Emotion Analysis of Tweets: Cross-Lingual Transfer, Message Propagation and User Profile

Ling Liu
CSCI5502 (in-class)

Bhavani Rishitha Ravipati
CSCI5502 (in-class)

Binpeng Wu
CSCI4502 (in-class)

Victoria Everts
CSCI5502 (distance)

1 INTRODUCTION

Emotion and sentiment analysis has been a well-studied topic in NLP and other areas[26]. However, the intensity of emotions and sentiment has been drawing research attention recently and less work has been conducted about it[20]. From 2016 until the present, Mohammad and a group of fellow researchers[18, 19] have constructed an incredibly useful dataset of labeled tweets with emotion categories and intensities. They have made their lalebed data public, which include tweets in three languages: Arabic, English, and Spanish. Two shared tasks have been organized using the dataset on affect intensity and attracted wide participation[17, 20].

For our project, we propose three subtasks concerning emotion intensity. First, we will do the standard task as the subtask on emotion intensity regression (or classification) of SemEval-2018 Task 1: Affect in Tweets, and extend it by doing emotion classification and emotion intensity prediction together. Second, we will explore the cross-lingual transfer effect as to with models created in the first part of work, i.e. to investigate whether it is possible to use the labeled data in one language to train models for another language, how to do this, and whether the degree of similarity between languages influences the cross-lingual transfer effect. Third, we will analysis the correlation between Tweets' emotion categories and intensity, message propagation, and socioeconomic status which we can get from user's profile.

In the following sections of the proposal, we will first provide a review of existing work related to the three subtasks we plan to do in Section 2. Section 3 will explain in details each of the three subtasks, the methods to do it, and the datasets the work will be conducted on. Evaluation metrics and methodologies will be described in Section 4, followed by milestones we plan to achieve step by step in Section 5. The proposal ends with a summary as Section 6.

2 RELATED WORK

2.1 Emotion analysis

A lot of work has been conducted on emotion and sentiment analysis (e.g.[2, 3, 14, 21, 25]). Yadollahi et al.[26] provide a comprehensive overview of sentiment analysis, in which emotion analysis is one part with opinion analysis being the other. Emotion analysis and mining is further divided into four groups: emotion detection, emotion polarity classification, emotion classification and emotion cause detection.

However, relatively much less research has been conducted on affect intensity. SemEval organized a shared task on affect intensity in 2018.[1] Mohammad et al.[17] detail the findings of a shared task completed by 75 teams with 200 total members. The task was to perform these 5 subtasks: 1. emotion intensity regression, 2. emotion intensity ordinal classification, 3. valence (sentiment) regression, 4.valence ordinal classification, and 5. emotion classification with labeled data created by the authors of this paper and study. Through watching 75 teams complete the tasks, the authors of this paper provide a summary of methods, tools and a focus on techniques and resources. The data is openly provided. The authors explain that they want to continue the exploration into how people convey emotion through language. The teams that performed the best used deep neural network representations of tweets with sentence embedding and emotion lexicons and sentiment lexicons (such as: Affect Intensity Lexicon (Mohammad,2018b), ANEW (Bradley and Lang, 1999), and NRC Valence-Arousal-Dominance Lexicon (Mohammad, 2018a)) that have already been created.

Preceeding the 2018 SemEval Shared task, Mohammad and Kiritchenko [18] created the data, The Affect in Tweets Dataset(11,000 tweets).Previous work on emotion and sentiment analysis mainly focused on whether the emotion was positive, negative or neutral. In this dataset, they identify sad, fearful, angry and joyful as the four main emotions and categories for classifying tweets. Their goal was to create a dataset annotated for many different emotions and intensities which was composed of tweets. Features of the dataset include tweets that contain the specific word associated with one of the main emotions while others did not and all of the emotions also have an intensity rating. They intentionally limited their database to tweets that were associated with an emotion by searching for all of the synonyms associated with an emotion. The tweets were annotated through crowd-sourcing by asking a crowd-source annotator to choose the word and a brief clarification of associated words that was best associated with a tweet.

In 2017, Mohammad and Bravo-Marquez [20] had their inaugural shared task event[2] that was followed in 2018 by the SemEval event. Like the 2018 event, the researchers' intent was to summarize the tools, techniques, resources and machine learning environments that were implored to better understand the most successful setups. Similar to the 2018 event, calculations were evaluated using the Pearson Correlation Coefficient and Spearman Rank Coefficient. They used a Weka machine learning workbench as the baseline system for automatically determining tweet emotion intensity. The winning team of the competition implemented a feed-forward neural network in two separate models that were then made into an ensemble and increased the Pearson Coefficient by two points. The

[1]https://competitions.codalab.org/competitions/17751
[2]http://saifmohammad.com/WebPages/EmotionIntensity-SharedTask.html

second winning team applied a random forest regression model. The third winning team created an ensemble of regression algorithms.

Preceeding the 2017 shared task event, Mohammad and Bravo-Marques record the creation of the dataset[19].They claim that this is the first dataset of tweets annotated for the intensity of angry, sad, fearful and joyful emotions. For each of the four emotions, angry, sad, fearful and joyful, 50-100 associated words are selected and used to find tweets containing them. Limits on tweets from particular Twitter users and from particular word searches were implemented to insure that the dataset wasn't skewed to a particular Twitter user or emotion. Best Word Scaling (BWS) was used in the 2017 dataset creation as it was in the 2018 dataset creation. Differing from the 2018 dataset, there was a focus on emotion hashtags. They showed that this dataset was useful in determining emotion intensity.

2.2 Neural cross-lingual transfer in NLP

Cross-lingual transfer has been found to be helpful for NLP tasks like speech recognition [8], machine translation [5, 9, 28], dependency parsing [6, 7] and morphological inflection and tagging [4, 11]. The philosophy of cross-lingual transfer is to use high-resource language data to facilitate NLP tasks for low-resource language.

Cross-lingual word embedding representations make it possible for us to deal with word meanings in multilingual contexts. They are the current state-of-the-art representations for developing natural language processing models for low-resource languages [24]. Downstream tasks have also been used as a way to evaluate the quality of cross-lingual word embeddings. For example, Klementiev et al. [13] trained a document classifier to predict topics on the document representations derived from word embeddings in the source language, and then tested the model on the documents in the target language. Mogadala and Reittinger [16] used the sentiment analysis of the multilingual Amazon product review database as an evaluation of their multilingual embeddings.

Mogadala and Reittiger's work [16] is related to cross-lingual transfer for emtion and sentiment analysis. However, we haven't found work directly exploring cross-lingual transfer effect on emotion intensity analysis.

2.3 Emotion, message propagation, and user profile

There have been few prior works which analyzed role of tweet's emotions in various tweeting behaviours. Kim and Yoo's work [12] concludes that "degree of emotion expressions in twitter messages can affect the number of replies generated as well as retweet rates." This paper examines the role of emotion in message propagation in a political setting. Along with building a model to predict the width-depth of a post to propagate, Kanavos et al.[10] examined the influence of emotion in the propagation, concluding that "posts containing negative emotional states are more likely to be retweeted than tweets with positive or neutral emotional states."

Preotiuc-Pietro et al.[23] uncovered findings such as "higher income users express more fear and anger, while posting less subjective content both positive and negative whereas lower income users express more of the time emotion and opinions. Users perceived as religiously unaffiliated and less anxious appear to have higher earnings. These users have more followers and get retweeted more, albeit following similar number of persons, tweeting less and with fewer URLs."

Some previous research works built predictive models on socioeconomic features like occupation,income,gender and the textual way of tweeting their opinion suggesting there is some relation between the features mentioned above. PreoÅčiuc-Pietro et al.[22] focuses on the prediction of the occupational class for a public user profile of twitter. Similarly Aletras and Chamberlain[1] predicts the occupational class and the income of Twitter users It did excellent work on extracting and preprocessing user profiles which have occupation. It is kind of highlighting the fact that the occupation,income of a user influences text use. They also concludes that "information extracted from the user's social network and their language use are complementary." There are some papers which examined social influence on tweets. For example, Ye and Wu [27] analyzed the correlation between various social influences and their relation with message propagation. It defines social influence as a phenomenon which "occurs when an individual's thoughts or actions are affected by other people. "They examine the metrics like Follower influence, Reply influence, ReTweet influence for social influence.

Lerman et al.[15] is the closest work to the question we are trying to answer for subtask 3. We are planning to use the results of this paper to evaluate ours. It's findings are discussed in evaluation of subtask 3.

3 PROPOSED WORK

In this part we will describe the three subtasks we propose to work on, followed by information about the data that we will use for the tasks.

3.1 Subtasks

3.1.1 Subtask 1: emotion classification and intensity prediction. For this subtask, the data we use will be the English data for emotion intensity regression provided by the SemEval-2018 Task 1: Affect in Tweets.

First, we will build a neural network model to do the standard emotion intensity regression task of the SemEval-2018 Task 1: Affect in Tweets, i.e. given a tweet and its emotion type, the task is to predict the intensity of the emotion expressed in the tweet.

Second, we will extend the task by making the model to classify the type of emotions expressed in a tweet and predict the intensity of the emotion.

3.1.2 Subtask 2: cross-lingual transferring effect for emotion classification and intensity prediction. For this subtask, the data we use will be the data for emotion intensity regression for all three languages, i.e. English, Spanish and Arabic, from the SemEval-2018 Task 1: Affect in Tweets.

In this part, We ask and try to answer two questions:

- Is it there cross-lingual transferring effect as to emotion classification and intensity prediction?
- If there is cross-lingual transferring effect, is it stronger between languages that are similar than between languages that are less similar?

To answer the questions, we will combine the English training and development data with those of Spanish and Arabic respectively, and then test the performance of the models (one for predicting emotion intensity, and the other for doing emotion classification and intensity prediction together) on the English test set. If the performance of the models trained on combined bilingual data set is better than the one trained on monolingual data, it will provide positive answer to our first question. The difference in the performance of the models trained on English and Spanish, and those trained on the combination of English and Arabic will provide answer to the second question.

To combine data in different languages, we plan to explore three different ways:

(1) We can directly add Spanish or Arabic data to English data, and train the model with the two languages as we train the model with just one language.
(2) We can first use an out-of-box machine translation system to translate Spanish or Arabic text into English, and then add the translated text with their original emotion category and intensity labels to the English to train the model as we will do for subtask 1.
(3) We can first get cross-lingual word embeddings for the three languages, and cross-lingual word embeddings as representations for tweets in each languages. We have found two related out-of-box sources which may provide us with cross-lingual word embeddings.[3] [4]

3.1.3 Subtask 3: correlation analysis between tweets' emotion types and intensity, message propagation and user's economic data. In this subtask, we try to see how strong or weak is the relationship between three different aspects of social media interactions: messages' emotion types and intensity, message propagation and user's socioeconomic data. The motivation behind exploring this subtask is to understand various factors that could possibly be the reasons for a user's opinion on a subject. This analysis will also partly give an idea about how communities grouped either by age, occupation or place form an opinion on a topic. This kind of investigation can be achieved by performing statistical analysis so as to find the correlation between all three combinations of the variables mentioned above. There are three major steps involved in performing this task: Firstly, we collect tweets as well as user's information (who tweeted it) from Twitter API (as described in 3 in data) and twitter archive (as described in 2 in data) as needed. This data contains information such as user's data-location-description-age-tweet count-time zone, tweet's text-retweet count-source-reply's-geographical data. In this step, it is important to preprocess the data to remove users who are not active enough. Activity of users is measured based on the number of tweets they post and number of days they post. Also, the occupation and age are optional while creating the user profile on twitter. Hence, the data will be skewed. Secondly, we apply the models that are developed for subtask 1 to the tweets obtained from the previous step and get their emotion type and their intensity. Thirdly, we perform statistical analysis for all three combination of variables.

[3]https://github.com/facebookresearch/fastText
[4]https://github.com/Babylonpartners/fastText_multilingual

3.2 Data

- Data collection: our data are from three main sources.
  (1) Shared SemEval-2018 data:The Arabic, English and Spanish tweets labeled with emotion types and intensity in this data.[5]
  (2) A online Twitter archive:a collection of JSON grabbed from the general twitter stream from 2012-2018. [6]
  (3) (optional)We will collect data directly from Twitter API by using the python library Tweepy[7].
- Overview and understanding of the data
  (1) The Shared SemEval-2018 data for 3 language all contain the training set, development set and test set. these set all include the data about user's ID, tweet text, its emotion type and intensity score. test set not include the intensity score.
  (2) The JSON online database have the detailed tweet data. for user data, it only displays user's name or id. but it collect the precise tweet information like geo data and count of retweet or favorites. although the data were not well organized, it contains a large amount of data from 2011 to 2018 that we could use as historical data in subtask
  (3) The online Twitter API data is more workable and easy to select some targeted user or compare the data by various occupation. however, it took more energy to preprocess the data and the data itself were limit in several days.
- Preprocessing of data
  - We will use Python libraries for data preprocessing and data visualization. Then,
  - We will use the python library Tweepy for collecting data from Twitter API.
  - We will re-organized and extracted the online tweet database .
- Data management
  - We will use PyTorch (or other deep learning frameworks) for building neural network models.
  - For our subtask one and two, We will use the training and development set in the semEval-2018 to train the model, and evaluate its performance on the test set for both models.
  - We will use the created neural models or other existing work to dual with the online historical tweet data to get the emotion classification, emotion intensity scores and even attention intensity level.
- Initial analysis of the data
  - The SemEval data have the 4 data type, user id, tweet text, emotion label and emotion intensity in all set, the emotion intensity score in test set is showed as NONE. Table 1 summarizes the number of tweets in the dataset as to its division of training, development and test sets.
  - The data from Twitter API contain two parts of inforamtion: User data and Tweet data. User data can include user name, location, created date, number of favorites, URL, profile image URL, language, protected, description(occupation),

[5]https://competitions.codalab.org/competitions/17751#learn_the_details-datasets
[6]https://archive.org/details/twitterstream?sort=-publicdate
[7]https://github.com/tweepy/tweepy

Table 1: The number of tweets for the emotion intensity subtask in the SemEval-2018 Affect in Tweets Dataset [17]

Dataset	Train	Dev	Test	Total
English				
anger	1,701	388	1,002	3,091
fear	2,252	389	986	3,627
joy	1,616	290	1,105	3,011
sadness	1,533	397	975	2,905
Spanish				
anger	1,166	193	627	1,986
fear	1,166	202	618	1,986
joy	1,058	202	730	1,990
sadness	1,154	196	641	1,991
Arabic				
anger	877	150	373	1,400
fear	882	146	372	1,400
joy	728	224	448	1,400
sadness	889	141	370	1,400

age, verified, tweet count, and time zone. Some of the information like age, description are optional, so it's possible to get such information but we may not be able to have it for every user. Tweet data include text, URL, retweet count, date, source, favorite count, hashtags, mentioned users, in reply to screen name, and geo data.

4 EVALUATION

4.1 Evaluation for subtasks 1 and 2

For both subtasks 1 and 2, the test data are the same, i.e. the English emotion intensity test set provide by the SemEval 2018 Task 1: Affect in Tweet. Gold-standard intensity is provided and the amount of the data is provided in Table 1

The shared task provides a baseline system for emotion intensity prediction. [8] We will use this system as our baseline as well and compare the result of emotion intensity prediction from our own system with this baseline.

To evaluate the emotion intensity prediction result, we will use the Pearson Correlation Coefficient with the Gold ratings as the official shared task competition metric[17]. For the emotion classification result, we will use evaluation metrics like accuracy, precision, recall, F1 score or AUC (area under the curve).

4.2 Evaluation for subtask 3

Since this subtask is about understanding the correlation between three variables discussed in subtask 3, we compare our results with a recent paper which worked on this topic[15]. This paper concludes that "social media users engage more deeply with less diverse social contacts are those where they express more negative emotions, like sadness and anger. Demographics also has an impact: these places have residents with lower household income and education levels. Conversely, places where people engage less frequently but with diverse contacts have happier, more positive messages posted from them and also have better educated, younger, more affluent residents".

5 MILESTONES

By the project checkpoint in Week 12 (i.e. Nov 13, Nov 15), we have two goals:

(1) Finish subtask 1, i.e. have the models ready for emotion intensity prediction and for doing emotion classification and intensity prediction together.
(2) Finish data preparation for subtask 3, i.e. decide what factors we will correlate the emotion analysis with, collect and preprocess the data either from Twitter API or the Tweet archive by selecting the tweets which fit our purpose, taking out the text, age, occupation, retweets count, favorite count information and putting their into a format ready for further analysis.

After the checkpoint, we will spend two to three weeks on the following two parts of work:

(1) Explore the second subtask.
(2) Apply the model developed for subtask 1 to the text data collected for subtask 3, and analyze the correlation between the emotion categories and intensity of the tweets with message propagation (as indicated by retweet count and favorite count) and user's socioeconomic status (as indicated by age, occupation, etc.)

We target at finishing all experiments around Week 15 and will spend one to two weeks writing up our work. In Week 15, We will also work on preparing for project presentation for the last week.

6 SUMMARY

We propose to classify tweets as to their emotion categories of anger, fear, joy and sadness, and predict the intensity of the emotions. Knowing the emotions users expressed in their tweets can be for tasks like making recommendations and evaluating their psychological health.

Emotion classification and intensity prediction requires labeled data for training and test. Labeling data is expensive. If the labeled data in one language can be used to train the models for another language, it will reduce the hunger for label data in such tasks. Therefore, we will also explore the cross-lingual transfer effect for tweet emotion classification and intensity prediction as a way of applying labeled data in one language to another language.

As a further investigation of how emotion can be related to other factors, we will analyze the correlation between the emotion categories and intensities of tweets, and message propagation, as well as user's socioeconomic status. The analysis between emotion and message propagation will provide insights about how to better manipulate the propagation of message, and the analysis between emotion and user's socioeconomic status can contribute to our understanding of people's emotions.

REFERENCES

[1] Nikolaos Aletras and Benjamin Paul Chamberlain. 2018. Predicting Twitter User Socioeconomic Attributes with Network and Language Information. *arXiv preprint arXiv:1804.04095* (2018).

[8] https://github.com/felipebravom/AffectiveTweets

[2] Johan Bollen, Huina Mao, and Alberto Pepe. 2011. Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. *Icwsm* 11 (2011), 450–453.

[3] Carlos Busso, Zhigang Deng, Serdar Yildirim, Murtaza Bulut, Chul Min Lee, Abe Kazemzadeh, Sungbok Lee, Ulrich Neumann, and Shrikanth Narayanan. 2004. Analysis of emotion recognition using facial expressions, speech and multimodal information. In *Proceedings of the 6th international conference on Multimodal interfaces*. ACM, 205–211.

[4] Ryan Cotterell and Georg Heigold. 2017. Cross-lingual, Character-Level Neural Morphological Tagging. *arXiv preprint arXiv:1708.09157* (2017).

[5] Orhan Firat, Kyunghyun Cho, and Yoshua Bengio. 2016. Multi-way, multilingual neural machine translation with a shared attention mechanism. *arXiv preprint arXiv:1601.01073* (2016).

[6] Jiang Guo, Wanxiang Che, David Yarowsky, Haifeng Wang, and Ting Liu. 2015. Cross-lingual dependency parsing based on distributed representations. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, Vol. 1. 1234–1244.

[7] Jiang Guo, Wanxiang Che, David Yarowsky, Haifeng Wang, and Ting Liu. 2016. A Representation Learning Framework for Multi-Source Transfer Parsing.. In *AAAI*. 2734–2740.

[8] Georg Heigold, Vincent Vanhoucke, Alan Senior, Patrick Nguyen, MarcâĂŹAurelio Ranzato, Matthieu Devin, and Jeffrey Dean. 2013. Multilingual acoustic models using distributed deep neural networks. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*. IEEE, 8619–8623.

[9] Melvin Johnson, Mike Schuster, Quoc V Le, Maxim Krikun, Yonghui Wu, Zhifeng Chen, Nikhil Thorat, Fernanda Viégas, Martin Wattenberg, Greg Corrado, et al. 2016. Google's multilingual neural machine translation system: enabling zero-shot translation. *arXiv preprint arXiv:1611.04558* (2016).

[10] Andreas Kanavos, Isidoros Perikos, Pantelis Vikatos, Ioannis Hatzilygeroudis, Christos Makris, and Athanasios Tsakalidis. 2014. Modeling retweet diffusion using emotional content. In *IFIP International Conference on Artificial Intelligence Applications and Innovations*. Springer, 101–110.

[11] Katharina Kann, Ryan Cotterell, and Hinrich Schütze. 2017. One-shot neural cross-lingual transfer for paradigm completion. *arXiv preprint arXiv:1704.00052* (2017).

[12] Jihie Kim and Jaebong Yoo. 2012. Role of sentiment in message propagation: Reply vs. retweet behavior in political communication. In *Social Informatics (SocialInformatics), 2012 International Conference on*. IEEE, 131–136.

[13] Alexandre Klementiev, Ivan Titov, and Binod Bhattarai. 2012. Inducing crosslingual distributed representations of words. *Proceedings of COLING 2012* (2012), 1459–1474.

[14] Mark E Larsen, Tjeerd W Boonstra, Philip J Batterham, Bridianne OâĂŹDea, Cecile Paris, and Helen Christensen. 2015. We feel: mapping emotion on Twitter. *IEEE journal of biomedical and health informatics* 19, 4 (2015), 1246–1252.

[15] Kristina Lerman, Megha Arora, Luciano Gallegos, Ponnurangam Kumaraguru, and David Garcia. 2016. Emotions, Demographics and Sociability in Twitter Interactions.. In *ICWSM*. 201–210.

[16] Aditya Mogadala and Achim Rettinger. 2016. Bilingual word embeddings from parallel and non-parallel corpora for cross-language text classification. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 692–702.

[17] Saif Mohammad, Felipe Bravo-Marquez, Mohammad Salameh, and Svetlana Kiritchenko. 2018. Semeval-2018 task 1: Affect in tweets. In *Proceedings of The 12th International Workshop on Semantic Evaluation*. 1–17.

[18] Saif Mohammad and Svetlana Kiritchenko. 2018. Understanding Emotions: A Dataset of Tweets to Study Interactions between Affect Categories.. In *LREC*.

[19] Saif M Mohammad and Felipe Bravo-Marquez. 2017. Emotion intensities in tweets. *arXiv preprint arXiv:1708.03696* (2017).

[20] Saif M Mohammad and Felipe Bravo-Marquez. 2017. WASSA-2017 shared task on emotion intensity. *arXiv preprint arXiv:1708.03700* (2017).

[21] Tony Mullen and Nigel Collier. 2004. Sentiment analysis using support vector machines with diverse information sources. In *Proceedings of the 2004 conference on empirical methods in natural language processing*.

[22] Daniel Preoţiuc-Pietro, Vasileios Lampos, and Nikolaos Aletras. 2015. An analysis of the user occupational class through Twitter content. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, Vol. 1. 1754–1764.

[23] Daniel Preoţiuc-Pietro, Svitlana Volkova, Vasileios Lampos, Yoram Bachrach, and Nikolaos Aletras. 2015. Studying user income through language, behaviour and affect in social media. *PloS one* 10, 9 (2015), e0138717.

[24] Sebastian Ruder, Ivan Vulić, and Anders Søgaard. 2017. A survey of cross-lingual word embedding models. *arXiv preprint arXiv:1706.04902* (2017).

[25] Duyu Tang, Furu Wei, Nan Yang, Ming Zhou, Ting Liu, and Bing Qin. 2014. Learning sentiment-specific word embedding for twitter sentiment classification. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Vol. 1. 1555–1565.

[26] Ali Yadollahi, Ameneh Gholipour Shahraki, and Osmar R Zaiane. 2017. Current state of text sentiment analysis from opinion to emotion mining. *ACM Computing Surveys (CSUR)* 50, 2 (2017), 25.

[27] Shaozhi Ye and Felix Wu. 2013. Measuring message propagation and social influence on Twitter. com. *International Journal of Communication Networks and Distributed Systems* 11, 1 (2013), 59–76.

[28] Barret Zoph and Kevin Knight. 2016. Multi-source neural translation. *arXiv preprint arXiv:1601.00710* (2016).