An adaptive sentimental analysis using ontology for retail market

J. Mannar Mannan[1]*, Jayavel .J[2]

[1]Associate Professor, Dept. of Information Technology, Karpagam College of Engineering, Coimbatore.
[2]Associate Professor, Dept. of Electronics and Telecommunication, Bharati Vidyapeeth College of Engineering, Kharghar, Navi Mumbai.
*Corresponding author E-mail: endeavour6381@yahoo.co.in

Abstract

The growth of digital documents on web becomes the massive sources for online market analyzing at broad level. The study of market research over online incorporating new parameter called sentiment analysis. The sentiment analysis plays a crucial role for identifying behavior of customers by means of natural language processing from customer feedback about product or services. The opinion mining have done from the user data over web related activities such as search history, blog activities, forums, comments on the social network, express the opinion about the concept/product and suggestion or recommendations. The present system is non-adaptive relation identification system works on existing, predetermined set of relations and it cannot identify the new type relation for opinion mining. The existing system are also neglected the static sentiments of users. This paper proposed ontology based adaptive sentiment analysis system for extracting new features added on the user space. In our work, the ontology and 3D space clustering framework which allows incorporation of domain knowledge for predicting sentimental analysis via opinion mining.

Keywords: *Ontology, opinion mining, sentimental analysis.*

1. Introduction

In the present globalization era, the increased web users are highly contributed to growth of digital documents on the web. The new dimension of market analysis on the web is sentiment analysis now becomes key research area. The sentiment analysis plays a crucial role in market analysis. Present globalization era, market and business play an important role development of a Nation. The day to day activities and statistical data analysis about the market makes impact on economy of a Country and everyone's life. The market analysis is a key challenging issue for business analysts. The essentials of market analysis are the raw data about the company, brand, product, availability, competitor and behavior of customers. Apart from these raw data, other factors are also highly influenced. One such a thing is sentiment. The sentiment analysis is the major research area of business data processing. The simple example for sentiment analysis is the "black color". Some people dislike the black color products. Similarly emotional, behavior, addiction, cultural and sentimental nature of the consumers is influenced the buying behavior. In the present market research, identifying, gathering and analyzing these types of customer data to predict the sentiment analysis is the key challenging task. The Information technology provides solution for analyzing using data mining concepts.

The sentimental analysis is varying from different culture, language, religion and some time influenced by certain events. The market can be influenced by many events like political statements, year-end profit of a company, average monsoon, climate and culture of the people. The psychological aspect also diverts the customer behavior. Example, the PepsiCo's blue Pepsi was fail in Indian market due to its color that looks like kerosene. In general, Indians are highly sentiment on Wednesday and Friday for doing good events and similarly Tuesday and Saturday not their preference.

Data mining is the process of applying mathematical function called algorithm for extracting common pattern or relation occurred on huge data set. It is defined as knowledge discovery. The knowledge discovery from data mining starts from warehousing, cleaning, selection, integration and transformation. The data cleaning is a process of eliminate incompatible data pairs. The data integration means, different data sources are mapped (combined) for knowledge extraction. The data selection is a process of selecting data for knowledge extraction using mining process. Whereas the opinion mining is a process of identify customer feedback about the product or services via their textural representation on social Medias like blogs, Facebook, Twitter or customer feedback system. It is a natural language processing used to semantically analyze the textual representation of customer feedback to analyze the pros and cons of product or service for improving business. The data mining provides the considerable solution for extracting opinion, but it is needed to be more optimized. The drawback on the data mining for processing opinion is its requirement of structured data whereas the feedback of the customer is the paragraph or textual represents in the natural language.

1.1. Ontology

The sentimental analysis is key research area to analyze the customer behavior through their opinion that has been collected as feedback from customers about the product or service. Ontology is a concept that defined as perception of shared conceptualization that describes an object or thing from different dimensions. In market analysis, each customer feedback is independently analyzed

with the help of extended keyword based search using Ontology. The WordNet ontology is used for working as background knowledge, and used to extend the important words into meaning full semantically related words for better understanding of user opinion. The ontology can play crucial role in sentimental analyze. Ontology is an specification of contributed concepts, defined as O = (I,C,P,O), where I as Individuals, C as classes , P as properties and O as object. The ontology is a knowledge base constructed with set of classes and relationship between classes using properties. The ontology is a hierarchical representation of knowledge about the particular domain.

The sentiment analysis otherwise called opinion mining, which is mandatory for retail market analyze and understand customer behavior for providing better service and rectify the drawback with the existing product or service. The customer feedback can be analyzed into positive, negative and neutral opinions. It can be identified, grouped and classified using natural language processing. The ontology provides the complete model of sentiment analyze.

1.2. WordNet

WordNet is an Ontology developed by Princeton University; it is a lexical analyzer used in natural language processing (English) contains around 1,50,000 'synsets' and their semantic relations. It also provides many meaning full information about the domains such as 'synonym' , 'coordinated terms', 'hypernyms' , 'hyponyms', 'holonyms', 'meronyms', 'domain' and 'domain terms'. The WordNet ontology used to compare the two terms based on the semantic imminence using Wu and Palmer semantic distance measurement techniques. The similarity between two concepts is measured using Wu and Palmer similarity measurement techniques using the following equation

$$Wu - Palmer \ \sigma(A, B) = \frac{2 * \ \delta(A \wedge B, \rho)}{\delta(A, \ A \wedge B) + \ \delta(B, \ A \wedge B) + 2 * \ \delta(A \wedge B, \ \rho)}$$

Where ρ – is the root concept of of the hierarchy, $\delta(A,B)$ is the number of intermediate edges between a concept A and B, $A \wedge B = \{ C \in O; A \leq C \wedge B \leq C \}$. WordNet with Wu-Palmer measurement techniques returns the value for two same terms as '1', semantically intimate terms return nearby to '1' and move closer to '0' otherwise.

1.3. Data Mining

The data mining is the process of extracting pattern of knowledge existing in the data set gathered from different sources. These data set may be domain specific, which deals with that the data is collected and that describe the specific domain such as medical data set, super market data set or customer data etc., The data mining works based on various clustering algorithms which is mathematically proven theory, such that strong occurrences of patterns among data. The other parameters of data mining are path analysis, clustering, classification. The data warehousing and mining play a major role in business development by predicating customers buying behavior.

2. Review of Literature

The sentimental analysing is used to determine a state or opinion of a person about the product or services. This information can be retrieved from feedback of social websites, news articles or comments. [1] The proposed a centroid-based clustering algorithm which is not mandatory as member of data set. To find the k means centroids and assign the observation to the nearest squared distance from cluster observations. This technique attempted to utilize the sentiment analysis for transportation system. [2] have introduced a new model called sentiment ontology to analyse context based online opinion posts in stock markets. The methodology integrates popular sentiment analysis into machine learning approaches based on support vector machine and generalized autoregressive. The sentiment analyse requires strong background knowledge for natural language processing. WordNet has been used [3] to semantic analysis concepts represented as words. [4] have experimented sentiment analysis on TREC data set and articles collected from web blogs and obtained the progressive curve. He ensured that the WordNet works better on semantically inspect terms. [5] have handled semantic orientation approach for sentiment classification on Chinese movies from viewers feedback. [6] Detailed survey on sentimental analysis and its key challenges. The sentimental analysis classified into structured, semi-structured and unstructured.

The sentimental information found on social Media like Facebook, Twitter and WatsApp. Reaching, collecting, and analyse the content of these social Medias and segregate the sentiments from their feedback. [8] have represented a detailed survey on sentiment analyse using various techniques like Support Vector Machine (SVM), Naïve Bayes, Maximum Entropy and K-nearest Neighbour methods. All these methods applied to large data sets from forums, blogs, Twitter, Facebook and other web based user activities for analysing customer behaviour. [9] The described that the opinion classified into three category such as Regular opinion, Comparative Opinion and Part of speech Tagger for identify opinion. All these methods are extracted in form of text classification, entity level, sentence level, and semantic level.

[10] Presented a scientometric mapping for sentiment analysing over the WoS data set. Their work on computationally process the unstructured data mainly to extract opinions and identify their sentiments. [11] Used the opinion mining concept for managerial decision making and design a tool that will give opinion of the manager with respect to their attributes. Here relevant features of managers such as decision making, knowledge, communication skill, humility, inspiration and integrity are considered with weighted value. [12] Opinion mining done on Twitter data set. In their work, a two-step approach is proposed. Firstly identify the sentiments and secondly algorithm to find the emotion values of the opinion words. [13] Conditional Random Fields (CRFs) model is used to to perform the opinion mining tasks and compared with lexicalized Hidden Markov Model (L-HMMs) and achieved improvements. [14] Surveyed about various text extraction and classification for supporting sentiment analyse. [15] Identify the role of pre-processing over Twitter. In Twitter, the slang words are used for sentiment classification. The key challenges this area is to identify the noise, emotions and folksonomies. Here, the Conditional Random Field (CRF) is used to find out the significant of the word co-occurrence. [16] The sentiment analysis is done on Twitter data set for measuring brand reputation based on customer feedback and satisfaction. The Naïve Bayes, Support Vector Machine and Decision Tree classifier methods are used to measure the brand reputation. [17] Proposed ontology based negative sentiments of customers from Twitter data which consists of two processes such as build the ontology model using the extracted data and the second process is to retrieve the problem area from the negative sentiments associated with a tweet using a previously built ontology model. [18] Have provided a detailed study on temporal sentiment analysis and Sentiment causal relation. The temporal is used to summarize events based sentiment date and time and causal is causes and effects of events. Using these analyses, it is possible to identify events prediction and sentiments of upcoming events. The performance of the temporal sentiment is done at using precision and recall whereas the accuracy of causal rule prediction is evaluated using parameters Mean Absolute Error (MAE) and the Root Mean Squared error (RMSE). [19] Proposed a aspect classification followed by polarity classification and also provided a detailed survey on how to measuring sentiments on social network. Here, they used lexical method, which is pre-processing of the text to be analysed on the social network data set. Here, they applied baseline approach such as stemming, part of speech tagging and conjunction rules. Another important approach for measuring sentiment is machine learning. The machine learning, machine learning methods, Support Vector Machines (SVM) and Naïve Bayes methods are used for measuring sentiments.

3. Ontology for Sentiment Analysis

The sentiment analysis is a process of knowledge extraction using Natural Language Processing (NLP) from user feedback to identify the pros and cons of services or products. Customer feedback analysis is important for business processing to enhance the service. The growth of web based applications provide the solution for easily handle the customer satisfaction through their feedback system over online. The feedback collected and evaluated from massive collection web data to identify the level of satisfaction of customers by identifying customer feelings and satisfactions. In the ontology based sentimental is the process of detecting the semantic level of the text related to particular context. It describes as positive, negative or neutral. The sentiments can be found in recommendations. Highest recommendation provides positive sentiments.

Apart from the customer feedback, the customer's static sentiments highly influenced, a market study says. In Indian market, the Friday gets first preference of the consumer for buying a luxury product and 'Wednesday' follows next. The religious and cultural aspects are also influencing for selecting the day for buying new products. In Tamilnadu, South India, 'Thai' month (mid of January to Mid of February), included maximum sentimentally good days. These days, includes birth of Christ, New Year and 'Pongal' festival all are comes in the shorter period. The study said that the sales increased during this month of all the year.

Unlike western countries, people from India, China and Middle Eastern stats are embracing more religious and cultural formalities. These types of ethics influenced by the religious, cultural and socio-economical factors are not yet included in the sentiment analysis for market research. These static sentiments are included to mining the opinion about the product on the market. The defining the static sentiments, the ontology are used which reflects knowledge about individual religious, cultural aspects of people (sentiments) which is incorporated along with web based sentiment analysis.

The ontology have constructed as classes of vocabularies based on sentiment analyses in certain order to represent hierarchy of consumer buying behaviour. This knowledge retrieved through data mining process collected from different data sources. The ontology is built with pattern of knowledge retrieved from the data mining along with static behaviour of consumers. In this proposed method, the consumer behaviour can be identified using customer's textual representation of opinion about the product or services. The set of pre-defined positive and negative feedback terms are collected for analyse the consumer buying behaviour.

The sentimental analysis is classified into two ways. One such a way is natural language processing to analyze negative words such as 'not good', 'bad', 'worst', 'waste', 'not worth', 'pathetic' can be extracted the positive, negative and neutral words. The second one is to classify symbols representing feelings or opinions. The positive words are 'good', 'better', 'worth', 'super' , 'fantastic', 'nice', 'excellent'. The neutral words are 'Ok', 'no problem', 'no-comments' etc., The user feedback or opinion can be represented in the pictorial manner such as providing levels of stars, symbols, representing happy, sorrow, and cry in the form of smiley (a pictorial representation of opinion or sentiments) on Social network like Facebook and Twitter. Every smiley have its background text that represents the opinion about the product or concept etc., From the smiley it is possible to get the customers behaviour about the concept or a product. Text pre-processing is the first step of detecting sentiments from the customer feedback. A careful stop-word removal is mandatory to corner the sentiments from the user feedback. The sentiment analysis is processing on three categories such as Machine learning approach, Natural Language Processing approach and lexicon based methods[7].

Let FB is set of feedback FB = {fb1,fb2,fb3….fbn}, the NLP is used to identify positive feedback (PF) based on the list of common

positive feedback, negative feedback and neutral feedback. The 3D space defines the customer feedback, symbolic representation and ontological representation of static sentiments. The set of feedback represented on the following table 1.

Table 1. Customer Feedback Representation

	Feedback 1	Feedback 2	Feedback 3
Good	1		
Bad			
Super			
Superb			
Fantastic			
Best		1	1
Nice			
Worst			
Waste			
Pathetic			
Useless			

This feedback collected as samples from various sources and its minimal samples taken for positive sentiment analysis. The different positive terms are only represented in the table 1. From this table, the total number of positive feedback can be identified from set of all feedback. The outcome of this matrix has been additionally included in the ontology based semantic sentiment analysis. The customer search history like date, day, and time and customers previous history has been collected on the other side. These customers data base been verified with statically contracted ontology for validate the customer behaviour.
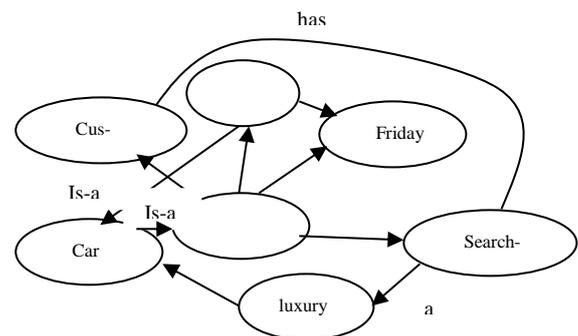
Fig. 1. Ontology for Sentiment Analysis

The Figure 1 shows that the part of the ontology which included the day, customer, searching, car, luxury. From this ontology, it is possible to predict the customer buying behaviour whether the customer interesting to buy new product. Here, the 'customer' class- > property 'has' reached another class concept 'searching'. It imply that the customer started to search something. Thereafter, the links extended up to the 'car' class via 'luxury', which implies that the customer searching to buy the luxury car. Finally, the day that the customer searches the car is 'Friday'. The 'Friday' gives the positive sentiments of customers. So, it is possible to predict that the customer ready to buy a luxury car. The day, month and festival season are influenced in this way for sentiment. This static information constructed with ontology and the rest of the details are retrieved from the customer feedback about the brand or product. The customer feedback history is a set of word represented by the different customers as opinion about the product or service. Here, the set of positive, negative and natural terms are used to identify the probability level of feedback about the product. The total number of feedback is collected and out of all, the positive, negative and natural feedbacks are clustered based on the set of pre-determined terms.
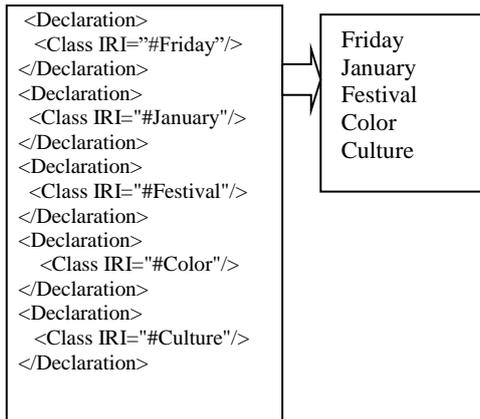
```
<Declaration>
  <Class IRI="#Friday"/>
</Declaration>
<Declaration>
  <Class IRI="#January"/>
</Declaration>
<Declaration>
  <Class IRI="#Festival"/>
</Declaration>
<Declaration>
  <Class IRI="#Color"/>
</Declaration>
<Declaration>
  <Class IRI="#Culture"/>
</Declaration>
```

| Friday |
| January |
| Festival |
| Color |
| Culture |

Fig. 2 Sentiment.owl – Extraction of concepts from ontology

The figure 2 shows that the ontological classes are extracted as concepts of terms. These conceptual terms are used to semantically determine the sentiment value of customers about the product. Here, the WordNet ontology used as background knowledge to identify the level of semantic relation between concepts represented in the ontology and customer search and other history. The Figure shows the static ontology, that represent cultural, social and sentimental behaviour of customers of one particular large region of geographically distributed, linguistically, culturally, environmentally united.
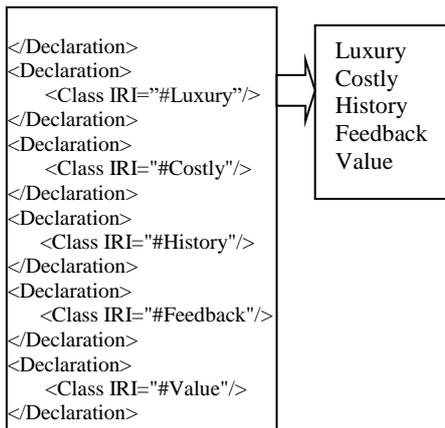
```
</Declaration>
<Declaration>
  <Class IRI="#Luxury"/>
</Declaration>
<Declaration>
  <Class IRI="#Costly"/>
</Declaration>
<Declaration>
  <Class IRI="#History"/>
</Declaration>
<Declaration>
  <Class IRI="#Feedback"/>
</Declaration>
<Declaration>
  <Class IRI="#Value"/>
</Declaration>
```

| Luxury |
| Costly |
| History |
| Feedback |
| Value |

Fig. 3. Brand.owl – extraction of concepts form brand

The Figure 3 shows the ontology about the product and customer feedback. These two ontological concepts are semantically mapped using WordNet ontology for measuring semantic relevance value.

$$WordNet(x,y) == z$$

Here, 'z' is the level of relevance fall between 0 to 1. If the relevance value defined as '1', shows the strong semantic relation and moving weak if it moves towards 0.

The ontology developed using protégé tool representing the static behaviour of the consumer which is developed based on the socio-cultural factors of people of one particular region. The region defines the state, language broadly speaking, economical background, cultural, religious and literature rate. From the static perception, if the customer started to search reviews of product on-line on Friday, and the month is festival season, it implies 50 % positive behaviour of customer willing to buy a product as the earliest. The studies confirms that the birthday date of a family members influencing the buying behaviour of customers. Customers buying behaviour can be predict using their search history and compared with ontology developed dynamically based customer static nature discussed

earlier. The comparison begins from user review about the product or service and regular expression matching of substrings by comparing existing positive set of string with on-line user review. The outcome of the regular expression has been verified with classes of concepts defined in the Ontology. Finally the outcome of the ontology comparison reaches 90% threshold values, confirmed the prediction that the valid customer willing to buy a product.

3.1. Precision and Recall

The efficiency of sentimental analysis system is prominently measured using a precision and Recall. Precision is parts of retrieved content that are relevant where as the recall is a part of relevant content that is retrieved successfully, represented in the Table 2 given below.

Table .2 Retrieval Performance Parameters

	Retrieved	Not Retrieved
Relevant	True Positive (**TP**)	False Positive (**FP**)
Irrelevant	False Negative (**FN**)	True Negative (**TN**)

They are shown in Table 2. such as, True Positive (TP) that the IR system retrieved relevant documents, False Positive (FP), Relevant Document not retrieved, False Negative (FN), irrelevant documents retrieved by IR system and True Negative (TN) irrelevant document not retrieved by IR system. The precision and recall are calculated using these parameters.

The precision and recall is calculated using the following equation.

$$Precision = (Relevant\ Retrieved)/Retrieved * 100$$

$$Precision = TP/(TP + FN) * 100$$

$$Recall = (Relevant\ Retrieved)/Relevant * 100$$

$$Recall = TP/(TP + FP) * 100$$

Venn diagram in Figure 2 shows the intersection of retrieved documents and relevant documents which contributes to the retrieval performance of IR system. There are four basic parameters used to calculate the performance of IR system.
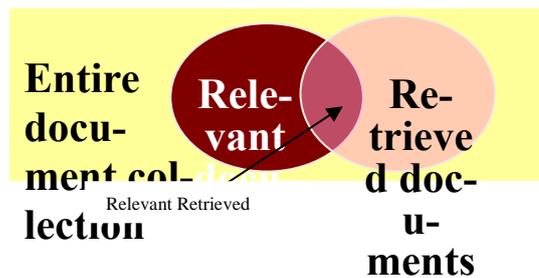

Relevant Retrieved

Fig. 2: Retrieval Performance

4. Conclusion

The sentiment analyse applied over digital document on web for extracting sentimental behaviour of the user or a customer. All these sentiments analyse are done over the user's textual representation on web using various algorithms. Apart from the user textual representations, the static and textually undefined sentiments are highly influenced the user by default. Socio, cultural and religious sentiments are not yet considered for sentiment analyse so far. To address this issue, this research work attempted the static sentiment analyse using ontology along with user's textual representation of sentiments over online is considered for analyse. These static

sentiments are highly influenced than the textual representation of words in web. The ontology based semantically analyse the sentiments perform better than the existing textual analysis.

References

[1] Jianping Cao, Ke Zeng, Hui Wang, Jiajun Cheng, Fengcai Qiao, Ding Wen and Yanqing Gao, "Web-Based Traffic Sentiment Analysis:Methods And Applications", *IEEE Transactions On Intelligent Transportation Systems*, Vol. 15, No. 2, April 2014.

[2] Desheng Dash Wu, Lijuan Zheng, and David L. Olson, "A Decision Support Approach for Online Stock Forum Sentiment Analysis", *IEEE Transactions On Systems, Man, And Cybernetics: System,* Vol. 44, No. 8, August 2014.

[3] Jaap Kamps, Maarten Marx, Robert J. Mokken and Maarten De Rijke, "Using Wordnet To Measure Semantic Orientations Of Adjectives", *Language & Inference Technology Group, University Of Amsterdam*, 2004.

[4] Lun-Wei Ku, Yu-Ting Liang And Hsin-Hsi Chen, "Opinion Extraction, Summarization And Tracking In News And Blog Corpora", *Department Of Computer Science And Information Engineering, National Taiwan University*, Taipei, Taiwan, 2006.

[5] Qiang Ye, Wen Shi and Yijun Li, "Sentiment Classification for Movie Reviews In Chinese by Improved Semantic Oriented Approach", *Proceedings of the 39th Hawaii International Conference on System Sciences*, 2006.

[6] Doaa Mohey El-Din and Mohamed Hussein, "A survey on sentiment analysis challenges",

[7] Emma Haddi, Xiaohui Liu and Yong Shi(2013), "The Role of Text Pre-processing in Sentiment Analysis", *Information Technology and Quantitative Management (ITQM2013),* Procedia Computer Science 17, 26 – 32.

[8] G.Vinodhini and RM.Chandrasekaran(2012), " Sentiment Analysis and Opinion Mining: A Survey", *International Journal of Advanced Research in Computer Science and Software Engineering*, Volume 2, Issue 6.

[9] Vishakha Patel, Gayatri Prabhu and Kiran Bhowmick (2015)," A Survey of Opinion Mining and Sentiment Analysis", *International Journal of Computer Applications*, Volume 131, No.1. 0975 – 8887.

[10] R. Piryani, D. Madhavi and V.K. Singh (2017), "Analytical mapping of opinion mining and sentiment analysis research during 20 0 0–2015", *Information Processing and Management 53 (2017) 122–150.*

[11] G. R. Brindha and B. Santhi(2012), "Application of opinion mining technique in talent management", *International Conference on Management Issues in Emerging Economies (ICMIEE)*

[12] Akshi Kumar, Prakhar Dogra and Vikrant Dabas (2015), "Emotion analysis of Twitter using opinion mining", *Eighth International Conference on Contemporary Computing* (IC3), 20-22 Aug. 2015.

[13] Luole Qi and Li Chen (2011), "Comparison of Model-Based Learning Methods for Feature-Level Opinion Mining", *IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*, 22-27.

[14] S Padmaja and S Sameen Fatima (2013), "Opinion Mining and Sentiment Analysis-An Assessment of Peoples' Belief: A Survey", *International Journal of Ad hoc, Sensor & Ubiquitous Computing (IJASUC)*, Vol.4, No.1.

[15] Tajinder Singh and Madhu Kumari(2016), "Role of Text Pre-Processing in Twitter Sentiment Analysis", *Twelfth International Multi-Conference on Information Processing-2016 (IMCIP-2016)*, Procedia Computer Science 89, 549 – 554.

[16] Nur Azizah Vidya, Mohamad Ivan Fanany and Indra Budi (2015) , "Twitter Sentiment to Analyze Net Brand Reputation of Mobile Phone Providers", The Third Information Systems International Conference, Procedia Computer Science 72 , 519 – 526.

[17] Pratik Thakor, and Dr. Sreela Sasi (2015), Ontology-based Sentiment Analysis Process for Social Media Content", 2015 INNS Conference on Big Data, Volume 53, 2015, Pages 199–207.

[18] P.G Preethi ,V. Uma and Ajit kumar (2015), "Temporal Sentiment Analysis and Causal Rules Extraction from Tweets for Event Prediction", International Conference on Intelligent Computing, Communication & Convergence, Procedia Computer Science 48 (2015) 84 – 89.

[19] Chetashri Bhadane,Hardi Dalal and Heenal Doshi (2015), "Sentiment analysis: Measuring opinions", International Conference on Advanced Computing Technologies and Applications, Procedia Computer Science 45 (2015) 808 – 814.