From "Half Empty" to "Half Full"

Paraphrasing pessimistic tweets to optimistic tweets

Kushal Kantharajappa & Wan-Hsuan Chiang
University of Twente

Group 15
Natural Language Processing
Email: k.kantharajappa@student.utwente
Email: w.chiang@student.utwente.nl

UNIVERSITY OF TWENTE.

Introduction

Positive thinking involves re-framing one's thoughts about a situation and can improve mental health [7]. Previous researches explored the model of classifying optimists and pessimists [8, 4, 2]. We believe this could be taken a step forward - one could actually rephrase a pessimistic statement into it's optimistic variant, while preserving the meaning.

Paraphrasing pessimistic sentences to optimistic sentences is a difficult problem as it mainly involves replacing the word/words with pessimistic connotation with its optimistic equivalents without changing the contextual meaning. For example: "The glass is half empty" to "The glass is half full"

In this project we try to solve this problem by using a general purpose paraphraser. The paraphrasing tool created 10 sentences for optimism/pessimism classifier to select from. The optimistic sentence will be the output of our optimistic paraphraser.

Workflow

1. Input the pessimistic tweets from the data set.
2. The tweets are paraphrased using a t5 language model. Multiple outputs are generated while preserving the semantic meaning of the sentence.
3. A classifier developed using XLNet checks picks out an optimistic tweet among the paraphrased tweets.
4. This optimistic tweet replaces the old pessimistic tweet

Dataset

The dataset we are working on is the OPT, the Optimism/ pessimism Twitter dataset generated by Ruan, Xianzhi et al. (2016) [8]. Each of the tweet is assigned an average annotation score from -3 (very pessimistic) to 3 (very optimistic) from crowd sourcing platform Amazon Mechanical Turk. We use threshold 1/-1 to split optimistic and pessimistic tweets as the previous researches [4, 2, 3] shows better model performances.

Methodology

The idea of creating the optimistic paraphrasing tool is to randomly generate 10 sentences rephrased by the well-trained paraphrasing tool. After that, the classifier classified and picked up the optimistic sentence.

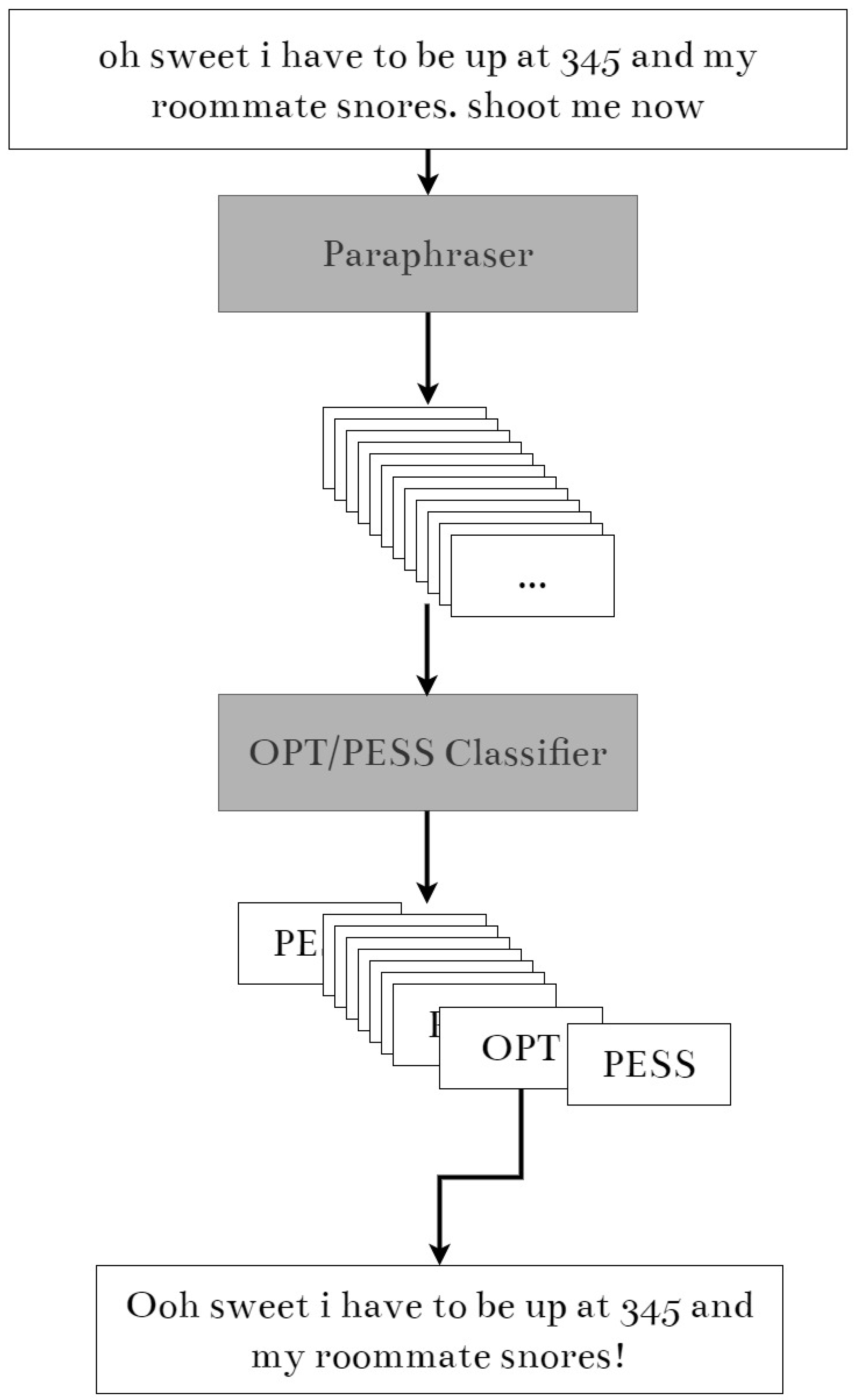

Figure 1: The idea of the paraphraser

Paraphraser

The paraphraser is developed using a t5 model(text-to-text-transfer-transformer) which takes text strings as inputs and returns text strings back. This model is pre-trained with c4 data set and fine-tuned using the google PAWS[1] dataset to generate the paraphrases with same meaning.

XLNet Classifer

The classifier is adapted from Alshahrani, Ali, et al. (2020) [2]. The research used the pretrained XLNet model [11], fined tuned by OPT and achieved 96.45% of the accuracy at tweet level and and 100% at user level through XLNet-Large model. Yet, due to the limited resource and project scale, our project applies a smaller XLNet-Base pretrained model, which contains 12 layers, 768 hidden size, and 12 heads.

The XLNet classifier follows the structures from Alshahrani, Ali, et al. (2020) [2] but simpler by simplifing the hyper parameters. We did five trainings and obtained 95.32% as an average accuracy.

The classifier can then be applied to the optimistic paraphrasing tool. Noted that the optimism/pessimism classifier is difference from the traditional sentiment classifier, as Caragea, Cornelia et al. (2018) already tested the difference.

Results

Input tweet:

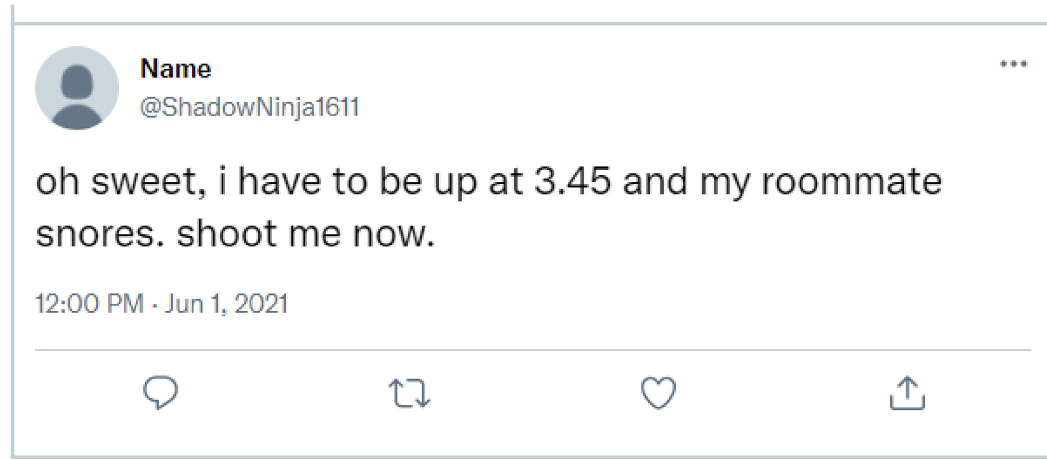

Figure 2: Input tweet

Paraphrased Tweet:

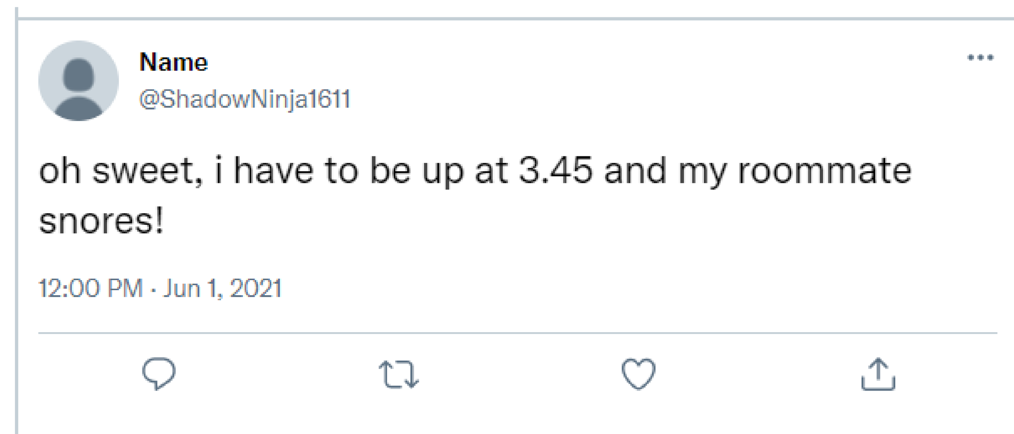

Figure 3: Paraphrased tweet

Conclusions

Paraphrasing pessimistic sentences using a straight forward approach has been brought forward and implemented. The results are satisfactory, however the paraphraser has been trained on a general dataset which is not completely specific to our use case. Im

Forthcoming Research

1. A dataset which includes pessimistic sentences and it's equivalent optimistic sentences as ground truth is required to train the paraphrasing network to improve the accuracy.
2. The annotation score in OPT dataset is measured by the public instead of psychologists. Although the researchers provide clear definition in advance, it might have the bias.
3. More evaluation should be involved to show more robust result.

References

[1] Google paws:. https://github.com/google-research-datasets/paws. Accessed: 2021-10-27.

[2] Ali Alshahrani, Meysam Ghaffari, Kobra Amirizirtol, and Xiuwen Liu. Identifying optimism and pessimism in twitter messages using xlnet and deep consensus. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2020.

[3] Ali Alshahrani, Meysam Ghaffari, Kobra Amirizirtol, and Xiuwen Liu. Optimism/pessimism prediction of twitter messages and users using bert with soft label assignment. In *2021 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, 2021.

[4] 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*, pages 652–658, 2018.

[5] Jonathan Mallinson John Wieting and Kevin Gimpel. Learning paraphrastic sentence embeddings from back-translated bitext. In *Proceedings of Empirical Methods in Natural Language Processing*, 2017.

[6] A. B. Jones and J. M. Smith. Article Title. *Journal title*, 13(52):123–456, March 2013.

[7] Zarghuna Naseem and Ruhi Khalid. Positive thinking in coping with stress and health outcomes: Literature review. *Journal of Research & Reflections in Education (JRRE)*, 4(1), 2010.

[8] 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)*, pages 320–325, Berlin, Germany, August 2016. Association for Computational Linguistics.

[9] J. M. Smith and A. B. Jones. *Book Title*. Publisher, 7th edition, 2012.

[10] John Wieting and Kevin Gimpel. Pushing the limits of paraphrastic sentence embeddings with millions of machine translations. In *arXiv preprint arXiv:1711.05732*, 2017.

[11] Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Russ R Salakhutdinov, and Quoc V Le. Xlnet: Generalized autoregressive pretraining for language understanding. *Advances in neural information processing systems*, 32, 2019.