3rd National Conference on Innovative Research Trends in Computer Science and Technology (NCIRCST 2018) ISSN: 2454-4248
Volume: 4 Issue: 103 – 105

Improving The Classification Of Social Tension Using Sentiment Analysis

A.Raeshman Nisha
UG Student
Department of CSE
Kongu Engineering College
Erode,Tamilnadu
+919994634487
raeshlebesten@gmail.com

S.Sharmiya
UG Student
Department of CSE
Kongu Engineering College
Erode,Tamilnadu
+919840064378
shar.sharmy@gmail.com

P.B.Pushpanath
UG Student
Department of CSE
Kongu Engineering College
Erode,Tamilnadu
+919952404389
pushpanathyogesh@gmail.com

D.Deepa
Assistant Professor
Department of CSE
Kongu Engineering College
Erode,Tamilnadu
+919952784664
deepabruce@gmail.com

***Abstract*:-**Social networks provide a powerful reflection of the happenings in the modern society. Using social networks an approach is made to measure social tension which analyses the dissatisfaction of users with the existing situation. The content analysis from the social network determines the changing moods in different social groups. This project proposes an approach to measure social tension in India based on the analysis of user's post. We developed a work that uses Twitter as the social network addressing on the popular social issues such as Politics, unemployment and GST in India. We applied sentiment analysis technique to improve the evaluation of social tension.

Keywords: social network; social tension; sentiment analysis; twitter

*****

1. INTRODUCTION:

The growing phenomena of social media, such as: Facebook, Twitter, LinkedIn, and Instagram, with each one has its own characteristics and its usages, are constantly reflecting the happenings in the societies. Twitter is a massive social networking site tuned towards fast communication. More than 140 million active users publish over 400 million 140-character "Tweets" every day. Twitter's speed and ease of publication have made it an important communication medium for people from all walks of life. Twitter has played a prominent role in socio-political issues. The massive information provided by twitter such as tweet messages, user profile information, and the number of followers/ followings in the network play a significant role in data analysis, which in return make most studies investigate and examine various analysis techniques to grasp the recent used technologies. In this paper we attempted to assess the level of social tension various regions using the content analysis of user's post in twitter.

2. BACKGROUND:

The concept of "social tension" expresses a certain state of social groups, their dissatisfaction with the existing situation, the attitude towards what is happening and to other groups. Analysis of social tension helps to evaluate the level of social disorder; it is one or more forms of disturbance caused by a group of people. Social disturbance is typically a symptom of, and a form of protest against, major socio-political problems; the severity of the action coincides with public expression(s) of displeasure. The social disorder includes illegal parades; riots; and other forms of crime. These activities demonstrate the dissatisfaction of public to the government.

Social dissatisfaction is the dominant factor in determining the level of social tension; therefore, the assessment of social tension is usually based on determining the measure of respondents' satisfaction with the various aspects of their life activity and the conditions of the external social environment. This provision is based on the basic principles of the theory of deprivation, according to which, the increase in dissatisfaction is accompanied by an increase in preparedness for action, including destructive.

Directly, the social tension problem is investigated in the work of V.A. Bykovsky, who examines statistical indicators: the monetary incomes of the population per / person); provision of housing for 1 person (square meters); total % of provision of residential premises with hot water and sanitation; population (thousand people capita per month in thousand rubles, the subsistence minimum; coefficient of the multiplicity of income to the subsistence level; budgetary provision per capita (thousand rubles); the number of registered crimes (per 1,000 people); population density; number of unemployed per 1000 people; expenditure on health, thousand rubles per one inhabitant; expenditure on education, thousand rubles per one inhabitant; expenditure on culture (thousand rubles per capita) .

Both in Russia and abroad, research is conducted to identify the hidden patterns of development of social processes and factors that determine the change in moods in different social groups. For example, solution of the problem of analyzing social tension reported in social media is presented in the article "Social Tension Detection and Intention Recognition Using Natural Language Semantic Analysis (on the material of Russian-speaking social networks and web forums)". The authors developed a technique based on Natural Language Semantic Analysis, which was used to analyze social tension in reality and in networks in a certain period of time

In "Analysis of Comments of Users of Social Networks to Assess the Level of Social Tension" [1], the authors have researched on the changing moods in different social groups and proposed an approach to measure social tension in certain regions of Russia based on the analysis of user's posts in the social network Vkontakte. They developed a program that obtains posts reflecting negative attitudes of VK users on popular social issues such as level of unemployment, corruption crimes, and inflation. This program is used to collect statistics about the number of such posts in certain periods and to analyze common information about users who post them. The authors analysed the following patterns after analysis of posts;

- The frequency of discussions of topics may vary greatly depending on the time of year, as well as on global Russian news, which are relevant not only for regions, but also widely discussed throughout the country. This correlates with the growth of protest moods in Russia, which led to the protest rallies in the second half of the March
- Analysis of cities, which account for the largest number of posts with a negative tonality for selected topics, showed differences with the considered rating of social tension. So, the greatest number of discussions falls on Moscow and St. Petersburg. However, these cities are quite prosperous in terms of the standard of living of the population in comparison with the regions listed in the rating.
- This difference can be due to the large spread of social networks among the population of large cities as well as greater social activity of citizens.
- Age and sex distribution of the users' activity is also monitored. The analysis showed that the age category from 17 to 30 years is the most labile in terms of potential protest activity. Men users are much more active in discussion of topics related to social tension problem regardless of the age group.

3. METHODS

3.1. Choosing Twitter for research

Twitter is an online news and social networking service where users post and interact with messages, known as "tweets". These messages are restricted to 140 characters which is useful for data analysis and text classification. Twitter allows the users to extract tweets easily without having an account. Twitter usage statistics-24% of all Internet male users use Twitter, whereas 21% of All Internet Female users use Twitter.

3.2. Data Set Collection and Pre-processing

The dataset contains tweets which were collected from twitter homepage using hash tags. For example when #GST is given all tweets related to GST and most recently posted posts are obtained. Thus the required number of tweets can be extracted according to the user needs. These datasets act as training datasets. Data is collected in the following three issues: Politics in India, Unemployment in India, GST in India. In pre-processing work Stop word removal, Stemming, Synonym word replacement and Abbreviation word replacement process are done.

3.2.1. Stop word removal

Stop words usually refers to the most common words in a language, there is no single universal list of stop words used by all natural language processing tools. The words related to Politics, Unemployment, GST are collected and the unwanted words are removed and replaced in the posts.

3.2.2. Stemming

In this process all the continuous verbs are removed and the actual verbs are considered from the dataset. For example, "act" is alone taken from the word "acting".

3.2.3. Synonym word replacement

A file containing a list of synonyms is prepared. For some complex words in the dataset, its relevant synonym is replaced from the file.

3.2.4. Abbreviation word replacement

In this process all the acronyms in the dataset are replaced with its abbreviation for easy understanding and analysis. For example, "GST" is replaced with the abbreviation "Goods and Services Taxes".

3.3. Sentiment File Preparation

In this process a sentiment file is prepared which contains a list of words and its relevant scores. As there are numerous number of words, scores cannot be assigned manually. Thus an online tool named "AFINN" is used. AFINN is an dictionary which contains sentiment lexicon with 2477 English words (including a few phrases) each labeled with a sentiment strength and targeted towards sentiment analysis on short text as one finds in social media. This tool can be used to analyse the polarity and for sentiment analysis.

3.4. Positive Words and Negative Words Processing

The Sentiment Analysis work is done where positive and negative word count is calculated from the sentiment value of each word. As the word count is computed for politics, unemployment and GST related posts. Once the word count is known, the total score for each tweet is analysed. With the help of total score the number of positive posts and negative posts are computed. Then the Positive/Negative Percent Calculation is performed for each issue.

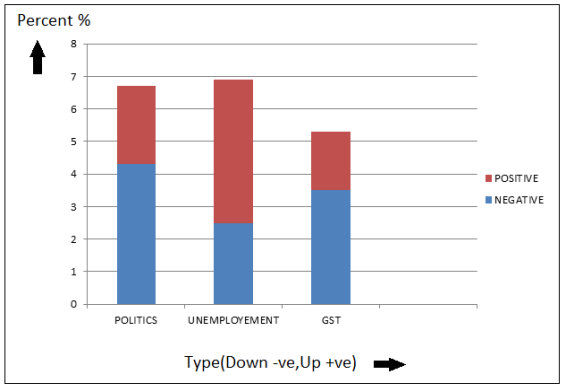
3.5. New Data Set Collection

A large number of data is collected from the twitter as new dataset. This new dataset acts as the test data. Then SVM classification is performed where the post that contains 'n'

machine learning, support vector machines (SVMs, also support vector networks) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. SVM(Support Vector Machine) are best for text classification.

Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall. Then Sentiment Analysis is performed for the newly collected dataset.

The below figure shows the results of the analysis.



4. CONCLUSION & FUTURE WORK

We applied sentiment analysis for analysing the polarity of the tweets and support vector machine for text classification for the posts collected from twitter. Analysing the user's posts shows that people's mood change frequently and most of the population are against the issues: politics, unemployment, GST.

Planned future work includes the following statements:

- Applying age classification to social tension analysis
- More detailed analysis on measuring the social tension level and its issues
- Applying this technique to more number of issues

5. REFERENCES

[1] Dmitry Donchenko, Nadezhda Ovchar, Natalia Sadovnikova, Danila Parygin, Olga Shabalina, Danish Ather (2017) "Analysis of Comments of Users of Social Networks to Assess the Level of Social Tension" 6th International Young Scientists Conference in HPC and Simulation.

[2] Donchenko, D., Sadovnikova, N., Parygin, D., Shabalina, O. (2016) "Promoting urban projects through social networks using analysis of users influence in social graph." Information Technologies in Science, Management, Social Sphere and Medicine, Atlantis Press, Tomsk: 162–165.

[3] Rukavishnikov, V. (1992) "Social tension: diagnosis and prognosis." Sociological Research 3: 3–23.

[4] Baranova, G., Alekhin, E. (2007) "Formation of social tension in the regions of the Russian federation. Methods of analysis and prediction of social tension." Vestnik VSU 2: 149–154.

[5] Ustugova, S., Parygin, D., Sadovnikova, N., Finogeev, A., Kizim, A. (2016) "Monitoring of social reactions to support decision making on issues of urban territory management." 5th International Young Scientist Conference on Computational Science, Elsevier BV, Krakow: 243– 252.