., Martínez, R., Métais, E., Meziane, F. (eds) Natural Language Processing and Information Systems. NLDB 2022. Lecture Notes in Computer Science, vol 13286. Springer, Cham.

Data

- Only employ data provided in competition: no augmentations

- Experimented with user vs instance level classification

	Subtask	Classes	UPC	TPU
(i)	*Profiling*	5	32	10
(ii)	*Interest Id.*	5	64	1
(iii)	*Intent Id.*	4	64	1

Method 1: Reshape through Metric Learning

- Use InstructOR [10] embeddings for the user profile, concatenating tweets into [AUTHOR-PROFILE] with this prompt:

Represent the cryptocurrency Tweet posts for profiling their authors into the classes nano micro macro mega non-influencer; Input: [AUTHOR-PROFILE]

- Definitions of each class, [USER-DEFINITION] were taken into account as prototypical points:

Represent the cryptocurrency influencers definition for comparing it with the representation of their tweets; Input: [USER-DEFINITION]

[10] Su, H., Kasai, J., Wang, Y., Hu, Y., Ostendorf, M., Yih, W. T., ... & Yu, T. (2022). One embedder, any task: Instruction-finetuned text embeddings. arXiv preprint arXiv:2212.09741.

Method 1: Reshape through Metric Learning

- Optimize a transformation of the embedding space
    - Samples from the same class are closer together (wrt squared euclidean distance)
    - From different classes are farther apart
- InstructOR profile encoder frozen
- Try dual encoder and triplet encoder with online computation

- For pairs or triplets:
    - Positive samples selected randomly within the same class
    - Negative samples selected as closest elements from a different class
    - Hypothesis: robustness to other negative samples by having more distance between class clouds is more important than keeping the positive class “compact”

Method 2: Parameter Efficient Fine-Tuning

- Fine-tune CryptoBERT [11] and BLOOM [12] with LoRA [13]
  - CryptoBERT: BERTtweet [14] trained on crypto tweets
  - Also tried without by doing conventional fine-tuning of classification head
- More standard and expected to perform better
- Expected: robust to distributional shifts (very low-resourcet training)

[11] https://huggingface.co/ElKulako/cryptobert
[12] Scao, T. L., Fan, A., Akiki, C., Pavlick, E., Ilić, S., Hesslow, D., ... & Manica, M. (2022). Bloom: A 176b-parameter open-access multilingual language model. arXiv preprint arXiv:2211.05100.
[13] Hu, E. J., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., & Chen, W. (2021, October). LoRA: Low-Rank Adaptation of Large Language Models. In International Conference on Learning Representations.
[14] Nguyen, D. Q., Vu, T., & Nguyen, A. T. (2020, October). BERTweet: A pre-trained language model for English Tweets. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations (pp. 9-14).

Experiments

- Metric learning only for subtask 1
  - Lack of sufficient hardware resources for InstructOR
  - Use contrastive and triplet loss
  - Use only definitions vs all training instances as prototypical samples of a class
  - Online update every 5 epochs
- PEFT for all subtasks
  -
- 4-fold CV on all subtasks, mean Macro-f1
  - Robustness

Results: Metric Learning

- Ins: all data as prototypes
- Def: definitions as prototypes

Strategy	Macro-F1
Contrastive + Ins.	0.332
Contrastive + Def. + Ins.	0.356
Triplet + Ins.	0.449
Triplet + Def. + Ins.	**0.461**

- Only definitions did not perform well

- Best system was submitted
  - Only subtask 1

Results: PEFT

- CryptoBERT performs better
- All results better than metric learning approach for subtask 1
- Couldn’t do conventional fine-tuning for BLOOM-7B
- Submitted systems in bold

Strategy	Subtask 1	Subtask 2	Subtask 3
CryptoBERT	0.554	0.405	0.647
CryptoBERT + LoRA	**0.68**	0.514	**0.608**
BLOOM-1b1	0.610	0.504	0.551
BLOOM-1b1 + LoRA	0.623	**0.528**	0.588
BLOOM-7b1 + LoRA	0.651	0.520	0.608

Official results

- Subtask 1: As expected, PEFT approaches performed worse than metric learning.
- CryptoBERT - LoRA: 46.5 in test but 68 in 4-fold CV
- Triplet+Def+Ins: 50.8 in test, 46.1 in 4-fold CV
    - Much more robust to potential distributional shifts given low resource
- Other subtasks w/best PEFT systems:
    - Subtask 2: 51.7
    - Subtask 3: 52.6

Conclusions

- Metric learning can be more robust than PEFT approaches
- Much more could be tried
  - Maybe PEFT models were not adequate and others could perform much better?
  - Little training data processing was done
    - We believed usernames, links, hashtags, etc. would better profile users
    - Could be incorrect assumption
  - We did no data augmentations, and believe this would massively boost our scores
- Cross validation can be finicky: best system for subtask 1 was not good in 4-fold CV
- Low resource tasks are difficult and require more care

Reshape or Update? Metric Learning and Fine-tuning for Low-Resource Influencer Profiling

Profiling Cryptocurrency Influencers with Few-shot Learning

Roberto Labadie Tamayo
rlabtam@posgrado.upv.es

Areg Mikael Sarvazyan
arsar1@inf.upv.es