

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 5, May 2016

Fuzzy Aspects in Sentiment Analysis and Opinion Mining

Dr. Ranjana Rajnish

Assistant Professor, Amity Institute of Information Technology, Amity University, Lucknow, Uttar Pradesh, India

ABSTRACT: “What other people think” has always been an important piece of information for most of us during the decision-making process. Dependence on internet based reviews for this decision making has become important in this era. Long-time back before emergence of Web 2.0 when an individual wanted to buy any product or utilize any service, he used to ask or discuss with the friends/relatives about the same & then take decision accordingly. But, nowadays Social media provides Terabytes of information about every product or service. With the growing availability and popularity of opinion-rich resources such as online review sites and personal blogs, new opportunities and challenges arise as people now actively use information technologies to seek out and understand the opinions of others. Opinion Mining and Sentiment Analysis have thus become more important and significant. Opinions are fuzzy in nature and can be expression between beyond true and false or 0 or 1. Fuzzy sets constitute the oldest and most reported soft computing paradigm [16]. They are well-suited to modelling different forms of uncertainties and ambiguities, often encountered in real life. Integration of fuzzy sets with other soft computing tools has led to the generation of more powerful, intelligent and efficient systems.

This paper discusses about different levels of opinion mining, sentiment lexicon, issues and challenges in Opinion Mining & Sentiment Analysis, concept of Fuzzy logic and how well it suits to the problem of OM & SA.

KEYWORDS: Opinion Mining, Sentiment Analysis, Sentiment Lexicon, Fuzzy Logic & Fuzzy Sets.

I. INTRODUCTION

Emotions are the complex state of feelings that that can not be represented by traditional boolean logic. Emotion is a subjective, conscious experience characterized mainly by psycho-physiological expressions, biological reactions, and mental states. Emotion is commonly associated significant with mood, nature, personality, disposition, and motivation. Emotion is a positive or negative experience that is associated with a particular pattern of physiological activity. Humans carry lot of emotions like happiness, sadness, angry, disgust, surprise, panic, etc.

Term Sentiment Analysis first appeared in (Nasukawa& YI 2003) and the term opinion mining first appeared in (Dave,Lawrence & Pennock 2003).Opinion Mining also called as sentiment analysis is a field that is used to extract the users opinions about any product, service or any topic. Long time back before emergence of Web 2.0 when an individual want to buy any product or utilize any service, he asks or discuss with the peer/friends about particular product & then take decision accordingly. But nowadays Social media emerge as a boom with Terabytes of data available about every product or a service. Opinion mining main objective is to classify the reviews expressed by users into positive or negative & then present it in summarize form that is understood by user. Different Data Sources like Blogs, Social network Site, E-commerce sites are available where users can post their reviews about particular topic(Like Modi's campaign of Swach Bharat) or any product like Mobiles, cameras etc. In order to promote marketing large companies and business people are making use of Opinion Mining[1],thus we can say that opinion mining and Sentiment Analysis had tremendous opportunity in every domain.

The growing importance of sentiment analysis coincides with the growth of social media such as Reviews, Forums, discussion groups, chatting, blogs, micro-blogs, twitter and social networks.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 5, May 2016

II. OPINION MINING & SENTIMENT ANALYSIS

Opinion Mining is the field of study that analyzes people's opinion, sentiments, evaluations, attitudes and emotions from written text. Opinion Mining is one of the most active research areas in Natural Language Processing and is also widely studied in data mining, web mining and text mining this research has spread outside of computer science to the management science and social science due to its importance to business and society.

Opinion Mining is used to extract information from various documents at different levels. According to Bing Liu, sentiment analysis has been investigated mainly at three levels.

- a. Document level: The task at this level is to classify whether a whole opinion document expresses a positive or negative sentiment (Pang et al., 2002; Turney, 2002). For example, given a product review, the system determines whether the review expresses an overall positive or negative opinion about the product. This task is commonly known as *document t level sentiment classification*. This level of analysis assumes that each document expresses opinions on a single entity (e.g., a single product). Thus, it is not applicable to documents which evaluate or compare multiple entities.
- b. Sentence level: At this level the task goes to the sentences and determines whether each sentence expressed a positive, negative, or neutral opinion. Neutral usually means no opinion.

This level of analysis is closely related to *subjectivity classification* (Wiebe et al. 1999), which distinguishes sentences (called *objective sentences*) that express factual information from sentences (called *subjective sentences*) that express subjective views and opinions.

However, subjectivity is not equivalent to sentiment as many objective sentences can imply opinions, e.g., “*We bought the car last month and the windshield wiper has fallen off.*”

- c. Entity and Aspect level: Aspect level performs finer-grained analysis. Aspect level was earlier called *feature level (feature-based opinion mining and summarization)* (Hu and Liu, 2004).

Instead of looking at language constructs (documents, paragraphs, sentences, clauses, or phrases), aspect level directly looks at the opinion itself. It is based on the idea that an opinion consists of a *sentiment* (positive or negative) and a *target* (of opinion). An opinion without its target being identified is of limited use. Realizing the importance of opinion targets also helps us understand the sentiment analysis problem better. For example, although the sentence “*Although the service is not that great, I still lovethis restaurant.*” clearly has a positive tone, we cannot say that this sentence is entirely positive. In fact, the sentence is positive about the *restaurant* (emphasized), but negative about its *service* (not emphasized).

This is unlike the document-level and sentence-level analyses that do not discover what exactly people liked and did not like.

In many applications, opinion targets are described by entities and/or their different aspects. Thus, the goal of this level of analysis is to discover sentiments on entities and/or their aspects. For example, the sentence “*The iPhone's call quality is good, but its battery life is short.*” evaluates two aspects: *call quality* and *battery life*, of *iPhone* (entity). The sentiment on iPhone's *call quality* is positive, but the sentiment on its *battery life* is negative. The *call quality* and *battery life* of iPhone are the opinion targets. Based on this level of analysis, a structured summary of opinions about entities and their aspects can be produced, which turns unstructured text to structured data and can be used for all kinds of qualitative and quantitative analyses.

Both the document-level and sentence-level classifications are already highly challenging.

To make things even more interesting and challenging, there are two types of opinions, i.e., *regular opinions* and *comparative opinions* (Jindal and Liu, 2006b). A regular opinion expresses a sentiment only on a particular entity or an aspect of the entity, e.g., “*Coke tastes very good,*” which expresses a positive sentiment on the aspect *taste* of Coke. A comparative opinion compares multiple entities based on some of their shared aspects, e.g., “*Coke tastes better than Pepsi.*” which compares Coke and Pepsi based on their tastes (an aspect) and expresses a preference for Coke.

III. ISSUES RELATED TO SENTIMENT LEXICON

Not surprisingly, the most important indicators of sentiments are sentiment words, also called opinion words. These are words that are commonly used to express positive or negative sentiments. For example, good, wonderful, and amazing are positive sentiment words, and bad, poor, and terrible are negative sentiment words. Apart from individual words,

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 5, May 2016

there are also phrases and idioms, e.g., cost someone an arm and a leg. Sentiment words and phrases are instrumental to sentiment analysis for obvious reasons. A list of such words and phrases is called a sentiment lexicon (or opinion lexicon). Researchers have designed numerous algorithms to compile such lexicons.

Although sentiment words and phrases are important for sentiment analysis, only using them is far from sufficient. The problem is much more complex. In other words, we can say that sentiment lexicon is necessary but not sufficient for sentiment analysis. Below, we highlight several issues.

1. A positive or negative sentiment word may have opposite orientations in different application domains. For example, "suck" usually indicates negative sentiment, e.g., "*This camera sucks.*" but it can also imply positive sentiment, e.g., "*This vacuum cleaner really sucks.*"
2. A sentence containing sentiment words may not express any sentiment. This phenomenon happens frequently in several types of sentences. Question (interrogative) sentences and conditional sentences are two important types, e.g., "*Can you tell me which Sony camera is good?*" and "*If I can find a good camera in the shop, I will buy it.*" Both these sentences contain the sentiment word "good," but neither expresses a positive or negative opinion on any specific camera. However, not all conditional sentences or interrogative sentences express no sentiments, e.g., "*Does anyone know how to repair this terrible printer?*" and "*If you are looking for a good car, get Toyota Camry.*"
3. Sarcastic sentences with or without sentiment words are hard to deal with, e.g., "*What a great car! It stopped working in two days.*" Sarcasms are not so common in consumer reviews about products and services, but are very common in political discussions, which make political opinions hard to deal with.
4. Many sentences without sentiment words can also imply opinions. Many of these sentences are actually objective sentences that are used to express some factual information. Again, there are many types of such sentences. Here we just give two examples. The sentence "*This washer uses a lot of water.*" implies a negative sentiment about the washer since it uses a lot of resource (water). The sentence "*After sleeping on the mattress for two days, a valley has formed in the middle.*" expresses a negative opinion about the mattress. This sentence is objective as it states a fact. All these sentences have no sentiment words.

These issues present major challenges. One more very challenging issue is of Sentiment analysis is that it is also a problem of NLP. It touches every aspect of NLP, e.g., coreference resolution, negation handling, and word sense disambiguation, which add more difficulties since these are not solved problems in NLP. However, it is also useful to realize that sentiment analysis is a highly restricted NLP problem because the system does not need to fully understand the semantics of each sentence or document but only needs to understand some aspects of it, i.e., positive or negative sentiments and their target entities or topics.

IV. FUZZY BASED ASPECTS OF OM & SA

A. What is fuzzy logic?

Fuzzy logic is an approach to computing based on "degrees of truth" rather than the usual "true or false" (1 or 0) Boolean logic on which the modern computer is based. It is thus superset of conventional (Boolean) logic that has been extended to handle the concept of partial truth -- truth values between "completely true" and "completely false". It was introduced by Dr. Lotfi Zadeh of UC/Berkeley in the 1960's as a means to model the uncertainty of natural language [Rajnish et al 2015].

Zadeh says that rather than regarding fuzzy theory as a single theory, we should regard the process of —fuzzification as a methodology to generalize ANY specific theory from a crisp (discrete) to a continuous (fuzzy) form [16]. It is a new way of representing vagueness in everyday life.

Fuzzy approach is based on the premise that key elements in human thinking are not just numbers but can be approximated to tables of fuzzy sets, or, in other words, classes of objects in which the transition from membership to non-membership is gradual rather than abrupt. Much of the logic behind human reasoning is not the traditional two-valued or even multivalued logic, but logic with fuzzy truths, fuzzy connectives, and fuzzy rules of inference. This fuzzy logic plays a basic role in various aspects of the human thought process [17].

Fuzzy Subsets:

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 5, May 2016

Just as there is a strong relationship between Boolean logic and the concept of a subset, there is a similar strong relationship between fuzzy logic and fuzzy subset theory. In classical set theory, a subset U of a set S can be defined as a mapping from the elements of S to the elements of the set {0, 1},

$$U: S \rightarrow \{0, 1\}$$

This mapping may be represented as a set of ordered pairs, with exactly one ordered pair present for each element of S. The first element of the ordered pair is an element of the set S, and the second element is an element of the set {0, 1}. The value zero is used to represent non-membership, and the value one is used to represent membership. The truth or falsity of the statement x is in U is determined by finding the ordered pair whose first element is x . The statement is true if the second element of the ordered pair is 1, and the statement is false if it is 0. Similarly, a fuzzy subset F of a set S can be defined as a set of ordered pairs, each with the first element from S, and the second element from the interval [0,1], with exactly one ordered pair present for each element of S. This defines a mapping between elements of the set S and values in the interval [0,1].

The value zero is used to represent complete non-membership, the value one is used to represent complete membership, and values in between are used to represent intermediate DEGREES OF MEMBERSHIP. The set S is referred to as the UNIVERSE OF DISCOURSE for the fuzzy subset F. Frequently, the mapping is described as a function, the MEMBERSHIP FUNCTION of F. The degree to which the statement

$$x \text{ is in } F$$

is true is determined by finding the ordered pair whose first element is x . The DEGREE OF TRUTH of the statement is the second element of the ordered pair.

In practice, the terms "membership function" and fuzzy subset get used interchangeably.

Fuzzy expert systems are the most common use of fuzzy logic. They are used in several wide-ranging fields such as Linear and Nonlinear Control, Pattern Recognition, Data Analysis, Variety of applications ranging from consumer products such as cameras, camcorders, washing machines, and microwave ovens to industrial process control, medical instrumentation, decision-support systems, and portfolio selection have been built based on Fuzzy Logic.

B. Fuzzy logic in OM

Opinion words are fuzzy in nature. For example, the words "Nice", "Good", and "Very Good" and the boundaries among them are not clear. Hence, Fuzzy logic can easily represent these types of subjective words and assign to classes with some degree of membership. Thus these words are already in fuzzification stage and defining fuzzy sets for such words needs to be done on the basis of some expert opinions. Since opinions are fuzzy in nature and meaning of opinion words can be interpreted differently, Fuzzy logic can be seen as an effective technique to be considered here to properly extract, analyze, categorize and summarize opinions. This due to the following reasons:

- Fuzzy logic is conceptually flexible, easy to understand and it is build to handle imprecise data like opinion words (Alina& Sabine 2006).
- Fuzzy logic is based on natural language and hence very suitable to resolve the fuzziness in human expressed phrases (Animesh and Deba et. al. 2011)
- Sentiment classification in many recent works employs supervised machine learning techniques like SVM and Naïve Bayesian (NB). Though these methods showed some good performance on topic-based text classifications; however, results obtained in sentiment classification are far from satisfactory. The traditional machine learning methods cannot perform well enough in sentiment analysis as concluded by most recent studies (Wilson et al 2004). This is because the opinion text does not clearly show or indicate which polarity classes they belong to. Moreover, sometimes sentiment orientation of the subjective text is dependent on context or domain for which opinions are expressed. (Pang and Lee 2008), Sometimes this might be due to the fact that subjective text is very vague and it is very difficult to make a clear boundary between positive and negative sentiments (Guohong and Xin 2010). This shows that we need more effective tools and techniques in addressing and better understanding such unclear (fuzzy) texts.
- Fuzzy logic is an intelligent control Technique which relies on human-like expert knowledge using IF-THEN reasoning rules. Such rules are based on sets that have flexible membership functions rather than just the normal crisp binary logic. Moreover, Fuzzy set theory offers a better straightforward and a simpler ways to present the intrinsic fuzziness in sentiments (Subasic, and Huettner 2001).
- Existing opinion mining techniques and approaches can classify opinions as positive, negative and neutral classes only. Opinion mining approaches like holistic lexicon approach does not allow classifying reviews granularity in

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 5, May 2016

order to determine the strength of each opinion. There is a need to increase the classifications of opinions and assign weightages for different opinion words. Only Fuzzy logic (via Fuzzy sets and Fuzzy rules) can add such a dimension to properly analyze opinions and classify them at different strengths. Very few researches have been done on using Fuzzy logic to classify sentiments and their strengths (Samaneh et. al. 2010).

- Fuzzy logic allows better classifications of sentiments with proper strength assigned to each opinion level. This will help to increase the accuracy of classifications (Animesh and Deba et. al. 2011).
- Subjective words are fuzzy in nature especially when it comes to opinion mining. Because opinions are always expressed in a fuzzy manner for example nice food, nice video ,huge building and so on and in such cases it becomes difficult to understand the level of fuzziness whether it is too much nice or too nice or too huge. The concept of a Fuzzy Logic is one that it is very easy for the ill-informed to dismiss as trivial and/or insignificant. It refers not to a fuzziness of logic but instead to logic of fuzziness, or more specifically to the logic of fuzzy sets.

V. ISSUES IN FUZZY BASED APPROACH

The most important indicators of sentiments are the words expressing sentiments, i.e. *sentiment words*, also called *opinion words*. These words are commonly used to express positive or negative sentiments. For example, *good*, *wonderful*, and *amazing* are positive sentiment words, and *bad*, *poor*, and *terrible* are negative sentiment words. Apart from individual words, there are also phrases and idioms, e.g., *cost someone an arm and a leg*. Sentiment words and phrases are instrumental to sentiment analysis for obvious reasons. A list of such words and phrases is called a *sentiment lexicon* (or opinion lexicon).

Some of the challenges and issues that need to be addressed to enhance fuzzy logic based systems (Sheroz et. al. 2007, Bakhtawar&, Farouque 2012; Vinodhini&Chandrasekaran 2012)

- Calculating opinion strength more accurately is an important area that needs more research. SentiWordNet provides score for subjectivity, positive and negative dimension of a subjective word. However, how to combine these values to arrive to more realistic value reflecting the strength an opinion word is still needs to be improved. One effective way is to use Fuzzy logic as the opinion words are fuzzy in nature and by using the defuzzification process a better crisp value can be arrived at to mathematically represent the opinion strength.

- Existing Fuzzy-based systems used only adjectives and adverbs as opinion words.

Nouns and verbs can also express opinion. Scores assigned to opinion words are manual and does not have proper justification behind it and these weights are not based on scientific methods. This may result in wrong opinion calculations and rankings.

- Thesaurus and ontologies like WordNet and SentiWordNet are essential tools to be used as integral components to enhance opinion extraction and scoring.

- Existing Fuzzy-based systems do not address properly problems like negations, vague words and ambiguous words.

- Features that are semantically similar need to be grouped together. This is one important area which needs to be looked more closely with effective solutions.

One way is to use Fuzzy sets and ConceptNet here to group semantically related features. No opinion mining research has used ConceptNet, SentiWordNet with fuzzy Logic so far.

- None of the existing fuzzy logic works have defined multi-level sentiment analysis covering the following levels of opinion strength: Excellent, Very good, Good, Average, Below average, Poor and Very Poor.

This paper recommends the use of fuzzy logic to address the challenges mentioned in the paper mainly because the use of fuzzy logic provides a more straightforward way to describe the intrinsic fuzziness in the quality of reviews and opinions. Fuzzy set theory provides excellent means to model the fuzzy boundaries of linguistic terms by introducing gradual memberships. In contrast to classical set theory, in which an object or a case either is a member of a given set (dened, e.g., by some property) or not, fuzzy set theory makes it possible that an object or a case belongs to a set only to a certain degree, thus modelling the penumbra of the linguistic term describing the property that defines the set. However there still exists many challenges in Fuzzy logic of opinion mining and semantic aspects which have been described in the paper earlier. Integration of soft computing techniques in Semantic web methodologies in the near future is one of the best possible solutions.

We expect fuzzy systems technology to play a prominent role in the quest to meet these challenges mentioned above.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 5, May 2016

VI. CONCLUSION

In this paper, we have discussed the Opinion Mining levels & various tools of mining reviews from different data sources. Nowadays Opinion Mining is an emerging field in decision making process and is developing fast. Researchers have also explored various milestones of opinion mining in different papers. However, the field of opinion mining is not well developed to provide user with a powerful opinion and sentiment mining systems. Resolving semantic problems of Opinion Mining using Fuzzy approach will enhance and improve the extraction, summarization and presentation of opinions with their weight-ages and strengths. Fuzzy logic is a powerful tool and it is built to resolve those problems that have fuzzy input parameters to arrive to most closest real and crisp figures. Opinion mining is among those domains of problems that can be effectively resolved using Fuzzy logic (Animesh and Deba et. al. 2011; Samaneh el. al. 2010).

REFERENCES

- [1] Alina., Sabine B., (2006), "Mining WordNet for Fuzzy Sentiment: Sentiment Tag Extraction from WordNet Glosses", pp. 209-216.
- [2] AnimeshKar, Deba P. Mandal, (2011), "Finding Opinion Strength Using Fuzzy Logic on Web Reviews", International Journal of Engineering and Industries, Vol. 2, No. 1, March 2010. Pp. 37-43.
- [3] BakhtawarSeerat, FarouqueAzam (2012), "Opinion Mining: Issues and Challenges (ASurvey)", International Journal of ComputerApplications (0975 – 8887) Volume 49–No.9, July 2012 42.
- [4] SolankiYogesh Ganesh Bhumika K. Shah, "Feature based OM A Survey", IEEE-2015.
- [5] B Pang L Lea & S. Vaithyanathan, "Thumbs up? Sentiment Classification using machine learning techniques", Proceeding of the conference on empirical methods in NLP- pp 79-86.
- [6] Guohong Fu, Xin Wang (2010), "Chinese Sentence-Level Sentiment ClassificationBased on Fuzzy Sets", Coling 2010: PosterVolume, Pages 312-319, Beijing, August2010.
- [7] Nidhi R Sharma, "OM Analysis & Its challenges", International Journal of innovation & advancement in computer science IJIACS. ISSN2347- Q616 vol 3, Issue 1,April 2014
- [8] AsmitaDhokrat, Sunil khellare, C. Namrata, "Review on techniques & tools used for OM", International Journal of Computer Application Technology & Research- vol 4 issue 6 page 419-424.
- [9] Book, Sentiment Analysis & OM- BING LIU April 2012.
- [10] S. Padmaga& Prof. S. Sane Fatima , "Opining Mining & SA, An Assessment of people's belief - A survey" ,International Journal of AD hoc, Sensor & Computing (IJASUE) vol4, feb-2013.
- [11] Book, "Introduction to Data Mining- Pang", Ning Tan, Vipin Kumar.
- [12] G. Angulakshmi& D R. MancikaChezian, "An Analysis on Opinion Mining- Techniques & Tools", IJARCCE-2014- Vol 3 Issue 7.
- [13] VandanaJha, LM Patnaik, "HOMS- Hindi OM System", IEEE 2015 / 2nd International Conference on Recent Trends in Info system.
- [14] S. Vasanthraj, "A Survey on SA applied in OM", JNCET March 2015.
- [15] Ghose A, "Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics", Knowledge & data Engineering, IEEE Transaction on (Vol 23,Issue:10)
- [16] Dr. RanjanaRajnish, Dr. ParulVerma, "Review of Soft Computing Techniques: Exploring Scope", International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 11, November 2015, Pg 256-261.
- [17] SushmitaMitra, Sankar K. Pal, "Fuzzy sets in pattern recognition and machine intelligence", Fuzzy Sets and Systems 156 (2005) 381–386, www.elsevier.com/locate/fss.
- [18] Subasic, P. and Huettner, A. (2001). 'Affectanalysis of text using fuzzy semantic typing'.IEEE-FS, (9):483–496.
- [19] Samaneh N., Masrah A. Murad, RabiahAbdul Kadir (2010), 'SentimentClassification of Customer Reviews Basedon Fuzzy Logic', Information Technology(ITSim), 2010 International Symposium, pp.1037-1040.
- [20] Sheroz Khan, I. Adam, A.H. M. Zahirul,Mohammed R. Islam, Othman O., Khalifa,(2007), 'Rule-based Fuzzy Controller withAdaptable Reference', World Academy of Science engineering and Technology 35, pp.305-310.
- [21] Vinodhini G., RM.Chandrasekaran (2012), "Sentiment Analysis and Opinion Mining: ASurvey", International Journal of AdvancedResearch in Computer Science and SoftwareEngineering, Volume 2, Issue 6, June 2012,ISSN: 2277 128X, Available online at:www.ijarcse.com.
- [22] Wilson T., Wiehe J. &Hwa R. (2004), 'Justhow mad are you? Finding strong and weakopinion clauses', in Proceedings of theAAAI-04, 21st Conference of the AmericanAssociation for Artificial Intelligence,AAAI-04, San Jose, US, pp. 761-769.