Social CRM: A Survey on Various Techniques for Text Analysis in Social Media

Mrs. Uma.N[1]

[1]Department of CSE, New Horizon College of Engineering, India

[1] umamam@gmail.com

Abstract— Customer relationship management (CRM) refers to practices, strategies and technologies that companies use to manage and analyse customer interactions and data throughout the customer life cycle, with the goal of improving business relationships with customers, assisting in customer retention and driving sales growth. CRM systems are designed to compile information on customers across different channels which includes the company's website, telephone, live chat, direct mail, marketing materials and social media. CRM systems can also be analysed to get informations like the purchase history, buying preferences and concerns. In this era where everyone is connected through social media, the relevance of improvement over CRM by analysing the opinions expressed by the customers can provide useful insights. Social networking websites, such as Facebook, Twitter etc. are rich in texts that enable user to create various text contents in the form of comments, wall posts, social media, and blogs. Due to ubiquitous use of social networks in recent years, an enormous amount of data is available via the Web. Application of text mining techniques on social networking websites can reveal significant results related to person-to-person interaction behaviours. In this paper an effort is made to understand the necessity of social CRM, which enables us to find general opinion about any specific subject, human thinking patterns, and group identification in large-scale systems .The various techniques used for text analysis are discussed.

Keywords— CRM, Social CRM, Text mining, Machine learning, Natural Language Processing, Sentiment Analysis

I. INTRODUCTION: WHAT IS CRM

Customer Relationship Management (CRM)[1] refers to the methodologies and tools that help businesses organizations manage customer relationships in an organized way. CRM systems are designed to compile information on customers across different channels ,or points of contact between the customer and the company which could include the company's website, telephone, live chat, direct mail, marketing materials and social media. Using the CRM model, a business organization can turn various analyzed data into useful business insights, which can help the organization to retain old customers, to attract new customers, and to maintain excellent relationship with customers which helps in the growth of the organization. In this era, where people are digitally connected, the social CRM model becomes one of the valuable asset as well as a necessity for an organizations.

Types of CRM

Operational CRM: has focused on creating a customer database that prevents a consistent picture of customer's relationship with the company, and providing the information in specific applications.

Analytic CRM: Here we use data mining strategies which will allow us to discover patterns and relationships between data and make accurate predictions like what products to be offered to which customers etc

Social CRM: With the recent advances in information technologies and web 2.0, people express their opinions about the products and services in the social media.Social CRM refers to the methodologies and techniques to use social media to tailor the needs of organization.

In this paper, more emphasis is given on the social CRM and it is discussed in detail in the following paragraphs.

Social CRM, a necessity, a threat or pool of opportunities

Social CRM(SCRM)[2] is referred as the use of social media services, techniques and technology to enable organizations to better connect with their customers, understand their opinions about the products or services. Social CRM adds a deeper layer of information onto traditional CRM by adding data derived from social networks like Facebook, Twitter, LinkedIn or any other social network .Using Social CRM, the companies can track a customers social influence and source data from the conversations through social media. With the recent advancement in web2.0 technologies, where people are digitally connected, it is advisable for the organizations to quickly engage in social CRM process without any delay.

The companies too has to face various social risks like employee misconducts, product incidents, reputational risks, Legal threats etc. To cope up with such risks, the organizations has to have an advanced social detection solution that delivers streaming big data processing, a concept model which can analyze threats from different directions, and requirement of deep intelligence experts. The mortin mom[] incident clearly shows the need for a social CRM.

The Social CRM approach rests on four pillars: engagement, conversation, participation and content distribution. There are a wide range of opportunities for the organization like they can create a win-win situation with dissatisfied customers, can user the customer likes to enrich on the loyality program, can attract new customers based on the positive comments in the social media

II. TEXT MINING

With the growth of the prominent social medias like facebook, twitter, google+, wikipedia, LinkedIn, PinInterest etc, it becomes a challenge to analyze the information shared by the public. In this paper a sinciere effort is made to identify techniques for text mining out of social media. There are different ways to analyze the datasets with varying difficulties ranging from techniques such as sentiment analysis applied from the feedback from customers(Unstructured,Non-categorized) to semisupervised learning like handling the CRM based on indirect aspects like claims, product return etc.

As the social media are flooded with information, among the social media analytics methods, text analytics finds the foremost place as it is highly challenging to derive useful information from the data available. The process of text mining involves data collection as the first step which requires authorization and handshake for the specific respective access, the data size, periodicity and the relevance of key phrases to be finalized. The various steps in text mining are summarized below**.**

A)Text Pre-Processing: Pre-processing guarantees successful implementation of text analysis, but may consume considerable processing time. It involves various substeps like text cleanup,tokenization(breaking up stream of text into words,phrases,symbols,or other meaningful items known as tokens.

B)Attribute Generation: In this step the text document is anaton analyzed for words and their no of occcurences. Two commonly used approaches to generate attribute generation are the BOW erae genibutmodel(Bag of Word) which uses frequency of word as attribute generation and the VSM(Vector Space Model) which uses cosine similarity, a number to represent the similarity of words in a document.

C)Attribute Selection: This phase helps in further reduction of high dimensionality. It selects only some features to represent the document by eliminating stop words,etc

D)Data Mining: In this phase various data mining techniques like classification, clustering, rule mining, sequential patterns etc are applied to conclude valuable information.

E)Interpretation /Evaluation: In this phase, the results are analyzed ,if satisfactory, the algorithm is terminated. If the desired results were not obtained again it is iterated.

III. Text Analysis and Natural Language Processing(NLP)

Text mining tries to apply these same techniques of data mining[4] to unstructured text databases. To do so, it relies heavily on technology from the sciences of Natural Language Processing (NLP)[5], and Machine Learning to automatically collect statistics and infer structure and meaning in otherwise unstructured text. Feature extraction, where in it extracts only a subset of features is a crucial step in text mining.

The customers generally can post about their opinions as text, images or videos. But since majority of the customers use text to express their opinion, it makes sense to analyse the text and try to do sentiment analysis[6].

A)Sentiment Analysis

Sentiment Analysis is the process of determining whether a piece of writing is positive, negative or neutral. It is also known as opinion mining, deriving the opinion or attitude of a speaker. A common use case for this technology is to discover how people feel about a particular topic. There is a widespread assumption that

sentiments are generally analyzed by artificial intelligence ,but the truth is that it is analyzed via a systematic process that uses sentiment lexicon. Using this lexicon we assign a degree of positivity or negativity to the words and then by analyzing the inherent positivity or negativity of each word, we determine whether it is a positive or negative opinion. Words like "happy","like","love" etc are deemed as positive degree words, whereas "hate","dislike" etc are assigned a negative degree.

However **word level** polarization ie giving positive or negative degree might not be able to capture the completely capture the essence of the opinion. Consider an opinion review about a movie like this-"The movie had a great plot,veteran actors,excellent supporting crew, but ultimately it sucks. Given the word level analysis, it has more positive words than Negative words,and hence it should be a positive review, whereas actually it is a negative one which can be interpreted only if we analyze again at a **sentence level**. Consider again a movie review like this -"All the movies Mr.X stars are great and interesting. But this one is a set back.Here the exact emotion can be captured only by analyzing the complete document ie **document level analysis** required.

IV. **Techniques of text analysis**

 There are various techniques for the text analysis. In this paper an effort is made to summarize the various techniques used for text analysis.

A)Machine Learning based Text Classification (MLTC)[7]: It comprises of quantitative approaches to automate NLP that uses machine learning algorithms. Some of the algorithms which comes under the supervised learning category are Rocchio Algorithm,Instance Based Learning Algorithm,and case based Learning algorithm.

Rocchio Algorithm (RA) is an implementation of the relevance feedback method and is mainly used for document refinement .In relevant feedback method.system generates results for a simple query submitted by the user and the user has to mark whether the results are relevant or not.This acts as an input to the Rocchio Algorithm.

Instance based Learning Algorithm, also known as lazy algorithims are based on the comparision between new problem instances with that of the instances stored during training. On arrival of any instance ,related instances are retrieved from memory, compared and classified accordingly. Some of the algorithms which exhibit instance based learning approach are the K-Nearest Neighbour (K-NN) algorithm, Case Based Reasoning(CBR) algorithm etc

B)Decision trees and Support Vector Machine

A decision tree is essentially a hierarchical decomposition of the (training) data space, in which a predicate or a condition on the attribute value is used in order to divide the data space hierarchically. In the context of text data, such predicates are typically conditions on the presence or absence of one or more words in the document. Decision tree is a method to semantically describe the concepts and similarities between the concepts.

Support Vector Machines (SVMs)[8] have been proven as one of the most powerful learning algorithms for text categorization. In contrast to other classification methods, SVM algorithm uses both negative and positive training datasets to construct a hyper plane that separates the positive and negative data. The document that is closest to decision surface is called support vector.

316

C)Artificial Neural Networks

In Neural Network that address classification problems, training set, testing set, learning rate are considered as key tasks. That is collection of input/output patterns that are used to train the network and used to assess the network performance, set the rate of adjustments. There are various algorithms like back propagation algorithm or Multilayer Perception (MLP) algorithm to do the classification[9].

D) Hybrid Approach

In this approach, a combination of various classification algorithms are used to provide better results iike increased text categorization, increased performance etc.

E)Text mining using clustering

Clustering is a technique used wherein we use algorithms and techniques on unsupervised document and the number, properties and memberships of the classes were not known in advance. In clustering ,documents are grouped together by identifying similar groups of data. The clustering techniques are again broadly classified into hierarchical clustering, partional clustering and semantic based clustering etc.

In **hierarchical clustering** ,the documents are organized into tree like structure where parent/child relationships can be viewed as a topic or subtopic relationship .Hierarchical clustering can be performed by different methods like agglomerative or divisive methods. Studies have proved that hierarchical agglomerative clustering is one of the most suggested technique for text analysis. In divisive method, we recursively split the cluster into smaller clusters until each document is a classified cluster.

In **partitional clustering**, we use a feature vector matrix, where in features of objects are compared with objects of similar patterns, and further classified through multiple iterations till the cluster is stabilized. Some of the partition clustering approaches are **K-Mean, K-Medoid, C-mean and C-medoid** [10].In the K-Mean approach data is divided into K-cluster and each cluster is represented by the means of points termed as the centroid. The advantage of K-means algorithm is that there is no prerequisite for data ordering. and works for all numerical attributes. In K-medoid algorithm, we select an object closest to center of the cluster which will act as the center of the cluster. Choosing any random value ,based on the distance it computes the object closest to the center. The K-medioid algorithm is further improved as PAM(Partition around Medioid),CLARA(Clustering large applications) and CLARANS(Clustering large Applications Based on Randomized Search).C mean algorithm is again a variation of K-mean which generates a given number of cluster based on fuzzy logic.

In **semantic based clustering**, structured patterns are extracted from an unstructured natural language. Various algorithms like Resnick and Lin algorithms[] tries to find the semantic similarity of text in a specific taxonamy.

V. Challenges in Opinion Mining

1)**Detection of spam and fake reviews:** The social media contains both authentic and spam contents. For effective Sentiment classification, this spam content should be eliminated before processing.

2)**Challenges in the development of opinion mining software:** The opinion mining software development is very challenging

3)Domain-independence: The biggest challenge faced by opinion mining and sentiment analysis is the domain dependent nature of sentiment words. One features set may give very good performance in one domain, at the same time it perform very poor in some other domain.

4)Natural language processing overheads: The natural language overhead like ambiguity, co-reference, Implicitness, inference etc. created hindrance in sentiment analysis too.In social networks, textual data may be large, noisy, and dynamic. Moreover, interpreting emoticons (Smile,Sad)for expressing any specific concept or emotion is still a challenging issue. Privacy and trust in online communication is also a major issue. Application of ethical values, such as integrity, veracity, in online communication is the only effective way to build trust online.

To overcome the challenges, researchers need to apply different text mining techniques in social networks that can filter out relevant information from the large text information. However, determining whether to use clustering or classification approach for text analysis in social networks is still a challenging task that totally depends on the dataset and the nature of problem being investigated. In future, text mining tools can also be used as intelligent agent that can mine users personal profiles from social networks and forward relevant information to the users without requiring an explicit request.

VI. Conclusion

Sentiment Analysis or Opinion Mining can be defined as a sub-discipline of computational linguistics that focuses on extracting people's opinion from the web. Whether the organizations like the idea or not, they should be available where their customers are available .The recent expansion of the web encourages users to contribute and express themselves via blogs, videos, social networking sites, etc. All these platforms provide a huge amount of valuable information that we are interested to analyze. CRM models have contributed a lot to improve customer relationships, which will attract new customers as well as retain existing customers. In a analyzing the social media, text mining is required and in this paper a sincere effort is made to understand various text mining techniques to do sentiment or opinion analysis.

REFERENCES

[1]   Chris Rygielski, Jyun-Cheng Wang,David C. Yen "Data mining techniques for customer relationship management",Technology in Society in 2002.

[2]   Timothy Baldwin, "Social Media: Friend or Foe of Natural Language Processing?" in the 26th Pacific Asia Conference on Language,Information and Computation in 2012

[3]   Xia Hu,Huan Liu "Mining Text Data" chapter 12:" Text Analytics in Social Media"

[4]   Gaurav Gupta and Himanshu Aggarwal "Improving Customer Relationship Management Using Data Mining" in the International Journal of Machine Learning and Computing,2012

[5]   Mr. A. A. Sattikar and Dr. R. V. Kulkarni "Natural Language Processing For Content Analysis in Social Networking" in International Journal of Engineering Inventions 2012.

[6]   Haseena Rahmath P " Opinion Mining and Sentiment Analysis - Challenges and Applications" in International Journal of Application or Innovation in Engineering & Management in 2014

[7]   Aurangzeb Khan, Baharum Baharudin, Lam Hong Lee*, Khairullah khan   "A Review of Machine Learning Algorithms for Text-Documents Classification" in the journal of advances in information technology, vol. 1, no. 1, February 2010

[8]   Zhijie Liu ;Xueqiang Lv ; Kun Liu;Shuicai Shi "Study on SVM Compared with the other Text Classification Methods" Published in:Education Technology and Computer Science (ETCS), 2010 Second International Workshop on (Volume:1)

[9]   Luiz Carlos Camargo;Dept. de Inf. (INF), Univ. Fed. do Parana (UFPR), Curitiba, Brazil ;Hegler Correa Tissot;Aurora Trinidad Ramirez Pozo "Use of Backpropagation and Differential Evolution Algorithms to Training MLPs" published in 31st International Conference Chilean Computer Science Society (SCCC), 2012

[10] R i z w a n a i r f a n , C h r i s t i n e k. k i n g , a h m e d s a e e d a l z a h r a n "A Survey on Text Mining in Social Networks" published in the The Knowledge Engineering Review in 2004