Journal of Theoretical and Applied Information Technology
20th May 2014. Vol. 63 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

SEMANTIC AND FUZZY ASPECTS OF OPINION MINING

[1] MAQBOOL AL-MAIMANI , [2] NAOMIE SALIM, [3] AHMED M. AL-NAAMANY

[1] IT Department, Oman Air, Muscat, Sultanate of Oman
[2] Faculty of Computer Science and Information Systems, Universiti Teknologi Malaysia, Johor Bahru, Malaysia
[3] Modern College of Business and Science, Muscat, Sultanate of Oman
E-mail: [1]maqbool.almaimani@omanair.com , [2]naomie@utm.my , [3]naamany@mcbs.edu.om

ABSTRACT

In recent years the amount of opinions and reviews available on the web has grown tremendously and therefore the role of sentiment analysis is even more crucial than before. Given the very ambiguous and imprecise nature of sentiments and of their expressions, this study focuses on the review of semantic and fuzzy aspects of opinion mining. Opinions are fuzzy in nature and dealing with the semantic part of the expressed sentiments possesses many challenges and require effective techniques to properly extract and summarize people's views. Previous studies have often reflected that the use of fuzzy logic techniques have proved to be beneficial in mining reviews. This paper presents a review covering the semantic and Fuzzy-based logic techniques and methods in sentiment analysis and challenges appear in the field.

Keywords: *Opinion, Opinion Mining, Sentiment Analysis, Fuzzy Logic, Semantic Orientation, Subject And Objective Information*

1. BACKGROUND

Thousands of products and services are being advertised daily over the web. It is estimated that 75,000 new blogs emerge daily with 1.2 million new posts each day covering many consumer opinions on products and services. Over 40% of people in the modern world depend on opinions and reviews over the web to buy products and apply for various services and express their opinions on diverse issues (Kim 2006; Khan et. al. 2009). As a result of such information wealth, consumers are faced with tough choices to make a rationale decision to select products with good features and competitive prices. In additions suppliers, service providers and manufacturers view customer reviews and opinions about their products and services in order to enhance their products and services. Hence, there is a big demand to develop automated opinion retrieval systems in order to allocate, mine, and classify these thoughts and opinions in some reasonable and presentable fashion (Yao , et al. 2008).

This paper is structured as follows. Section II briefly explains the difference between objective and subjective information and overall techniques used in both areas. In section III opinion mining approaches are explained. The semantic dimension of opinion mining is discussed in section IV. This section covers the followed techniques, approaches, challenges and open issues. The Fuzzy aspects covering techniques, approaches, challenges and open issues are explained in section V followed by concluding remarks in section VI.

2. OBJECTIVE AND SUBJECTIVE INFORMATION

Information can be classified as Objective and Subjective. Objective information are facts which everybody agree on. Subjective information are opinions that can be classified as positive, negative or neutral for some people and partially/totally different or opposite for others. A sentence is considered to be subjective if it contains subjective words i.e opinions of people. In contrast, a sentence is considered to be objective if there are no subjective clues and all information and details are factual (Pang & Lee, 2005).

A. Opinion Mining

Opinion Mining (OM) is a field that offers a number of tools and techniques that are used to find people's/customer's opinion on certain products, services, event, occasions etc (Mejova, 2009; Khan et. al. 2009).

Journal of Theoretical and Applied Information Technology
20th May 2014. Vol. 63 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Finding and extracting opinions is very essentials for various reasons (Pang and Lee 2008, Rainie and Horrigan 2007; Zabin & Jefferies 2008).

- To scientifically record customers' feelings and opinions on a particular product/services in order improve the quality and delivery of such goods and services.
- To improve social services provided to public by governments and social organizations by understanding their demands and suggestions.
- Companies, supplier and manufacturer firms can utilize online reviews to respond to their consumer insights and provide required enhancements.

The mining process can be as simple as learning polarity (positive or negative) and sentiment of the words, or as complicated as performing deep parsing of data to identify grammar and structure of the sentences (Pang & Lee, 2005,Gamgarn and Pattarachai, 2010; Pavol et. al. 2010). Given an object and a collection of reviews on it, the task in opinion mining process usually consists usually of the following tasks:

- Identify and extract object features that have been commented on in each review.
- Determine the orientation and strength weather the opinion is positive, negative, neutral.
- Provide summary of opinion in textual or in a visualization way

3. Opinion Mining Approaches

Research on opinion mining intends to serve both customers and companies/firms. Opinion mining systems helps customers to retrieve those opinion phrases on specific features matching their requests (Jain et. al. 2006). In addition, opinion mining helps companies, firms and various bodies to know people opinions about their products and services in order to enhance them.

Opinions can be classified at different levels: document, sentence and phrase levels. When reviewing various literatures on opinion mining, one can notice that there are many approaches for sentiment extraction and analysis. Major ones are (Vinodhini & Chandrasekaran 2012, Khan et. al. 2009; Mejova, 2009):

- At document level by classifying an opinionated document as expressing a positive or negative opinion. Here techniques like terms presence and NLP approaches like POS are followed just to get opinion orientation of the document. Analyzing sentiment at document may not give accurate results as for example when using term-frequency/presence method does not reflect the relationship between a feature and its related opinions; it only analyse presence of used opinion terms in the document.
- At a sentence level by classifying a sentence / clause as subjective or objective, and for a subjective sentence or clause classifying it as expressing a positive, negative or neutral opinion
- At a feature level by identifying object features that have been commented on and determining whether the opinions on the features are positive, negative or neutral. Although classifying opinionated texts at the document level or at the sentence level is useful in many cases, they do not provide the necessary detail needed for some other applications. A positive opinionated document on a particular object does not mean that the author has positive opinions on all aspects or features of the object. Likewise, a negative opinionated document does not mean that the author dislikes everything. In a typical opinionated text, the author writes both positive and negative aspects of the object, although the general sentiment on the object may be positive or negative. Document-level and sentence level classification does not provide such information.

Feature-based approach presents many advantages like (Bakhtawar 2012):

- Feature-based approach allows analyses of sentences/phrases. Opinion detection on sentence level can produce better results on opinion extraction and orientation rather than opinion detection on document level. Moreover, sentence level-analysis allow to combine more

Journal of Theoretical and Applied Information Technology
20th May 2014. Vol. 63 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

than one technique to improve sentiment analysis

- Feature-based approach produces better quality results when summarizing opinions as it analyses opinions on features level by analyzing phrases / words
- Feature-based sentiment analysis is very important as customers who usually do not express product opinions holistically but separately according to product/service individual features.
- Feature-based approach makes more sense as usually people may like some features and dislike some others. Some people give their opinions on few features that are of their interest and hence don't comments on other features.

4. SEMANTIC ASPECTS OF OPINION MINING

Linguistically, the word semantics refers to the study of understanding human expression through language. It focuses on the relation among words, phrases, signs, symbols, and what they stand for. Semantic aspect of a term, sentence or a text linguistically means the orientation or direction in which the related concept is interpreted (Dave et al. 2003; Esuli & Sebastiani 2005). For example, the word 'good' has a positive orientation; whereas, the word 'bad' or 'not good' has a negative orientation. *Semantics is a branch of linguistics which relates with meaning. Semantics is considered as a study of meaning in language. It deals with the expression of linguistic objects such as word, phrases and sentences. It does not pay attention to the syntactical arrangement or pronunciation of linguistic object. As states by Katz (1972: 1), "Semantics is the study of linguistic meaning. It is concerned with what sentence and other linguistics object express, not with the arrangement with their syntactic parts or with their pronounciation."*

Existing approaches to automatic identification and extraction of opinions from text can be grouped into three main categories: keyword spotting, in which text is classified into categories based on the presence of fairly unambiguous affect words (Ortony et al., 1998;Wiebe, Wilson, and Claire 2005), lexical affinity, which assigns arbitrary words a probabilistic affinity for a particular opinion (Wilson et al.,2005; Somasundaran, et al., 2008), and statistical methods, which consist in calculating the valence of keywords, punctuation and word co-occurrence frequencies based on a large training corpus (Hu and Liu 2004;Pang and Lee 2005;Abbasi et al., 2008).

These approaches mainly rely on parts of text in which opinions are explicitly expressed such as positive terms (e.g. good, nice, excellent, fortunate, correct, superior, best) and negative terms (e.g. bad, nasty, poor, unfortunate, wrong, inferior, worst). This is where the role of Semantic aspects comes into play. Semantic aspect for opinion mining built exploiting common sense reasoning techniques, such as blending and spectral activation, together with an emotion categorization model and an ontology for describing human emotions.

A. Techniques for the semantic aspects

Existing literature on OM presents a variety range of techniques, tools and methods to achieve the required objectives. These techniques can be grouped as follows (Mejova, 2009, Pang & Lee 2008; Lei et. el. 2008):

- NLP covering POS, tokenizing, N-Gram, term presence, stemming, and lexicon-based methods using lexical based dictionaries like WordNet.
- Machine learning which can classified as
  - Supervised learning like SVM and BN classifiers
  - Unsupervised (not widely used) like PMI-IR classifier
- Rule-based techniques
- Statistical techniques
- Semantic using semantic web techniques and resources like ontology, OWL and RDF
- Statistical methods using techniques like K-Nearest, SVD and WF

The below subsections summarizes few researches and techniques used for mining opinion from reviews:

1) Techniques to extract opinions and features

Extracting part of the speech from news corpus was proposed by Berland. He used possessive constructions and prepositional phrases (Berland

Journal of Theoretical and Applied Information Technology
20th May 2014. Vol. 63 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

& Charniak 1999). On another research, Hu proposed to use association rule mining to find out the frequent features appearing in product reviews (Hu and Liu 2004). Yi et al proposed a way to extract features from online reviews (Jeonghee et. al. 2003). Popescu used PMI method to evaluate each candidate feature after extracting explicit features for a given product for parsed review (Popescu and Etzioni, 2005). Jian focused on a better way of retrieving user opinions to the searcher. The methodology consisted of two steps. First they index all opinions and produce opinion tuples as (product, feature, sentiment). Then they used the tuples to retrieve those opinions that match users' retrieving interests. PMI technique was used here (Jian et. al. 2006). This approach was largely preferred by many researchers for sentiment analysis. The approach was largely based on mining rules and these approaches are being slowly enhanced and other methods of analysis are also being introduced by expert researchers.

Liu's approach was to extract features from reviews by using association rule mining. Liu's system measured similarities between features and frequently selected terms. The drawback of this method is that there is no relationship between a feature and its related opinion. The system only considers the information from the term itself like term frequency without linking this to the associated opinion on the term (Ding et. al. 2008). Another feature extraction method was proposed by Ding (Ding et. al. 2008). This is a rule-based approach which extracts a large number of features compared to the amount of review data. For example from 50 reviews it will extract around 300 features. This is due to the fact that the system considers words that have the same or similar meaning as different features. This has resulted inability to provide proper summary details. Another feature-based extraction system was proposed by Aciar (Aciar S. et al. 2007). This system used an ontology and semantically shown good results. However, the continuous growth of review data requires a frequent maintenance of the ontology. This was the major problem of this system. In order to overcome the above limitations, Jeong et al. came up with a new system called FEROM (Features Extraction and Refinement for Opinion Mining) which extract correct features by considering both semantic and grammatical properties of feature terms. Similar features are merged in order to reduce number of features and in order to produce better opinion summarization results. FEROM has number of drawbacks. First the process of extracting the features from reviews is lengthy. The proposed solution for negation was not very effective. In addition to this, FEROM can only group similar features based on synonym opinion words. FEROM cannot do this for antonym words which can also express opinion information for homogenous features (Hana Jeong et. al. 2011).

2) Techniques To Identify Semantic Orientation (SO)

Semantic orientation is a basic task in Sentiment Orientation and it refers to the polarity and objectivity of a given word, a sentence or a document. The polarity process identifies whether a given opinion is positive, negative or neutral (Jian et. al. 2006, Mejova, 2009, Kushal, 2003, Shitanshu & Pushpak, 2008; Lei et. al. 2008). Variety of techniques and systems has been developed for this purpose.

Few systems like Opinion observer and OPINE have been developed to determine the SO of a sentence (Pang Lee, 2008; Jian et. al. 2006). In opinion mining if a sentence has many positive opinions, the opinion sentence is regarded as a positive one. If number of positives is equal to number of negative opinions, the system predicts the orientation using the average orientation or orientation of the previous sentence (Hu and Liu 2004). Liu et al, proposed a method which search for opinion words near to their features and associated product. Then the system calculates the dominant orientation of these opinions and assigns it as the opinion orientation of the feature after taking consideration the distance of the opinion word to said feature and product (Jian et. al. 2006).

Many techniques are used to identify the SO of a given text like lexicons, statistical techniques which looks at occurrence of word compared to other words which known polarity. Other techniques use training documents, labelled or unlabelled as a source of knowledge. Suasic and Huettner manually developed a lexicon relating words with their effect words and associated intensity (Mejova, 2009). Many other lexicons have been built among which WorldNet is the largest and well-known lexicon. WordNet was enhanced by Esuli and Sebastiani who added polarity and objectivity labels for each word (Esuli & Sebastiani, 2006; Shitanshu & Pushpak, 2008). Later a more enhanced lexicon

Journal of Theoretical and Applied Information Technology
20th May 2014. Vol. 63 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

called SentiWordNet was built as an enhanced lexical resource explicitly devised for supporting sentiment classification and opinion mining applications (Pang and Lee, 2008). SENTIWORDNET 3.0 is freely available from http://sentiwordnet.isti.cnr.it/.

Besides Lexicon approaches, statistical techniques gain its own popularities to find sentiment orientation or words. Turney and Littman (2003) used Pointwise Mutual Information (PMI) and Latent Semantic Analysis (LSA) to infer semantic orientation of words. PMI queries a search engine in order to calculate word co-occurrence. LSA uses singular Value Decomposition (SVD) which is a matrix factorization technique to analyse the statistical relationship between words. The assumption used here is that the semantic orientation of a word tends to correspond to the semantic orientation of its neighbors.

Other ways to determine the Sentiment Orientation of an opinion is to use training labeled or unlabeled documents as a source of knowledge to know the polarity of a given text. Labeled data is done manually by using annotation or a start/point system. A range of manually labeled data is available via well-known endeavours like TREC, CLEF and NTCIR (Mejova, 2009). Once data set is decided, many of the existing machine learning techniques such as SVM and NV can be used to train sentiment classifiers (Pang Lee 2005; Mejova, 2009).

Once semantic orientation of the word is determined, it is extended to sentences and documents. One way do this is to take an average of all polarities of words in a sentence in order to arrive to sentence polarity. Same method used to determine the polarity of a document. For example, Popescu and Etzioni (2005) use a relaxation labelling which is an unsupervised classification method which extends sentiment of word to the sentence it appears in. Dave et al and Turney (2003) use averaging method to determine the polarity of documents.

B. Semantic based resources

There are many efforts made by different researches to build or refine resources (i.e. corpus, lexical dictionaries) for Opinion Mining. The following are few examples for corpus and Data sets (Pang Lee 2008):

- NTCIR multilingual corpus: The corpus for the NTCIR 6 pilot task consists of news articles in Japanese, Chinese, and English and formed the basis of the Opinion Analysis Task at NTCIR6. The training data contains annotations regarding opinion holders, the opinions held by opinion holder, and sentiment polarity, as well as relevance information for a set of predetermined topics. The corpus of the NTCIR Multilingual Opinion-Analysis Task (MOAT) is drawn from Japanese, Chinese, and English blogs (Pang Lee 2008)
- WordNet is an online English-based lexical database which groups English words into sets of synonyms called synsets, provides short, general definitions, and records the various semantic relations between these synonym sets. The purpose is twofold: to produce a combination of dictionary and thesaurus that is more intuitively usable, and to support automatic text analysis and artificial intelligence applications. Though WordNet contains a sufficiently wide range of common words, it does not cover special domain vocabulary. WordNet is the most commonly used computational lexicon of English for word sense disambiguation (WSD), a task aimed to assigning the most appropriate senses (i.e. synsets) to words in context.
- SentiWordNet is a lexical resource for opinion mining. SentiWordNet assigns to each synset of WordNet three sentiment scores: objectivity, positivity, negativity. The SentiWordNet interface now provides the possibly to accept user feedback on the values assigned to synsets. This feature is a first step towards building a community of SentiWordNet users that collaboratively improve SentiWordNet (Esuli & Sebastiani 2006).
- HowNet is an on-line common-sense knowledge base unveiling inter-conceptual relations and inter-attribute relations of concepts as connoting in lexicons of the Chinese and their English equivalents. The philosophy behind

Journal of Theoretical and Applied Information Technology
20th May 2014. Vol. 63 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

HowNet lay ground on its understanding and interpretation of the objective world. For example, space can be segmented into "up", "down", "left", "right" while Time can be seen from "the past", "the present" and "the future". Nothing can only function as a component and not a whole and the reverse is true. Depending on the system of reference, the same point of reference can either be regarded as a whole or a part. In HowNet, Part is taken as a constituent in a larger whole. The role and function of Part in whole is analogous to the human body, for instance, "hilltop", "hillside", "mountain foot", "table leg", "back of chair", "estuary". "door" and "window" of buildings are analogous to the relevant parts of the human body such as the eyes, the mouth etc (Lei et. al. 2008).

- ConceptNet is a freely available commonsense knowledgebase and natural-language-processing toolkit which supports many practical textual-reasoning tasks over real-world documents right out-of-the-box (without additional statistical training) including topic-listing (e.g. a news article), affect-sensing (e.g. this email is sad and angry), analogy-making (e.g. “scissors,” “razor,” “nail clipper”), text summarization, contextual expansion, causal projection, cold document classification and many other inferences (Liu & Singh 2004).

C. Semantic Problems and Open Issues

Semantic dimension of Opinion Mining carries many challenges like (Qingliang et al. 2008, Bakhtawar &, Farouque 2012, Mejova 2009, Pang & Lee 2008; Vinodhini & Chandrasekaran 2012):

- Certain phrases if said in one context can indicate positive opinion; whereas, if said in different context may mean negative opinion. For example, the phrase “go read the book” can mean positive if said in a book review; whereas, it will have a negative impact if said in review of a movie as it may indicate that the book is better than watching the movie.
- Emotional-based opinion remains a challenge. There is a need to develop powerful techniques in mining emotional related opinions like happiness, sadness, humour, anger etc
- There is a need to build standard measures for opinion words like for example “this pen is good” – how the word ‘good’ should be compared to the word ‘nice’ in the following sentence ‘this pen is nice.’
- There is a need to find effective ways to group semantically related words as different words and terms can be used to refer to the same thing. Jeong et al (2011). In their proposed system FEROM proposed a solution to this by reducing such terms. Words that have same of similar meanings are merged using semantic similarities between features. However, FEROM could not address this for antonym words which can also express opinion information for homogenous features (Jeong et. al. 2011).
- Word negation is another important challenge to resolve. The major challenge here is to know which words are negated by which negation term. Several efforts have been made to address this issue but there still enough room for improvement (Alexander et. al. 2011). Das and Chen (2001) suggested in their approach to append the negation word to the term (like ‘Like-not’ for sentence ‘I don’t like that movie’). But this approach has its own problems as many other forms of negations cannot be represented this way. Other basic efforts on negation focus only on reversing the sentiment of the sentence like the solution proposed by Jeonghee (Jeonghee et. al. 2003). Ding (Ding 2008) proposed a pattern-based approach to address the negation. However, the authors didn’t explain how such negation is handled and how different forms of negations are tackled. Jeong et al.2011), in their proposed FEROM system suggested a solution for this problem. The author proposed the negative word is replaced with its antonym (Jeong et. al. 2011). An Antonym word may in certain cases not equivalent to a negated word. For example, “Not bad” can mean “average’ or little above average but doesn’t fully equal in meaning to “good.” Alexander et. al. (2011) proposed a different method based on WordNet and SentiWordNet.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The proposed framework is based on a basic sentence-level sentiment analysis. A Word Level sentiment score in the range of is used and derived from SentiWordNet lexicon. When all word scores are calculated, sentences can be marked as negative or positive. If the sum of scores of word-level sentiments in a sentence is less than 0, then sentence is classified as negative otherwise positive.

- Comparative opinion is another challenge that needs to have effective solutions. A comparative sentence is a sentence that compare two products based on similarities and differences as define by liu (Liu 2006). Comparison can be gradable and non gradable. Gradable can uses equal, greater than or less than operations. In non-gradable comparison features are compared without ranking (e.g. Coke tastes differently from Pepsi). (Mejova 2009). Few attempts made to extract such opinions. A data mining solution was followed by Jindal and Liu (2006). They used WordNet and Class Sequential Rule (CSR) to identify comparative sentences. They could not achieve good precision (i.e. 32%). Another approach followed by Hou and Li (2008) used another data mining technique called Conditional Random Fields (CRF).
- Another challenge is that, people can be contradictory in their statements. Most reviews will have both positive and negative comments, which is somewhat manageable by analyzing sentences one at a time. However, the more informal the medium (twitter or blogs for example), the more likely people are to combine different opinions in the same sentence. For example: "the movie bombed even though the lead actor rocked it" is easy for a human to understand, but more difficult for a computer to parse. Sometimes even other people have difficulty understanding what someone thought based on a short piece of text because it lacks context. For example, "That movie was as good as his last one" is entirely dependent on what the person expressing the opinion thought of the previous film.
- Opinion Extraction can be further enhanced by applying SRL technique on sentence level by detecting basic event structures such as “who” did “what” to “whom”, “when” and “where”. From a linguistic point of view, the identification of such event frames holds potential for significant impact in many NLP applications, such as Opinion Extraction, Information Retrieval and Question Answering (Lluis . et al. 2008).
- Opinions are composed by people from different ages, cultures, religions. Hence, proper English structure and vocabulary are not followed Sometimes slangs, short cuts and emotions are used to express opinions. This makes it difficult to extract opinions properly (Esuli & Sebastiani 2006). Different people have different writing styles. Each person expresses his/her opinion in his/her own way and style. Moreover, some people use abbreviations while expressing their views. Due to this sometimes it is difficult to identify boundary of sentences.
- Certain opinions are time bound and opinion may change over time due to many reasons like improvement of products and services. How such change can be controlled and from which cut off date certain opinions should not be considered for certain services or products. Moreover, people’s mood may change overtime and this affect the previously expressed opinions (Pang and Lee 2008)
- Sarcastic and Ironic statements can result in misleading opinions. In such cases positive opinions may appear negative. Moreover, people may mix positive and negative opinions in the same sentences a case which will complicate the sentiment extraction process.
- Reviews are written in different languages has created a challenge in opinion mining. The main problem becomes the time spent in reviewing all available data and resolves the language barrier. Can we have a language independent method that automatically analyze, extract and assign values for a given product or service? This method will present the polarity and orientation of other customers’ decisions covering percentage of positive, negative and

Journal of Theoretical and Applied Information Technology
20th May 2014. Vol. 63 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

neutral expressed opinions (Alexandra & Andrés 2008).

5. FUZZY-BASED ASPECTS OF OM SYSTEMS

In recent years, the number and variety of applications of fuzzy logic have increased drastically. 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.

Fuzzy Logic (FL) was founded by Lotfi A. Zadeh (1965) as an extension of the classical crisp set theory. Fuzzy is based on a theory which relates objects in a set with a degree of membership. An object may belong to different sets with different degree of memberships. Basically FL is a problem-solving control system methodology which provides a simple way to arrive at a definite conclusion based upon vague, ambiguous, imprecise, noisy, or missing input information. FL's approach to control problems mimics how a person would make decisions, only much faster. Fuzzy logic is well-known with its if-then rule, or simply fuzzy rule. Although rule-based systems have a long history of use in Artificial Intelligence (AI), what is missing in such systems is a mechanism for dealing with fuzzy consequents and fuzzy antecedents. In fuzzy logic, this mechanism is provided by the calculus of fuzzy rules (Kazuo and Niimura 1996).

A. Why Fuzzy logic?

Opinion words are fuzzy in nature. For example, the words “Nice”, “good”, and “delicious” 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. This means that these words are already in fuzzification stage. Defining fuzzy sets for such words needs to be based on some expert opinions Since opinions are fuzzy in nature and meaning of opinion words can be interpreted differently, Fuzzy logic is 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 (Samaneh el. al. 2010; & 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 order to determine the strength of each opinion. There is a need to increase the

Journal of Theoretical and Applied Information Technology
20th May 2014. Vol. 63 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

classifications of opinions and assign weightages for different opinion words. Only Fuzzy logic (via Fuzzy sets and Fuzzy and 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 el. 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.

B. Research on Fuzzy Logic

Research on Fuzzy-based opinion systems is still in its infancy. There are few researches on this field. Many of these fuzzy-based system have either not addressed many essential challenges of opinion mining or have not utilised the powerfulness of Fuzzy logic technique.

Fu and Wang (Guohong and Xin 2010) presented a fuzzy set theory based on framework for Chinese sentence-level sentiment classification. This framework consists of three major steps. The first step is on sentence-level sentiment intensity calculation is to develop a fine-to-coarse strategy. This strategy calculates sentiment intensities for morphemes, words and phrases by using Chi-square technique via different formals to arrive to the intensity scores for each sentence part. As a second step, they defined their fuzzy sets to represent sentiment polarities (i.e. positive, negative and neutral). Once the sentiment intensity of a sentence is calculated, a membership function (which is built for this purpose) is used to identify to which set a sentence belongs and then decide its polarity under the principle of maximum membership. This method achieved better results when compared with CUHK system which is the Chinese Opinion Mining system. However, the achieved Precision, Re-call and F-score are not high and are within mid-range scores. Moreover, the proposed system scored lower on Precision when compared to CUHK system. The system outperformed the CUHK system on F-Score only by 5%. The author did not show how to address many problems like negations, vague and ambiguous words, etc. In addition, the system could not properly aggregate multiple-granularity polarity within opinionated text and hence more tailored techniques need to be developed.

Animesh and Debael (2011) proposed an opinion mining systems called Fuzzy Opinion Miner (FOM). FOM is a supervised opinion orientation detection system that mines reviews using Fuzzy logic. FOM executes the following tasks:

- Extract product features on which customers have commented;
- Identify opinion sentences in each review and extract opinion phrases
- Measure the strength of opinion phrases and summarize the results.

This system has few drawbacks. FOM does not focus on all features mentioned in the review. It only collects important features whose frequency are 20% and above. Moreover, FOM focuses on adjectives and adverbs as opinion words. Verbs can also be opinions. Also FOM does not use full Fuzzy features like Fuzzy sets, rules and defuzzification process. It only uses Fuzzy weights which are assigned to opinion words. Additionally, FOM does not group features according to the strength of the opinions that have been expressed on them. This will help to show which features customers strongly like or dislike. In addition to the above, the system was not compared to other system to show its performance and advantages. Moreover, Precision, Re-call and F-score measures are not calculated to present system performance.

Samaneh (Samaneh el. al. 2010), proposes a Fuzzy logic system (FLS) which performs sentiment classification of customer reviews. Here customer reviews were classified into various sub classes (i.e. strongly positive (or

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

negative), moderate positive (or negative), weakly positive or negative and very weakly positive (or negative)) by using adjectives, adverbs and verbs as combinations following holistic lexicon approach. FLS used adjectives, adverbs, verbs and Nouns as opinion words. Special degree for each opinion words were assigned (i.e. excellent 6, good 3 like 4, very 5 etc). These degrees were assigned by human experts. FLS used three triangular membership functions which are low, Moderate, High. Boundaries for these sets were also assigned by human experts. Based on these fuzzy sets, Fuzzy rules were designed to address each case and accordingly find the orientation when a condition is met. Based on these rules minimum degree of membership function is selected for each rule. The output is computed by using the Mamdani's defuzification function (center of gravity). Such defuzzification function will find the crisp value of each membership degree. The authors have not reported any results. Precision, Recall and F-score were not calculated to see the performance of the proposed systems.

6. CHALLENGES FOR FUZZY-BASED APPROACHES

The following points highlights few 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.
- Fuzzy Light Ontology, which has been used in IR field, has not yet been applied onto Sentiment Analysis domain. (Dragoni et al. 2012) (Qi Zhang et.al. 2011). An ontology is a set of categories consisting of objects or ideas related with each other with a defined relationship. Ontology defines a common vocabulary in a specific domain for researchers to share information. Ontology contains machine-interpretable definitions of basic concepts in the domain and relations among them (Dragoni et al. 2012). WordNet and SentiWordNet are examples of light ontology.

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

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

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 denes 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

7. CONCLUDING REMARKS

Opinion mining is a vast research area which is developing quite fast. Till today the field of opinion mining is not well developed to provide user with a powerful opinion and sentiment mining systems. The above listed challenges needs to be researched in depth and addressed (Hana Jeong et. al. 2011).

Resolving semantic problems of Opinion Mining using Fuzzy approach will drastically 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] Aciar S. et al.,(2007) "Informed Recommender: Basing Recommendations on Consumer Product Reviews," IEEE Intell. Syst., vol. 22, no. 3, pp. 39-47.
[2] Alexander Hogenboom, Paul van Iterson, Bas Heerschop, Flavius Frasincar, and Uzay Kaymak (2011), "Determining Negation Scope and Strength in Sentiment Analysis", IEEE pp 2589-2594.
[3] Alexandra BALAHUR, Andrés MONTOYO (2008), "A Feature Dependent Method for Opinion Mining and Classification", 978-1-4244-2780-2/08, IEEE.
[4] Alina A., Sabine B., (2006), 'Mining WordNet for Fuzzy Sentiment: Sentiment Tag Extraction from WordNet Glosses', pp. 209-216.
[5] Animesh Kar, 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.
[6] Bakhtawar Seerat, Farouque Azam (2012), "Opinion Mining: Issues and Challenges (A survey)", International Journal of Computer Applications (0975 – 8887) Volume 49– No.9, July 2012 42.
[7] Berland M. and Charniak E.(1999), "Finding parts in very large corpora". In Proceedings of the 37th ACL Conference, pages 57–64.
[8] Dave K., Lawrence S., and Pennock D. M. (2003), "Mining the peanut gallery: Opinion extraction and semantic classification of product reviews," in Proceedings of WWW, pp. 519–528.
[9] Ding X., Liu B., and Philip Y., (2008) "A Holistic Lexicon-Based Approach to Opinion Mining," Proc. Int. Conf. Web Search Web Data Mining, 2008, pp. 231-240.
[10] Dragoni, M., et al. (2012), 'A conceptual representation of documents and queries for information retrieval systems by using light ontologies', Expert Systems with Applications, doi:10.1016/j.eswa.2012.01.188.
[11] Eiman AI-Shammari (2010), "Improving Arabic Text Processing via Stemming with Application to Text Mining and Web Retrieval", PhD Dissertation, George Mason University, Virginia, USA.
[12] Esuli A. & Sebastiani F. (2005), 'Determining the Semantic Orientation of Terms through Gloss Classification', in Proceedings of 14th ACM International Conference on Information and Knowledge Management, CIKM 05, Bremen, DE, pp. 617-624.
[13] Esuli A. & Sebastiani F. (2006), 'Determining Term Subjectivity and Term Orientation for Opinion Mining', in Proceedings of the ACL-97, 35th Annual Meeting of the Association for Computational Linguistics, CL-06, Madrid, ES, pp. 174-181.
[14] Esuli A. & Sebastiani F. (2006), 'SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining', in Proceedings of LREC 2006, LREC, pp 417-

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

422.

[15] Gamgarn S., Pattarachai L., (2010), "Extracting Product Features and Opinions from Product Reviews Using dependency Analysis", 2010 7th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD) pp. 2358-2362.

[16] Guohong Fu, Xin Wang (2010), 'Chinese Sentence-Level Sentiment Classification Based on Fuzzy Sets', Coling 2010: Poster Volume, Pages 312-319, Beijing, August 2010.

[17] Hana Jeong, D. Shin, J. Choi (2011), "FEROM: Feature Extraction and Refinement for Opinion Mining", ETRI Journal, Volume 33, No 5, Oct 2011, pp 720-730.

[18] Hou, F. and Li, G.-H. (2008). "Mining Chinese comparative sentences by semantic role labeling|, Proceedings of the Seventh International Conference on Machine Learning and Cybernetics.

[19] Hu M. and Liu B.. (2004), "Mining opinion features in customer reviews", In AAAI, pages 755–760.

[20] Jeonghee Yi , Tetsuya Nasukawa, Razvan Bunescu, and Wayne Niblack (2003) "Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques", IEEE Intl. Conf. on Data Mining (ICDM).

[21] Jian Liu , Gengfeng Wu and Jianxin Yao (2006), "Opinion Searching in Multi-product Reviews", Proceedings of The Sixth IEEE International Conference on Computer and Information Technology (CIT'06).

[22] Jindal, L. and Liu, B. (2006). "Identifying comparative sentences in text documents", Proceedings of the 29th annual international ACM SIGIR conference on Research and Development in Information Retrieval.

[23] Khan K., B. Bahaeudin, A. Khan, F. Malik (2009), "mining Opinion from Text Document: A Survey", 3rd IEEE International Conference on Digital Ecosystems and Technologies, pp 217-222.

[24] Kazuo Tanaka and Niimura (1996) "An Introduction to Fuzzy Logic for Practical Applications" (1996), textbook November 15,1996.

[25] Kim P. (2006), "The forrester wave: Brand monitoring, Q3 2006," Forrester Wave (white paper).

[26] Kushal Dave, Steve Lawrence, David M. PennocK (2003), "Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews", Copyright is held by the author/owner(s). WWW2003, May 20–24, 2003, Budapest, Hungary. ACM 1581136803/03/0005.

[27] Lei Yu, Jia Ma; Tsuchiya, S.; Ren, F. (2008), "Opinion mining: A study on semantic orientation analysis for online document", Intelligent Control and Automation, WCICA. 25-27 June 2008, 7th World Congress, pp. 4548-4552.

[28] Liu H. and Singh P., (2004) "ConceptNet: A Practical Commonsense Reasoning Toolkit", BT Technology Journal 22, 4, pp. 211-226, Oct. 2004, Kluwer Academic Publishers.

[29] Mejova Y. (2009), "sentiment Analysis: An Overview", Computer Science Dept., Univ. of Iowa, Nov 2009.

[30] Lluis Marquez, Xavier Carreras, Kenneth C. Litkowski, Suzanne Stevenson (2008), "Semantic Role Labeling: An Introduction to the Special Issue", Association for Computational Linguistics, Volume 34, No 2, pp. 145-159.

[31] Pang B. and Lee L. (2008), "Using very simple statistics for review search: An exploration," in Proceedings of the International Conference on Computational Linguistics (COLING), (Poster paper).

[32] Pang B. and Lee L. (2004), "A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts," in ACL,pp. 271–278.

[33] Pang B. and Lee L. (2005), "Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales," in Proceedings of the Association for Computational Linguistics (ACL), pp. 115–124.

[34] Pavol Navraj, Anna Bon, Lukas slizik, (2010), "Mining Overall Sentiment in Large Sets of Opinions", Advances in Intelligent Web Mastering – AISC 67, pp. 167-174.

[35] Popescu, A.-M. and Etzioni, O. (2005). Extracting product features and opinions from reviews. Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing.

[36] Qi Zhang, Yuanbin Wu, Yan Wa, Xuanjing (2011), 'Opinion Mining Sentiment Graph', 2011 IEEE/WIC/ACM International Conferences on Wed Intelligence Agent

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

technology, pp. 249-252.

[37] Qingliang Miao, Qiudan Li, Ruwei Dai (2008), "A Unified Framework for Opinion Retrieval", IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology.

[38] Rainie L. and Horrigan J. (2007), "Election 2006 online," Pew Internet & American Life Project Report, January.

[39] Samaneh N., Masrah A. Murad, Rabiah Abdul Kadir (2010), 'Sentiment Classification of Customer Reviews Based on Fuzzy Logic', Information Technology (ITSim), 2010 International Symposium, pp. 1037-1040.

[40] Sheroz Khan, I. Adam, A.H. M. Zahirul, Mohammed R. Islam, Othman O., Khalifa, (2007), 'Rule-based Fuzzy Controller with Adaptable Reference', World Academy of Science engineering and Technology 35, pp. 305-310.

[41] Shitanshu Verma Pushpak Bhattacharyya (2008), "Incorporating Semantic Knowledge for Sentiment Analysis", Proceedings of ICON-2008: 6th International Conference on Natural Language Processing Macmillan Publishers, India. Also accessible from http://ltrc.iiit.ac.in/proceedings/ICON-2008.

[42] Subasic, P. and Huettner, A. (2001). 'Affect analysis of text using fuzzy semantic typing'. IEEE-FS, (9):483–496.

[43] Turney P. (2002), "Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews," Procs. of the 40th Annual Meeting of the Association for Computational Linguistics.

[44] Turney, P. D. and Littman, M. L. (2003). Measuring praise and criticism: Inference of semantic orientation from association. ACM Transactions on Information Systems (TOIS), 21(4):315–346.

[45] Vinodhini G., 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, June 2012, ISSN: 2277 128X, Available online at: www.ijarcsse.com.

[46] Vivek Kumar Singh, Rakesh Adhikari, Debanjan Mahata (2010), "A Clustering and Opinion Mining Approach to Socio-political Analysis of the Blogosphere", IEEE.

[47] Wilson T., Wiehe J. & Hwa R. (2004), 'Just how mad are you? Finding strong and weak opinion clauses', in Proceedings of the AAAI-04, 21st Conference of the American Association for Artificial Intelligence, AAAI-04, San Jose, US, pp. 761-769.

[48] X. Ding, B. Liu, and Y. Philip (2008) "A Holistic Lexicon-Based Approach to Opinion Mining," Proc. Int. Conf. Web Search Web Data Mining, pp. 231-240.

[49] Yao T.F., Cheng X.W., Xu F.Y., Hans S., Wang R. (2008)." A Survey of Opinion Mining for Texts", [J]. Journal of Chinese Information Processing,22(3) 71-80.

[50] Yee W. LO, and Vidyasagar POTDAR (2009), "A Review of Opinion Mining and Sentiment Classification Framework in Social Networks ", 3rd IEEE International Conference on Digital Ecosystems and Technologies.

[51] Zabin J. and Jefferies A. (2008), "Social media monitoring and analysis: Generating consumer insights from online conversation," Aberdeen Group Benchmark Report, January 2008.