A study on the use of emoticon expressions on SNS and its relation with social trends

Kazuhiro Yamaguchi

School of Knowledge Science
Japan Advanced Institute of Science and Technology
March 2013

Keywords: Data Mining, Emoticon, SNS, Social Trend.

1 INTRODUCTION

1.1 back ground

SNS services, such as Twitter or Facebook became popular, so they enable us to daily communicate with many people at same time on Internet. We can understand about society with gathering widely the ideas what people think. We also can make ourselves more influential each other by using SNS. Because of this, it occurs many new trends and phenomena, which started from SNS. For example, Arab Spring in Arab states and a demonstration parade to stand against nuclear energy in Japan. Nowadays, it is impossible to see society without knowing what people are doing on SNS.

There are many reports about the relationship between SNS and society.

In the politics field, President Obama's Democratic National Convention speech sets new Twitter politics record with 52,757 tweets per minute. On the other hand, at the case of Romney, who run for Obama, the peek of tweets per minute rate was 14,299. This is very interesting, but there are few researches refering to the relationship between politics and SNS.

In the medical field, there is a report to estimate the number of influenza patients from tweets which include the " flu ". The tasks are done by following steps. First, the tweets were collected include the keyword such as " flu ". Then, decide whether the user is really in sick. After apply data to the infection model, finally the number was estimated.

In an economic field, there is a report to predict the box-office revenue from tweets. Another report said to predicted the Dow John's Industrial Average(DJIA) from the mood of twitter. In this report, the fluctuation of DJIA 3 days after was estimated using the data of DJIA and twitter with the precision 86%.

As we saw, it is actively reported about the investigation into the society using Big Data from SNS and Data Mining approach.

To understand the society, it is important to investigate the social trend. It is conceivable that people's emotion is a factor to affect to the society. The communication on SNS is text-base. So to extract emotional information from these texts, Natural Language Processing(NLP) methods was used.

But in SNS people use new words, such as emoticon, abbreviation, and some words we haven't seen until now. And it is very difficult to analyze these words with NLP. The methodology which is to designed analyze these words, is called Un-Natural Language Processing, and in particular, the method which analyze emoticon is now widely noticed.

There is a report shows that the emoticon affect the others emotion, and we can see emoticons appearing in tweet as the one's emotion.

Basically, the work to push tweet into Twitter is done by each person, who is a member of society. So we can assume that SNS and society are linked by human communication. If this is true, we have a question whether people affect each other when some event happens.

If SNS and society are linked by any mapping function, we can recognize society by investigating SNS. It means that we can use SNS as some index instead of understanding society doing hard work such as a census of the population, consumer price index, and an opinion survey.

In this reason, it is meaningful to investigate the relationship between SNS and society. This research field is very new, so we find interesting knowledge from many reports.

1.2 purpose

So in this study, I have two purpose. First, I try to analyze emotion using only emoticon. Second, I investigate the relationship between emotion on SNS and social trend.

2 METHOD

2.1 sentimental analysis

I extracted emoticons from tweets using regular expression, and I could distinguish whether tweet has emoticon in accuracy of 87%. After dividing a emoticon into morphemes, such as eyes and a mouth, then presume feeling from a morpheme. I created a single dimension model and a multi dimension model. I estimated emotion from morphemes in each model using decision tree analysis.

2.2 result

As a result, in case of single dimension model, I found that it is very difficult to get a good accuracy with only emoticon. In case of multi dimension model, I estimated feeling score with using database, which is made by questionnaire method. That database has data set of emoticon and emotional scores, such as laugh, cry, angry and so on. I estimated feeling score from emoticon using decision tree analysis and these data set as the supervised data. As a result, I found that ”glad”, ”grieving”, ”pleasure”, ”getting impatient”, and ”surprised” feeling between ordinal emotion score and estimated feeling score have a strong Peason’s correlation . And the power of test was also enough.

However, there were few kinds of emoticons used as training data, so there is a problem that I couldn’t analyze all of emoticons. Actually, I couldn’t estimate feeling score from 15 emoticons of 30 high-frequency of appearance.

3 INVESTIGATE BETWEEN SNS AND SOCIAL TREND

3.1 data

Using tweet data collected by Twitter Streaming API, I expressed the feeling of the emoticon by the multi-dimension model. And by totaling, it was regarded as the amount of feeling and I analyzed about the relationship between SNS and social trend.

Through this analysis, it became clear that the ratio of the count of tweets which include an emoticon to all tweets was stable at about 21%. So the analysis for an emoticon is stabilized.

3.2 sentimental analysis using multi dimention model

The feeling score of the emoticon was computed by the multi-dimension model, and analysis was conducted from the viewpoint of the direction of a time series.

In comparison of a weekday and weekend, public holiday, it became clear that transition of the amount of feeling expressed has a difference.

Moreover, it was seen that the maximum of the amount of feeling expressed as Christmas rather than non-event time. And after 23o'clock, I could see the downward tendency of the amount of feeling on non-Christmas, but on Christmas, it was not clear.

In the earthquake occurrence day, the rapid increase in the amount of feeling was checked immediately after the earthquake occurrence.

In comparison of distributions of all tweets count par minutes and tweets which include emoticons count par minutes, it became clear following. The amount of tweets which include emoticons relatively increase than that of all tweets around 8 A.M.

In the analysis calcurating the ratio of "fun" and "sad" using estimated feeling score, it became clear that many people feel relatively fun in week end and holidays more than week days. And in the earthquake occurrence day, it showed that many people feel relatively sad soon after the earthquake occurred.

4 CONCLUSION

To investigate the relationship between SNS and society, I analyzed using Twitter data and created two models. In case of single dimention model, I found that this is not suitable if I use only emoticon. In case of multi dimention model, I got good estimation values. But there is a problem that is a few emoticon I can estimate feeling score.

In a analysis on time series, in comparison of weekday and weekend, public holiday, it became clear that transaction of the amount of feeling expression has difference.

In comparison of Christmas day and non-Christmas day, there are difference on the maximun of the amount of feeling expression.

In the days when earthquake occurred, there was the rapid increase in the amount of feeling after the earthquake occurred. And the ratio of the amount of people who feel fun to that of sad showed a special feature. This is that an adrupt event in society affects to the emotion on SNS.

The method I created has some problems, but in analyses, it could show some special features. So I concluded that the analysis using emoticon on SNS is useful for understanding society.