6Q-05

Emotion recognition of Arabic tweets on Saudi Arabia social issue

Ghadeer Attar   **Toru Sugimoto**
Graduate school of electrical engineering and computer science,
Shibaura Institute of Technology

1.   INTRODUCTION

Language is a very powerful tool human has not only to communicate with each other but to understand each other's emotions as well.

People use different kinds of communication media to deliver their emotions, wither their facial expression, speech or text. Nowadays, the main focus is on textual communication, specifically the usage of social media, such as twitter and Facebook. By pressing a button, people can share their emotions and opinions regarding social issues, political issues and review. Moreover, with the increase of twitter users, shared emotions are increasing as well. Due to this increase, many researchers have been targeting emotions as the main focus of text mining research.

Text mining is a process of mining that derives important and useful data from large data [1]. By collecting the data, scientist, marketers and decision makers can benefit from analyzing it in order to conduct studies or to make a decision. For example, when a user writes a review about iPhone's camera and post it on twitter, the maker can know what the public's opinion is like. Moreover, they can invest in what their customers like and eliminate what they dislike.

In this study, the author conducted an experiment to recognize emotions in Arabic text in social media regarding a social issue.

2.   METHOD

In this section the data went through several steps, starting with collecting the data, annotation and creating emotion dictionaries, pre-processing and finally applying keyword spotting technique on the collected dataset. Collecting the data from Twitter using Twitter API and some certain keywords related to a social issue in this case the words were related to women driving in Saudi Arabia such as قيادة المرأه and لن تقودي which are translated to "women driving" and "you will not drive" sequentially is the first step.

Since the source of the data is Twitter that result to the data containing a lot of unwanted noise that we had to clean. The next step is annotating the data based on total of 7 different emotions: Happy, anger, fear, disgust, joy and surprise [2] in addition, no-emotion label was added. After annotating the dataset

manually by 2 Arabic speakers, keyword spotting technique was applied to recognize emotions form the text. Figure 1 shows the workflow of the process.
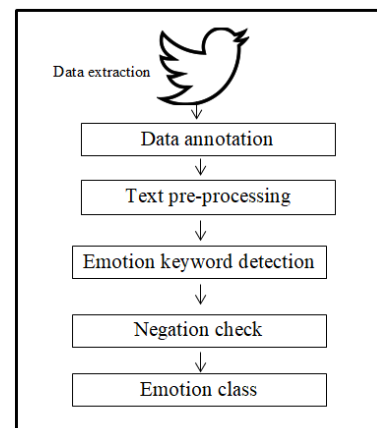

Figure 1: Visualizing the work

The data were collected from Twitter using Twitter API. By identifying a keyword which is about a social issue in Saudi Arabia of women driving cars "قيادة المرأة" and "لن تقودي". The amount of collected data is 1000 tweets which contain different emotions and opinions about the matter. The data then were divided into 2 sections; training set and testing set. The training set contained 300 tweets and testing set 700 tweets.

The data set has been annotated manually by 2 Arabic speakers. They were asked to annotate tweets to 6 basic emotions and a no-emotion label based on their knowledge and the daily usage of the language.
Since the keyword spotting technique depends on a keyword matching from emotion dictionaries. And the data is in Saudi dialect Arabic. In this study dictionaries were created by examining each tweet and depending on the annotation result the keyword was picked and assigned to its emotion dictionary. In this experiment 7 emotions dictionaries were created each contained keywords that represent the label of emotion.

2.1 PRE-PROCESSING

Pre-processing the collected data is an important step to apply before working on the recognition of the emotion [3]. The pre-processing applied on the data was tokenization.

The sentence was split depending on the white

space between words therefore each word was separated. Table 1 shows an example before and after tokenization.

Table 1: text data before and after tokenization

Data before processing	Data after processing
قيادة المرأه ل السياره من ضروريات الحياه وتحتاجها كما يحتاجها الشاب	قيادة المرأه ل السيارة من ضروريات الحياة وتحتاجها كما يحتاجها الشاب
Women driving is one of the necessities of life and she needs it same as the men needs	

2.2 EXPEREMENT AND RESULT

We created a keyword spotting technique system which is a keyword pattern matching and it is based on predefined keywords from the original data, these keywords were divided into different emotion labels. After Extracting and pre-processing the data, the result shows each tweet with its associated emotion based on the keyword matching. In addition, before assigning the emotion class the system will check for negation existence and handles it.

The data then inputted to the system where the keyword spotting would scan for a keyword matching within the emotion dictionaries. Follows, the system would check for the existence of negation word with the emotion keyword to handle the negation.

From the evaluation of the data from the 300 data set and after the experiment was conducted the major issue was negation errors.
Table 2 shows the result from the first trial

Table 2: The result of emotions from the experiment

happy	anger	surprised	disgust	fear	joy	no-emotion	error
20	78	14	27	31	11	110	27

From the result it shows that no-emotion class was the highest meaning that no emotion keywords were found in the context. After that was the anger class with 78 data count containing keywords such as نهب "looting", تافه "trivial" and حمقى "fools" which are keywords that indicates non acceptance of the situation. Fear was next with 31 data count with strong fear emotion words such as "be careful" وانتبهوا and "conspiracy" مؤامرة. Disgust contained keywords referred to "lying" كذب and "pathetic" مثير للشفقة. After disgust was happy which in general tweets that showed acceptance and positive feedback about the situation using keywords such as فرحان "happy", انا "I support" مؤيد and ممتاز "excellent".

The error shown in the experiment result was caused by negation existence. In result, the negation handling was added before assigning the emotion class. It was done based on multiple rules that are suitable for the data. The reason being that after grouping the negated data it showed that the major

case would result to the opposite emotion of happy which is anger. Even though the data contained different cases of negation we found it to be difficult to handle each and every case. It might not be ideal way however, it worked with the data. Therefore, not happy label was created which is a label that contains happy keywords with negation words. For instance, below is an example of negation + happy keyword = not happy.

لا تضحكون وكأنه امر بسيط هذه مو من عاداتنا والله
Don't laugh as if it is a simple matter I swear this is not our culture
هذا مو انجاز اللي قاعدين تسوونه ترا غيركم وصل القمر و انتم توكم في السواقه
What you all are doing is not an achievement people went to the moon and you all are still in the driving issue.

If the negation word was found, second scan was conducted in order to check for another emotion keyword within the same sentence to assign it to the emotion class. Lastly, the default case of negation handling was to scan for negated emotion keyword from both emotion dictionaries and negation list when found the emotion will be assigned to no-emotion.

3. CONCLUSON

To conclude, in this paper we recognize emotions from Arabic text. Data collection and annotation was the first step by collecting data from Twitter and then pre-processing the text so that it prepares the text for the next step which is emotion recognition by keyword spotting technique which will matched keywords existed in the data with different emotion dictionaries, handled to negation cases and then show the associated emotion.

4. REFERENCES

[1] R.SonyKrishna, L. Priyanka, K. VijayaLakshmi, M.Sowmya. A Text Mining Application of Emotion Classification of Twitter's Users , SSRG International Journal of Computer Trends and Technology (IJCTT)-Special issue-, pp.203-207,2017

[2] S. Dhawan, K. Singh, D. Sehrawat. Social Mining Techniques in Social Networking Sites, International Journal of Information & Computation Technology.vol.4, pp.1145-1153, 2014

[3] S. Xylina Mashal, K. Asnani. Emotion Analysis of Social Media Data Using Machine Learning Techniques. IOSR Journal of Computer Engineering (IOSR-JCE),pp.17-20, 2017