Literature Review

Jiaying Guo, Ang Li, Yuan Li, Yujie Sun, Shiping Yi

1 Problem Definition

Sentiment analysis is the extraction and classification of the emotions of a sentence or text that has wide applications in various areas. There have been many breakthroughs with Natural language processing techniques that have paved the way for commercial advertising, social media monitoring of elections, as well as detecting criminal activity from user tweets.

In this project, we aim to explore various models ranging from naive bayes to CNN and LSTM to train a set of processed dataset on Kaggle, and compare them with the baseline in literature. To extend the scope, the best model will be applied to the task of monitoring the overall sentiment of twitter users during the time of the Covid-19. We can also use this to analyze specifically which regions of the United States were affected the most or how deeply people felt through the negativity level of their tweets during this time period. To narrow down the scope further, news sentiment during the virus period can be gathered and analyzed to get insights on the virus outbreak and its trend with respect to time.

2 Literature Reviews

This paper (Allem et al., 2017) evaluates the importance of debiasing twitter data when extracting sentiments. The author uses data pulled from the Twitter API using the root terms related to "Hookah" and discriminates between the tweets of automated social bots and real human writers by comparing three main features –timing of tweets, spam labels, ratio of tweets from mobile versus laptop1. Some other features used are information diffusion pattern, friends features, content, and sentiment features2. These features were then fed into a model combining Support Vector Machine algorithm and rule-based reasoning and automated inference logic to classify sentiments. Grammatical structures such as the modifier "very" and "but" are used to adjust sentiment scores. The model is trained on 4 SemEval and the ISEAR emotion datasets (Mohammad et al., 2016) and an emotion-tagged tweet (Gou et al., 2014) corpus.

Pagolu, V. S.et. al.(Pagolu et al., 2016) manually built twitter text dataset for sentiment analysis by applying techniques like tokenization, stop-words removal and regex matching for special character removal for raw data pre-processing, and classified tweets into three categories: positive, negative and neutral. For feature extraction and text representation, they selected N-gram and 300d word2vec representations. They experimented with three types of machine learning models, random forest, logistic regression and SMO, for analyzing the correlation between stock market movements of a company and sentiments in tweets. On the whole, they found out that models trained with word2vec representation gives more sustainable and promising performance over large datasets. Among the models, the random forest models with word2vec representations achieved the highest accuracy of 70.2%.

S Rosenthal et. al.(Rosenthal et al., 2019) automatically filtered the tweets for duplicates and removed those for which the bag-of-words cosine similarity exceeded 0.6, retaining only the topics for which at least 100 tweets remained. Then they used CrowdFlower to annotate the new training and testing tweets. In terms of methods, the use of deep learning stands out in particular, including neural network methods such as CNN and LSTM networks. Supervised SVM and Liblinear were also very popular, e.g. combining SVM with neural network methods or SVM with dense word embedding features. Classifiers such as Maximum

Entropy, Logistic Regression, Random Forest, Naive Bayes classifier, and Conditional Random Fields were also used by other teams. Their primary measure for sentiment analysis is AvgRec (average recall), which is averaged across the POSITIVE (P), NEGATIVE (N), and NEUTRAL (U) classes.

Severyn, A et. al.(Severyn and Moschitti, 2015) first gained their word embeddings for tweets through training a neural language model on unsupervised tweets dataset, follow a 3-step process to initialize the weights of their network: 1) use word2vec model to learn the word embeddings on an unsupervised tweet corpus 2) use a distant supervision approach with their convolutional neural network to refine the embeddings 3) use the parameters obtained from previous step to initialize the network. Their neural network architecture for sentiment classification is a single convolutional layer followed by a non-linearity, max pooling and a soft-max classification layer. They use stochastic gradient descent (SGD) to train the network and backpropagation algorithm to compute the gradients. For tuning the learning rate, they apply the Adadelta update rule. To avoid overfitting, they augment the cost function with l2-norm regularization terms and apply dropout when computing the activations at the softmax output layer. Their approach sets a new state-of-the-art on the phrase-level and ranks 2nd on the message-level subtask of Semeval-2015 Twitter Sentiment Analysis (Task 10) challenge.

The author in this paper (Kim, 2014) used word embeddings trained from unsupervised learning models–word2vec, to train a series of Convolutional neural Networks with some additional modification such as having separate channels for static word vectors and task-specific word vectors (fine-tuned via backpropagation). The model consists of one layer of convolutional layer convolved with multiple filters and a max-over-time pooling operation for different feature extraction, followed by a fully-connected softmax layer to output probability distribution over the labels. The embeddings used are vectors trained by (Mikolov et al., 2013) on 100 billion Google News words. L2-norm regularization method was used to prevent overfitting. The model is tested on several datasets including movie reviews, Stanford Sentiment Treebank 1 and 2, as well as TREC (question dataset), and CR (customer reviews). The results show that models fed with the pre trained word embeddings outperform the state-of-the-art models on 4 out of 7 tasks including sentiment analysis and question classification.

Rojas-Barahona, Lina Maria (Rojas-Barahona, 2016) applied non-recursive neural networks, recursive neural network and combination of both methods for sentiment analysis, movies reviews corpus (Pang Lee, 2005) were used as training and testing dataset. Among the 11 methods, DCNN (Kalchbrenner et al., 2014) reached the highest accuracy of 86.8% for binary classification. CTree-LSTM(Tai et al., 2015) reached the highest fine-grained of 51.0 with an accuracy of 88.0. Sentiment analysis in twitter was also done. Tweets were annotated with the polarities positive, negative, and neutral. The combination of sentiment-specific word embeddings and NRC-ngram reached the highest F-1 score of 86.48%.

The paper (Stojanovski et al., 2016) proposed a model consisting of two neural networks: a convolutional neural network(CNN) with a single filter and a gated recurrent neural network(GRNN). Words are mapped to word embeddings using GloVe embeddings pretrained on the Common Crawl dataset with a dimensionality of 300. The convolutional operation is applied to every possible window with window size of 3, followed by max-over-time pooling operation. GRNN makes use of sequential data and modulates the flow of information inside the gating unit. Both networks output a fixed size vector. They are concatenated to a single feature vector, which is fed into a softwax layer. The softmax regression classifier gives probability distribution over labels in the output space. The label with highest possibility is the final prediction. Dropout regularization with a dropout parameter of 0.25 is used because too many parameters are being learned. Results are evaluated using subtasks from (Nakov et al., 2019). Different evaluation metrics are used for different tasks. Subtask B has an F1 of 0.748, subtask D has the best KLD with 0.034 and task E has an EMD of 0.316.

The authors (Ramadhani and Goo, 2017) are Adyan Marendra Ramadhani and Hong Soon Goo from Dong-A University. The title of the paper is Twitter Sentiment Analysis using Deep Learning Methods and it is published in 2017 7th International Annual Engineering Seminar (InAES). The paper used the dataset from the stream API with 1000 good dataset (both Korean and English), 1000 bad dataset (both Korean and English) as training dataset (without words Themes) and 2000 test dataset (both Korean And English). The Deep Neural Network used has the following parameters: Feedforward Neural Network, 3 Hidden Layer, Using the Rectified Linear Unit (ReLU) and the sigmoid function activation, The input is 100 neurons, Using Mean Square Unit and the Stochastic gradient descent. The first hidden layer is filtering the words, the second layer is filtering based on the sentence and the third layer is based on the popularity of the words based on the online dictionary. Based on the graph the training model accuracy is going up constantly and the test gets the convergence on the 55%. The table shows the accuracy of the train and test set. With the result of 77.45 % and the 75.03 %.

3 Evaluation Metrics

F-1 score would be selected as the main evaluation metrics to score system outputs. Precision and Recall would also be taken into consideration as secondary evaluation metrics. F-1 score would be calculated based on the predicted labels (0 = negative, 2 = neutral, 4 = positive) and the actual labels.

4 Dataset to Evaluate

Twitter for Sentiment Analysis(T4SA) dataset has been used for training and testing dataset generation. It contains about 1 million tweets, each one labeled according to the sentiment polarity of text(negative = 0, neutral = 1, positive = 2). These tweets were selected from a larger dataset of 3.4 million tweets by most confident textual sentiment predictions. With 884967 tweets used as training dataset, 235992 tweets as test dataset and the predicted labels would be evaluated using macroaverage F-1 score.

References

Jon-Patrick Allem, Jagannathan Ramanujam, Kristina Lerman, Kar-Hai Chu, Tess Boley Cruz, and Jennifer B Unger. 2017. Identifying sentiment of hookah-related posts on twitter. *JMIR public health and surveillance*, 3(4):e74.

Liang Gou, Michelle X Zhou, and Huahai Yang. 2014. Knowme and shareme: understanding automatically discovered personality traits from social media and user sharing preferences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 955–964.

Yoon Kim. 2014. Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119.

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

Preslav Nakov, Alan Ritter, Sara Rosenthal, Fabrizio Sebastiani, and Veselin Stoyanov. 2019. Semeval-2016 task 4: Sentiment analysis in twitter. *arXiv preprint arXiv:1912.01973*.

Venkata Sasank Pagolu, Kamal Nayan Reddy, Ganapati Panda, and Babita Majhi. 2016. Sentiment analysis of twitter data for predicting stock market movements. In *2016 international conference on signal processing, communication, power and embedded system (SCOPES)*, pages 1345–1350. IEEE.

Adyan Marendra Ramadhani and Hong Soon Goo. 2017. Twitter sentiment analysis using deep learning methods. In *2017 7th International Annual Engineering Seminar (InAES)*, pages 1–4. IEEE.

Lina Maria Rojas-Barahona. 2016. Deep learning for sentiment analysis. *Language and Linguistics Compass*, 10(12):701–719.

Sara Rosenthal, Noura Farra, and Preslav Nakov. 2019. Semeval-2017 task 4: Sentiment analysis in twitter. *arXiv preprint arXiv:1912.00741*.

Aliaksei Severyn and Alessandro Moschitti. 2015. Twitter sentiment analysis with deep convolutional neural networks. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 959–962.

Dario Stojanovski, Gjorgji Strezoski, Gjorgji Madjarov, and Ivica Dimitrovski. 2016. Finki at semeval-2016 task 4: Deep learning architecture for twitter sentiment analysis. In *Proceedings of the 10th International workshop on semantic evaluation (SemEval-2016)*, pages 149–154.