ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 4, April 2017

Fuzzy Opinion Mining for Product Recommender System

Aarti Bandodker
Department of Information Technology,
Padre Conceicao College of Engineering,
Goa University, Goa, India

***Abstract*— Over the last few years, opinions are voiced in the form of blogs, tweets, social media and review over the web. These opinions do matter, as, we always consider other persons view in our decision making process. Proposed fuzzy technique finds out how positively or negatively the masses have responded to the launched product, based on voiced opinions. Using a supervised rule based technique, fuzzy values are assigned to opinions, to know the measure of its positivity or negativity. Steps involve, first mining the opinion about the product from ecommerce sites. Then extraction of the opinion phrases as per pre-defined phrase patterns. Next, assigning of fuzzy weights to the opinion words using SentiWordNet and measuring the strength of opinion phrases. And, finally summarization of people's views to yield product recommendation. Proposed fuzzy technique is an enhancement over the existing opinion mining techniques which does binary polarity determination.**

***Keywords:*-- Fuzzy weights; Polarity strength; Opinions; Opinion Mining; Rule based Technique; Phrase pattern.**

I. INTRODUCTION

With the unstoppable growth of web over the last few years, online opinions, customer reviews, blogs, tweets, social media, discussion forums have become the part and parcel of every individual's life. Especially the customer reviews have helped the online customers to a great extent in choosing the right product that suits their need. The reviews present on these sites are appreciations, comparisons or criticism about a certain product.

Our research focuses on categorizing these reviews to identify what degree of appreciation or criticism they have. The core goal of opinion mining systems is to identify pieces of text that express opinions and then find polarity and polarity strength of the expressed opinions. But the existing Opinion Mining Systems cannot differentiate between the two positive opinions. For example, given two opinions, "nice camera" and "great camera", current systems classifies both these opinions into positive polarity opinions. But in true scenario of natural language, both these have varying degree of positivity, i.e. they depict different polarity intensity which the normal opinion mining systems cannot identify. Hence we propose a fuzzy opinion mining system.

This system will perform following tasks: 1) Identify the opinionated sentences from the customer review dataset. 2) Find various opinions expressed on the given product. 3) Identify opinion polarity and measure polarity strength of an opinion using SentiWordNet and fuzzy set.4) Summarizes the opinion scores to get final product recommendation as very bad, bad, good or very good product to buy.

As the fuzzy logic theory is quite effective in processing natural languages, and measure its vagueness, it is also effective in analyzing customer reviews, which are generally in natural languages. Basically Fuzzy Logic (FL) is a problem-solving system method which provides a simple way to arrive at a definite conclusion, based upon, vague, ambiguous or imprecise input information [1]. Opinion words are fuzzy in nature. For example, the words "bad", "very bad", and extremely bad" have the boundaries among them, that are not clear. Hence, Fuzzy logic can easily represent these types of subjective words and assign to classes with some degree of membership.

II. RELATED WORK

Some of the work done in fuzzy opinion mining area in the past include research on sentiment analysis and opinion mining by Bing Liu. [2]. Sentence level opinion mining done by A. Pimpalkar [3]. Affect analysis of text using fuzzy semantic typing by Subasic, P [4] which is a novel approach to text analysis that combines semantic typing techniques from natural language processing with fuzzy techniques, under the common framework of fuzzy semantic typing. Samaneh N. [5], presented a supervised method for mining customer reviews that combines existing text mining approaches with fuzzy approximation techniques. Mita K. Dalal [6], proposed a technique that extends the feature-based classification approach to

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 4, April 2017

incorporate the effect of various linguistic hedges by using fuzzy functions to emulate the effect of modifiers, concentrators, and dilators, Maqbool-Al-Maimani [1] presented a research work on semantic and fuzzy aspects of opinion mining. Animesh and Deba [7] has done research on proving that Opinion mining is among those domains of problems that can be effectively resolved using Fuzzy logic. Ohana, B. & Tierney [8], presents some results that indicate SentiWordNet could be used as an important resource for sentiment classification tasks. V. N. Huynh [9], proposes a model for the parametric representation of linguistic hedges. Liu, Bing. [10], gave an introduction to sentiment analysis and subjectivity in its NLP Handbook. Amit Pimpalkar [11], proposes a technique that uses SentiWordNet and smiley's dictionary to determine the scores of each word present in the comment. S.Nadali [12], proposed a Fuzzy Semantic Classifier (FSC) to identify the strength levels of customer reviews in smaller categories. Nidhi Mishra et al [13 present the insights into opinion mining at different levels. BlessySelvam ,A. Abirami [14], presented research work that focuses on the frame work of opinion mining and survey on some of the tasks which have been done in each phases.

III. FUZZY OPINION MINING SYSTEM MODEL

The proposed fuzzy opinion mining system works as a recommender system for any online product reviews. In our research we are making use of the Camera reviews to show the system behavior. We have collected reviews for five different cameras from various e-commerce sites like Amazaon.com, Flipkart.com and Snapdeal.com. The five camera's selected to build our reviews database are, Nikon D3100 DSLR,Sony SLT-A58K DSLR,Canon EOS 1100D DSLR, Kodak Astro Zoom AZ521 and Fujifilm S4500.We have collected about 50 reviews for each of the product from the above mentioned e-commerce sites.

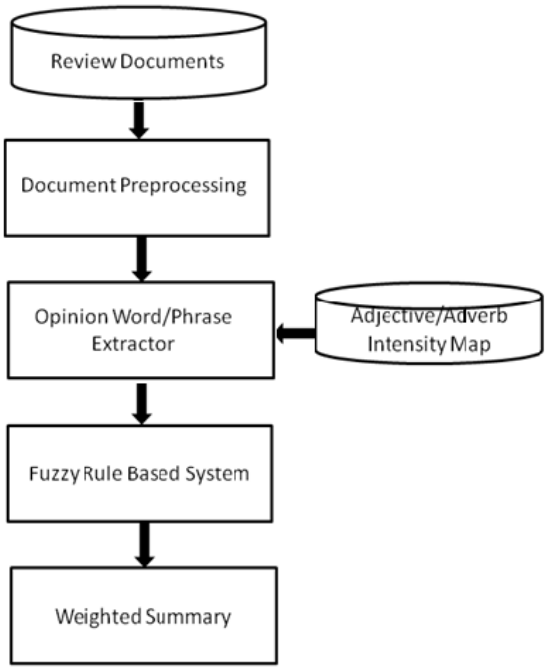

Fig. 1 Fuzzy Opinion Mining System Architecture

Our system is made up of four major activities which are, Document Pre-processing, Opinion Word/Phrase Extraction, Application of Fuzzy Rule based technique and Weighted Summary Processing. The review corpus created by fetching reviews from ecommerce sites. Each review from each user is put in a separate document to create reviews database and are then pre-processed using various pre-processing techniques. Once pre-processing is done the next activity is to identify opinion words from each review using Part of Speech tags and identify opinionated phrases using a pattern matching approach. Once done, we apply fuzzy rule based technique to find opinion intensity of each review. Finally we find the weighted summary of all the reviews to yield a product recommendation as either very bad, bad, good or very good camera to buy based on the final weighted score. Now we look at all the System activities in detail.

3.1 Document Pre-processing

In this activity we remove all the logos, images, and other graphics from the review document, split review into sentences to create a plain text file of reviews and do Part Of Speech (POS) tagging of reviews. This activity also involves two sub activities: Stop Words Removal: Here we eliminate words like prepositions, digits, articles

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 4, April 2017

and nouns like name of product etc. as their existence are meaningless in system for us. It assists in better extraction of opinion phrases/words from the tagged file.

Part of Speech tagging: Here each sentence is parsed to yields the part of speech tag of each word whether the word is a noun, verb, adjective, adverb, etc. and identifies simple noun and verb groups.

3.2 Opinion Word/Phrase Extraction

In this activity, using the POS tags of the words in the reviews, we identify the opinion words which are normally adjectives and adverbs. For these opinion words, we than find an opinion score or strength using SentiWordNet tool and create an Adjective Score file. SentiWordNet is an open source natural language processing tool available on the web. We then extract opinion phrases using a pattern matching technique, where the phrase patterns are predefined.
Opinion Phrase Patterns used in our research are:

Table 1. Opinion Phrase Patterns

Phrase Patterns	Example
Adjective	Great, good, bad, worst
Adjective, Noun	Nice camera, beautiful Picture
Noun, Adjective	Camera great, lens powerful
Adverb, Adjective, Noun	Very bad picture, not great focus
Noun, Adverb, Adjective	Camera very nice
Adverb, Adjective	Very costly, extremely heavy

3.3 Fuzzy Pattern based Technique

Once the opinionated phrases are extracted and opinion scores are found using SentiWordNet, we proceed to find the strength of entire opinion phrase using the rules defined below:

Pattern 1:
For Phrase Pattern (Adjective), (Adjective Noun), OR (Noun Adjective)
Opinion Score=SentiScore (Adjective)
Example1: Great
OpinionScore= SentiScore (Great) =0.7

Pattern 2:
For Phrase Pattern (Adverb) (Adjective) (Noun) OR (Noun) (Adverb) (Adjective) OR (Adverb) (Adjective)
If SentiScore (Adjective)>=0
Opinion Score=Sqrt (SentiScore (Adjective))
If SentiScore (Adjective) <0
Opinion Score= - (SentiScore (Adjective)) ^2

Example1: extremely Great
SentiScore (Great) = 0.7 which is greater than 0
Opinion Score= Sqrt (SentiScore (Great)) =Sqrt (0.7) =0.83

Example2: extremely bad
SentiScore (bad) = - 0.45 which is less than 0
Opinion Score = - (SentiScore (bad)) ^ 2 = - (0.45) ^ 2= - 0.20

Pattern 3:
For Phrase Pattern (Not) (Adjective)

If SentiScore (Adjective)>0
Opinion Score= - 1 + SentiScore (Adjective)
If SentiScore (Adjective) <0
Opinion Score= 1 + SentiScore (Adjective)

Example1: not Great
SentiScore (Great) =0.7
Opinion Score= - 1+ SentiScore (Great) = -1 + 0.7= - 0.3

Example2: not bad
SentiScore (bad) = - 0.45
Opinion Score=1+ SentiScore (bad) =1+ (- 0.45) = 0.55

Pattern 4:
For Phrase Pattern (Not) (Adverb) (Adjective)

If SentiScore (Adjective) >0
Opinion Score= Sqrt (x*y)
If SentiScore (Adjective) <0
Opinion Score= - Sqrt (x*y)

Where x=Opinion score of phrase (Adverb) (Adjective)
y=Opinion score of phrase (Adjective)

Example1: not extremely bad
SentiScore (bad) = - 0.45 which is less than 0
Opinion Score of x = - (SentiScore (bad)) ^2= - (0.45) ^2= - 0.20

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 4, April 2017

Opinion Score of y= SentiScore (bad) = - 0.45
Opinion Score= - Sqrt (x*y) = - Sqrt (- 0.20 * - 0.45) = - 0.3

Example2: not extremely Great
SentiScore (Great) =0.7 which is greater than 0
Opinion Score of x = Sqrt (SentiScore (Great)) =Sqrt (0.7) = 0.83
Opinion Score of y= SentiScore (Great) = 0.7
Opinion Score= Sqrt(x*y) = Sqrt (0.83 * 0.7) = 0.76

3.4 Weighted Summary Processing

Once the review strengths are found, the overall review corpus summary is found by summing all the review scores and dividing by the total number of reviews.

Summarized Score=∑All Review Scores ÷ Total Number of Reviews in the corpus

The final score obtained will be mapped to fuzzy data as shown below.
If score>=-0.5 and score<0 recommendation will be Very Bad product to buy
If score>=-1 and score<-0.5 recommendation will be Bad product to buy
If score>0 & <=0.5 recommendation will be Good product to buy
If score>0.5 & <=1 recommendation will be Very Good product to buy

IV. RESULTS

The proposed technique was applied to five different camera review databases. Result set shows the summarized Fuzzy scores of review corpuses and also the respective recommendation for each camera.

Table 2. Fuzzy Scores and Recommendation for Products under consideration

Product Name	Fuzzy Scores	Product Recommendation
Nikon D3100 DSLR	0.3245	Good Product
Sony SLT-A58K DSLR	0.5699	Very Good Product
Canon EOS 1100D DSLR	-0.2377	Very Bad Product
Kodak Astro Zoom AZ521	0.6766	Very Good Product
Fujifilm S4500	-0.5601	Bad Product

As compared to the binary polarity determination of the customer reviews where user have option of just knowing whether the camera is good or bad, having a slightly big range of values to choose from is an advantage provided by our fuzzy technique.

V. CONCLUSION

The fuzzy opinion mining technique proposed in this paper helps in overcoming the drawback of existing opinion mining systems. Existing systems only focuses on determining the binary polarity of the text which is either positive or negative. The fuzzy technique used in this paper associates some strength score to the opinion words using the SentiWordNet Tool and Pattern based method. These scores are then used to assign the opinion phrases to either positive or negative polarity class with some degree of membership. In our research we only focus on adjectives and adverbs as opinion words. Future scope of this research may involve adding nouns and verbs to the list of opinion words. It may also involve dynamically finding the features on which opinions have been expressed and also handling of implicit opinion extraction challenge. Work can also be done in the area of making opinion words more domain dependent.

REFERENCES

1. Maqbool Al-Maimani,Naomie Salim Ahmed M.Al-Naamany," Semantics and Fuzzy Aspects of Opinion Mining" Journal of Theoretical and Applied Information Technology ,20th May 2014. Vol. 63 No.218

2. Bing Liu. Sentiment Analysis and Opinion Mining, Morgan & Claypool Publishers, May 2012.

3. Amit Pimpalkar, "A System for Sentimental Analysis of Movie Reviews Involving Rule-Based and Fuzzy Measure", International Journal of Artificial Intelligence and Knowledge Discovery (IJAIKD), ISSN 2231-0312,Vol.3, No.2, 2013.

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

5. 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.

6. Mita K. Dalal1 andMukesh A. Zaveri2" Opinion Mining from Online User Reviews Using Fuzzy Linguistic Hedges"Applied Computational Intelligence and So Computing Volume 2014, Published 20 February 2014

7. Animesh Kar, Deba Prasad Mandal, "Finding Opinion Strength Using Fuzzy Logic on Web Reviews",International Journal of Engineering and Industries,volume 2, Number 1, pp 37-43, March, 2011.

8. Ohana, B. & Tierney, B. (2009) Sentiment classification of reviews using SentiWordNet. 9th. IT&T Conference, Dublin Institute of Technology, Dublin, Ireland, 22-23 October.
9. V. N. Huynh, T. B. Ho, and Y. Nakamori, "A parametric representation of linguistic hedges in Zadeh's fuzzy logic," International Journal of Approximate Reasoning, vol. 30, no. 3,pp. 203–223, 2002.

10. Liu, Bing. Sentiment Analysis and Subjectivity, in Handbook of Natural Language Processing, Second Edition, N. Indurkhya and F.J. Damerau, Editors. 2010.