mattica@SMM4H'22: Leveraging sentiment for stance & premise joint learning

Oscar Lithgow-Serrano[1]†, **Joseph Cornelius**[1], **Fabio Rinaldi**[1], **Ljiljana Dolamic**[2]
[1]IDSIA, Lugano, Switzerland
[2]armasuisse S+T, Thun, Switzerland
†Corresponding author: oscar@idsia.ch

Abstract

This paper describes our submissions to the Social Media Mining for Health Applications (SMM4H) shared task 2022. Our team (mattica) participated in detecting stances and premises in tweets about health mandates related to COVID-19 (Task 2). Our approach was based on using an in-domain Pretrained Language Model, which we fine-tuned by combining different strategies such as leveraging an additional stance detection dataset through two-stage fine-tuning, joint-learning Stance and Premise detection objectives; and ensembling the sentiment-polarity given by an off-the-shelf fine-tuned model.

1 Introduction

The Social Media Mining for Health Applications (SMM4H) shared task 2022 (Weissenbacher et al., 2022) is aimed to apply Natural Language Processing to address different challenges of using social media for Health research. Specifically, we participated in Task-2, which consisted of detecting the stance of a tweet towards a given topic (subtask 2a), and detecting if a tweet contained or not a premise (subtask 2b). The organizers provided a manually labeled dataset (Davydova and Tutubalina, 2022) split into training (3,669 tweets) and validation (600 tweets). Each record in the dataset was labeled on both axis; the *Stance* classification included the labels FAVOR, AGAINST, and NONE; and the *Premise* axis with values 1 and 0 indicating the presence or absence of a premise respectively. For the evaluation phase, an unlabeled test set of 2,000 tweets was provided.

Our approach consisted of leveraging an in-domain Pretrained Language Model **(PLM)** and adapting it to the tasks through two consecutive fine-tuning steps. In the second, we used the predictions of an off-the-shelf model for sentiment polarity as additional features and applied joint learning of the stance and premise detection.

Related work (Sun et al., 2019; Li and Caragea, 2019; Fang et al., 2019) use sentiment analysis as an auxiliary task in an MTL setting. Whereas (Mohammad et al., 2017, 2016), like us, use the sentiment as an additional input feature for the stance classification. To our knowledge, no other approaches jointly-learned stance and premise detection leveraging sentiment polarity.[1]

2 Extra data and Pre-processing

We used COVIDLies (Hossain et al., 2020) as an extra training dataset. The dataset consists of 6,761[2] COVID-19-related tweets, each paired with a misconception and annotated with the concerning tweet's stance. In COVIDLies, the misconceptions are misinformation statements that were manually rephrased as short and positive expressions (e.g., "Coronavirus is caused by 5G"). Thus, different from a topic stance, in this dataset, the annotated standpoint is with respect to a specific statement. To combine both datasets, we migrated the provided data from topic-stance to the COVIDLies statement-stance format by manually formulating a positive statement from each related topic (see table 1). We also pre-processed the tweets by delexicalizing user mentions and URLs and replacing them with *@user* and *URL* correspondingly; striping out the hash character from hashtags, removing extra spaces; and replacing emojis with their corresponding short-code aliases (i.e., demojize[3]).

Topic	Statement
face masks	Face masks help to protect us.
stay at home orders	Stay at home is a needed measure.
school closures	Schools need to remain closed.

Table 1: Map from *topic* to *statement*.

[1]Code available at https://github.com/OWLmx/ws_ssm4h22
[2]Only 3,256 could be recovered.
[3]We used the package https://pypi.org/project/emoji

Proceedings of the 29th International Conference on Computational Linguistic, pages 75–77
October 12–17, 2022.

3 System description

The base of our approach is a transformer-based language model pretrained on a corpus of Twitter messages on the topic of COVID-19 (CT-BERT V2) (Müller et al., 2020).

We leveraged the COVIDLies dataset by an initial fine-tuning **(ft)** for stance-detection. Next, we performed a second fine-tuning on the data provided for this task. The second fine-tuning included a multitask-learning **(MTL)** setup where the objectives were Stance and Premise classification. Both losses were computed by cross entropy and combined using homoscedastic uncertainty to weight each task loss (Kendall et al., 2018). Also, in the second fine-tuning, the resulting logits of sentiment classification were appended directly to the pooled output of the encoder just before passing it to the classification heads[4]. The sentiment classification was obtained with an already fine-tuned roBERTa-base model for sentiment-analysis in tweets (Loureiro et al., 2022).

We trained the models with AdamW (Loshchilov and Hutter, 2019) optimizer without a warm-up period. A weighted random sampling with replacement was used in both fine-tuning steps (except in the multi-task setups), with learning rates of $\alpha = 10^{-4}$ and $\alpha = 10^{-5}$ respectively. For single task configurations, we used a training batch size of $bs = 8$ and a $bs = 4$ for MTL configurations; in both cases, the maximum sequence length was set to 160. An early stop with patience of 3 using losses scores of the held-out validation split ($bs = 16$) was applied during training. [5]

4 Results and Discussion

We used the COVIDLies dataset for the first fine-tuning. For the second **ft** we used the provided data as follows: we created a validation split with 5% of the training and 12% of the validation data, and used the rest of the training data as a training split and the rest of the validation data as a test split.

All our submissions were based on the two-stage fine-tuning **(2st-ft)** and what differed was the strategies combination in the 2nd ft. For subtask-2a, we submitted a run with a base 2st-ft, another including the logits from sentiment classification **(sent)**, another with the 2st-ft and MTL, and finally, one that combined all the strategies (2st-ft + sent + MTL). For subtask-2b, we used the two setups that included MTL. This is, our two submissions corresponded to the premise inference from the (2st-ft + MTL) and the (2st-ft + sent + MTL) runs.

Setting	Accuracy	F1-macro
2st-ft	0.840	0.836
2st-ft + sent	0.848	0.841
2st-ft + MTL	0.853	0.849
2st-ft + sent + MTL	0.857	0.853

Table 2: Stance detection results in our held-out test set.

Evaluating our test-set, we observed that the cumulative combination of the different strategies resulted in small but consistent gains for the stance detection task performance (see Table 2). We analyzed each strategy's impact on identifying the different stances (see Fig. 1). We observed that integrating the sentiment polarity gives an important boost to detecting the FAVOR stance, whereas jointly learning to predict the presence of Premises is more beneficial to recognizing the AGAINST stance. Combining all the strategies resulted in the best balance for both -stances.

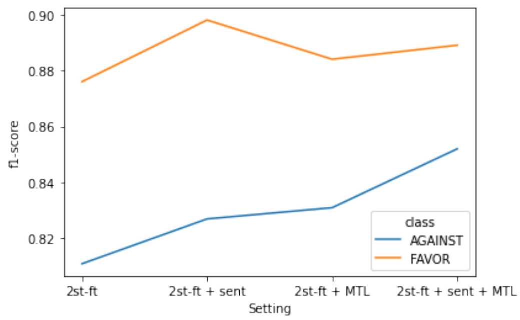

Figure 1: Stances' F1 score with the different strategies.

5 Conclusions

Our approach involved leveraging a related dataset by a preliminary fine-tuning, and combining sentiment analysis along with multi-task learning of premise and stance.

In the official test set, our best result for stance detection (subtask 2a) was 0.633, which is 14 percentage points (p.p.) above the mean and 8 p.p. above the median of all participants' submissions. For the premise detection (subtask 2b), our best score was 0.647, which is precisely the median but 7 p.p. above the mean of all submissions.

The results show that jointly learning to detect premise and stance is beneficial for both tasks. Combined with the tweet's sentiment polarity, the two-stage fine-tuned model gave the best results.

[4] Linear -> ReLU -> Dropout (0.1) -> Softmax

[5] The models were implemented using Pytorch (Paszke et al., 2019), Pytorch Lightning (Falcon and The PyTorch Lightning team, 2019) and Huggingface's Transformers (Wolf et al., 2020) library.

Acknowledgements

This research was partly supported by the project MisInfoFastCatch, funded by the armasuisse Science and Technology Swiss Federal Office (contract nr. 8203004722).

References

Vera Davydova and Elena Tutubalina. 2022. Smm4h 2022 task 2: Dataset for stance and premise detection in tweets about health mandates related to covid-19. In *Proceedings of the Seventh Social Media Mining for Health Applications (#SMM4H) Workshop Shared Task*, pages –.

William Falcon and The PyTorch Lightning team. 2019. PyTorch Lightning.

Wei Fang, Moin Nadeem, Mitra Mohtarami, and James Glass. 2019. Neural Multi-Task Learning for Stance Prediction. pages 13–19.

Tamanna Hossain, Robert L. Logan IV, Arjuna Ugarte, Yoshitomo Matsubara, Sean Young, and Sameer Singh. 2020. COVIDLies: Detecting COVID-19 Misinformation on Social Media.

Alex Kendall, Yarin Gal, and Roberto Cipolla. 2018. Multi-task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 7482–7491.

Yingjie Li and Cornelia Caragea. 2019. Multi-Task Stance Detection with Sentiment and Stance Lexicons. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 6299–6305, Hong Kong, China. Association for Computational Linguistics.

Ilya Loshchilov and Frank Hutter. 2019. Decoupled weight decay regularization. *7th International Conference on Learning Representations, ICLR 2019*.

Daniel Loureiro, Francesco Barbieri, Leonardo Neves, Luis Espinosa Anke, and Jose Camacho-Collados. 2022. TimeLMs: Diachronic Language Models from Twitter.

Saif M. Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. 2016. SemEval-2016 task 6: Detecting stance in tweets. *SemEval 2016 - 10th International Workshop on Semantic Evaluation, Proceedings*, pages 31–41.

Saif M. Mohammad, Parinaz Sobhani, and Svetlana Kiritchenko. 2017. Stance and Sentiment in Tweets. *ACM Transactions on Internet Technology*, 17(3):1–23.

Martin Müller, Marcel Salathé, and Per E Kummervold. 2020. COVID-Twitter-BERT: A Natural Language Processing Model to Analyse COVID-19 Content on Twitter.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc.

Qingying Sun, Zhongqing Wang, Shoushan Li, Qiaoming Zhu, and Guodong Zhou. 2019. Stance detection via sentiment information and neural network model. *Frontiers of Computer Science*, 13(1):127–138.

Davy Weissenbacher, Ari Z. Klein, Luis Gascó, Darryl Estrada-Zavala, Martin Krallinger, Yuting Guo, Yao Ge, Abeed Sarker, Ana Lucia Schmidt, Raul Rodriguez-Esteban, Mathias Leddin, Arjun Magge, Juan M. Banda, Vera Davydova, Elena Tutubalina, and Graciela Gonzalez-Hernandez. 2022. Overview of the seventh social media mining for health applications smm4h shared tasks at coling 2022. In *Proceedings of the Seventh Social Media Mining for Health Applications (#SMM4H) Workshop Shared Task*, pages –.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Perric Cistac, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander M Rush. 2020. Transformers: State-of-the-Art Natural Language Processing. pages 38–45. Association for Computational Linguistics.