The Analysis of Emotion Mining from Text In Online Social Network

Pawade Abhijeet, Pawade Amit, Safa Madki, Namrata Doijode, Imran Pathan

Pawade Abhijeet

Solapur University,India.

tpocsevvpiet@gmail.com

Safa Madki

Solapur University,India.
safa.madki *@gmail.com*

Namrata Doijode

Solapur University,India
doijodenamrata.7@gmail.com

Abstract— **Online social networks are so popular nowadays that they are major component of an individual's social interaction. They are also emotionally-rich environments where close friends share their emotions, feelings and thoughts. In this paper new framework is proposed for characterizing emotional interaction in social networks, and then using these characteristics to distinguish friends from acquaintances. The goal is to extract the emotional content of the texts in online social network. this interest is whether the text is expression of the writer's emotion or not .For this purpose ,text mining technique are performed on comments retrieve from social network. The framework includes a model for data collection, database schemas, data processing and data mining steps. In general the paper presents a new perspective for studying friendship relations and emotion's expression in online social networks where it deals with the nature of these sites and nature of language used.**

Keywords— **Online social Network, emotion mining, text mining, friendship, emotions.**

I. INTRODUCTION

The growth in popularity of online social networks has significantly affected the way people interact with friends and acquaintances nowadays. Indeed, interacting through online social networks and online chatting systems become ubiquitous, and a major component of a person's life. In this paper , interest is in mining emotions from texts shared in online social networks in the form of wall posts and comments. The purpose is not to identify specific emotions but rather to tell if the text contains emotions or not; In other words, if the text is subjective reflecting the writer's affect and emotional state or if it is factual and objective where the writer does not express any feelings. An application of this approach is to predict relationship strength between two individuals based on the affective content of the comments they share online. In this case, the primary interest is whether the individual communicates his emotions in the text or not. The rest of the paper will be divided as follows: Section II presents the literature review pertaining to emotion mining from texts. Section III presents the proposed framework. The evaluation of the model is the subject of Section IV and the conclusion is included in Section V.

II. LITERATURE REVIEW

A. Emotions

Emotions are mental states accompanied by physiological changes. Ekman identified six basic emotions: happiness, sadness, anger, fear, disgust and surprise. Other approaches do not search to categorize emotions in specific categories, but rather identify them on two scales: the valence of the emotion indicating if the feeling is positive or negative and the arousal level indicating the energy level associated with the emotion Although emotions are universal, there are huge differences between cultures and between individuals in the way and the extent in which these emotions are expressed. In general, women are more likely to share their emotions and their feelings than men; and this observation was also verified in online social networks In the following section, techniques of sentiment mining will be presented.

B. Emotion Mining

Emotion mining can be divided into three categories depending on the purpose for mining emotions. The first category aims at extracting the valence of the text, indicating if the text has positive or negative emotions associated with it.

The second category aims at identifying whether the text is subjective or factual (i.e. objective), thus the purpose is to find if the text is emotionally rich or not. The third category aims at recognizing not just the emotion but also its strength or arousal.

C. Language In Online Social Network

Texts in online social networks have their specificity that must be taken into account. Indeed, it is common in these sites for users to use an informal and less structured language to communicate with their friends. some features of this "online language". They are presented here in addition to other interesting features:

• Intentional misspelling, in particular the repetition of a letter in the same word, (e.g. "helloooooo").
• Interjections and lexical surrogates for vocalizations (e.g. "mwah" indicating a kiss or "hmmmm").

• Grammatical markers such as the use of upper-case letters and the excessive use of punctuation (e.g. repetition).
• Social Acronyms: Acronyms of popular expressions used in online chatting systems and online social networks. For instance "BRB" denotes the expression "be right back".

• Emoticons: visual arrangements of characters in order to form facial expressions conveying emotions. For instance ":)" indicates joy and ":(" indicates sadness. In addition to this informal language, sentences may also lack proper syntax structure and words may be misspelled.

Furthermore, specific to the non-English users is the use of non Latin-based languages transliterated into the English Alphabet; the use of other languages such as French and Spanish is also common.

III. PROPOSED FRAMEWORK

In this section, we first present the general architecture shown in Figure 1 and then elaborate on each part.

A. Architecture

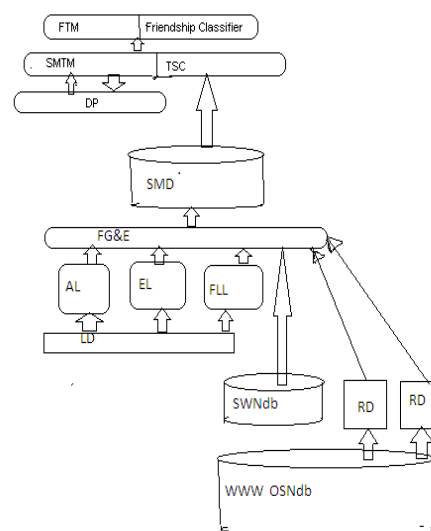The framework shown in Figure 1 is structured into six steps:

Fig. 1. Framework Architecture

Keywords:
1: Friendship Training Mode (FTM)
2: Friendship Classifiers (FC)
3: Sentiment Mining Training Models (SMTM)
4: Text Subjectivity Classifiers (TSC)
5: Data Pre-processing(DP)
6: Sentiment Mining Database (SMD)
7: Feature Generation And Extraction (FL&E)
8: Acronyms Lexicons (AL)
9: Emotions Lexicons (EL)
10: Foregin Language Lexicons (FLL)
11: Lexicon Development (LD)
12: SentiWord Net Database (SWN db)
13, 14: Raw Data(RD)
15: WWW Online Social Network Database (WWW OSN)

1. Raw data collection:

This step is concerned in collecting exchanged texts between users. This is done through a social network application which gathers data from the social network and stores them in a custom database.

2. Lexicons development:

This step deals with the informal language of online social networks. For this purpose, three types of lexicons have to be developed: lexicons for social acronyms, for emoticons and for expressions in foreign languages.

TABLE I
EMOTIONS

Emotions	Significance

:) ; :-) ; :> ; :]	Smiling
:(; :-(; :< ; :[	Sad
:* ; :-* ; :-x	Kissing
:P ; :-P	Joking

3. *Feature Generation:*

This step computes new features from available raw data collected in step 1 to assess subjectivity of text. It uses word-matches to existing affective lexicons and employs new lexicons developed in step 2 to handle social acronyms, emoticons and foreign languages transliterated into English. Accordingly, comments collected from step 1 along with the features computed in this step will be stored in the Sentiment Mining Database which is the database used for analysis.

4. *Data Pre-processing:*

This step is applied on extracted features. It involves removing redundant attributes, discretization by clustering and normalization using Min-Max.

5. *Crating a training model for text subjectivity:*

This step generates a model using k-means clustering algorithm with k=3 to categorize texts into three subjectivity levels: neutral, moderately subjective and subjective. The output of the model is the three centroids of the clusters.

6. *Text subjectivity classification:*

This step uses the centroids generated in the previous steps and employs the k-nearest neighbour algorithm with k=1 to classify all comments into one of three subjectivity levels.

7. *Friendship Classification:*

This step generates an SVM training model and then uses it to predict tie strength between online friends based solely on the subjectivity of the texts they share online. The classification is done by first classifying the subjectivity of the texts exchanged which is performed in step 6 and then taking an average measure. Online friends are classified in one of two classes: close friends and acquaintances.

IV. MODEL EVALUATION

This section presents the results and discusses the findings of the proposed method The evaluation of the model was done at two different levels. First, we tested how accurate the clustering algorithm was in recognizing three different classes of texts with respect to text subjectivity. Second, we tested the accuracy of the model in determining based solely on the subjectivity of the text, the relationship's strength between two Face book users.

A. *Text Subjectivity Classification*

This section tested the ability of the model to recognize three different classes of texts, based on the subjectivity of the writer. A sample of 850 comments were extracted from the database and categorized manually into the three classes by three different persons by majority vote. Table 7 shows three different comments, one from each class:

TABLE II
EXAMPLE OF COMMENTS

Comment Id	Comment	Class
1	caroooooooooooooooooooooooooo im going to kiiiilllll uuuuuuuuuuuuuuuuuu... n u know why! but i still looove u (a little bit:P) dont worry :P mwahhh	subjective
2	i love ur profile pic, its much better like this :) best	moderately subjective
3	86 and u how much did u get ?	Objective

B. *Friendship Classification*

Another way to evaluate the model was to test its results when it comes to using it for friends' classification. In other words, the model was used to predict relationship strength on Face book. Indeed, a way to evaluate friendship between two Face book users is to examine the comments shared on the site between these users. We expect that the relationship is stronger when the average subjectivity measure of the comments is greater.

V. CONCLUSION

This paper discusses a novel sentiment mining technique for texts in online social networks. It presents a new perspective for studying friendship relations and emotions' expression in online social networks where it deals with the specific nature of these sites and the nature of the language used. The purpose was to identify whether the writer conveys his emotions and feelings in his writings. The processed data was then used to identify tie strength between two persons based on the subjectivity of the texts they share online. The main challenge for the model proposed is the unstructured language of online social networks.

REFERENCES

- P. Ekman, "An argument for basic emotions," Cognition & Emotion, Vol. 6, No. 3, pp. 169-200, 1992.

- M. Thelwall, D. Wilkinson and S. Uppal. "Data mining emotion in social network communication: Gender differences in MySpace". In *Journal of the American Society for Information Science and Technology*, pp. 190-199, 2010.

- Esuli and F. Sebastiani, "SentiWordNet: a publicly available lexical resource for opinion mining," in *Proceedings of the 5th International Conference on Language Resources and Evaluation*, pp. 417–422, Genoa, Italy, May 2006

- Ma, H. Prendinger, and M. Ishizuka, "Emotion Estimation and Reasoning Based on Affective Textual Interaction," In *Proc. First Int'l Conf. Affective Computing and Intelligent Interaction* (ACII '05), pp. 622-628, 2005

- H. Liu, H. Lieberman, and T. Selker, "A Model of Textual Affect Sensing Using Real-World Knowledge," *Proc. Eighth Int'l Conf. Intelligent User Interfaces,* pp.369-374, 2003.