International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-6, April 2020

Web Interaction based Business Sentiment Analysis using Benchmark Classifier

Syed Salim, Dr. Madhu B K

Abstract*: Success of the business activities may showcase positive progress or developments of the country. These successes involve huge efforts with respect to analysis of the market, target customers, environment, availability of the resources and many more. Analysis would incur both the time and initial investment. Hence, most of the new business initiatives usually drop at the beginning stage. The proposed paper is trying to provide a solution to these cases by involving technological aspects and data from social media's. Social media has provided a platform for the internet users to share their views and opinions on some of the components. These could be the resources to generate an analytical report for new or existing business strategies.

***Keywords*: Sentiment Analysis, Machine Learning, Classification, Natural Language Processing**

I. INTRODUCTION

Business is the process of dealing with buying and selling of certain products based on the requirements or needs. Recent studies have shown that the online business may reach $20 billion by year 2020 [1]. Others studies have given that Indian online market may reach upto $200 billion by 2026 [2]. These developments are encouraging the budding entrepreneurs to think about new ideas / innovations towards innovative products. Numbers of government and private agencies are supporting with wide variety of opportunities for the budding entrepreneurs to look at new startups [3]. The success of e-commerce not only depends on the creative or innovative products but also relies on the need or requirement of the product. Data analysis exercises may bring certain useful conclusions to budding entrepreneurs to study and enter into the market for the success. This paper is proposing a Natural Language Processing and Machine Learning based methodology for drawing certain sentiments or moods of the online users. These sentiments may play a very important role in controlling the business strategies.

Revised Manuscript Received on April 30, 2020.
***** Correspondence Author
Syed Salim*, Research Scholar, Department of Computer Science & Engineering, Vidya Vikas Institute of Engineering & Technology, Mysuru, Visvesvaraya Technological University, Belagavi, India, Email: syed.vviet@gmail.com
Dr. Madhu B K, Professor and Head, Department of Computer Science & Engineering, Vidya Vikas Institute of Engineering & Technology, Mysuru, Visvesvaraya Technological University, Belagavi, India, Email:prof.madhubk@gmail.com

II. LITERATURE SURVEY

A. Business strategies in Social medias

Author [4] tried to design feature extraction frameworks for twitter posts. Here the author has tried to improve the performance of the classification and text normalization techniques on twitter posts. The highest efficiency of 95.7% is recorded and it is increased with the difference of 18.7% over Bag of Words. Author[5] has tried to show that the usage of twitter as a platform for business knowledge sharing purpose nearby South African places namely Western Cape. Author [6] tried to understand the positive or negative sentiments of the consumers on a particular food product by using clustering and classification techniques. Author [7] has tried to extract certain useful business information from social networking applications namely Google++, Facebook and Twitter by using Crawler and IE Techniques.

All these surveys indicate that the number of efforts taking place towards business activity analysis using social networking platforms. Here they tried to employ the current on-going buzz words clustering and classification techniques.

B. Sentiment Analysis

Sentiment is a kind of opinion, exhibited by the customer during interactions and it is used to determine applicability of the content [8]. Author [9] has tried to survey all the information available in the literature for the various developments in data analytics corresponding to Sentiment analysis in Tourism.

III. METHODOLOGY

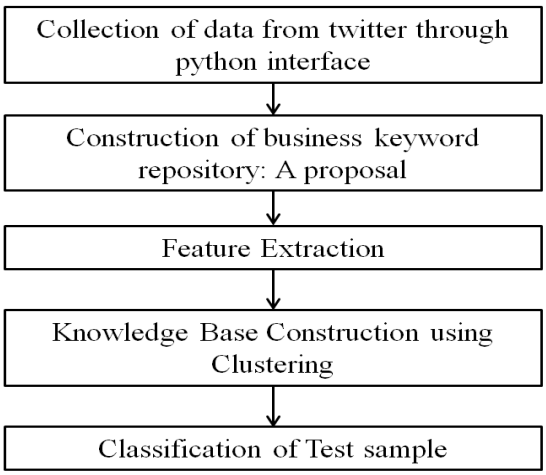

Fig. 1.Methodology of the proposed research work

Figure 1 depicts the methodology of the proposed paper. This methodology mainly focuses on the extraction of the features on data collected from twitter. Knowledge base construction using K-means and then classification using k-NN classifier.

A. Collection of Data from Twitter

Python has several packages, namely tweepy. This package is useful for accessing tweets from twitter for research purposes [10].

B. Construction of business keyword repository

The proposed paper is trying to analyse the sentiment of the tweet towards business that is buying or selling of any product. The estimation of type of sentiment, require the type of business sentiment. Table I shows certain samples identified for business sentiment keyword as a repository.

Table-I : Business Sentiment Keyword Repository

SN	Business sentiment keywords
1	Not happy, Not interested, Bad, useless, not applicable…….
2	Applicable, good, advantageous, good, acceptable, effective, Cost effective, less cost, affordable …

C. Feature Extraction and Selection

Clustering and classification models require the samples to be in the form of features. The collected tweets from twitter through tweepy are converted into samples by estimating the following features.

Word2vec [11]: Word2vec is able to calculate the context words through word embedding technique. The business sentiment keywords identified in the previous step are used here to design the word embedding layer.

Word2vec functions using the architectures shown in figure Fig 3. These are called as CBOW (Continuous Bag of Words) and Skip-gram models.

CBOW estimates the keyword using the context keywords provided to the model. Skip-gram model follows the inverse operation of CBOW model that is, it tries to identify context keywords using the given target keyword.

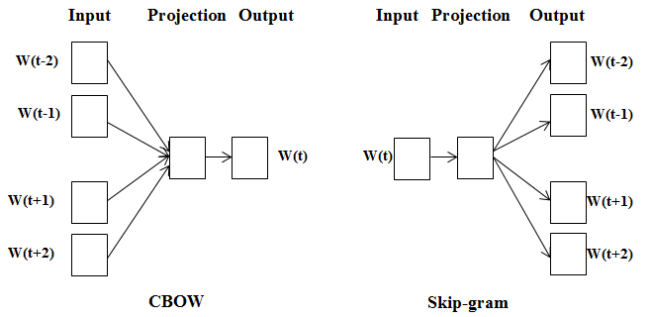

Fig. 2.Architecture of CBOW and Skip-gram

The probability of the context semantics corresponding to the given keyword is estimated using softmax evaluation.

i. Positive Sentiment Quotient (PSQ): softmax evaluation of all the keywords mentioned in the Table I and row 1 are estimated corresponding to given tweet. Principal Component Analysis [12] is applied to combine all these softmax evaluations and results PSQ.

ii. Negative Sentiment Quotient (NSQ): Similar steps are followed by considered row 2 of Table I. PCA is applied on these softmax evaluations to estimate NSQ of given tweet.

D. Knowledge Base Construction using Clustering or Classification model

K-means is a well-known clustering technique employed for clustering the given unsupervised samples [13]. This paper proposes, creating four clusters based on the assumption that an opinion may contain the percentage of both positive and negative shades of statements .

- Confused Little: Less Positive (LP) and Less Negative (LN) sentiments
- Clear Negative: Less Positive (LP) and More Negative (MN) sentiments
- Clear Positive: More Positive (MP) and Less Negative (LN) sentiments
- Confused More: More Positive (MP) and More Positive (MP) sentiments

Figure 3 shows the graphical representation of the proposed clustering model.

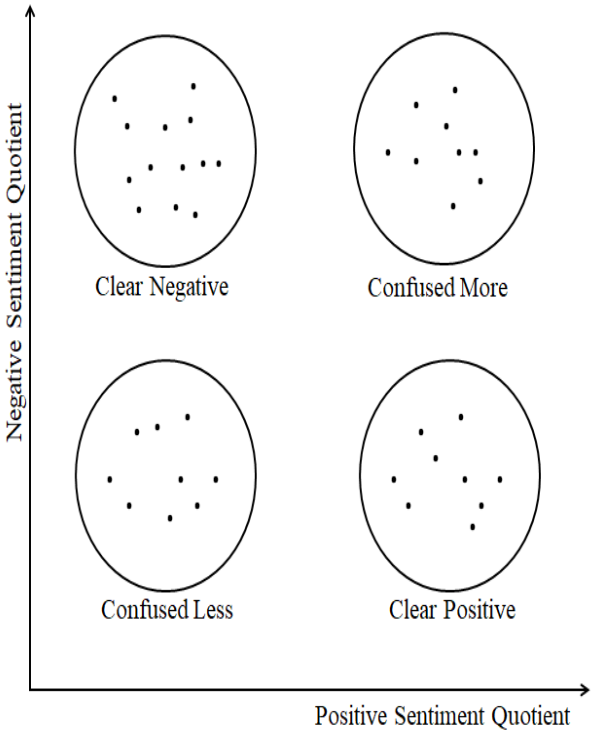

Fig. 3.Proposed knowledge base

E. Classification of Test sample

The major agenda of the paper is to witness the test sample and its evaluation corresponding to business decisions. k-NN [14] is applied on Test Sample along with Knowledge Base. The number of tweets and their labels may declare the type of strategy required for the successful business.

IV. RESULTS AND DISCUSSIONS

Around 104 tweets have been considered, softmax evaluations of all the negative and positive business sentiment keywords are estimated corresponding to a particular tweet. These estimations are shown in the Table II and Table III.

Table-II:Softmax evaluations of the certain negative business sentiment keywords

Tweets	Not happy	Not interested	Bad	Useless	Not applicable
BTweet1	0.3	0	0.4	0	0
BTweet2	0.4	0.81	0	0	0
BTweet3	0.2	0	0.7	0.1	0
BTweet4	0	0	0.2	0.1	0.5
.....					

Table-III :Softmax evaluations of the certain positive business sentiment keywords

Tweets	good	acceptable	less cost	More useful	affordable
BTweet1	0.3	0	0.4	0	0
BTweet2	0.4	0.81	0	0	0
BTweet3	0.2	0	0.7	0.1	0
BTweet4	0	0	0.2	0.1	0.5
.....					

Table IV depicting the PCA estimation using the Table II and Table III properties. Basically PCA is applied to reduce the number of features that minimizes the complexity of clustering and classification.

Table-IV: PCA estimations of both Table II and Table III features into PSQ and NSQ

	PSQ	NSQ
BTweet1	0.97	0.21
BTweet2	0.54	0.56
BTweet3	0.99	0.2
BTweet4	0.33	0.88
.....		

Figure 4 is the graphical representation of the knowledge base of behavioural sentiment classification model. It contains the labelled samples into four clusters and their centroids.

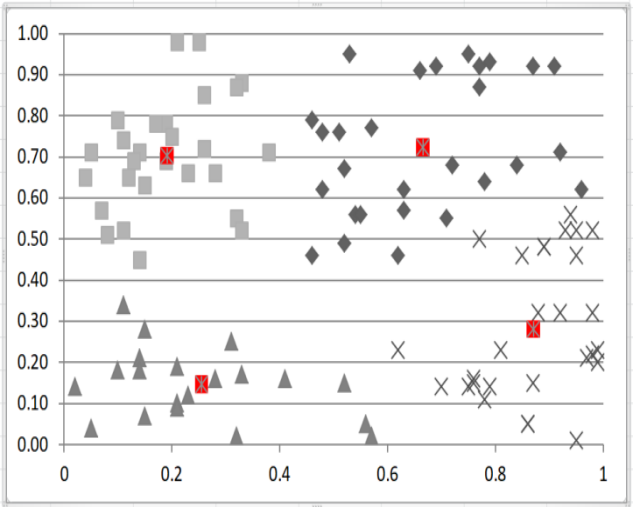

Fig. 4.: Knowledge base of behaviour sentiment classification model

Test sample 1 features that PSQ and NSQ are estimated, which is shown in Table V. Test sample 1 belongs to the class labelled Clear Negative. This indicates that test sample 1 expressed negative thoughts about the some product.

Table-V: Test sample and its label

	PSQ	NSQ	Label
Test sample1	0.14	0.89	Clear Negative

V. CONCLUSION

Current system is providing huge or abundant amount of data concerned to several aspects namely relations, business, matrimonials, entertainment and so on. The effective utilization of this data and creating social impacts would be the responsibility of research & developments domain. In the regard, the proposed paper is one such attempt to use the existing abundant data and create useful inferences to the society. This paper can be further extended with more number of test samples for efficiency verification.

ACKNOWLEDGMENT

Authors would like to thank the Management, Principal and all the teaching, Non-Teaching faculty members of *Vidya Vikas Institute of Engineering and Technology, Mysuru, India* for supporting and encouraging our research work undertaken.

And we gratefully thank the Visvesvaraya Technological University, Jnana Sangama, Belagavi for encouragement and support extended to this research work.

REFERENCES

1. Sujata, Joshi, Domb Menachem, and Thakur Viraj. "Impact of Flash Sales on Consumers & E-Commerce Industry in India." International Conference on Qualitative and Quantitative Economics Research (QQE). Proceedings. Global Science and Technology Forum, 2017.
2. E-commerce Industryin India. (2019, October). Retrieved from Indian Brand Equ ity Foundation: https://www.ibef.org/industry /ecommerce. aspx
3. https://innovate.mygov.in/ideas-2020/, Accessed on 26.02.2020
4. Thongsuk, Chanattha, Choochart Haruechaiyasak, and Somkid Saelee. "Multi-classification of business types on twitter based on topic model." The 8th Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI) Association of Thailand-Conference 2011. IEEE, 2011.
5. Heyns, Wiaan, and Sheryl Buckley. "A Twitter knowledge sharing model based on small businesses in the Western Cape." 2019 Conference on Information Communications Technology and Society (ICTAS). IEEE, 2019.
6. Halibas, Alrence Santiago, Abubucker Samsudeen Shaffi, and Mohamed Abdul Kader Varusai Mohamed. "Application of text classification and clustering of Twitter data for business analytics." 2018 Majan International Conference (MIC). IEEE, 2018.
7. Khan, Asif Uddin, and Bikram Kesari Ratha. "Business data extraction from social networking." 2016 3rd International Conference on Recent Advances in Information Technology (RAIT). IEEE, 2016.
8. Ribeiro, F.N., M. Araujo, P. Goncalves, M.A. Goncalves, and F., Benevenuto. 2016. “A benchmark comparison of state-of-the-practice sentiment analysis methods.”, 2015arXiv151201818N.
9. Alaei, Ali Reza, Susanne Becken, and Bela Stantic. "Sentiment analysis in tourism: capitalizing on big data." Journal of Travel Research 58.2 (2019): 175-191.
10. Roesslein, Joshua. "Tweepy." Python programming language module (2015).

11. Rong, Xin. "word2vec parameter learning explained." arXiv preprint arXiv:1411.2738 (2014).
12. Wold, Svante, Kim Esbensen, and Paul Geladi. "Principal component analysis." Chemometrics and intelligent laboratory systems 2.1-3 (1987): 37-52.
13. Jain, Anil K. "Data clustering: 50 years beyond K-means." Pattern recognition letters 31.8 (2010): 651-666.
14. Peterson, Leif E. "K-nearest neighbor." Scholarpedia 4.2 (2009): 1883.

AUTHORS PROFILE

Syed Salim working as Assistant Professor in the Department of Computer Science & Engineering at Vidya Vikas Institute of Engineering & Technology Mysuru, which is affiliated to Visvesvaraya Technological University, Belagavi, India. Also Salim is pursuing Ph.D. degree under the guidance of Dr. Madhu B K, Professor and Head, Dept of CSE, VVIET, Mysuru. Currently, Salim is working on Business Sentiment Analysis, Machine Learning using Python programming language. He has published several papers in National and International publications.

Dr Madhu B K, is at Technocrat with exposure to academics, Software Industry and Research. He is M.Tech and PhD in Computer Science & Engineering, with specific research interest in the field of Agile Software Testing. He is Advisor and Consultant for Performance Testing, Optimizing SDLC and Implementation of CMMI Levels. Member of Vocational Training Group. He is also Visiting Professor, Guest Faculty and Editorial Member and Reviewer of IEAE, IJCEE, GJCSM, JIS, IEEE International Conference and published 40+ research papers in IEEE and many reputed journals. He has 2 students completed PhD under his guidance and presently guiding 2 PhD research scholars. At present he is Professor and Head of CSE and with additional charges as Head, Placements at Vidya Vikas Institute of Engineering & Technology Mysuru, which is affiliated to Visvesvaraya Technological University, Belagavi, India.