Detecting Optimism in Tweets using Knowledge Distillation and Linguistic Analysis of Optimism

Ștefan Cobeli*, Bogdan Iordache†, Shweta Yadav*, Cornelia Caragea*, Liviu Dinu†, Dragoș Iliescu†

*University of Illinois at Chicago, Illinois, US & †University of Bucharest, Bucharest, Romania

13th Edition of the Language Resources and Evaluation Conference

LREC, June 2022





Overview

1. Introduction
2. Method
3. Datasets
4. Results
5. Analysis
6. Conclusions

Introduction

- We analyze the *optimism* and *pessimism* expressed in Twitter posts;
- We present a multi-task knowledge distillation framework to simultaneously learn the target task of optimism detection with the help of the auxiliary task of sentiment analysis and hate speech detection;
- We analyzed the robustness of predictions under human limited corrections;
- We also explored the linguistic peculiarities of optimism and pessimism.

Related Work

- There were multiple studies on automatic optimism and pessimism detection [RWM16; CDD18; Als+20; Als+21];
- Knowledge Distillation used as a model compression technique as well as for multi-task learning [HVD15; Sun+19; Cla+19];

Motivation

Optimism-pessimism is connected to **happiness**, [ABK13], **stress management**, **mental and physical well-being** [RSG09].

Knowledge Distillation (KD)

Knowledge Distillation Setting

Given a dataset $\mathcal{D}$, a student model $\mathcal{S}$ learns only to mirror the logits generated by a teacher $\mathcal{T}$, minimizing the cross entropy loss between the outputs of the student model and the outputs of the teacher model:

$$\mathcal{L}_{\text{KD}} = \sum_{(x,y)\in\mathcal{D}} \ell(f(x;\theta_{\mathcal{S}}), f(x;\theta_{\mathcal{T}})).$$

Multi-Task Knowledge Distillation (MTKD)

Loss with respect to a Single Teacher:

We augmented the patient knowledge loss to use *Teacher Annealing*:

$$\mathcal{L}_{\mathcal{T}} = (1 - \alpha) * \mathcal{L}_{01} + \alpha * (\mathcal{L}_{\text{KD}} + \beta * \mathcal{L}_{\text{PKD}}).$$

α linearly decreases from 1 towards 0.

Multi-Task Knowledge Distillation (MTKD)

Loss with respect to Multiple Teachers

Given the tasks $\{\mathcal{T}_1, \ldots, \mathcal{T}_t\}$, we define the multi-task loss as:

$$\mathcal{L}_{\text{MTKD}} = \sum_{i \in \overline{1,t}} \mathcal{L}_{\mathcal{T}_i}.$$

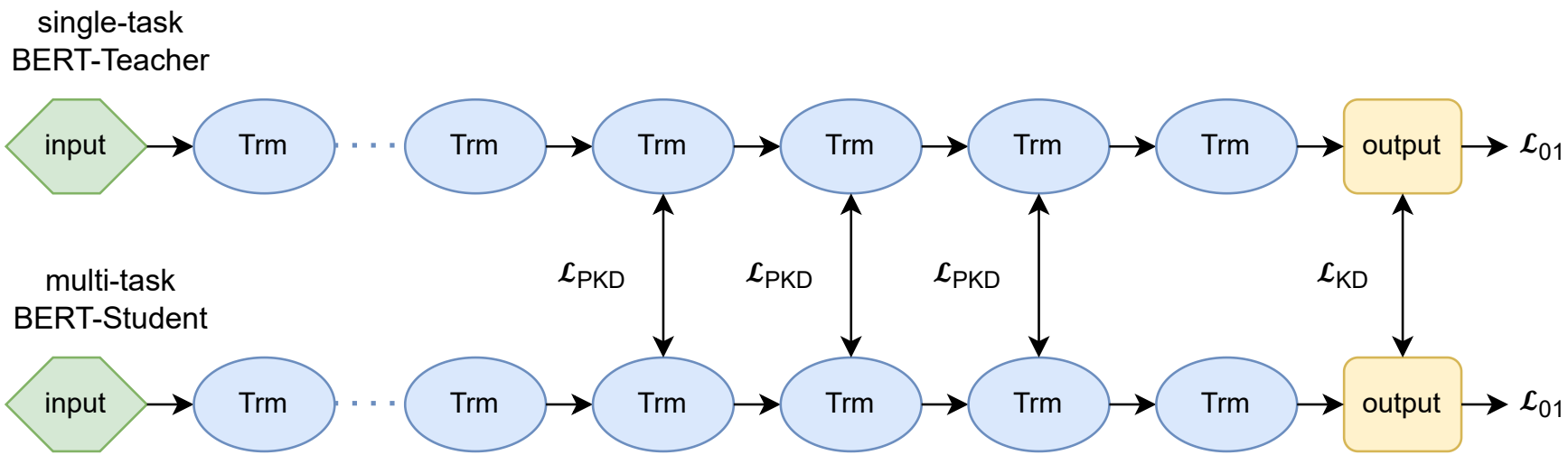


Figure: Multi-Task Knowledge Distillation (MTKD) architecture.

Optimism dataset

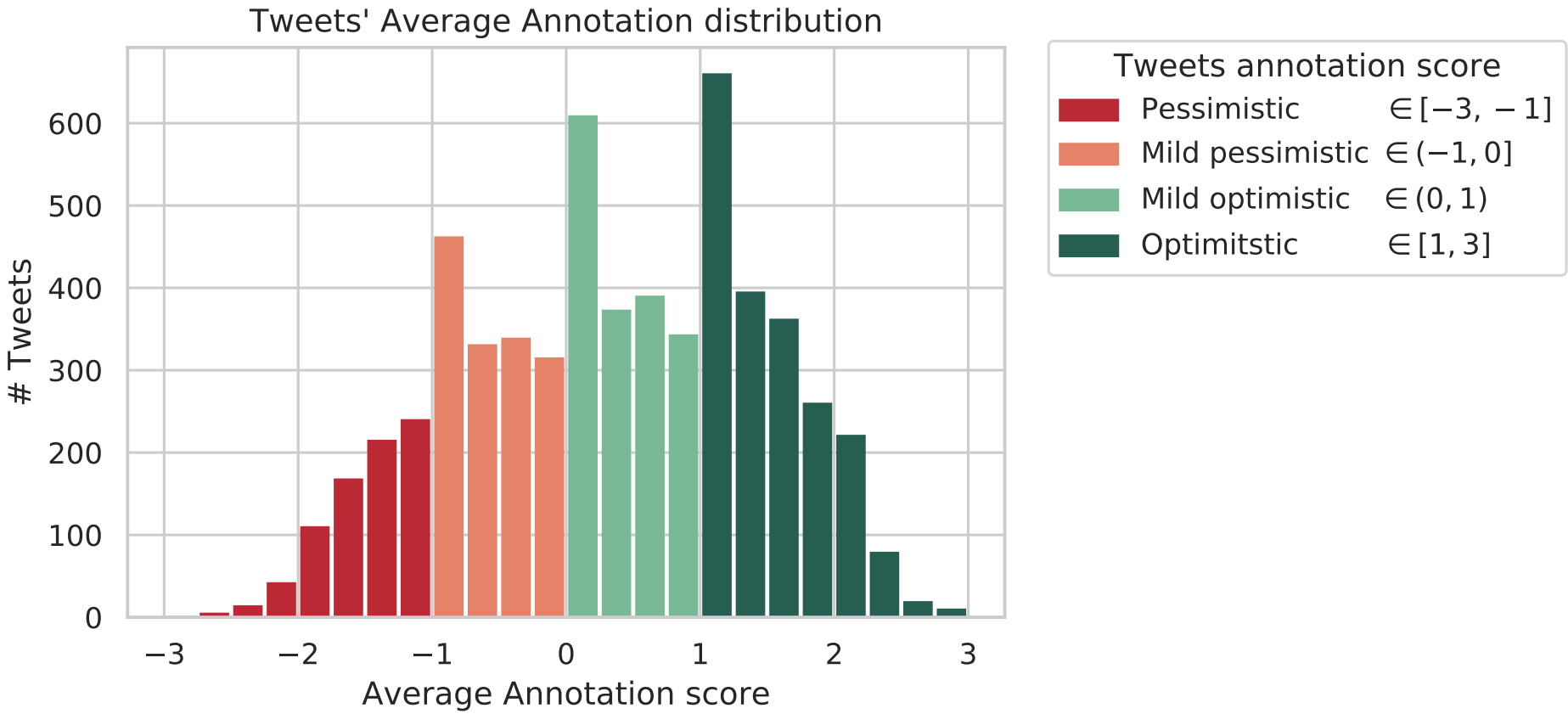


Figure: Each tweet was rated by 5 different annotators with a score $\in \{-3, \ldots, 3\}$, according to the tweet's optimism level. The average annotation score was obtained by averaging the 5 individual scores [RWM16].

Optimism dataset

# Tweets, setting 0	
Pessimist	**Optimist**
2,796	4,679
Total	
7,475	

# Tweets, setting $1/-1$	
Pessimist	**Optimist**
1,340	2,507
Total	
3,847	

Table: Number of tweets in OPT dataset [RWM16].

Auxiliary Tasks

Hate Speech Datset (Hate)

Hate was labeled using an iterative procedure, in multiple rounds. There are $80,000$ tweets each one labeled as either *normal* (59%), *spam* (22.5%), *abusive* (11%) or *hateful* (7.5%).

Sentiment Polarity Dataset (Sent)

We used a twitter sentiment dataset (Sent) proposed at the SemEval competition in 2017 The Sent dataset is composed of $50,333$ tweets annotated with one of the three labels: *negative* (15.57%), *neutral* (44.89%) or *positive* (39.54%).

Distillation Technique Selection

Model	Test Acc.	Dev. Acc.
BERT Base	83.90	82.62
BERTweet	84.84	84.58
KD	85.21	83.94
PKD	85.21	84.79
PKD and T.A.	85.45	**84.85**
GRUStack baseline [CDD18]	80.19	–

Table: Accuracy of BERT models combined with distillation techniques on the OPT dataset compared with the best non-BERT baseline.

Results on the 0 setting

Model	Test Acc.	Dev Acc.
BERTweet	84.84	84.58
MTKD OPT + Hate	86.52	85.30
MTKD OPT + Sent	86.23	85.44
MTKD OPT + Hate + Sent	**86.60**	85.14
MTKD no KD	82.11	81.82
MTKD vanilla-BERT	85.64	84.71
MTKD downsampled	86.19	85.23
XLNet Base [Als+20]	84.25	–
BERT Base with SLA [Als+21]	85.69	–

Table: Best models' performances on optimism prediction. We can see that all the components of the best model are relevant. The combination between BERTweet, PKD and both intermediate tasks provides the highest accuracy on the test set.

Models Accuracies Comparison on the 0 setting

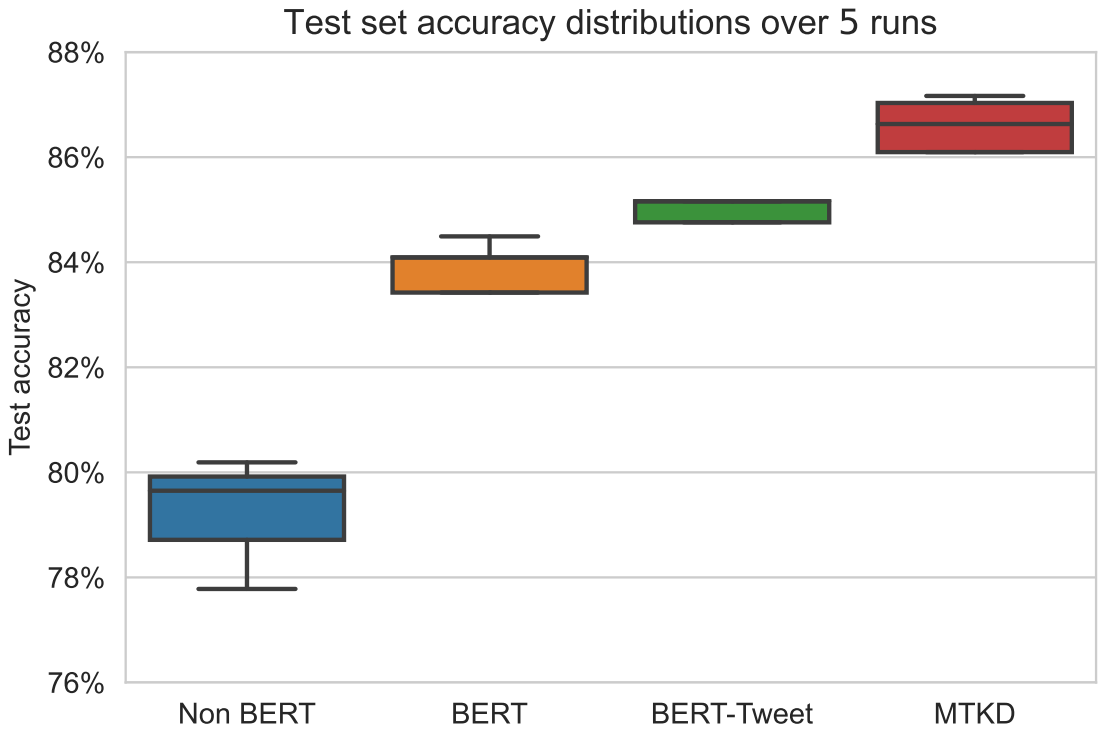


Figure: Accuracy distribution over five independent runs for optimism prediction. We can see that attention-based models BERT and BERTweet clearly outperform previous baselines. By using multi-task knowledge distillation we further improve models' performance.

Results on the 1/ − 1 setting

Model	Test Acc.	Dev Acc.
BERT	95.29	95.97
BERTweet	95.78	96.84
MTKD OPT + Hate + Sent	96.57	**97.24**
BERT Base with SLA [Als+21]	**97.10**	–
XLNet Base [Als+20]	96.16	–

Table: Model accuracies for predicting optimism in the setting 1/-1.

Linguistic Analysis of Optimism

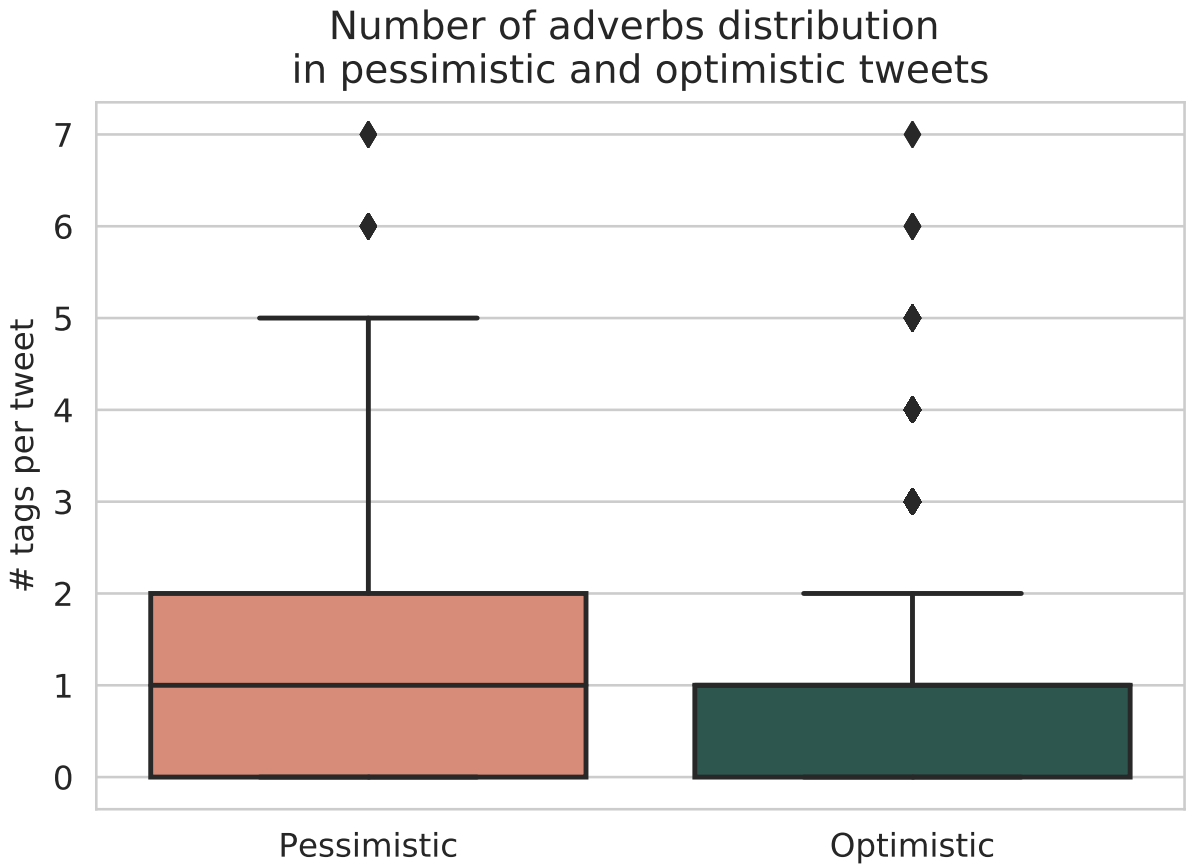


Figure: Distribution of the number of adverbs used in pessimistic and optimistic tweets. We can see that pessimistic tweets tend to have more adverbs than optimistic ones.

Linguistic Analysis of Optimism

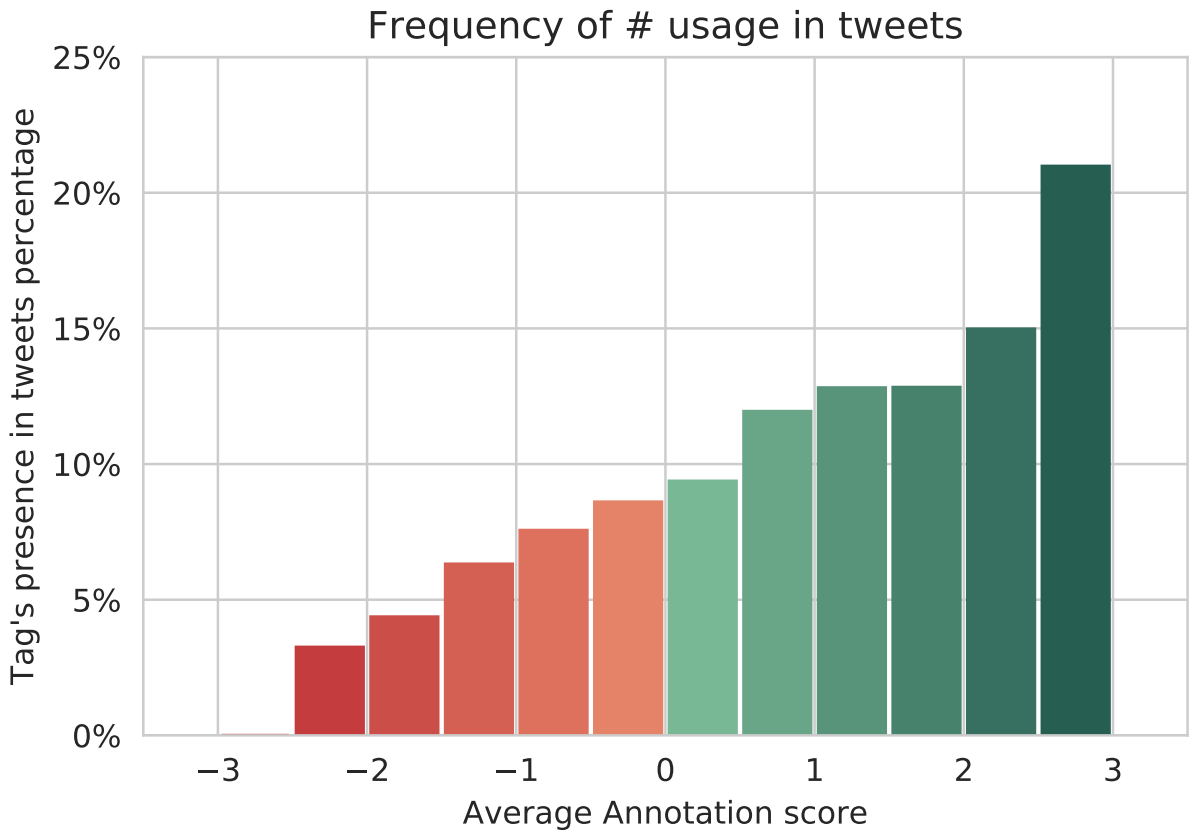


Figure: Frequency of hashtags usage in tweets as the optimism polarity increases from -3 to $+3$.

Linguistic Analysis of Optimism

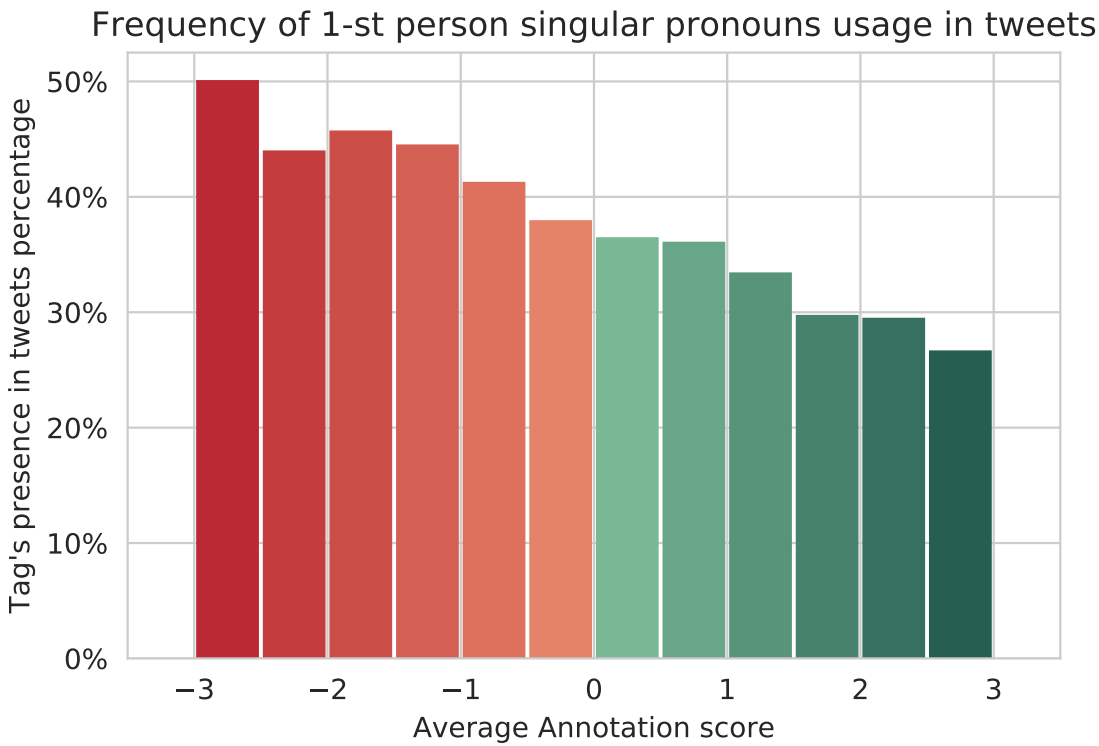


Figure: Frequency of first person singular pronouns usage in tweets as the optimism polarity increases from -3 to $+3$.

Error Analysis

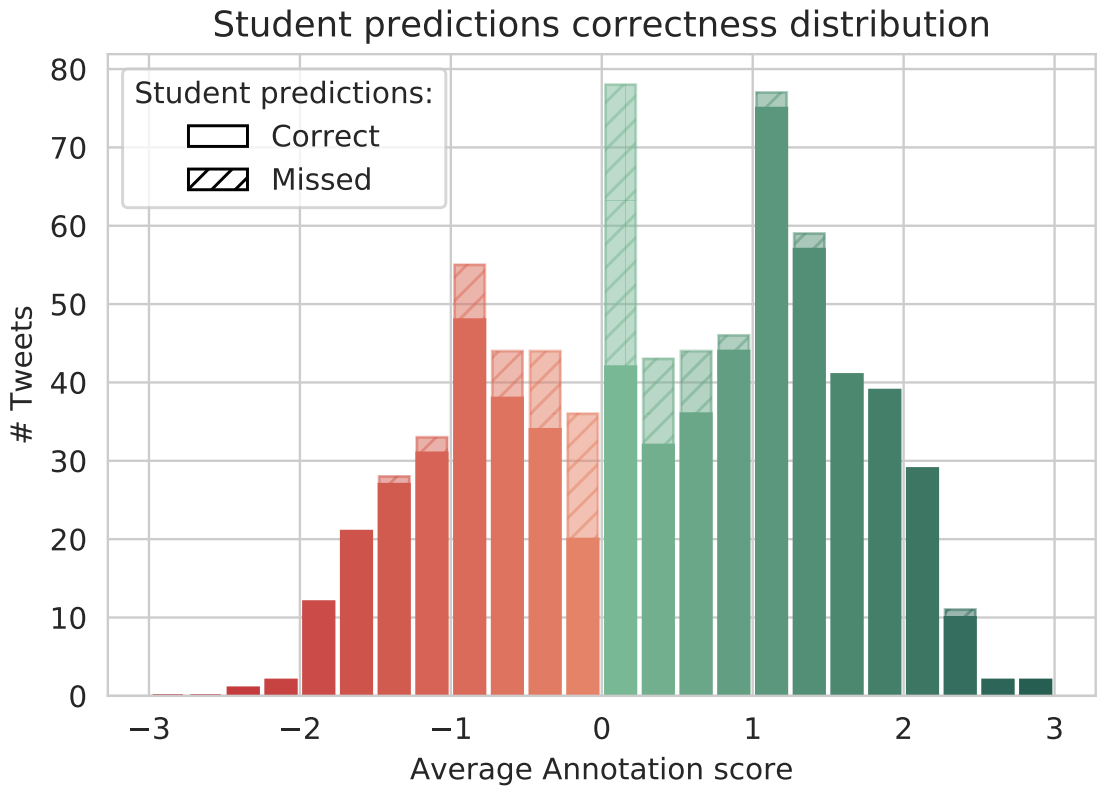


Figure: Distribution of the tweets correctly classified by the best model in comparison with the entire dataset. We can see that most missclassified tweets are in the annotation range $(-1, 1)$, where the differences between pessimism and optimism may be subjective.

Teacher vs. Student Errors

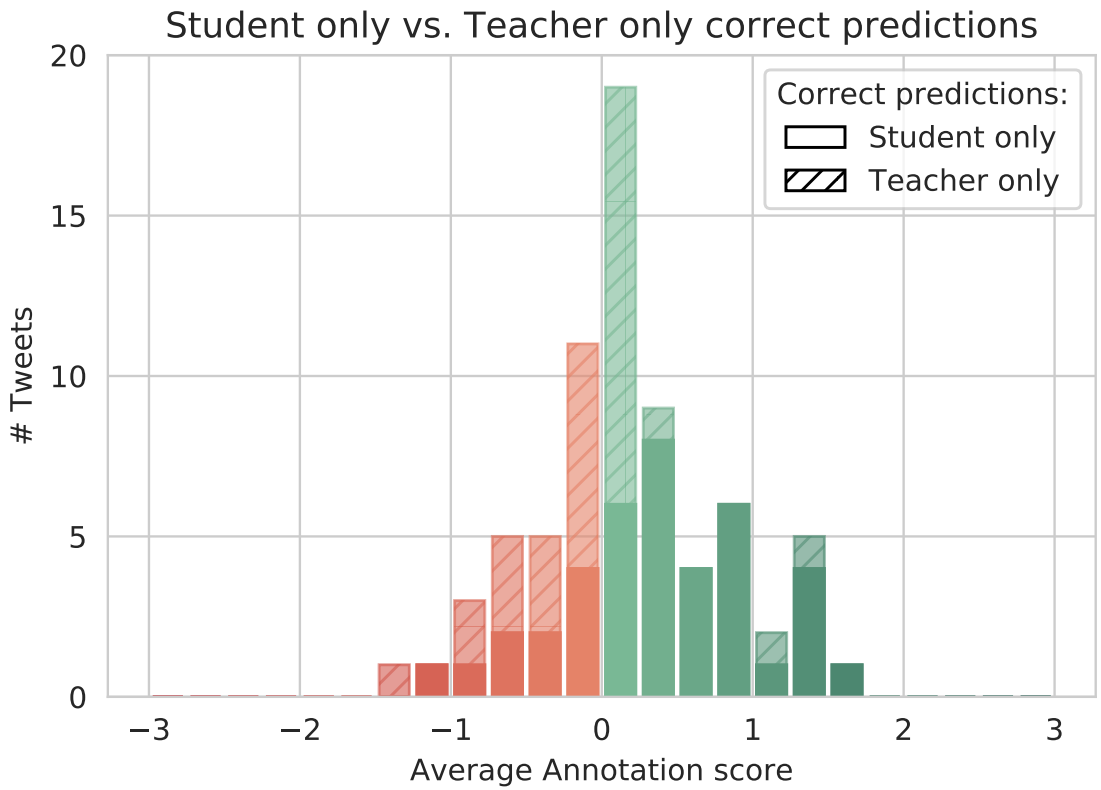


Figure: Distribution of tweets correctly classified by only the best teacher or the best student. We can see that the student outperforms its teacher on mildly optimistic tweets. The teacher has a better performance on the mildly pessimistic tweets.

Limited Corrections

	Tweet	Average Annotation	Prediction Confidence
Original tweet	i don't know how they can be so perfect	−0.2	55.03%
Pessimistic correction	i don't know how they can be so perfect liars!	−1.2	77.01%
Optimistic correction	flawless! i don't know how they can be so perfect.	1.0	85.03%

Table: Tweet modified to be more optimistic and pessimistic. Whilst the original tweet was missclassified by our best model, after limited clarifying corrections the model predicts accurately the pessimism and respectively the optimism of the tweet.

Frame Title

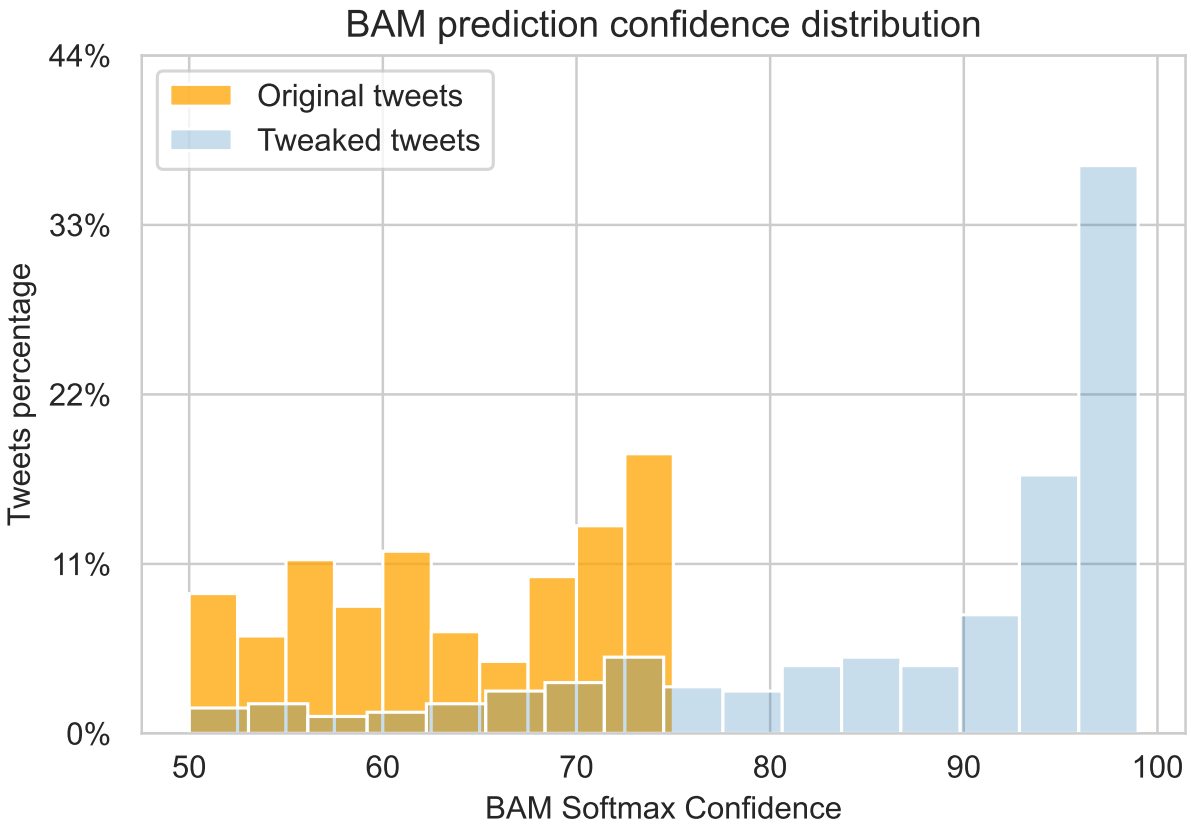


Figure: Softmax confidence distribution over the selected tweets before and after limited corrections. We can notice the prediction confidence increases significantly when tweets become more clear.

Conclusions

- We have presented a multi task learning methodology for optimism detection;
- We have observed a relationship that can be explored between optimism, sentiment and mental health;
- We have explored linguistic characteristics expressed in optimism and pessimism;
- The relationship between optimism-pessimism and mental health constructs can be further explored;
- A larger and better documented dataset for optimism/pessimism classification would also be beneficial.

Thank you! ☺



Credit: Getty Images/iStockphoto

Reference I

[ABK13] Gene M Alarcon, Nathan A Bowling, and Steven Khazon. "Great expectations: A meta-analytic examination of optimism and hope". In: *Personality and Individual Differences* 54.7 (2013), pp. 821–827.

[Als+20] Ali Alshahrani et al. "Identifying Optimism and Pessimism in Twitter Messages Using XLNet and Deep Consensus". In: *2020 International Joint Conference on Neural Networks (IJCNN)*. IEEE. 2020, pp. 1–8.

[Als+21] Ali Alshahrani et al. "Optimism/Pessimism Prediction of Twitter Messages and Users Using BERT with Soft Label Assignment". In: *2021 International Joint Conference on Neural Networks (IJCNN)*. IEEE. 2021, pp. 1–8.

Reference II

[CDD18] Cornelia Caragea, Liviu P Dinu, and Bogdan Dumitru. "Exploring optimism and pessimism in twitter using deep learning". In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. 2018, pp. 652–658.

[Cla+19] Kevin Clark et al. "BAM! Born-Again Multi-Task Networks for Natural Language Understanding". In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Florence, Italy: Association for Computational Linguistics, July 2019, pp. 5931–5937. DOI: 10.18653/v1/P19-1595. URL: https://www.aclweb.org/anthology/P19-1595.

[HVD15] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. "Distilling the knowledge in a neural network". In: *arXiv preprint arXiv:1503.02531* (2015).

Reference III

[RSG09] Heather N Rasmussen, Michael F Scheier, and Joel B Greenhouse. "Optimism and physical health: A meta-analytic review". In: *Annals of behavioral medicine* 37.3 (2009), pp. 239–256.

[RWM16] Xianzhi Ruan, Steven Wilson, and Rada Mihalcea. "Finding optimists and pessimists on twitter". In: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. 2016, pp. 320–325.

[Sun+19] Siqi Sun et al. "Patient Knowledge Distillation for BERT Model Compression". 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)*. Hong Kong, China: Association for Computational Linguistics, Nov. 2019, pp. 4323–4332. DOI: 10.18653/v1/D19-1441. URL: https://www.aclweb.org/anthology/D19-1441.