

An Opinion Mining and Sentiment Analysis Techniques: A Survey

S. Kasthuri¹, Dr. L. Jayasimman², Dr. A. Nisha Jebaseeli³

¹Department of Computer Science, Srimad Andavan Arts & Science College, Trichy, India.

² Department of Computer Science, Srimad Andavan Arts & Science College, Trichy, India.

³Department of Computer Science, Bharathidasan University Constituent College, Lalgudi, Trichy, India.

Abstract - *Opinion Mining is a type of natural language processing for tracking the mood of the people about any particular product by review. The Opinions are collected from public, it considered as most valuable data. The opinions are reviews from customers; comments are collected from web sites and user groups. The collected opinions are manipulated by various techniques, methods, algorithms and software tools to get the opinions from them. This process is also called as Opinion Mining or Sentiment Analysis. Opinion analysis is very interesting research topic in both extraction of information and discovery of knowledge. Opinion mining can be used in many new applications. This paper discusses an overview of the topic data sources used for Opinion Mining, basic components of Opinion Mining, different levels of Sentiment Analysis and tools used for Sentiment Classification.*

Key Words: Data Mining, Opinion Mining, Sentiment Analysis, NLP.

1. INTRODUCTION

Natural Language Processing (NLP) discusses with actual text element processing. The text element is transferred into machine format by NLP. An Opinion Mining is a type of natural language processing for tracking the mood of the people about any particular product. The Opinion Mining is also called as many different names such as Sentiment Analysis, Sentiment mining, Opinion extraction, Review mining etc. [1]. The Opinion Mining is not an important thing for a user but it is necessary for an organization / company growth. An Opinion Mining plays an important and crucial role in business sector. A manufacturer can improve and modify their product design, quality, sales and service by the use of customer reviews and opinions [2]. Opinion Mining is a method, which is used to collect and study necessary information's from Internet forums, search engines, web blogs and social networks about any particular product of the company or organization.

The main challenge in this area is the classification in which the Opinion / Review may be a judgement, mood or evaluation of an object namely camera, laptop, book, etc which can be in the form of document or sentence or feature that can be labelled as positive or negative or neutral.

1.1 Overview: Basic Components of an Opinion:

The Opinion Mining is mostly integrated with the topic Information Retrieval (IR). The Information Retrieval works

on actual data but the Opinion Mining works on subjective data [3].

The basic components of an Opinion Mining are,

1. Opinion: It is a view, suggestion or sentiment about the specific object done by the user.

2. Opinion holder: This is the person who gives a specific opinion about an object.

3. Object : It is an entity, which is felt by the user.

1.2 Evaluation of Opinion:

The Evaluation of Opinion is divided into two types.

1. Direct Opinion: It talks about only a single object. It gives positive or negative expressions about an object, product, topic or person directly [4]. For example, "This Bike has poor mileage" expresses a direct opinion about a single object ie. Bike.

2. Comparison Opinion: It talks about multiple objects. This type of opinion expresses similarities or differences between more than one object. For example, "Car X is better than Y" expresses a comparison opinion between two objects ie. Car.

2. Data Source:

People and companies across good trainings exploit the high and unique sources of data for various purposes. User Opinion is an important criterion for the progress of the quality services. Blogs, review sites, Data set and micro-blogs furnish a good understanding for the deliverable view of the products and services provided to the customers / viewers [5].

2.1 Blogs :

With an growing usage of the Internet, Blog pages and blogging are mostly used. The names connected to universe of all blog sites are called blogosphere [6]. Blog pages are used to express one's personal opinions about any product or topic. People like to share their opinions, ideas or suggestions with others on a blog.

Blogging is a occurrence thing because of its simplicity of creating blog posts and reviews, its free form and unedited nature. Blogs are used as a source of opinion in many of the studies linked with sentiment analysis.

2.2 Review Sites:

For any user in deciding a purchasing decision, others opinion is an important factor. The user generated reviews

and suggestions are largely available on the Internet. The reviews for products or services are available as opinions in unstructured format. The reviewers data are used in Sentiment classification studies are collected from the e-commerce websites like www.yelp.com (restaurant reviews), www.amazon.com (product reviews), www.flipkart.com (product reviews), which hosts millions of product reviewed by customers.

2.3 Data Set :

Many works in the field uses movie reviews data for classification. The Multi-Domain Sentiment (MDS) contains different types of product reviews taken from Amazon.com, Flipkart.com including Books, dresses, Kitchen appliances and Electronics things, with many positive and negative suggestions / reviews for each territory.

2.4 Micro-blogging:

Micro-blogging is the well-known communication tool for internet users. A large number of messages appear daily in web-sites for micro-blogging such as Twitter, Tumblr and Facebook. Twitter is very popular micro-blogging service where users express messages called "tweets". These Tweets are used to express their own opinions/suggestions about different topics. Sometimes these Twitter messages are also used as data source for Sentiment Classification.

3. Sentiment classification:

Sentiment classification is the binary classification task of justifying an opinionated document as expressing either an overall positive or an overall negative opinion. The machine learning algorithms used in sentiment classification methods.

3.1 Machine Learning :

A system capable of getting, integrating and analyzing the knowledge automatically is known as Machine Learning. This type of Machine Learning method is mostly used in Opinion mining and Sentiment Analysis. In Sentiment analysis, machine learning approach is belongs to text classification techniques and generally used in supervised classification technique. Thus, it is also known as 'Supervised Learning'. In this machine learning based classification, two sets of documents are needed. They are training set and test set. The training set is used to learn the differentiating characteristics of documents by an automatic classifier, and a test set is used to check the performance of the automatic classifier. Machine Learning techniques like Naive Bayes (NB), K-Nearest Neighbor (KNN) have obtained great success in text categorization.

Naive Bayes (NB) is a simple but effective Learning & Classification algorithm. It is mostly used in Text Classification. The Classification method is based on theory of probability. It plays a vital role in probabilistic classification. It is also used in statistical method for

classification and Supervised Learning method. When Bayesian classifiers applied to large databases, it has exhibited high accuracy. Naive Bayes Classification method is easy to implement. It requires only a small set of practical training data to judge a standard quantity which satisfies a particular set of equations. In most of the cases, good results are acquired through this classification method.

K-Nearest Neighbor (KNN) is a simplest algorithm of all machine learning algorithms. It is also referred as Lazy Learning, Case-based Reasoning or Memory-based Reasoning. KNN is simple, it yet able to solve most complicated problems. It is a non-parametric method used for classification [7].

3.2 Different Levels of Sentiment Analysis :

Sentiment Analysis tasks are mainly divided into the polarity of a given text at the document, sentence and feature level / attribute level / aspect level / phrase level to find whether it give positive opinion, negative opinion or neutral. This is also referred as 'Sentiment Polarity Prediction' [8]. The Sentiment Analysis performance is carried out into three levels,

- i) The document level
- ii) The Sentence Level
- iii) The Feature Level

3.2.1 Document Level Sentiment Classification:

It is about classifying the overall opinionated text presented by the authors in whole document as positive, negative or neutral about a certain subject or object. Therefore subjectivity / objectivity classification is important in this type of Sentiment Classification [9]. The main challenge in this classification is to extract informative text for deducing sentiment of the entire document.

3.2.2 Sentence Level Sentiment Classification:

In this type of classification, the polarity of each sentence is calculated. It is a fine-grained level than the document level sentiment classification. The sentence level sentiment classification is connected with two jobs. First one is to recognize whether the given sentence is objective or subjective opinionated. The Second one is to discover opinion of an opinionated sentence as positive, negative or neutral. Like the document classification, the sentence classification does not think about object features that have been commented in a sentence [10].

3.2.3 Feature Level Sentiment Classification:

This level of sentiment classification is a much more pinpointed method to opinion mining. This type of classification considers the opinions on features of particular objects. Features of the product are defined as attributes, components and other aspects of the product, Analysis of such features are recognizing sentiment of the document is called as Feature based Sentiment Analysis [11]. The task of Feature Level sentiment classification is to extract the

features of the commented object and after that conclude the opinion of the object. Positive or negative and then group the feature synonyms and make the summary report.

4. Tools:

The tools are used to trace the opinion or polarity from the user generated texts. They are,

1. Red Opal : It is a tool that makes able the users to find products based on attributes / features. It assigns the points / ranks to each product based on their features, which are extracted from the customer generated reviews. The extracted features are displayed in graph format. The extracted features / attributed are assigned rank by Naive Bayes Classifier as positive and negative review. The results are displayed in the form of attributes and its score.

2. Web Fountain: It is used to create a simple web interface. It uses the beginning definite Base Noun Phrase (bBNP) heuristic method to extract the features of the product.

3. Review Seer Tool: It is used to automate the work performed by aggregation sites. The Naive Bayes Classifier method is used to collect positive and negative opinions from customer reviews for assigning a rank to the extracted features [12].

4. Opinion Observer: This Opinion Mining system is used for analyzing and comparing customer generated opinions on the Internet. This system displays the results of a product feature by feature in a graph format.

5. Conclusion:

This paper provides about an overview of opinion mining and sentiment analysis in detail with the data sources, components and tools.

In the name of Opinion Mining and Sentiment Analysis the large number of tasks are used, various techniques and methods are being followed by many researchers based on domains and new applications.

The challenges in Sentiment Analysis the opinion words / sentences are also occur in objective sentences or subjective information. It is difficult to distinguish between objective sentences and subjective informations. The techniques and many algorithms applied for Opinion Mining are very fast, and many of the studies are remain unsolved. More future research works could be committed to these difficulties and challenges.

References:

- [1] A. Nisha Jebaseeli, Dr. E. Kirubakaran "M-Learning Sentiment Analysis with Data Mining Techniques", International Journal of Computer Science And Telecommunications, Volume 3, Issue. 8, Aug. 2012.
- [2] Arti Buche Dr.M.B. Chandak, Akshay Zadgaonkar, "Opinion Mining And Analysis : A Survey", International Journal on Natural Language Computing (IJNLC), Volume 2, No. 3, June 2013.
- [3] "Data Mining Concepts and Techniques" Jiawei Han, Micheline Hamber Morgan Kaufman Publishers, 2003.
- [4] G. Angulakshmi, Dr. R. Manickachezian, "An Analysis on Opinion Mining : Techniques and Tools", International Journal of Advanced Research in Computer and Communication Engineering, Volume 3, Issue. 7, July 2014.
- [5] Nidhi Mishra, Dr. C.K. Jha, "Classification of Opinion Mining Techniques", International Journal of Computer Applications, Volume 56, No.13, October 2012.
- [6] Sumathi.T, Karthik.S, Marikannan.M "Performance Analysis of Classification Methods for Opinion", International Journal of Innovations in Engineering and Technology (IJJET), Volume 2, Issue 4, August 2013.
- [7] G. Vinodhini, RM.Chandrasekaran, "Sentiment Analysis and Opinion Mining: A Survey", International Journal of Advanced Research in Computer and Communication Engineering, Volume 2, Issue 6, June 2012.
- [8] K.G. Nandakumar, Dr.T.Christopher "Opinion Mining: A Survey", International Journal of Computer Applications, Volume 113, No.2, March 2015.
- [9] Zhu Zhang, 2008 Weighing Stars, "Aggregating Online Product Reviews For Intelligent E-Commerce Applications", IEEE Intelligent Systems, 42-49.
- [10] Ziqiong Zhang, Qiang Ye, Zili Zhang, Yijun Li, "Sentiment Classification of Internet Restaurant Reviews written in Cantonese", Expert Systems with Applications, 2011.
- [11] B.Liu. 2010 "Sentiment Analysis and Subjectivity", Second Edition, The Handbook of Natural Lanugage Processing.
- [12] N.Mishra and C.K.Jha 2012, "An Insight into task of opinion mining", Second International Joint Conference on Advances in Signal Processing and Information