Reading Assignment

CSCE 638 NLP, Spring 2021

Tian Liu, UIN: 525004380

Mar 24, 2021

The following 5 papers are specifically chosen because they are closely related to our group project, which focuses on humor detection from sentences.

1 Paper 1

1.1 Title and authors

ColBERT: Using BERT Sentence Embedding for Humor Detection[1]
Author: Issa Annamoradnejad

1.2 Task

The paper addresses a sentiment analysis task in NLP.

1.3 Problem

Detection of humor in text is difficult because of its subjectivity, ambiguity and semantic intricacies. The paper aims to propose a model that can automatic detect the presence of humor in texts.

1.4 Solution

The solution of this paper is the proposed ColBERT model, which uses the BERT model to encode text into a few sentences embeddings which are fed into an eight-layered neural network. Then the model will output the classification results. The model is trained and tested with a new dataset which contains 200k short texts.

1.5 Novelty

The novelty of this paper lies in its assumption on the linguistic structure of humor which contains several normal sentences concluded with a punchline. And the paper proposed using separate lines of parallel hidden layers to extract the underlying structure of each sentence in a proper way.

1.6 Evaluation

The trained model was evaluated on 20% of the dataset and compared with 5 baseline models, including Decision Tree, SVM, Multinomial Naive Bayes, XGBoost, and XLNet. Results show that the proposed ColBERT method achieves much better performance. The ColBERT model has accuracy and F1 score of 98.2%, while other baselines range from 72% to 91%.

1.7 Analysis

The proposed ColBERT model outperform the XLNet with much less number of layers and parameters due to the analysis on the structure of humor.

1.8 Thoughts

The paper contributes a great idea on utilizing the structure of humor sentences for detecting the presence of humor. The results outperformed baselines. However, the paper did not compare its solution to any other state-of-art deep learning models. Also, the great performance of the proposed method comes with a trade-off in increasing training time of the model.

2 Paper 2

2.1 Title and authors

A Long Short-Term Memory Framework for Predicting Humor in Dialogues[2]
Authors: Dario Bertero and Pascale Fung

2.2 Task

Sentiment Analysis

2.3 Problem

The paper aims to predict the humor in dialogues. Previous works show that using a bag-of-ngram representation over a sliding window or a simple RNN to capture the contextual information of the setup was not ideal.

2.4 Solution

Propose a method based on LSTM, where each sentence was encoded through a CNN and then passed into the LSTM. Finally, the output of LSTM was incorporated with a set of high-level features to predict the humor presence in an utterance.

2.5 Novelty

The first-ever attempt that the LSTM is applied to humor response prediction or general humor detection tasks.

2.6 Evaluation

The proposed LSTM based framework is particularly effective in increasing the F-score to 62.9% over a Conditional Random Field baseline of 58.1%. And the LSTM is more effective in obtaining an higher recall with fewer false positives compared to simple n-gram shifting context window features.

2.7 Analysis

The CNN-LSTM network seems to overcome the many false-positives. Because the CNN stage is better in modeling the lexical and semantic content of the utterance and the LSTM allows to put each utterance in relation with the past context.

2.8 Thoughts

A good point of this paper is to use the LSTM to allow the forgetting of unrelated/unimportant information between different time steps. This helps in reducing many false-positives.

3 Paper 3

3.1 Title and authors

Predicting Humor by Learning from Time-aligned Comments[3]
Authors: Zixiaofan Yang, Bingyan Hu, Julia Hirschberg

3.2 Task

Sentiment Analysis

3.3 Problem

Most previous works on humor detection were done on text alone. Very little has been done on multimedia humor including text and speech information. This is mainly because of the lack of multimedia data annotated with humor.

3.4 Solution

Proposed a novel approach using time-aligned user comments on videos to generate unsupervised humor labels. With the generated labels, CNN model was trained to predict the humor in audio data. The CNN model was compared with the baseline Random Forest classifier.

3.5 Novelty

A novel approach to generate unsupervised humor labels for videos. And also this is the first study predicting perceived humor in large-scale audio data.

3.6 Evaluation

The unsupervised labels generated by the proposed method is validated with the gold labels by annotators. Results show a high correlation between them. The classifier trained on speech and text-based features obtained an accuracy as high as 0.751 on the gold test set.

3.7 Analysis

To identify the most important features related with the humor, the feature importance from the RF model was calculated. It was found that the most important features in comedy movies are related to MFCCs, indicating that patterns in specific cepstral components contribute to the humor expression in the movies.

3.8 Thoughts

This paper presents a new idea on generating labels for massive video data using the time-aligned comments. However, this approach has limited applications as not all video website supports the feature for time-aligned comments.

4 Paper 4

4.1 Title and authors

Humor Recognition using Deep Learning [4]
Authors: Peng-yu Chen, Von-Wun Soo

4.2 Task

Sentiment Analysis

4.3 Problem

Automatically detecting humor in text is challenging because the extent to which a person may sense humor depends on his/her personal background. Also, many types of humor require substantial external knowledge such as irony, wordplay, metaphor and sarcasm.

4.4 Solution

This paper proposed a convolutional neural network (CNN) with augmentation of both the filter sizes and filter numbers. Used the architecture called high-way network to implement a much more proficient model for humor recognition.

4.5 Novelty

Used the CNN with more extensive filter sizes and filter numbers to explore the best parameter settings. To ease the training of deeper neural networks, the paper used the concept of highway network which allows shortcut connections with gate functions.

4.6 Evaluation

The model was evaluated on four datasets which consists of different joke types, sentence lengths, data sizes and languages. Results show that the CNN model increases F1-Score from 0.859 to 0.903 on the first dataset, and 0.864 to 0.901 on the second dataset, compared with previous study. Also, for the other two datasets, the model achieves 0.924 and 0.943 F1-scores.

4.7 Analysis

The paper showed some examples in each category (true positive, false positive, true negative, and false negative) to get a sense of what kinds of sentences are predicted correctly and incorrectly by the model. Although it shows that the model can catch the humor in some sentences but still there are more room to be improved.

4.8 Thoughts

Personally speaking, the paper does not give much novelty. Simply trying CNN with varies filter sizes and filter numbers is more like hyperparameter optimization problem. However, the optimal parameters found are not shown in the paper.

5 Paper 5

5.1 Title and authors

Humor Recognition and Humor Anchor Extraction[5]
Authors: Diyi Yang, Alon Lavie, Chris Dyer, Eduard Hovy

5.2 Task

Sentiment Analysis

5.3 Problem

Humor recognition is challenging because of several aspects. First, different people hold different understandings of even the same sentence. Second, understanding the humor sometimes requires a lot of external knowledge. Third, there are different types of humor, such as wordplay, irony and sarcasm. But there exist few formal taxonomies. On the other hand, identifying the anchors is important to understanding the humor. But few studies have analyzed this perspective.

5.4 Solution

The paper explored the semantic structure behind humor from four perspectives: incongruity, ambiguity, interpersonal effect and phonetic style and then design a set of features to capture the potential indicators of humor. With the extracted sets of features, the paper used Random Forest to classify the humor presence in the texts. In addition, to extract the humor anchors in sentences, the paper proposed an easy Maximal Decrement algorithm.

5.5 Novelty

The paper first designed 4 types of latent structure of humor sentences and then design a set of features to characterize these structure. Moreover, the paper proposed an easy Maximal Decrement algorithm to extract the humor anchors by finding the subset that result the largest decrement in the predication score.

5.6 Evaluation

The proposed method is tested on two datasets and compared with various baselines. It was found that the Word2Vec and HCF gives the best classification performance because it takes into account both latent structures and semantic word meanings. The F1-score on two datasets are 0.859 and 0.805 which are much better than the baseline models. On the other hand, in terms of the humor anchor extraction, the proposed MDE method performs much better than the random sampling baseline methods.

5.7 Analysis

For the humor anchor extraction, in addition to the quantitative evaluation, the paper also showed qualitative evaluation. A table showing examples of good and bad extracted humor anchors was presented. The results show that the extracted humor anchors are quite reasonable in explaining the humor causes or focuses.

5.8 Thoughts

This is a great paper on humor detection. Although this paper dates back to 2015, it gives clear instruction on detecting humors from texts. It would be interesting to see how to incorporate these features with modern deep learning techniques.

References

[1] Issa Annamoradnejad. Colbert: Using bert sentence embedding for humor detection. *arXiv preprint arXiv:2004.12765*, 2020.

[2] Dario Bertero and Pascale Fung. A long short-term memory framework for predicting humor in dialogues. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 130–135, 2016.

[3] Zixiaofan Yang, Bingyan Hu, and Julia Hirschberg. Predicting humor by learning from time-aligned comments. In *INTERSPEECH*, pages 496–500, 2019.

[4] Peng-Yu Chen and Von-Wun Soo. Humor recognition using deep learning. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 113–117, 2018.

[5] Diyi Yang, Alon Lavie, Chris Dyer, and Eduard Hovy. Humor recognition and humor anchor extraction. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 2367–2376, 2015.