International Journal For Technological Research In Engineering
Volume 8, Issue 7, March-2021 *(Online): 2347 - 4718*

TOPIC BASED SENTIMENT ANALYTICS ON CHINESE PRODUCT BOYCOTT FROM TWITTER

Deepak Mane
Senior Data Scientist, Analytics and Insights, Tata consultancy Services
Dr. Sirbi Kotrappa
Senior Professor, Computer Science Department, KLE's College of Engineering & Technology
Kiran Shibe
Solution Architect Analytics and Insights, Tata consultancy Services

Abstract: - In May 2020, people started movement against china products after Galwan's border crisis. Twitter is become popular social media platform which is used by millions of peoples across worlds to share their feeling and opinions about Chinese products. Topic based sentiment analysis will help to process data and automatically analyze data. This paper will contribute on Topic based sentiment analysis on Tweets related Chinese product boycott. In this paper, we conduct topic modeling based on subject, theme, rank etc. This methods allows to describe different methods for topic based sentiment analysis

1. INTRODUCTION

A Trade is a important part between two countries . It's the most instrument of trade where the traders from India and China send trades from India to china or vise-versa. This helps to satisfying requirements of countries, helps each growing economy. In May 2020, There was Galwan boundary issue happened between India and China, then There was movement started to boycott Chinese product in India. Twitter is online available platform where peoples started to post their opinions, images and videos about Chinese product boycott. Twitter's flow of data may allow monitoring events which are occurring and understand people's feelings. In order to automatically process Twitter data, several data analysis methods, such as sentiment analysis and topic modeling, can be applied. The outcomes of these analyses may be used by several applications, such as event monitoring, and opinion mining about products or brands. Indeed, companies always need fast and accurate information in order to be able to react the market trends. This paper proposes a topic based sentiment analysis method: First, topics are extracted from the training dataset. Then we can develop sentiment based algorithm based on different topic themes extracted hen this paper will provide topic based sentiment analysis based on Ranking, Topic Subject, Topic Theme and Positive/Neutral and Negative based)

2. POSSIBLE IMPACTS BY BOYCOTTING CHINESE PRODUCT

- *Trade Impact*

As per WTC's Bilateral trade agreement form between India and China. After Galwan's Incident, India put ban 53 Chinese Mobile application and stop all investment directly from china. This kind of decisions will be create negative impact on China and India for long term. China's president is very close friend of India PM but it will impact on Chinese trade to India

- *Pharmacy Industry Impact*

In india , 60% of Pharmacy industry exports raw material from China because of low price , low transportation cost , Due to Chinese product boycott movement , Pharmacy industry need to exports all raw material from Europe which will cost higher. In 2020 , Make-in-India policy , Raw material export decreased by 20% . It will be good sign for Pharmacy industry

- *Chinese Investment*

Indian Government cancelled all tenders where Chinese industry had involvement in various government project like metro rail project, BSNL 4G network. It was big loss for Chinese industry

3. LITERATURE REVIEW

In research done by Yang et al (2018) on paper entitled "A survey on Sentiment Analysis by using Machine Learning methods" mainly describes the popular Sentiment Analysis Technique based on topic modeling such as : SVM, NB, ME, ANN method. They evaluate the performance on the basis of Accuracy, Precision, Recall and F1 Score metrics. SA analysis methods are in English language. There is a lot of problem if we apply this method to other languages, such as no open source material is available, only paid version is present which is very costly. Some classifier provides high success rate but accuracy percentage is not been achieved.

In research done by Aloui et al (2018) on paper entitled "A novel adaptable approach for sentiment analysis on big social data" provides an adaptable sentiment analysis approach is proposed. Dynamic dictionary of word is constructed here. Classification of tweets is on the basis of 6 different parameters: highly +ve, moderately +ve, lightly +ve, highly –ve, moderately –ve & lightly –ve

In research done by Shaya et al (2018) on paper entitled "Sentiment Analysis of Big Data: Methods, Applications,

and Open Challenges" presents a systematic comprehensive literature survey that discuss both technical (based on the techniques used) and non-technical sentiment analysis based on topic

In research done by Mohey et al (2016) on paper entitled "Survey on sentiment analysis challenges", mainly presents the SA challenges of different classifiers. Challenges are Negation Handling, Feature extraction, NLP overheads. Second comparison relies on Accuracy rate. Here, Comparison is made on the basis of latest techniques used to analysis Sentiment Analysis.

In research done by Jianqiang et al (2017) under the paper "Comparison Research on Text Preprocessing Methods on text Pre-Processing Methods on Twitter Sentiment Analysis", they are mainly focusing on text pre-processing method of data. Six pre-processing methods they apply: Replacing negative mentions, removing url's link, Reverting words that contain repeated letters, Removing numbers, Removing stop words and Expanding acronyms. T

4. RESEARCH METHODOLOGY

We have created a accounts for each sources and access data from respective sources. For Example We have created a twitter account and to access the tweets, we got the credentials in form of consumer key, consumer secret, access token, access token secret.

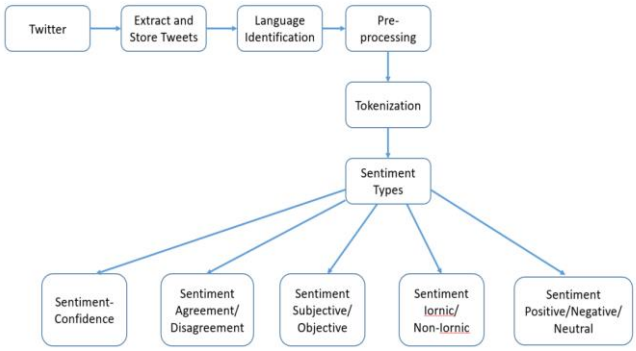


Figure 1 Research Methodology Approach

In topic bases sentiment analysis process , We developed sentiment analysis function which covers following information

- Language identification
- Topic type
- Topic – Ranking
- Sentiment analysis based on Topic Subject
- Sentiment analysis based on Topic Theme
- Sentiment analysis based topic (Positive/Negative/Neutral)

5. DETAILED ANALYSIS

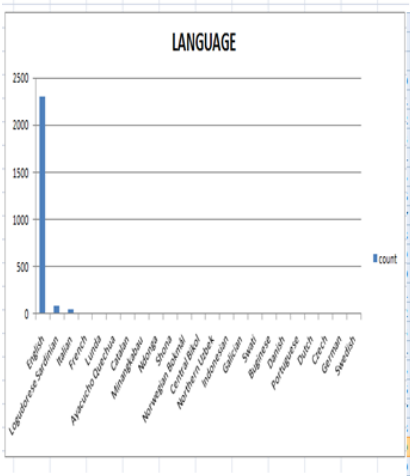
5.1. Feature Extraction

Performing feature extraction on Twitter messages raises new challenges:

- Short messages: Sentiment analysis is usually performed on longer text. Because of the text limitations, Twitter messages are short, and the algorithm has fewer features available for analysis.
- Internet language: Twitter users adopt the "internet language" when writing their messages. This language differs from the traditional English: new words, repeated letters emoticons
- Twitter characteristics: Twitter allows users to add three specific entities to their messages: hashtags, user references, and URLs. These entities require to be processed differently than common words.

5.2. Language Identification

In given dataset it contains different types of languages such as English, French Dutch, to identify type of language we used Lang detect is a re-implementation of Google's language-detection library from Java to Python. Which provides highest confidence factor? All detailed analysis given below (In Graph and Table format)

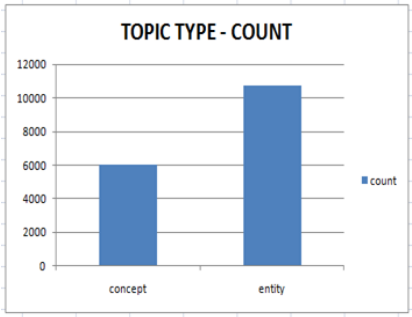


Laungauge Name	count
English	2302
Logudorese Sardinia	86
Italian	49
French	13
Lunda	2
Ayacucho Quechua	8
Catalan	8
Minangkabau	2
Ndonga	2
Shona	7
Norwegian Bokmål	3
Central Bikol	3
Northern Uzbek	4
Indonesian	1
Galician	1
Swati	6
Buginese	5
Danish	4
Portuguese	2
Dutch	8
Czech	2
German	8
Swedish	6

Figure and Table: 2 Language Identification

5.3. Topic Type - count

In this experiment we done sentiment analysis based Topic type (Concept and Entity).

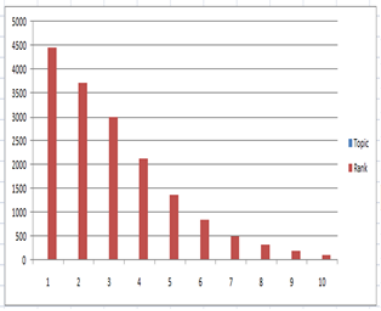


Topic-type	count
concept	6022
entity	10759

Figure and Table : 3 Topic Type –count

5.4. Rank – category

In this experiment we calculated ranking based on topic counts.

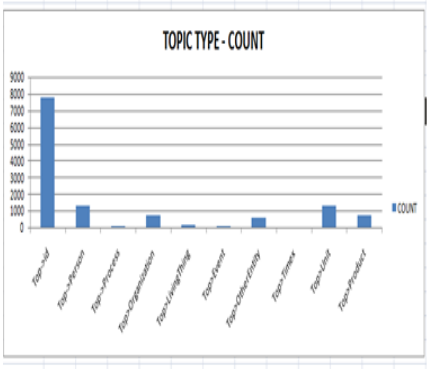


Topic	Rank
1	4462
2	3719
3	3004
4	2134
5	1372
6	845
7	503
8	318
9	200
10	118

Figure and Table: 4 Rank category

5.5. Topic – Subject

In this experiment, we completed topic sentiment analysis based on subject, we found that topic based identify posted high number of requests

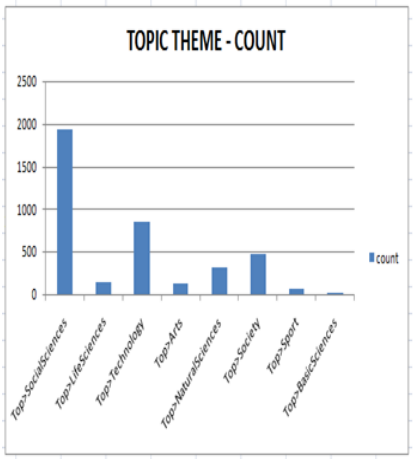


Topic	COUNT
Top->id	7753
Top->Person	1294
Top->Process	62
Top>Organization	739
Top>LivingThing	131
Top>Event	70
Top>OtherEntity	589
Top>Timex	8
Top>Unit	1336
Top>Product	701

Figure and Table: 5 Topics - Subject

5.5. Topic – Theme wise

In this experiment, we found that, Peoples are posted their views under topic category – social sciences.



Topic Theme	count
Top>SocialSciences	1937
Top>LifeSciences	138
Top>Technology	853
Top>Arts	121
Top>NaturalSciences	311
Top>Society	474
Top>Sport	52
Top>BasicSciences	11

Figure and Table: 6 Theme wise

5.6. Sentiment – Positive/Negative/Neutral

In detailed analysis, we classified sentiment into 5 types, Positive+, Positive, Neutral, Negative and Negative+

Step (1): In this steps , we are classifying words into 5 different groups , positive+(P+) , positive , Neutral , Negative and Negative+ based on the frequency and rank and polarity of each tweets

Step (2): after selecting each feature of tweet, the polarity of tweet determined using following formula

$$pol(T) = \sum_{i=1}^{k} pol(m_i) \qquad T=\{m_1 \ldots m_k\}$$

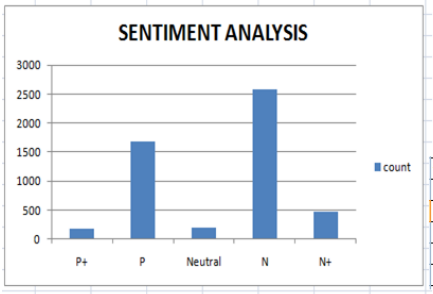
Tweet — microphrase

$$pol(m_i) = \sum_{j=1}^{n} score(t_j) \qquad M_i=\{t_1 \ldots t_n\}$$

term

•Step (3): We established following rules based on tweet's feeling.

Sentiment Type	Start	End
P+	0.6	1
P	0	0.59
Neutral	0	0
N	0	-0.59
N+	-0.6	-1



type	count
P+	183
P	1684
Neutral	197
N	2584
N+	481

Figure and Table : 7 Sentiment analysis – Positive/Negative/Neutral

5.7. Tag Cloud

As a part of detailed analysis, we need to understand popular keywords based on posts. It can be interpreted that by showing tag cloud of "Chinese Product Boycott" from different posts

Figure 8 - Tag cloud

6. CONCLUSION AND FUTURE SCOPE

This paper's first contribution to the sentiment analysis research field is to detail the preprocessing steps which have to be applied to extract bags of words from Twitter data. The second contribution is to propose a topic based sentiment analysis approach. This approach relies on the fact that the complexity of an analysis can be reduced when the algorithm focus on a smaller range of topic. The algorithms have to deal with less vocabulary and its estimation can be more accurate. Hence, the experiments have shown that, applying this method performs better than the classical sentiment analysis model. The proposed approach can be extended to other sentiment analysis problems. The method can be improved. The paper has focused on exploiting the results of the default parameter for the topic modeling method. More research on the topic extraction may allow having more distinct topics; hence the estimation should be more accurate. Moreover, the method proposed here cannot handle neutral data. In order to apply sentiment analysis on real application cases, neutral data need to be handled. Therefore, future works should focus on these neutral data. Two methods may allow detecting neutrality

REFERENCES

1] Gupta A., S. Joshi, P. Gadgul, and A. Kadam (2014). Comparative Study of classification algorithm used in sentiment analysis. International Journal of Computer Science and Information Technologies. 5(5), 6261-6264

[2] Jadav B., V. Kumar, and B. Vaghela (July 2016). Sentiment Analysis using Support Vector Machine based on feature Selection and Semantic Analysis. International Journal of Computer Application. 146-No. 13, 26-30

[3] Mohey D., and M. Hussain (2016). A survey on sentiment analysis challenges. Journal of King Saud University- Engineering Science. 1-9

[4] M. Trupthi, S. Pabboju, and S. Narasimha (2017) Sentiment Analysis of twitter Data using Streaming API. IEEE 7th International Advance Computing Conference. 915-919

[5] Zha J. , and G. Xiaolin (2017). Comparison Research on Text Pre-processing Methods on text Pre-Processing Methods on Twitter Sentiment Analysis. IEEE, Translations and Content Mining. vol 5,2870-2879

[6] Yang P. , and Y. Chen (2018). A survey on Sentiment Analysis by using Machine Learning methods. IEEE 2nd Information Technology, Networking Electronic and Automation Control Conference ITNEC. 117-121

[7] Aloui I. , Y. Gahi, R. Messoussi, Y. Chabbi, and A. Alexi (2018). A novel adaptable approach for sentiment analysis on big social data. Journal of Big Data, Springer Open. 3-18

[8] Shaya S., N. Ismawati, S. Bahri, A. Sulaiman, P. Seuk Wai, and Y. Wai Chung (2018). Sentiment Analysis of Big Data: Methods, Applications, and Open Challenges. IEEE Access, Magazine and Journal. 6, 37807-37827.

[9] Park Won C. , and D. R. Seo (2018).Sentiment Analysis of Twitter Corpus Related to Artificial Intelligence Assistants. 5th International Conference on Industrial Engineering and application.495-498

[10] Jianqiang Z., G. Xiaolin, and Xuejun Z (2018).Deep Convolution Neural network for Twitter Sentiment analysis. IEEE, Translation and Content Mining.23253-23259

