



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirccce.com

Vol. 5, Issue 7, July 2017

Domain Based Sentimental Analysis using Naïve Bayes Classification over Twitter

Gargi ¹, Tapas Kumar ²

Student of Masters in Technology, Dept. of CSE., Lingayas University at Faridabad, Haryana, India

Professor & Head of Department, Dept. of CSE., Lingayas University at Faridabad, Haryana, India

ABSTRACT: Societal medium provide a podium where users distribute a profusion of information on anything and everything. The information may consist of users' emotions, feedbacks, review, and special familiarity. In this research a narrative Domain-based Sentiment Analysis Process for Social Media content with unconstructive sentiments is obtainable. The social media content is automatically extracted from the twitter messages. An domain-based process is designed to retrieve and analyze the customers' tweet with negative sentiments. This idea is demonstrated with the identification of customer dissatisfaction of the delivery service issues of the various products based on automobile sector for national and international market. The tweets related to the products and its services include various different aspects in respect the automobile intricacies and customer services and requirements. A combination of technologies for twitter extraction, data cleaning, subjective analysis, domain model building, and sentiment analysis are used. The results from this analysis could be used by the company to take corrective measures for the problems as well as to generate an automated online reply for the issues. A rule based classifier could be used for generating the automated online replies.

KEYWORDS: Naive Bayes, Sentimental Analysis, Twitter4J, WorldNet.

I. INTRODUCTION

The sentiment analysis of customer's social media data is very important in the present day business scenarios. Customers share information about products, services and their experiences on social media. This information can be used for market research, product feedback and analyzing customer service effectiveness. Bing Liu explained that opinions are subjective expressions. Opinions describe people's sentiments or feeling towards entities, events or their properties [1]. The sentiment analysis of the opinions could lead to many interesting results. The dynamically expanding web and social media are generating huge amount of opinion data. People's opinion about any product or service on social media is a very valuable asset for any organization. The organizations can generate information on customer's response, or its behavior for any product or service, by doing the sentiment analysis of these social media data.

Mike Thelwall, Kevan Buckley, Georgios Paltoglou, and Di Cai presented an algorithm for understanding the sentiment associated with the short text [2]. The users' shared informal text or sentences convey different emotions on 'likes' and 'dislikes'. They developed algorithms to identify the sentiment and the sentiment strength in these informal communications. Janyce Wiebe and Ellen Riloff showed a technique for creating subjective and objective sentence classifiers from un-annotated texts [3]. They tried to develop a system to analyze text and find out opinion or emotions associated with it. They suggested a method for classifying subjective and objective information from a sentence. This subjective and objective information can be used to develop different models, which can be utilized to answer a customer based on their behavior.

In this research, a process for sentiment analysis of customers' social media data using an domain model is presented. This process would help in identifying the problem area associated with the customers' social media data that contains positive, negative or neutral sentiments. The remaining portion of the paper is organized as follows: Section II



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirccce.com

Vol. 5, Issue 7, July 2017

provides the related work research used for building the domain model, Section III provides information on building an domain model and the use of domain model to for sentiment analysis, Section IV focuses on discussion about different attempts taken during this research, their limitation and usefulness, Section V provides the Simulation Results and Section VI briefs on conclusions and future work.

II. RELATED WORK

“Domain” represents the context knowledge as a hierarchy of concepts. It includes machine-interpretable definitions of basic concepts of the domain and the interrelation of these concepts. Entities (classes), objects, object properties and the relation between them could be incorporated in to an domain model. Also, it defines a common vocabulary for researchers who need to share information in a domain. Defining domain helps in sharing a common understanding of knowledge among people and software agents, and makes domain knowledge reusable. The domain model could be stored in dictionary using key value with polarity based aspects or Domain-based Sentiment Analysis Process for Social Media Content Pratik Thakor and Sreela Sasi Pratik Thakor and Dr. Sreela Sasi / Procedia Computer Science 00 (2015) 000–000 tabular format. It is possible to merge two domain models with the same knowledge domain. We can retrieve the information organized in domain model by querying it. Sayed Zeesan Haider and Yaakub, R. M, Li. & Feng, Y. developed an domain to do feature based sentiment analysis of customer’s review on smart phones, and simple mobile phones [4, 5]. K.M Sam and C. R. Chatwin proposed a relatively similar model for doing sentiment analysis of customer reviews on electronic products [6]. Larissa A. de Freitas and Renata Vieira had also built and used an domain model to do feature based opinion mining of Portuguese movie reviews [7].

All these research works used the information on categories and features of the products to build the domain model. Later the domain model was used to identify the most popular products by analysing the customer reviews. Tim Finin, Li Ding and Lina Zou shared an interesting concept of developing an domain based intelligent application [8]. Matteo Baldoni, Cristina Baroglio, Viviana Patti and Paolo Rena presented an domain based approach for sentiment analysis of social semantic web [9]. The idea was to share information and knowledge in an effective way with the help of domain. They showed the use of domain’s to enable computer programs to read, to publish or to exchange information and knowledge by enhancing interoperability. Natalya F. Noy and Deborah L. McGuinness from Stanford University gave a perfect insight into the process of building domain from the scratch [10]. They provided excellent information about the various components of the domain model. They explained the idea of identifying the important building components of an domain mode in a text by applying OOP concepts.

In this research an domain model based process for identifying customer dissatisfaction with any postal service from their tweets is developed. The subjective analysis of tweet text is done to retrieve the information on sentence builders like nouns and verbs. By using the OOP concepts, the noun and verb can be used as object and object property respectively. This information on objects and object properties were used to build the domain model. Mike The wall, Kevan Buckley, Georgios Paltoglou, and Di Cai used Senti-Strength tool to identify sentiments associated with random text extracted from social media [2]. The SentiStrength can calculate the strength of the sentiments associated with the text. Only negative tweets are analysed in this research to find the problem area associated with it.

III. PROPOSED ALGORITHM

The proposed approach uses an domain which add features and meaning to the tweets and WordNet which is an online semantic dictionary, electronic lexical database of nouns, verbs, adverbs, and adjectives. WordNet helps in extraction features from tweets. In our proposed approach we are considering automobile domain. The proposed approach has the following steps:-

1. Tweet Extraction: Tweets are extracted from Twitter4J. Twitter4J is referred as an unofficial library of Java for the Twitter API. With Twitter4J, you can easily integrate your Java application with the Twitter service.

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 7, July 2017

2. Tweet Repository: A repository is a central place in which an aggregation of data is kept and maintained in an organized way. The repository is made using SQL server management.

3. Sentence segmentation: Sentence segmentation is the process of determining the longer processing units consisting of one or more words. This task involves identifying sentence boundaries between words in different sentences. Since most written languages have punctuation marks which occur at sentence boundaries, sentence segmentation is frequently referred to as sentence boundary detection, sentence boundary disambiguation, or sentence boundary recognition. All of these terms describes the same task i.e. determining how a text should be divided into sentences so as to use for further processing.

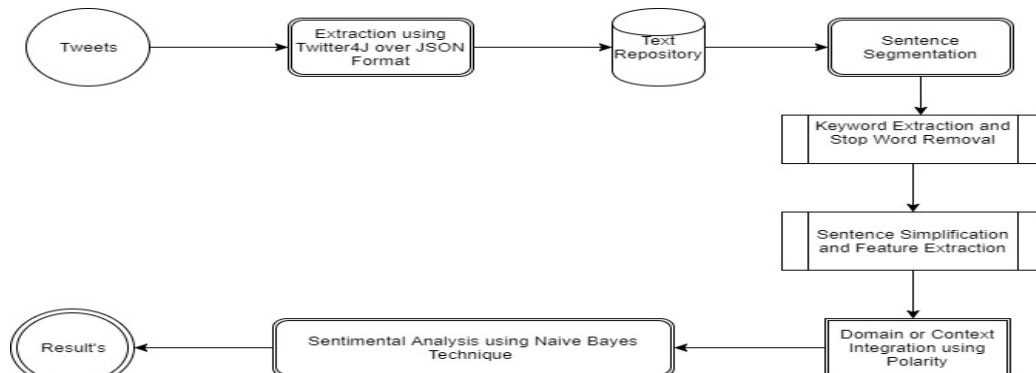
4. Keyword extraction: Keyword extraction is a subtask of information extraction. The goal of terminology extraction is to automatically extract relevant terms from a given set of tweets.

5. Sense matching : Overlap value is obtained by matching the sense of the words with the help of WordNet.

6. Domain Context: Domain is created which involves relation between the tokens. Domain helps in adding meaning to the tweets and relevant analysis can be obtained by removing redundant information and irrelevant information from the raw and JSON data.

7. Domain Mapping and Validation: The only process Database Schema the only the steps of the SDM is create

9. Analysis: In this stage, comparative analysis is done using naïve Bayesian classification algorithm. Whereas the results or sentiments can be predicted using negative, positive or neutral using comparative approach with the help of attributes and polarity.



The above discussed architecture is shown as figure 1.

Algorithm: Domain Creation

- Step 1. Ont_input: concept (c)
- Step 2. Ont_var: T- Free set of tweets
- Step 3. P- Free set of objects
- Step 4. Q- Free set of attributes
- Step 6. T<- ret_tweets(s)
- Step 7. foreach t ∈ T do
 - p<- ret_obj (t); if p ≠ Null then
 - p:= P U {p}
 - Q'<- ret_attr (t)
- Step 8. foreach q ∈ Q'; if (p,q) ≠ Φ do
 - Q: = Q U {q}
 - Table<- (p, q)
 - return Table

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirccce.com

Vol. 5, Issue 7, July 2017

IV. PSEUDO CODE

Consider the concept as Car

$C = \{Car\}$

$P = \{\text{set of objects}\}; \{Maruti, Tata, BMW\}$

$Q = \{\text{set of attributes}\}; \{\text{Engine Power, Torque, Fuel Capacity, Max Speed etc.}\}$

In formal concept analysis the operator “” is used as an object for derivation.

For a set of object P, P' can be defined as

$P' = \{\text{all attributes in } Q \text{ shared by the objects } P\}$

For a set of attributes Q, Q' can be defined as

$Q' = \{\text{all objects in } P \text{ that have all the attributes of } Q\}$

The pair sets (P, Q) of objects and attributes can be given as $P'=Q$ and $Q'=P$ which is called as formal concepts.

Generating the formal concept analysis

- (i) A set of objects P is picked
- (ii) Attributes of P, (i.e.) P' are derived
- (iii) (P')' is derived
- (iv) (P'', P') is a formal concept analysis.

V. SIMULATION RESULTS

User opinion analysis system that predicts the opinion of user whether the user is in drifting mode, positive or negative on the basis of the tweet_id of user on live social twitter data. Also to predict the general opinion of users in different locations in particular time stamp in a certain context and depicted in graph form using domain.

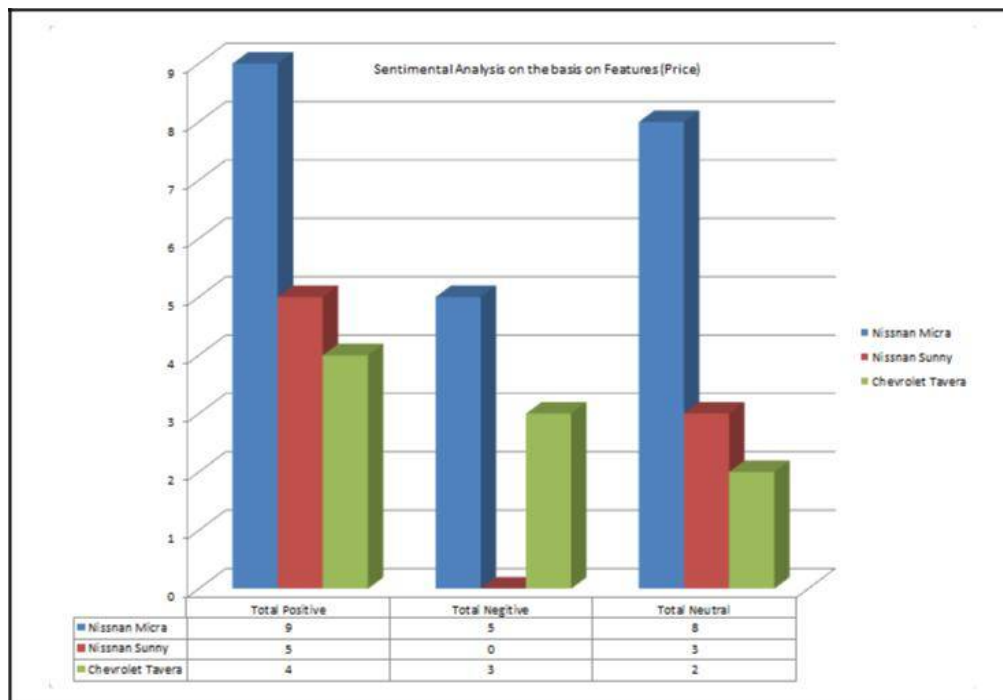


Figure 2: Statistical Data

Figure 2 shows the statistical data of the tweets used to analyse opinion of user corresponding to user_id



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirccce.com

Vol. 5, Issue 7, July 2017

VI. CONCLUSION AND FUTURE WORK

For future aspect analyzing the sentimental analysis we will integrate the same with Big Data Model, first of all twitter data is extracted using flume. The data extracted is available in unstructured (JSON) format. The data is integrated with SQL and interrelated with domain. Using hive it is given a tabular form i.e. a structured form of data is obtained. Maven framework is used to get the executable jar to integrate eclipse and domain. Data needs to be filtered before analyzing. Data is cleaned by removing stop words. For classification, using support vector model will be inculcaed

Contemplating over support vector model technique, we can use a dictionary which stores a list of words that are positive, negative and neutral. Lastly, data is imported to excel to give a graphical form and to get the results. In the scheme, we can identify the user sentiments with the help of user_id whether the user is positive, negative or in drifting mode. The system tells the general behavior of users country-wise as well as city-wise for a particular topic. The system will 90% to 99% accurate. In future, the data can be from multiple sources at the same time. Also various different tools like R, tabluue can be integrated, also we can continue with domain in it. Finally, multiple topics also can be taken into consideration. Further works can be done to improve the efficiency and accuracy.

REFERENCES

- [1] Bing Liu, "Sentiment Analysis and Subjectivity", from Handbook of Natural Language Processing, Second Edition, (editors: N. Indurkha and F. J. Damerau), 2010.
- [2] Mike Thelwall, Kevan Buckley, Georgios Paltoglou, and Di Cai, "Sentiment Strength Detection in Short Informal Text", Statistical Cybermetrics Research Group, School of Computing and Information Technology, University of Wolverhampton, Wulfruna Street, Wolverhampton WV1 1SB, UK, 2010.
- [3] Janyace Wiebe and Ellen Riloff (2005). "Creating subjective and objective sentence classifiers from unannotated texts." Sixth International Conference on Intelligent Text Processing and Computational Linguistics (CICLing-2005).
- [4] Sayed Zeesan Haider, "Domain-based sentiment analysis case study", a case study for Master degree project, University of Skovde, pp. 05-67, 2012.
- [5] Yaakub, R. M, Li, & Feng, Y., "Integration of Opinion into Customer Analysis Model", proceedings of Eighth IEEE International Conference on e-Business Engineering, pp. 90-95, 2011.
- [6] K.M Sam and C.R. Chatwin, "Domain-Based Sentiment Analysis Model of Customer reviews for Electronic Products", Proceedings of International Journal of e-Business, e-Management and e-Learning, Vol. 3, No. 6, Dec.2013.
- [7] Larissa A. de Freitas and Renata Vieira, "Domain-based Feature Level Opinion Mining for Portuguese Reviews", PUCRS FACIN, Porto Alegre, Brazil, 2013.
- [8] Tim Finin, Li Ding and Lina Zou "Social Networking on the Semantic Web", Learning Organization Journal, special issue on Ubiquitous Business Intelligence, Miltiadis Lytras et al, 2005.
- [9] Matteo Baldoni, Cristina Baroglio, Viviana Patti and Paolo Rena, "From Tags to Emotions: Domain-driven Sentiment Analysis in the Social Semantic Web", Universit`a degli Studi di Torino, 2010.
- [10] Natalya F. Noy and Deborah L. McGuinness "A guide to creating your first domain", Stanford University. http://protege.stanford.edu/publications/domain_development/domain101-noymcguinness.html
- [11] H. Cunningham, et al. "Text Processing with GATE (Version 6)". University of Sheffield Department of Computer Science. Apr.2011. ISBN 0956599311. <https://gate.ac.uk/>