

A Review of Various Sentiment Analysis Techniques

Shubhashree Acharya
Department of Computer Engineering
Marwadi University
Rajkot-Gujarat, India
shubhu.achary3@gmail.com

Prof. Manali Modi
Assistant Professor, Department of Information Technologies
Marwadi University
Rajkot-Gujarat, India
manali.modi@marwadieducations.edu.in

Abstract—This paper focuses on the utilization of sentiment analysis techniques in various application domains. Here we present major part of the research work done in the field of sentiment mining or opinion mining using the techniques and tools of sentiment analysis. We get a brief idea regarding the comparison of the techniques and the importance of the data set in acquiring the desired outcomes. This paper gives a comparison on the solutions presented in the research paper.

Keywords-Sentiment Analysis, Education Data Mining, Data Mining

I. INTRODUCTION

Sentiment Analysis is a Detection of attitude. It is getting information about the feelings, emotions, sentiments, views or attitude of the person based on the comment he/she has written or uploaded on any data source (e.g. reviews, forum discussions, blogs, micro blogs, Twitter, and social networks). Others' opinions can be very important in making decisions between various alternatives [12] People depended on the communication with friends, relatives and other people like consultants for the opinion on the product they wanted to buy. But with the increase in the influence and use of social media and other e-commerce websites, people now rely upon the reviews of the products with the opinions of their friends and people. Now these reviews and opinions are analyzed and extracted using Sentiment Mining.

There is high availability of data from multiple sources which are of varying nature. These reviews are not in the structured format so more often users get confused and cannot get a clear idea about the particular product. So, we use sentiment analysis to analyses all the data and make them in a structured format so that users get a clear idea about the particular product. Opinions can be in classified among weighted polarity ex. positive, negative or neutral [8] For this classification, we use different machine learning techniques.

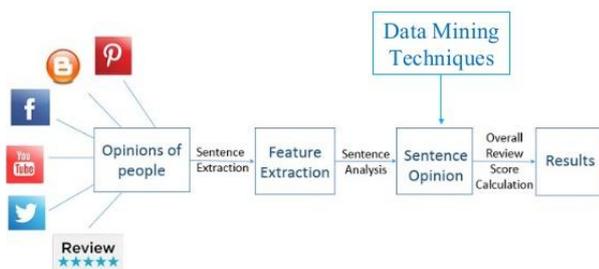


Figure 1 Sentiment Analysis Process

II. APPLICATIONS OF SENTIMENT ANALYSIS

With the growth of internet online market has also grown with a rapid pace. There are many alternatives available for

the same product or service available to the costumer and hence increases the confusion. Customers have to make a decision regarding the choice of product, choice of seller etc. One of the key factor that helps customer take this decision is the reputation of the product which is derived from sentiments/opinions of other users. Customer tends to rely more on the first-hand experiences of other users rather on the advertisements of the product. The reviews are technically termed as sentiments, which can reveal the general perception of the product in the market. Main application of the sentiment analysis is to give a general trend and information about the product based on the opinions of the users. Sentiment Analysis or Opinion Mining is a challenging Text Mining and Natural Language Processing problem for automatic extraction, classification and summarization of sentiments and emotions expressed in online text. [9,10,13]

Many customers also purchase products based on single aspect of the product, for example camera quality of a mobile phone or battery support for a laptop. These are technically termed as aspects of an opinion. Sentiment analysis can also regroup opinions based on these aspects and also give an aspect-based rating for the product, thus helping the user make an easy aspect-based decision.

Another Application of sentiment analysis is for industries who wants to research on the reviews or opinions of customers on their products. Industries tend to improve the product aspects that customers found unsatisfying. Sentiment analysis can also help in pin-pointing the specific problems(aspects) that are the main factors for the customer decisions and hence help industries to focus on those aspects and features of their product making them more likely for the decision.

Sentiment analysis is also used by the mass media industries to extract meaningful information. This information can be used by industries like film industry to track the trends and current status of their movies. Fashion industries can also regularly monitor the reviews and opinions to plan their trends. News industries can use these techniques to predict the outcomes of elections. Sentiment analysis also use in the field of Education. In Education mining we use sentiment analysis for predicting student's

future behavior, their performance, no. of admissions, predicting placement, studying the effects of educational support, and advancing scientific knowledge about learning, also use for teacher’s evolution purpose. This paper mainly focusses on the use of sentiment mining in this field and its practical application in the education sector.

III. TECHNIQUES OF SENTIMENT ANALYSIS ^[11]

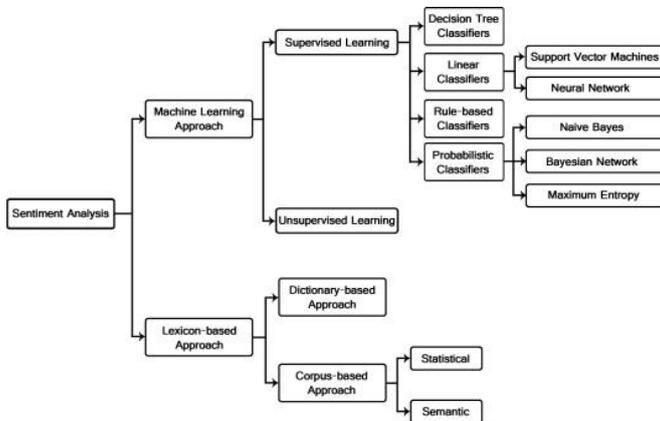


Figure 2 Sentiment Analysis Techniques

A. Machine Learning Approach

The machine learning approach is useful in opinion mining than other approach due to fully automatic completion and its capability to hold large collection of Web Data. Machine Learning-Based Sentiment classification techniques can be classified into two types: supervised, unsupervised learning techniques.

- **Supervised Learning.** Supervised learning is our prime and successful key in traditional topical classification and has been adopted and investigated for opinion detection with satisfactory outcome. Supervised classification techniques are: 1. A linear classifier achieves this by making a classification decision based on the value of a linear combination of the characteristics. The most commonly used linear classifier is Support Vector Machines (SVM). 2. Probabilistic classifier use combination models for classification. Commonly used probabilistic classifier are: a) Naïve Bayes classifier (NB), Bayesian Network (BN) and Maximum Entropy classifier (ME).
- **Unsupervised Learning.** It is anything but difficult to gather labeled and unlabeled data with help of this learning. LDA and pLSA are cases of unsupervised learning hence favorable position of unsupervised learning is to enormous measure of information for precisely preparing reason.

B. Lexicon-Based Approach

The lexicon-based approach based on sentiment lexicon which is integration of known and precompiled sentimental terminology, this is mainly classified dictionary-based approach and corpus-based approach. This approach based on to obtained opinion lexicon which are used in analyzing the text.

- **Dictionary-based Approach.** It depends on finding opinion and then searches dictionary of synonyms

and antonyms. It has prebuilt dictionary that defines semantic orientation of word such as SentiWordNet which is standard dictionary today

- **Corpus-based Approach.** It begins with seed list of opinion words than after find opinion of other word large corpus which help in finding opinion words in context orientation which can help of Statistical and Semantic approach. Statistical approach is for obtained polarity values be corpus which is specific dictionary creation to achieve adoptability to solve problem of unavailability of word uses by large corpus. Semantic approach is a based on a various principle calculating the similarities between show its provide sentiment values directly.

C. Hybrid Based Approach

It is a combination of machine learning and lexicon-based approach and is frequently use with sentimental lexicon express major role in different important method.

IV. LITERATURE REVIEW

K. Purthi and Dr.P. Bhatia [1] predict final placement of the student and classify them based on the prediction whether the student will be placed in a core IT company or a Consultancy firm. They applied 3 data mining techniques (NB, clustering, classification) on the data set acquired from the placement department of Thapar University and the student academic as well as personal details from the college. They were successful in predicting the outcome with 95% accuracy. This research help student analyzes their performance and also improve in the areas that can give them their desired outcome. Such research also helps to increase the placement for the institutes and hence it is also beneficial for the institute. However, this research was conducted on a limited scale of data set and hence may need to revise for the future use. They focused on the data only from the department of Computers and IT and hence they can scale up combining the data from all the branch students.

Nabeela Altrabsheh and et.all [2] learn sentiment from real student feedback using NB, CNB, ME, SVM compare their results using parameters such as Accuracy, Precision, Recall and F-score. They also experimented this with and without neutral class at University of Portsmouth. Here propose system automatically analyzes student’s feedback in real-time and presents it to lecturer whereas gives lecturers the most important information from the feedback. They suggest to use and analyze more preprocessing techniques and their impact on model performance, also use models with more real time collected data.

C. Lin and Y. [3] He proposes a novel probabilistic modeling framework based on Latent Dirichlet Allocation (LDA) [13] called joint sentiment/topic model (JST), which detects sentiment and topic simultaneously from a text. JST model is fully unsupervised, thus provides more flexibilities and can be easier adapted to other applications. It represents each document as a bag of words and thus ignores the word order. It will probably predict the sentiment of “not good movie” being positive and the sentiment of “not bad movie” being negative.

Bin liu [4] investigate power of topic model based on 2 multi-aspect sentiment analysis tasks: First aspect sentence labeling and second aspect rating prediction. Similar techniques are investigated by N. Jadeja and A. Pandya that demonstrate that weakly supervised topic models perform well on multi-aspect sentence labeling [14]. Multi-Aspect Sentiment Analysis, that aims to take into account various, potentially related aspects often discussed within a single review. Labeling propose approach with minimal prior knowledge in the form of seed words and get direct correspondence b/w topics and aspects & in prediction they find that overall ratings can be used in conjunction with our sentence labeling to achieve reasonable performance compared to a fully supervised baseline. They use LDA (model word cooccurrence at document level), local LDA (at sentence level), STM (segmented topic model extension of local LDA sentence level topic distribution. Features derived from unsupervised topic model provides substantial increase in performance but this diminished, when topic model is paired with more competitive supervised baseline. All this analysis applied on online available datasets (OpenTable.com, CitySearch.com, TripAdvisor).

Wayne Xin Zhao, et all [5] empirically compare the content of Twitter with a traditional news Medium (New York Times) using unsupervised topic modeling. They use a Twitter-LDA model to discover topics from a representative sample of the entire Twitter., also study the relationship between the proportions of opinionated tweets and retweets and topic categories and types. Twitter is a valuable source for entertainment and lifestyle topics. Retweets can be used to indicate trendy topics among Web users to help search engines refine their results. In future, they aim to summarize and visualize Twitter content in a systematic way. Method of associating tweets with different categories and types may also help visualization of Twitter content. Edinburgh Twitter Corpus as Twitter dataset was used for time frame of November 11, 2009 to February 1, 2010. New York Times (NYT) as source of news articles.

D. Mimno, A. McCallum [6] Present probabilistic model that ranks authors based on their influence in particular areas of scientific research. Combines several sources of information: citation information between documents as represented by PageRank scores, authorship data gathered through automatic information extraction, and the words in paper abstracts. compare the performance of a topic model versus a smoothed language model. This paper helps the researcher to find most influential authors based on the area they work in. however, they use a limited scale corpus and can widen their research using a larger scale corpus.

I. Titov and R. McDonald[15] present their work in identifying product aspects from online sources like social media sites and micro blogging sites. They provide us with a framework that extracts product aspects automatically from the user reviews. They argue that multi-grain models are more efficient for the task as standard models generates topics that represent the global properties of objects (e.g., the brand of a product type) rather than the aspects of an object that tend to be rated by a user.

A. Comparison for existing sentiment analysis techniques

The table below represents the overall comparison of various sentiment analysis or opinion mining techniques. Also, the table highlights various algorithms used by authors based on application domains of their research. Various parameters such as accuracy, entropy and recall ratio are also mentioned in the table to represent the need of specific algorithms.

Table 1 Comparison of Existing Solutions

Paper Name	Algorithm	Parameters	Dataset	Paper Name
Application of Data Mining in Predicting Placement of Students [1]	NB, clustering	Accuracy	Placement Details of Thapar University	NB, clustering
Learning Sentiment from Students' Feedback for Real-Time Interventions in Classrooms [2]	NB, CNB, ME	Accuracy Precision Recall Entropy	Lectures at University of Portsmouth (PG & UG student)	Lectures at University of Portsmouth (PG & UG student)
Joint Sentiment/Topic Model for Sentiment Analysis [3]	LDA	Accuracy	LingPipe Package (Movie Review)	LingPipe Package (Movie Review)
Multi-aspect Sentiment Analysis with Topic Models [4]	LDA	Accuracy	OpenTable.com, CitySearch.com, TripAdvisor (Restaurant)	OpenTable.com, CitySearch.com, TripAdvisor (Restaurant)
Comparing Twitter and Traditional Media using Topic Models [5]	LDA Twitter-LDA	Accuracy	Edinburgh Twitter Corpus as Twitter dataset in November 11, 2009 to February 1, 2010. New York Times (NYT) as source of news articles	Comparing Twitter and Traditional Media using Topic Models [5]
Mining a Digital Library for Influential Authors [6]	PageRank, LDA	PageRank	Rexa Digital Library	Mining a Digital Library for Influential Authors [6]

<i>Paper Name</i>	<i>Algorithm</i>	<i>Parameters</i>	<i>Dataset</i>	<i>Paper Name</i>
Aspect Extraction through Semi-Supervised Modelling [7]	SAS, ME-SAS	Entropy	Tripadvisor.com (Hotel Review)	Aspect Extraction through Semi-Supervised Modelling [7]

LDA: Assumes: data is normally distributed. All groups are identically distributed, in case the groups have different covariance matrices, LDA becomes Quadratic Discriminant Analysis. LDA is the best discriminator available in case all assumptions are actually met. QDA, by the way, is a non-linear classifier. Naive Bayes is a popular algorithm for classifying text. Although it is fairly simple, it often performs as well as much more complicated solutions.

Applying these techniques to the analysis of the sentiment analysis in education sector gives us some astonishing outcomes. This result is significant if we look into the overall scale of education as an industry and hence can become critical in the growth of it as a sector for research as well as a corporate sector overall.

V. Conclusions

Sentiment analysis is wide domain and get the relevant decision-making patterns or understanding the data generated in huge numbers can be tricky. But if right algorithm is applied with proper preprocessing then it can be very. This paper discusses about different sentiment analysis and opinion mining techniques keeping in mind various application domains. As a part of future work, we would like researchers to apply sentiment analysis techniques in education domain and specifically feedback systems to get better outcome for education institutes.

VI. REFERENCES

- [1] Karan Purthi, Dr. Parteek Bhatia “Application of Data Mining in Predicting Placement of Students”, IEEE 2015
- [2] Nabeela Altrabsheh, Mihaela Cocea, and Sanaz Fallahkhair, ” Learning Sentiment from Students’ Feedback for Real-Time Interventions in Classrooms”, Springer 2014
- [3] Chenghua Lin, Yulan He “Joint Sentiment/Topic Model for Sentiment Analysis”, ACM conference in Information and knowledge management,2009
- [4] Bin liu, “Multi-aspect Sentiment Analysis with Topic Models”, IEEE International Conference on Data Mining Workshops
- [5] Wayne Xin Zhao, Jing Jiang, Jianshu Weng, Jing He1, Ee-Peng Lim, Hongfei Yan and Xiaoming Li, “Comparing Twitter and Traditional Media using Topic Models”, Springer 2011
- [6] David Mimno, Andrew McCallum “Mining a Digital Library for Influential Authors”, ACM 2007
- [7] Bing Liu, Arjun Mukherjee, “Aspect Extraction through Semi-Supervised Modelling”, 50th Annual Meeting of the Association for Computational Linguistics
- [8] Bing Liu. “Sentiment Analysis and Opinion Mining”, Morgan & Claypool Publishers, May 2012
- [9] Indurkha N., Damereau F.J., (Eds). 2010. Handbook of Natural Language Processing. 2nd Ed., Chapman & Hall/CRC, Boca Raton
- [10] K. Khan, B.B. Baharudin, A. Khan, F. e-Malik, Mining opinion from text documents: a survey, Proc. of the 3rd IEEE International Conference on Digital Ecosystems and Technologies, pp.217-222, 2009.