Identifying Optimism and Pessimism in Twitter Messages Using XLNet and Deep Consensus

Ali Alshahrani, Meysam Ghaffari, Kobra Amirizirtol, and Xiuwen Liu
Department of Computer Science, Florida State University, Tallahassee, Florida
alshahra@cs.fsu.edu, ghaffari@cs.fsu.edu, ka18g@my.fsu.edu, liux@cs.fsu.edu

***Abstract*—Modeling optimism and pessimism accurately in social media has important applications to personal health individually and society wellness collectively. In this paper, we predict optimism and pessimism in Twitter messages by building multiple models on top of XLNet, an integrated model using multiple auto-regressive language models to capture left and right contexts jointly in sentences. Utilizing multiple-head self attentions via multi-layer transformers, XLNet models are able to model negations and other semantic relationships by paying attentions to crucial and important words, leading to more accurate predictive models for optimism and pessimism. For example, using XLNet models, we have improved the state of the art accuracy of 90.32% to 96.45%, a 63.32% error reduction on a benchmark dataset. Based on the observations that all deep models should generalize to new messages based on the same training samples, we train multiple predictive models and use the consensus to further improve the accuracy on subsets of the test samples. We also demonstrate that positive emotions and sentiments in optimistic messages are much more common while negative emotions and sentiments are more so in pessimistic ones using XLNet models finetuned for emotion classification and sentiment analysis. The proposed models could be used for understanding optimism and pessimism in social media.**

***Index Terms*—XLNet, outlook prediction, emotion classification, sentiment analysis, transformers**

I. Introduction

Within the last few years, there has been a significant increase in the number of social media users. There are more than 70 million daily active users on Twitter generating 500 million tweets and 2.45 billion monthly active users on Facebook in 2019; it has been growing [11], [42]. Freedom of accessibility and speech that social media provide, allow people to express their feelings and beliefs which make it possible to judge user's personality [32]. At the same time, usage of social media affects human's life and has increased the reported mental health issues [2], [7]. Identifying and analyzing optimistic and pessimistic users based on social media activities can help to deal with the related mental health issues. Self-reported methods such as life orientation test [37] are popular in psychological tests to identify optimism and pessimism (OP/PE). However, Barker and Wright have suggested that one should study people's behaviors in daily routines for accurate understanding [1]. Social media platforms make it possible to implement studies based on daily routines. Toward the goal of obtaining accurate models, it is important to predict OP/PE accurately so that further analyses and studies can be built on them such as how OP/PE affects the way people use social media. Thus identifying optimistic and pessimistic users and their outlook sentiments based on their tweets is important.

In this paper, we propose a method to identify OP/PE at the tweet and user levels by applying the XLNet language model [44] and the deep consensus algorithm. By pretraining multiple auto-regressive models via different factorization orders, the pretrained XLNet models are able to capture the left and right contexts of words jointly as such jointed and contextualized representations are important to characterize OP/PE in Twitter messages. In addition, the XLNet models are built using the extra long transformers [9], which are able to capture longer dependencies in sentences more efficiently. With multiple-head attentions in the transformers, the models are able to pay attention to crucial words (such as negation words) that are essential to predict OP/PE correctly. In this paper, we first fine-tune XLNet models to predict OP/PE in Twitter messages. With the contextualized representations provided by XLNet, we have improved the OP/PE classification accuracy significantly measured on a benchmark dataset. For example, on the Twitter messages identified using thresholds 1 and -1 (to avoid neural messages; see Section IV), among the five runs using the XLNet base model and five runs using the XLNet large model, we have improved accuracy up to 97.92% with an average of 96.16% and 96.45% respectively. More importantly, even the worst among all the ten runs has improved the state of the art accuracy, reducing the error by 54.24% (from 9.68% to 4.43%). In the next step, we apply consensus on the output of the XLNet models which is able to achieve 99.61% accuracy for 34.89% of the test set. We have also improved substantially when all the messages are classified using threshold 0. To the best of our knowledge, this work is the first one using the pretrained XLNet models for predicting OP/PE in tweets. As identifying optimistic and pessimistic people has multiple applications including detecting people having depression problems or at risk of committing suicide, the proposed method could be an important component in order to provide better social support and treatment for people at risk [4], [5], [20].

It has been proven there is a correlation between being optimistic/pessimistic and one's health [19], [38]. Thus, we extracted the most common emotions of the users in the OP/PE dataset using the same XLNet architectures but fine-tuned for emotion classification. Using our models, we were

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

able to obtain results that are consistent with psychological findings. For example, optimistic people experience much more joy and love while sadness and fear are the most common feelings among pessimistic people [3], [15], [29]. We also examine the sentiments in Twitter messages to examine the relationships between OP/PE and sentiments commonly defined in sentiment analysis datasets.

The rest of this paper is organized as follows. We explain and discuss the related works in Section II. In Section III we explain the proposed method based on XLNet. The datasets that we used in this paper are described in Section IV. We present experimental results and analysis in Sections V and VI. We conclude the paper in Section VII with a brief summary and future work directions.

II. Related Work

Sentiment analysis, also known as opinion mining has been widely studied recently. Largely driven by commercial applications, sentiment analysis methods are proposed to analyze product and reviews to identify positive, negative, and neutral sentiments [24]. Such sentiments are often indicated by polarity words [21], [26]. While related and relevant, sentiment analysis methods are not sufficient to predict optimism and pessimism accurately. There are several recently published papers trying to address classification of optimistic and pessimistic in social networks such as Twitter. Ruan et al. [35] classified the collected tweets to pessimistic and optimistic using several machine learning models such as naive Bayes [23], nearest neighbor [28], and gradient boosting classifier [14]. They showed that the gradient boosting classifier outperforms the other models. Since the bags of words are used to represent tweets, the methods could not capture contextual information well. In a follow-up study, Caragea et al. [4] applied several deep learning models such as Convolutional Neural Networks (CNN) [22], Bidirectional Long Short Term Memory networks (BiLSTMs) [17], and Recurrent Neural Network (RNN) [6] using Glove embeddings [30]. On the same dataset, they show that these models perform better than the traditional machine learning methods in [35]. The general training strategy used is to feed the embeddings of the words in an input sentence into an encoder, which produces a representation for the input sentence. Then classification is done using the resulting representation. Three classifiers are used in the paper. Each classifier includes three fully connected layers and a softmax layer on the top, while trained on the Optimism/Pessimism Twitter dataset (OPT). The classifiers are bidirectional LSTM, CNN, and stacked gated RNN. They show that their models perform better than the Naive Bayes and Support Vector Machines used in [35]. One limitation of their approach is that the Glove embeddings used are static, which makes the models less sensitive to the contexts of the words. Therefore, it can be challenging for such network architectures to handle words with multiple different meanings (i.e., polysemous words), and the syntactic relationships such as negation. XLNet on the other side is a contextualized language model that is capable of overcoming the challenges using self-attention mechanisms and learning bidirectionally, leading to more accurate representations and resulting in better prediction accuracy, which was the reason that we selected XLNet language model as the base of our methodology.

III. Methodology

With the massive text data that are available through social media platforms, it is difficult for humans to read and analyze them. Being able to capture the dependencies in text data is essential for accurate prediction models.

A. Language Modeling

Language modeling is an essential task for (NLP) and has played a significant role for other NLP tasks such as translation [39], [43] and speech recognition [27]. Better language models should capture the language patterns that depend on larger contexts. Recently, pre-trained models have shown remarkable improvements in many NLP tasks [8], [31], [33]. With pre-trained language models, fine-tuning and feature-based are two techniques to apply them to downstream tasks. The Generative Pre-trained Transformer (OpenAI GPT) [33] is an example of the fine-tuning method. One limitation of this model is that it can only attend to the left tokens while effective representations for many NLP tasks should depend on both directions. Thus, ELMo [31] uses two separate LSTM networks (one forward and one backward) and then concatenates the representations from both networks. It improved the state-of-the-art (SOTA) in many NLP tasks such as question answering [34], sentiment analysis [41]. However, there are no interactions between the left and right networks while producing their representations and therefore the joint relationships are not modeled.

More generally, auto-regressive language modelings are feed-forward or backward models which use all the preceded words (according to the orders given by the model) to predict the current word. In other words, they learn from the previous time steps to predict value at the current step iteratively. However, auto-regressive models can not model the right and left context jointly such as GPT [33] and ELMo [31].

B. XLNet

XLNet is a generalized auto-regressive model which achieved SOTA in many NLP tasks [44]. XLNet uses permutations which enables it to learn bidirectional contexts jointly with order-aware via positional encoding. Therefore, it overcomes a fundamental limitation common to all auto-regressive models.

XLNet uses an objective function that is defined as an expectation over all factorization orders. Specifically, given a length-T sequence W = $[w_1, w_2, .., w_T]$, there are $T!$ different orders to estimate the joint probability distribution defined on all such sequences. Consider the set of all possible permutations Z=$\{[1, 2, ..., T], ..., [T, ..., 2, 1]\}$, XLNet is an integrated model that includes all the individual auto-regressive models over all possible permutations via averaging (estimated via sampling). It optimizes the probability of token w, defined as

$$\max_\theta \ \mathbb{E}_{z\sim\mathbb{Z}_T} \left[\sum_{t=1}^{T} \log\ p_\theta(w_{z_t} \mid W_{z<t})\right].$$

Here z is a possible factorization order; given z, the inner term gives the log probability of the entire sequence. The purpose of permutations is to use different factorization orders which help the model to be bidirectional, without changing the order of the given sequence. For example, $S = \{we, all, deserve, happiness\}$, assume the objective is to predict the third token "deserve". In sequence S we have 4! different permutations. For simplicity we consider only two such orders $[3 \rightarrow 1 \rightarrow 2 \rightarrow 4]$ and $[1 \rightarrow 2 \rightarrow 4 \rightarrow 3]$. In the first factorization order, the target token appears as the first element in that sequence which means no preceded token to look at, then the probability of the 3rd can be expressed as $P(deserve)$. In the second order, all other tokens appear before the 3rd token, the model needs to attend to all others token when calculating $P(deserve \mid we,\ all,\ happiness)$.

In addition, XLNet uses the extra long transformers [9] to improve the efficiency of modeling long-term dependencies in the input. Together with the factorization orders, XLNet provides an effective pre-trained model for many NLP tasks.

C. Fine-Tuning

To adapt XLNet to a specific NLP task, a proper fine-tuning procedure is needed. We use XLNet (base and large) models and fine-tune them for OP/PE prediction. For fine-tuning procedure, XLNet follows the approach proposed in [10] by utilizing the representation associated with [CLS], which is a special token for classification. For classification, we let each input sequence start with [CLS] and use its final hidden state as the aggregate input sentence representation. We denote this vector as $V \in R^H$. We add a single layer on the top of the model as a sentence classifier. In other words, the classifier utilizes V as its input. The weight matrix of the classifier layer is $W \in R^{H\times N}$ where N indicates the number of classes which in our case two (OP/PE). Then we compute the probabilities P of each class by $P = softmax(VW^T)$, and use the class with the largest output as the classification output.

D. Deep Consensus

When training multiple deep learning models, they generate similar linear regions. At the same time due to initialization and other random factors, they will have different incorrectly classified samples. This is the main idea of using consensus models. When multiple models are trained, adversarial or falsely classified samples can be identified by considering aggregation among the models. For this purpose, having n deep learning models, a sample is classified when at least k of the models predict the same output. In other words, we can eliminate the majority of incorrectly classified samples in deep learning models. Algorithm 1 shows the process of $D(n, k)$ consensus algorithm.

Algorithm 1: D(n,k) consensus model.

Result: Prediction class (can be class 0, 1, or unknown
Get the prediction values of $M_1, M_2, ..., M_n$ for each record ;
1- Apply each model and get the prediction value for each record R_i;
2- For each record R_i, if the prediction value P_i of the record is higher than the threshold T, classify it as $C_{i,n}$, otherwise, unknown;
3- For each record, $C_{i,n}$, if at least k models, prediction value $C_{i,n}$ is the same, classify the record as C_i, otherwise unknown;

TABLE I: Dataset summary.'O' stands for optimistic tweets.

	Threshold 0	Threshold 1/-1
Number of Tweets	7475 (O:4,679)	3847 (O:2,507)
Number of Users	500	500

IV. Data sets and setting

In this section, we describe three data sets that we have used in this work. First, the optimism/ pessimism Twitter data set (OPT) published by [35] that contains 7,475 randomly selected tweets corresponded to 500 pessimistic and 500 optimistic users. Each tweet in the OPT data set was manually annotated by Amazon Mechanical Turk. Hence, five independent annotators were manually annotated each tweet using scale of 3 (very optimistic) to -3 (very pessimistic). For our evaluations, we followed the approach proposed by [4]. We considered (0 and 1/-1) as two different thresholds for labelling the tweets. For the zero threshold, if the averaged score of a tweet is zero or higher, then the tweet would be considered as optimistic, else it would be pessimistic. In (1/-1) threshold, a tweet with an averaged score of 1 or greater would be optimistic and if its averaged score is -1 or lower then it will be labeled as pessimistic. Table 1 shows a summary of OPT data set.

The second data set is an emotional data set which we used to identify emotions that are in the optimistic and pessimistic tweets, made available to us by [36]. This data set has 340,541 labeled tweets which contain six types of emotions: "joy", "anger", "love", "surprising"[1], "fear" and "sadness". This data set was collected by generating hashtags for each emotion and annotated via distant supervision [16]. Lastly, we also use the Twitter Sentiment Analysis (TSA) dataset [2] which has 1,578,627 tweets that are classified as 0 for negative sentiment and 1 for positive one.

V. Experiments and Results

In our experiments, we fine-tuned the XLNet (large and base) models [44] for OP/PE prediction. We divided OPT dataset into 80% for training, 10% for evaluation and 10% for testing. We repeated each experiment five times and reported

[1]We excluded surprising emotion since it is neutral emotion

[2]Obtained from http://thinknook.com/twitter-sentiment-analysis-training-corpus-dataset-2012-09-22/.

TABLE II: Accuracy of our method vs. previous studies on the OPT dataset (shown as percentage).

	Tweet Level		User Level	
Threshold	0	1/-1	0	1/-1
NB [4]	74.20	84.10	71.30	80.10
SVM [4]	67.80	83.30	64.70	81.80
BiLSTM [4]	79.65	87.24	76.65	90.72
GRUStack [4]	80.19	87.76	76.38	92.24
CNN [4]	77.78	90.32	73.55	91.68
XLNet-Base	**84.25**	**96.16**	**84.31**	**100**
XLNet-Large	**85.28**	**96.45**	**89.11**	**100**

the average results along with the minimum and maximum results from large and base models respectively. Our training procedure is as follows: First, we used the XLNet's tokenizer to convert the input sequence into tokens that corresponding to XLNet's vocabulary before feeding them into the model. XLNet large consists of 24 layers, 1024 hidden size, 16 heads while the base model has 12 layers, 768 hidden size, 12 heads. Both models are topped with an untrained classifier layer which is trained during fine-tuning in addition to tuning the transformers. Different combinations of hyper-parameters, including learning rate in [2e-4, 2e-5, 2e-6], batch size [16, 32, 64], and maximum input length [32, 64, 128], were tested on both XLNet Large and Base, and the setting with the most efficient and stable learning was chosen. We used AdamW optimiser [25].

A. Optimism and Pessimism Prediction

In the first set of experiments, we evaluate the model performance on the OPT dataset and compare the results to that from the previous studies [4], [35]. Ruan et al. used traditional machine learning such as naive Bayes (NB) [35]; Caragea et al. examined the ability of three different deep-learning models [4]: (1) Bidirectional Long Short Term Memory networks (BiLSTMs) [17] (2) Convolutional Neural Networks (CNNs) [22] (3) Stacked Gated RNNs [6]. Table II shows the results and comparisons of our model and all previous studies.[3]

As shown in Table II, our models outperformed all the previous models for both thresholds 0 and 1/-1. In fact, our models improve the performance by more than 6% at the tweet level and 15% at the user level. For instance, at the tweet level XLNet-Large model achieved 96.45% and 85.28% for 1/-1 and 0 threshold respectively. At the user level, both models achieve 100% for 1/-1 threshold as compared with 92% achieved via GRUStack. In the zero threshold case, it is not surprising to see a decrease in the model performance, since all the tweets that are closer to the decision boundary are included, possibly including ambiguous and neural ones. Still, our models are able to improve the accuracy by more than 5%. Additionally, as mentioned before each experiment was repeated five times, we also report the best and worst accuracy among the five runs in each of the models to show the robustness of the models. As shown in Table III, even the worse model among them outperforms the SOTA models. For example, for the zero threshold, the lowest accuracy is 83% which is much higher than that of the best previous model.

[3]Note that the results of the Table II are based on the results reported in [4], since they defined new thresholds for the experiments.

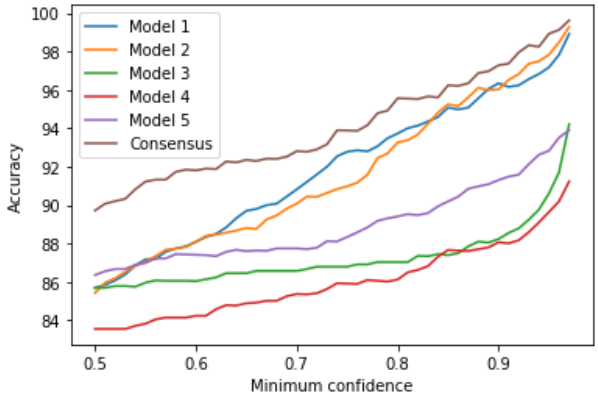


Fig. 1: The accuracy for 5 models and D(5,4) consensus model based on different confidences.

B. Consensus Results

Considering the confidence of each model can improve the accuracy further on subsets of the messages. Here the predictions of the models are reported for just the records that the models are confident about them and the prediction value exceeds a confidence threshold. Besides that, applying consensus model improves the accuracy of the prediction. We have performed several experiments for the 0 threshold case, that our models have to predict all records and it is more challenging than other cases. Figure 1 shows that by increasing the confidence threshold in all five models, the accuracy will be increased and at the same time a model will make a decision for a smaller subset of the samples. As expected, applying the consensus algorithm increases the accuracy for all different confidence thresholds. The results of D(5,4) consensus is significantly higher than any single model as shown in Figs. 1 and 2. Here we use three XLNet base models from the large XLNet case and two from the base XLNet case. XLNet Large has been shown to work better at all setting as table II illustrating that. Due to limited computational resources we have to balance the benefits of the Large model and having multiple models in the Consensus method, thus we use combination of Base and Large to take advantage of both.

As shown in the figure, the consensus model has higher accuracy than any single model for all confidence threshold. For the threshold of 0.98, it reaches 99.61% accuracy for 34.89% of the samples. Such accurate models are important as accurate results are necessary on large sets of available Twitter messages for certain applications.

C. Optimism and pessimism Emotions

In this section, we preformed OP/PE emotions analysis by extracting the most prevalent emotions among optimistic and

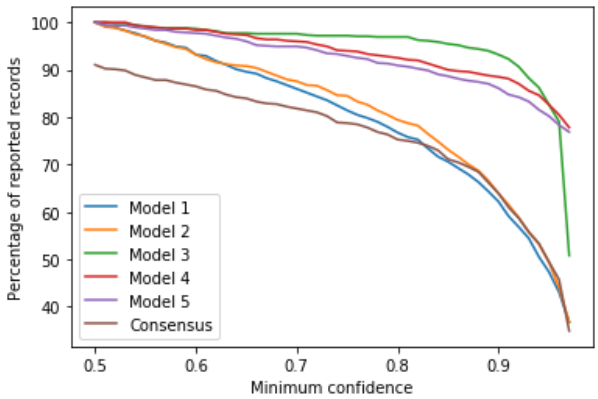


Fig. 2: Percentage of reported records for 5 different models and D(5,4) consensus model for different confidences.

TABLE III: Minimum and maximum accuracy of large and base XLNet model.

	Tweet Level	
Threshold	0	1/-1
XLNet-Base minimum accuracy	**83.82**	**95.57**
XLNet-Base maximum accuracy	**86.00**	**97.39**
XLNet-Large minimum accuracy	**84.00**	**95.83**
XLNet-Large maximum accuracy	**87.16**	**97.92**

pessimistic tweets and explain their relations and impact on health and well-being. For this experiment we used tweet level of OPT with 1/-1 threshold to study the emotions on clear OP/PE by avoiding ambiguous and neural messages. First, we train several models on the emotion datatset (see Section IV for more information), and then we evaluate the performance of this emotional model on OPT to find the most prevalent emotions in each of the two categories (OP/PE). One motivation behind this analysis is that many psychological studies have shown there are relations between being OP/PE and health [19], [38]. Thus, finding and explaining OP/PE emotions, and their influences on health could help to illustrate how being OP/PE impact our wellness and health.

Figure 5 illustrates the distributions of emotions over OP/PE tweets. It is clear that positive emotions like joy and love are experienced more by optimistic users. In comparison, negative emotions such as sadness, fear and anger are more popular among pessimistic users. That could be one reason of why optimistic people are healthier than pessimistic once since many psychological findings have shown that positive emotions like love and joy associated with healthier hearts, stronger immune system, and longer life expectancy [12], [13] while negative emotions like anger could be a cause of health problems such as coronary heart diseases and bulimic behaviors.

D. Optimism and pessimism sentiments

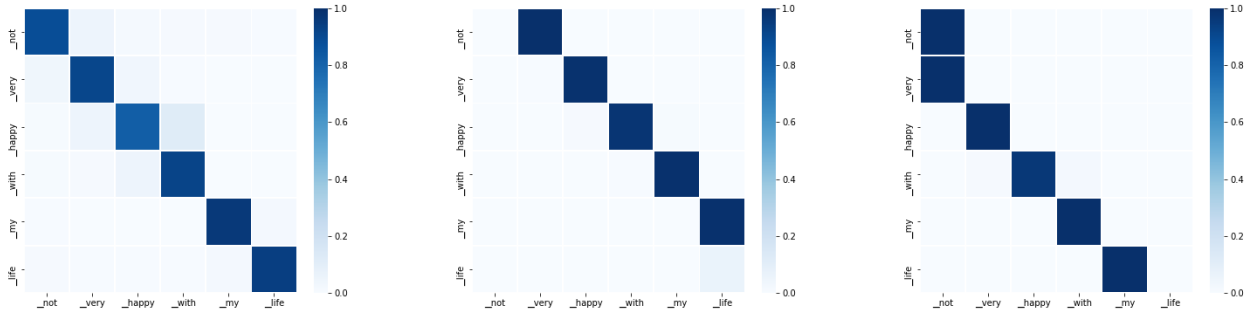
Next, we analyzed sentiments of OP/PE using the TSA dataset described in Section IV. First, we trained a classifier on a subset of TSA and evaluated its performance on OPT with both thresholds (0 and 1/-1). Figure 6 shows the results which suggest that sentiment analysis is not sufficient to identify OP/PE as Caragea et al. demonstrated in [4] using polarity word statistics. Clearly, in Fig. 6 positive and negative sentiments appear in both categories (OP/PE) with different percentages. For example, in the zero threshold case, 30% of optimistic tweets classified as negative tweets. We hypothesize that appearance or absence of polarity words plays an important role in sentiment analysis, but that is not enough to capture OP/PE. For instance, from OPT dataset we have this tweet "done holding grudges with people. life's too short to have hate in your life " consider as an optimistic tweet even though it has negative terms such as "holding grudges", "too short" and "hate".

VI. Analysis via Visualization

A. Multi-Heads Attention

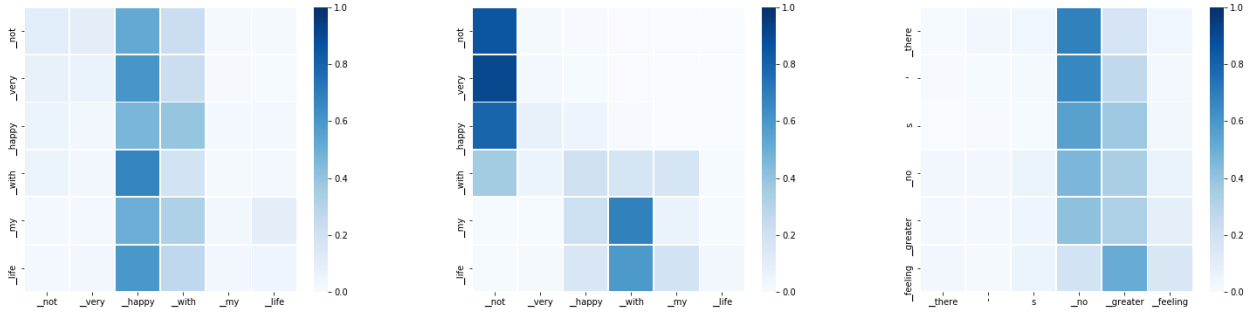
In this section, we explored the multi-head self-attention mechanism of XLNet base model by plotting the attention of all the heads in each layer. The goal is to understand their contributions and relative importance of the significant ones. Attention-mechanism has been popular and successful technique since it was introduced by [43]. Using 12 and 16 heads in the XLNet base and large model respectively in each layer that indicates how it is heavily involved in the models architecture. The main feature of multi-head self attention is to produce "context-sensitive" representation of the input tokens. When the model processes an input sentence, self-attention mechanism allows each input token to distribute its attention weights overall tokens in the current sequence in a row wise manner, base on how significant they are to the current token. After plotting the attention weights of all heads in the model we found some important and interesting patterns. First, many heads in the model direct the most of their attentions to the current token as in shown in Fig 3a, while some other heads paying more attention to next and previous tokens of their positions in the sequence as Fig 3b and 3c illustrate. This pattern of attention allows each token to include the surrounded tokens in its next representations. That is very important since many words have different meanings based on their context.

Second, negation is an essential and significant aspect in all humans language and detecting negation is very important especially for sentiments analysis since it affect contextual polarity [18], [40]. For example, in this tweet "not very happy with my life" the negation term here invert the polarity of the sentence. interestingly, in multi-heads attention some heads focus on negation words. For instance, Fig 4b shows heavy attention is directed to the token "not" while producing the next representations of "very happy". Similarly, Fig 4c shows the attention weights of the same head. It was able to capture



(a) Current token attention. (b) Next token attention. (c) Previous token attention.

Fig. 3: Positional attention.



(a) Crucial word attention. (b) Negation attention. (c) Negation attention.

Fig. 4: Attention of negation and crucial word (shown only part of the tweets for readability).

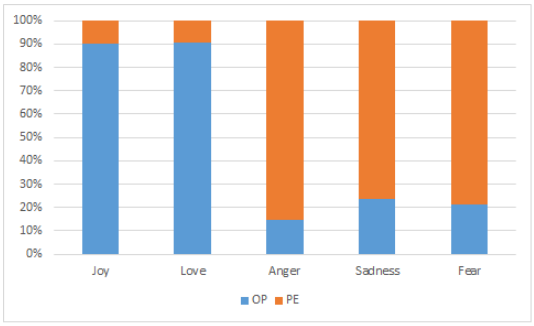


Fig. 5: Distribution of emotions among optimistic and pessimistic users (To be viewed in color).

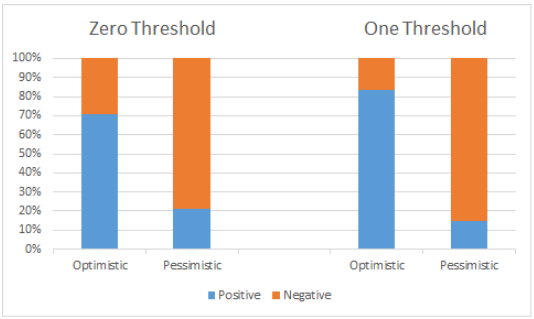


Fig. 6: Sentiments in OP/PE (To be viewed in color).

the negation even though it expressed in this example using different negation word "no".

Moreover, we find a few heads direct some of their attention to rare words which are usually the crucial words in the sentence. Fig 4a shows the attention of this type of pattern which is a significant pattern specially for sentiment analysis.

B. XLNet Tokenizer

A model Tokenizer is one of the initial steps as we mentioned in section V we use XLNet's Tokenizer to convert input sentences into corresponding XLNet's vocabularies. Also, it handles out of vocabulary (OOV) words by separating them into sub-tokens. It could also participate and help in the model's decision by correctly separating OOV tokens such as "misspelling words by connecting them" which seem to be a common practice in Twitter, due to the limited characters that are allowed in a tweet. Importantly, when the misspelling words are crucial words to the model's decision. For example, from OPT dataset we have this tweet "there's no greater

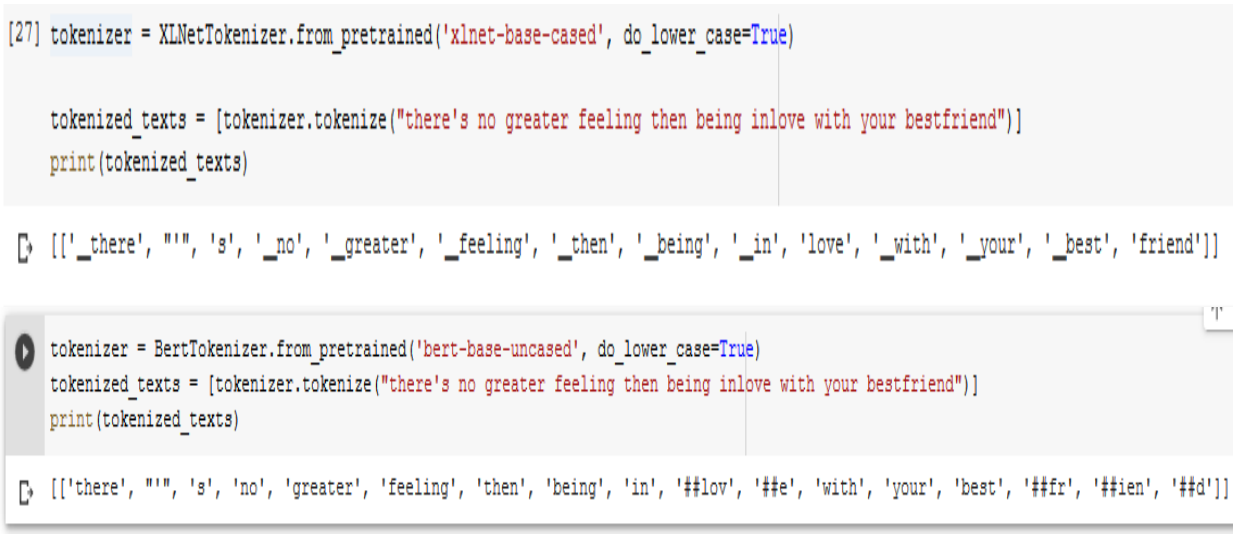


Fig. 7: XLNet Tokenizer (Top) and BERT Tokenizer (Bottom).

feeling then being inlove with your bestfriend" which has two misspelled important words "love" and " best friend", however, XLNet's tokenizer are able to separate them correctly and recover the correct words as shown in Fig 7(top). In comparison, we also use BERT's tokenizer to tokenize the same sentence. The result shown in Fig 7 (bottom) shows it fails to separate and retrieve the correct words that may mislead and affect the model's decision.

VII. Conclusion

In this paper, we have fine-tuned XLNet models to predict outlook sentiments in Twitter messages both at the individual message level and at the user level. As the XLNet models are able to capture left and right contexts jointly and compute contextualized representations using multiple-head attentions, our methods improved the state of the art on a benchmark dataset substantially. Furthermore, using a deep consensus algorithm, we can improve the accuracy of subsets significantly. Accurate models like ours may be necessary for applications where accuracy is important.

While our models gave the best accuracy on a benchmark dataset, as Twitter messages often include special characters for predefined meanings, these special terms need to be handled correctly in order to classify them well. Additionally, the effectiveness of the proposed models for applications such as suicide detection need to be further investigated. Furthermore, how to use the models like the proposed ones to study personality and other psychological traits needs to be studied as Twitter messages may well be location and position dependent (such as tweets at jobs and at homes could be very different for the same users).

References

[1] Roger G Barker and Herbert F Wright. *One boy's day; a specimen record of behavior.* Harper, 1951.

[2] Chloe Berryman, Christopher J Ferguson, and Charles Negy. Social media use and mental health among young adults. *Psychiatric quarterly*, 89(2):307–314, 2018.

[3] Peter Boman, Douglas C Smith, and David Curtis. Effects of pessimism and explanatory style on development of anger in children. *School Psychology International*, 24(1):80–94, 2003.

[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] Edward C Chang. *Optimism & pessimism: Implications for theory, research, and practice.* American Psychological Association, 2001.

[6] Junyoung Chung, Caglar Gulcehre, Kyunghyun Cho, and Yoshua Bengio. Gated feedback recurrent neural networks. In *International Conference on Machine Learning*, pages 2067–2075, 2015.

[7] Mike Conway and Daniel O'Connor. Social media, big data, and mental health: current advances and ethical implications. *Current opinion in psychology*, 9:77–82, 2016.

[8] Andrew M Dai and Quoc V Le. Semi-supervised sequence learning. In *Advances in neural information processing systems*, pages 3079–3087, 2015.

[9] Zihang Dai, Zhilin Yang, Yiming Yang, William W Cohen, Jaime Carbonell, Quoc V Le, and Ruslan Salakhutdinov. Transformer-xl: Attentive language models beyond a fixed-length context. *arXiv preprint arXiv:1901.02860*, 2019.

[10] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[11] Dsayce. https://www.dsayce.com/social-media/tweets-day/, 2020.

[12] Barbara L Fredrickson. Cultivating positive emotions to optimize health and well-being. *Prevention & treatment*, 3(1):1a, 2000.

[13] Barbara L Fredrickson. Positive emotions. *Handbook of positive psychology*, pages 120–134, 2002.

[14] Jerome H Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001.

[15] Shohreh Ghorbanshirodi, Javad Khalatbari, and Mostafa Akhshabi. Researching the positive and negative emotion and optimism with enduring in the scholar. *Life Science Journal*, 10(2s), 2013.

[16] Alec Go, Richa Bhayani, and Lei Huang. Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford*, 1(12):2009, 2009.

[17] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[18] Laurence Horn. *A natural history of negation*. CSLI PUBLICATIONS, 1989.

[19] Leslie Kamen-Siegel, Judith Rodin, Martin E Seligman, and John Dwyer. Explanatory style and cell-mediated immunity in elderly men and women. *Health Psychology*, 10(4):229, 1991.

[20] Hamed Khanpour, Cornelia Caragea, and Prakhar Biyani. Identifying emotional support in online health communities. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[21] Soo-Min Kim and Eduard Hovy. Determining the sentiment of opinions. In *Proceedings of the 20th international conference on Computational Linguistics*, page 1367. Association for Computational Linguistics, 2004.

[22] Yoon Kim. Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*, 2014.

[23] David D Lewis. Naive (bayes) at forty: The independence assumption in information retrieval. In *European conference on machine learning*, pages 4–15. Springer, 1998.

[24] Bing Liu. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1):1–167, 2012.

[25] Ilya Loshchilov and Frank Hutter. Fixing weight decay regularization in adam. *arXiv preprint arXiv:1711.05101*, 2017.

[26] Isa Maks and Piek Vossen. Sentiment analysis of reviews: Should we analyze writer intentions or reader perceptions? In *Proceedings of the International Conference Recent Advances in Natural Language Processing RANLP 2013*, pages 415–419, 2013.

[27] Tomáš Mikolov, Martin Karafiát, Lukáš Burget, Jan Černockỳ, and Sanjeev Khudanpur. Recurrent neural network based language model. In *Eleventh annual conference of the international speech communication association*, 2010.

[28] Marius Muja and David G Lowe. Fast approximate nearest neighbors with automatic algorithm configuration. *VISAPP (1)*, 2(331-340):2, 2009.

[29] Jari-Erik Nurmi, Sari Toivonen, Katariina Salmela-Aro, and Sanna Eronen. Optimistic, approach-oriented, and avoidance strategies in social situations: Three studies on loneliness and peer relationships. *European Journal of Personality*, 10(3):201–219, 1996.

[30] Jeffrey Pennington, Richard Socher, and Christopher Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543, 2014.

[31] Matthew E Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. Deep contextualized word representations. *arXiv preprint arXiv:1802.05365*, 2018.

[32] Lin Qiu, Han Lin, Jonathan Ramsay, and Fang Yang. You are what you tweet: Personality expression and perception on twitter. *Journal of Research in Personality*, 46(6):710–718, 2012.

[33] Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. Improving language understanding by generative pre-training. *URL https://s3-us-west-2. amazonaws. com/openai-assets/researchcovers/languageunsupervised/language understanding paper. pdf*, 2018.

[34] Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. Squad: 100,000+ questions for machine comprehension of text. *arXiv preprint arXiv:1606.05250*, 2016.

[35] 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, 2016.

[36] Elvis Saravia, Hsien-Chi Toby Liu, Yen-Hao Huang, Junlin Wu, and Yi-Shin Chen. Carer: Contextualized affect representations for emotion recognition. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3687–3697, 2018.

[37] Michael F Scheier and Charles S Carver. Optimism, coping, and health: assessment and implications of generalized outcome expectancies. *Health psychology*, 4(3):219, 1985.

[38] Michael F Scheier, Charles S Carver, and Michael W Bridges. Optimism, pessimism, and psychological well-being. *Optimism and pessimism: Implications for theory, research, and practice*, 1:189–216, 2001.

[39] Holger Schwenk. Continuous space translation models for phrase-based statistical machine translation. In *Proceedings of COLING 2012: Posters*, pages 1071–1080, 2012.

[40] Vaidehi Shah and Purvi Rekh. A survey: Importance of negation in sentiment analysis. *International Journal of Emerging Technology and Advanced Engineering*, 4(3):70–73, 2014.

[41] Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D Manning, Andrew Ng, and Christopher Potts. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 conference on empirical methods in natural language processing*, pages 1631–1642, 2013.

[42] Statista. https://www.statista.com/statistics/274564/monthly-active-twitter-users-in-the-united-states/, 2019.

[43] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008, 2017.

[44] Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Ruslan Salakhutdinov, and Quoc V Le. Xlnet: Generalized autoregressive pretraining for language understanding. *arXiv preprint arXiv:1906.08237*, 2019.