ENHANCING OPINION MINING CLASSIFICATION AND SCORING

Maqbool Al-Maimani[1], Naomie Salim[2], Ahmed M. Al-Naamany[3]
[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*: There is a need to find more effective methods and techniques to extract, classify and summarize customers' online opinions on products and services for better sentiment analysis. In order to achieve this, it is important to consider all opinion word types (Adjectives, Verbs, Adjectives and Noun – termed as AVAN) while analyzing the classification and scoring of opinions. SentiWordNet, which is the lexical resource built specifically for opinion mining, is a base resource for opinion classification and score assignment. Moreover, opinions can be expressed at different degrees and accordingly score of an opinion should vary based on the level at which an opinion is expressed. Opinion classification and scoring is enhanced by using AVAN and opinion degree approaches. Moreover and since opinions are fuzzy in nature, Fuzzy Logic is applied to further enhance opinion mining scoring. This paper addresses the issue of how to enhance opinion classification and scoring using SentiWordNet, AVAN, Opinion Degree and Fuzzy logic as classification features using Sequential Minimal Optimization (SMO) classifier. Result shows that AVAN, opinion degree and Fuzzy logic can drastically enhance SentiWordNet in terms of opinion classification. Accordingly, an accuracy of 92% is achieved using SMO Classifier to classify reviews as "Excellent", "Good", "Fair", "Poor" and "Very Poor."*

Keywords:, Classification, Opinion scoring, Opinion Degree, Fuzzy Logic

1.0 BACKGROUND

With the advent of Web 2.0, many new technologies and platforms have emerged such as blogs, discussion forums, e-commerce sites to enable people procure products and services and provide their opinions and feedbacks online. Consumers have at their disposal different types of information on the web which enable them to share their experiences and opinions (positive or negative) on any product or service [1]. One person may find a particular feature is interesting; whereas, it may not make sense for another.

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 [2,3,4]. Moreover, statistics show that more than 81% of Internet users do online research on a product at least once and this has a significant influence on their purchases [4, 5, 6]. Such an online wealth of information over the web has helped customers, firms, manufacturers, service providers, social and government bodies to take proper decision to procure and enhance various products and services. This has triggered the need to enhance existing methods and techniques to extract, classify and summarize opinions of different online reviews [3].

This paper is structured as follows. Section 2.0 sheds light on related work of opinion classification and scoring. Section 3.0 gives details on SentiWordNet; whereas, Section 4.0 introduces AVAN and the proposed approach. Section 5.0 describes opinion degree and its importance for better sentiment analysis. Section 6.0 presents details on how to enhance opinion scoring by using Fuzzy logic. Section 7.0 explains experiment setup, obtained results and discussion on opinion classification. Section 8.0 summarizes the work with concluding remarks.

2.0 RELATED WORK

Information can be classified as Objective or Subjective. Objective information are facts which people 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. Classification can be viewed also as categorizing opinions based assigned scores and with different levels and degrees. When reviewing literature, one can see a lot of efforts are being done to enhance sentiment analysis, opinion classification and opinion scoring. Kim and Hovy developed a system to get the pros and cons automatically extracted from online reviews [7]. The holder and the topic of the opinion are extracted by the system. However, the system was not able to quantify the strength of the extracted pros and cons and hence the overall rating of reviews could not be predicted. As another effort, Pang came up with a comparison among three machine language algorithms which were trained on the frequencies of positive and negative terms [3]. Based on such comparison, they found out that unigram-based SVM classifiers can be efficiently used when classifying the polarity of movie reviews. On the other hand, a similar approach was applied by Martineau and Finin using the same corpus [8]. Here a Delta TF-IDF function was used to score words before classifying the reviews into positive and negative. However, Brooke used three and five rating classes to classify reviews of different types of products using a set of linguistic features including modality, intensification, discourse structure and negation [9]. In their survey on sentiment analysis and opinion mining, Pang and Lee presented opinion oriented information access, challenges, opinion classification and summarization [4]. Mikalai and Themis made a survey which covered opinion mining, opinion aggregation and subjectivity analysis [10]. Their study mentioned different work performed on this issue covering different domain data such as Movie, Products, Restaurants, and Travel which were used in this experiment. Many researchers used machine learning methods for sentiment analysis. This involved training of classifier on datasets and used the trained model for new document classification. Other scholars suggested using another method such as dictionary of word lexicons. The Dictionary approach is based on a prebuilt dictionary that contains opinion polarity values of words [11, 12]. On the Fuzzy part, literature show very few researches which implement Fuzzy logic using the classic fuzzy concepts and fuzzy calculus. Among those are Kar and Mandal who proposed an opinion mining systems called Fuzzy Opinion Miner (FOM) [13]. FOM is a supervised opinion orientation detection system that mines reviews using Fuzzy logic. FOM Extracts product features, on which customers have commented, identifies opinion sentences in each review, extracts opinion phrases and finally Measures 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. 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, Recall and F-score measures are not calculated to present system performance. As another effort, Nadali proposed a Fuzzy logic system (FLS) which performs sentiment classification of customer reviews [14]. Here customer reviews were classified into various sub classes (i.e. strongly positive (or 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 level for each opinion words were assigned (i.e. excellent 6, good 3 like 4, very 5 etc). These levels 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 finds the crisp value of each membership degree. There are few points noticed in this approach. No membership functions were defined for positive, negative and opinion intensities (opinion degree). Values are predefined based on the classification module which was defined at the beginning. Moreover, fuzzy rules were based on predefined linguistic patterns which cannot be assured to be comprehensive to cover all cases in reviews. In Addition, the defuzzification crisp values are also predefined using a set of expected results. The authors have not reported any results. Precision, Recall and F-score were not calculated to see the performance of the proposed systems.

None of the above approaches considered applying other factors that influence the polarity of opinions especially using SentiWordNet with Opinion Degree and fuzzy logic as a combination to analyze and classify opinions covering all word types (Adjectives, Verbs, Adverbs and Nouns). This paper applies such a combination to enhance opinion classification and polarity.

3.0 SENTIWORDNET

This is a lexical resource built specifically for opinion mining and in line with WordNet. SentiWordNet is basically a databank which is used to obtain word polarity and evaluate the total polarity score of a given review, event, or any given situation from every part of speech. It includes scores for conditions extracted from WordNet 2.0. A semi-supervised process is used to construct this databank in order to obtain opinion sentiments (in form of scores) from a subset of seedling phrases that contain opinion polarity. Each phrase contains conditions that discuss exact significance or synsets linked with three numeric scores that range from 0 to 1. Each phrase suggests the objectiveness and negative and positive prejudice of synset. One significant feature of SentiWordNet is that its negative and positive marking is scored for any specified phrase, which can be used for a phrase with non-zero ideals for equally positive and negative scores [2]. For a synset, the following are determined:

- P(score) Positive score for synsets.
- N(score) Negative score for synsets.
- O(score) Objectiveness score for synsets.

Subsequently, the sum of these three scores is: P(score) + N(score) + O(score) = 1.

4.0 AVAN APPROACH

A lot of focus has been given to Adjectives and Adverbs as opinion words when conducting classifications, sentiment analysis and scoring opinions. Verbs and Nouns, which also can express opinions, are not given proper attention and as a result polarity and scores of opinion sentences or paragraphs are not properly measured. Hence, this paper considers all opinion word types as a combination covering Adjectives, Verbs, Adverbs and Nouns – termed as AVAN. AVAN uses SentiWordNet to classify sentences as objective and subjective. Moreover, AVAN enhances opinion scores extracted from SentiWordNet as follows:

- After the P-Score and N-Score are obtained from SentiWordNet, the Objectivity score element is reduced from the P-Score or N-Score (depending if the word is positive or negative). For example, the word "Good" has the following SentiWordNet values: P=0.72, O=0.28, N=0. To remove the Objectivity part from the positive score, 2.8% of the 72% is reduced from the positive score of 72. The final positive score is 70% or 0.70.
- If an opinion word appears in many domains in SentiWordNet, the final score is calculated by using the following formula ('n' is the number of domains):

((Sum(positive scores)-Sum(negative scores))/n.

Example: Five Senses are found in SentiWordNet for the word 'Like' in Verb Domain:

Sense 1: P=1, O= 0, N= 0 (No objectivity score)

Sense 2: P=0.125, O=0.875, N=0, (P = .114 after removing the objectivity score)

Sense 3: P=0.125, O=0.875,N=0, (P = .114 after removing the objectivity score)

Sense 4: P=0.375, O=0.625, N=0, (P = .316 after removing the objectivity score)

Sense 5: P=0.375, O=0.625, N=0 (P = .316 after removing the objectivity score)

So, final score = ((1+0.114+0.114+0.316+0.316) – (0 + 0 + 0 + 0 + 0))/5 = 0.372 (Score for 'like').

5.0 OPINION DEGREE

AVAN is further enhanced by introducing the concept of Opinion Degree as opinions are expressed in different degrees [14]. For example, the following three statements cannot have similar weights: “Wow! This house is extremely beautiful,” “This house is very beautiful,” and “This house is beautiful.” It is clear that the first statement holds the strongest opinion among the three sentences about the house. Hence, when assigning score to each sentence, it is clear that the third sentence will receive the lowest score compared to first and second sentences. The objective here is to increase or decrease the score of an opinion word according to the score of the degree word used before it. This paper introduces four levels as defined in table 1.

Table 1: Opinion Degree Levels

Degree	Degree Qualifiers	Example
Low Degree (LD)	Below Average, likely, looks, somehow, barely.	“This house looks beautiful”
Normal Degree (ND) –	Here the opinion word is not preceded with any qualifier. (No Degree)	“This is a beautiful house”
Medium Degree (MD)	very, Above average, especially	“This is a very beautiful house”
High Degree (HD)	Extremely, incredibly, extraordinarily, awfully, exceedingly, amazingly,	“This is an extremely beautiful house”

Opinion degree can be one of the following types and for each type a scoring method is suggested:

- The opinion word is positive and it is preceded with a non-negative degree word like “very”, “extremely” etc like “This garden is very beautiful.” In this case the score is calculated and added to opinion score as additional strength to the overall opinion. The score calculation is done as below:

Score = (Opinion Word Score * 100) + (Degree score% of opinion word score) /10

  - Example: “This garden is a very beautiful house”
    - Opinion Degree: Very - SentiWordNet Score : 50
    - Opinion Word: Beautiful - SentiWordNet Score : 75
    - Score for “very Beautiful” = 75 + (50% of 75)/10 = 75 + 3.75= 78.75
- The opinion word is negative and it is preceded with a non-negative degree word like “very”, “extremely” etc like “This is extremely bad view.” In this case the score will be calculated as follows:

Score = (Opinion Word Score * 100) - (Degree score% of opinion word score) /10

  - Example: “This is extremely bad view”
    - Opinion Degree: extremely - SentiWordNet Score : 62.5
    - Opinion Word: bad - SentiWordNet Score : 25
    - Score for “extremely bad” = 25 - (62.5% of 25)/10 = 25 – 1.56 = 23.44
- The opinion word is positive and it is preceded with a negative degree word like “less”, “abnormal” etc like “This picture is less clear.” In this case the score will be calculated as follows:

Score = (Opinion Word Score * 100) - (Degree score% of opinion word score) /10

  - Example: “this matter is less important”
    - Opinion Degree: Less - SentiWordNet Score : 12.5
    - Opinion Word: Important - SentiWordNet Score : 87.5
    - Score for “less important” = 87.5 – (12.5% of 87)/10 = 87 – 1.08= 85.92
- The opinion word is preceded with a negation like “not”, “dis”, “un” etc like “This picture is not clear.” In this case the score will be calculated as follows:

Score = (Opinion Word Score * 100) * Threshold value of (0.01)

  - The rationale for above threshold of (0.01) is to flip the score value to the opposite side and lower the score drastically in order to reflect the actual meaning of the opinion.
  - This will drastically reduce the overall opinion score. The threshold values can be changed based on preference.
  - Example: “this house is not big”
    - Opinion Degree: not - threshold value : 0.01
    - Opinion Word: big - SentiWordNet Score : 73
    - Score for “not big” = 73 * 0.01 = 0.73 (very low)
- The opinion word is positive and it is not preceded by any degree (normal degree level). In this case the degree is implicit and threshold value of 60 is assigned. This is due to the fact that a positive opinion carries little more weight when expressed as the degree is implied through not stated explicitly. In this case the formula will be:

Score = (Opinion Word Score * 100) + (Degree score% of opinion word score) /10

  - Example: “this matter is important”
    - Opinion Degree: normal degree - SentiWordNet Score : 60
    - Opinion Word: Important - SentiWordNet Score : 77
    - Score for “Important” = 77 + (60% of 77)/10 = 77 – 5= 82
- The opinion word is negative and it is not preceded by any degree (normal degree level). In this case the degree is implicit and threshold value of 1 is assigned. Here the reviewer meant that level as such; otherwise, s/he would have added a degree word. In this case the formula will be:

Score = (Opinion Word Score * 100) * threshold values (1)

  - Example: “this is a bad decision”
    - Score for “bad” = 20 (SentiWordNet Score) * 1 = 20
    -

6.0 FUZZY LOGIC

Fuzzy logic has gained popularity because many real world scenarios are fuzzy by nature. Among those are opinions which are expressed continuously by many people on various products, services and situations.

Fuzzy logic consists of the following three main components. First is Fuzzification which is a process of transforming crisp values to fuzzy terms using membership functions and fuzzy sets. Second is Fuzzy Rules. A fuzzy rule is defined as a conditional statement in the form: IF x is A. THEN y is B. where x and y are linguistic variables; A and B are linguistic values determined by fuzzy sets on the universe of discourse X and Y, respectively. Third is the Deffuzification process which converts the degrees of membership of output linguistic variables using Fuzzy rules into numerical crisp value [14, 15].

For the purpose of using Fuzzy Logic to enhance the scoring of opinion words, Table 2 defines Fuzzy sets. These sets are verified with experts to ensure that defined ranges are reasonable and acceptable.

Table 2: Fuzzy Sets

Set Name	Value Ranges	Description
Positive	{ 0 to 40 }	These two sets are used to map positive and negative scores. These sets are shown in Figure 1 (denoted here as "PN" graph). The X-axis denotes the set ranges for each polarity (positive and negative); whereas, the Y-axis demotes the membership values which is between 0 and 1 only.
Negative	{ all values >= 25 }	
Low Degree	{ 0 to 40 }	These sets are used to map opinion degree scores. The values for these sets are shown below. These sets are depicted in the degree graph (denoted herewith as "D") as shown in Figure 2. The X-axis denotes the set ranges for each degree; whereas, the Y-axis demotes the membership values which is between 0 and 1 only.
Medium Degree	{ 25 to 65 }	
High Degree	{ all values >= 50 }	

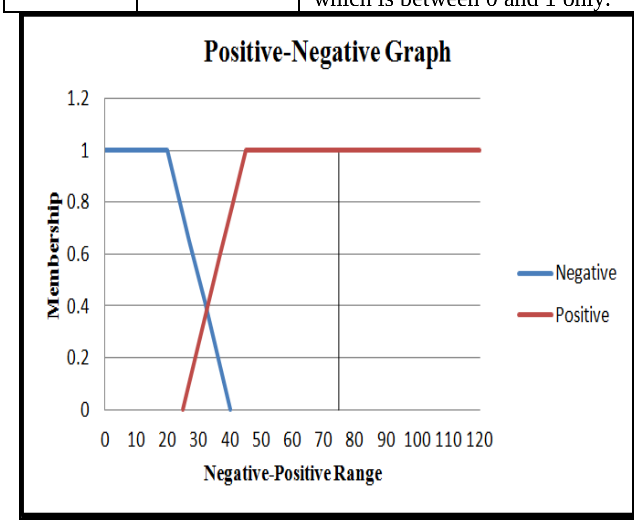

Figure 1: Positive-Negative (PN) Membership Graph

6.1 Fuzzy Domain Knowledge

In addition to above and for the fuzzification and Fuzzy rules process, opinions are broken down into 8 levels of opinions with their proposed scores. Table 3 shows these levels and scores which were verified with experts to ensure that these are logical.

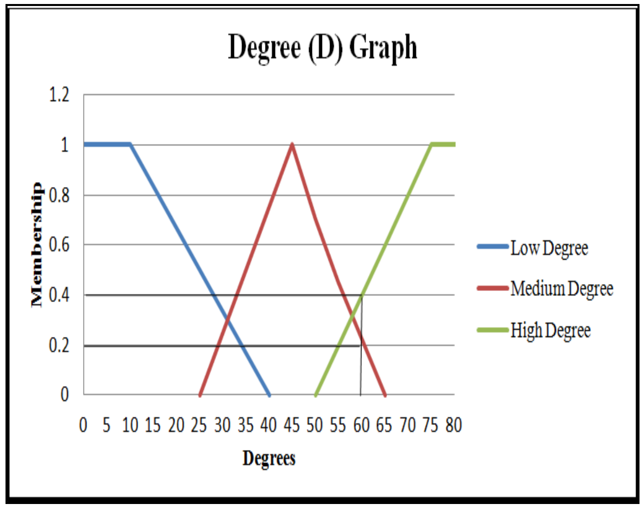

Figure 2: Opinion Degree (D) Graph

Table 3: Opinion Levels and Scores

Opinion	Code	Score
Ver Very Good	VVG	>= 85
Very Good	VG	65 - 90
Good	G	50 - 75
Above Average	AAV	35 - 60
Below Average	BAV	20 - 40
Bad	B	10 25
Very Bad	VB	5 15
Poor	PR	0 - 10
Note : No Average is used as it is in the middle between + and -. So it is neutral which is not considered here		

Based on the above levels, Table 4 shows the domain knowledge for the fuzzy rule and the deffuzification processes.

Table 4: Fuzzy Domain Knowledge

	ND	LD	MD	HD
Positive	AAV	G	VG	VVG
Negative	BAV	B	VB	PR

The proposed domain knowledge is used for the defuzzification process. Moreover, this domain mimics human logic. It can be explained as follows: If the opinion is Positive and the degree is low (LD), then the overall opinion is good (G). For example: "this seems a good house." The degree "seems" gives a low degree (LD) for the positive opinion "good," so the overall opinion will be Good (G). This builds such a logic in the Fuzzy process. The above eight levels of opinions represent the knowledge that is used when firing Fuzzy rules. These 8 levels are depicted in Figure 3. This graph is called "opinion Graph," (denoted herewith as "OP" graph) and is used during the defuzzification phase.

6.2 Applying Fuzzy Logic – An Example

Let us consider the following opinion sentence: "The served ice was very tasty." Let us assume that SentiWordNet extraction and AVAN process will produce a score of 60 for "very" and a score of 75 for "tasty." These scores are converted to percentages for graph plugging process.

- The opinion score (75) is plugged in the positive-negative (PN) graph as can be seen in Figure 1
- The degree score (60) is plugged in the Degree (D) graph as can be seen in Figure 2
- When we plug the score 60 in the (D) graph and the score 75 in the (PN) graph, a vertical line is drawn to see how many curves gets intersected (crossed) in each graph. Also, we should draw the horizontal line to see the value of y-axis which represents the membership value for both the (D) and (PN) graphs (i.e. how much that score belong into that graph or class)
- When the two values are plugged, it can be seen that the score 75 intersects the positive curve and returns Y-axis value of 1 (which means it belongs to this set with highest membership - Figure 1)
- On the other side, when the degree score of 60 is plugged in the degree graph, it shows two intersects: 1) with the medium curve and it returns a Y-axis values of 0.2 and 2) with high degree curve and it returns a Y-axis values of 0.4 (Figure 2).
- At this stage, we calculate number of Fuzzy rules. We have 2 fuzzy rules as (1 intersect x 2 intersects). The fuzzy rules will be as below
  - Rule1 : if (D.medium-degree(0.2) and PN.positive (1.0)) → OP.VG (0.2)
  - Rule2 : if (D.high-degree(0.4) and PN.positive (1.0)) → OP.VVG (0.4)
- The above rules are as per Fuzzy Logic followed calculus. The decisions of the above two fuzzy rules are obtained from the Fuzzy Domain Knowledge table (Table 4). The first rule gives 'Very Good (VG)' result as medium degree with positive opinion will result in VG decision. The second rule gives 'Very Very Good (VVG)' result as high degree with positive opinion will result in VVG decision.
- At this stage the defuzzification process starts.
  - The first Fuzzy rule puts a condition of Medium degree with positive and as a result the Fuzzy domain knowledge returns "Very Good" Answer with a level of 0.2 value which is the minimum of (.02 & 1)
  - The second Fuzzy rule puts a condition of High degree with positive and as a result the Fuzzy domain knowledge returns "Very Very Good" Answer with a level of 0.4 value with is the minimum of (.04 & 1)
- The above values for "Very Good" and "Very Very Good" (0.2 and 0.4 respectively) are plugged in the Opinion Graph (OP). This creates two areas under these two Graphs (VG & VVG) as shown in figure 3.

To calculate the crisp value for the above defuzzification process, the Centroid Function (Center of Gravity) used. This function is used to calculate the area under the shaded curve of Figure 3 and as can be seen a crisp score of 90 is an output of the defuzzification process. Hence the score for "very tasty" is 90 which is produced by Fuzzy Logic.

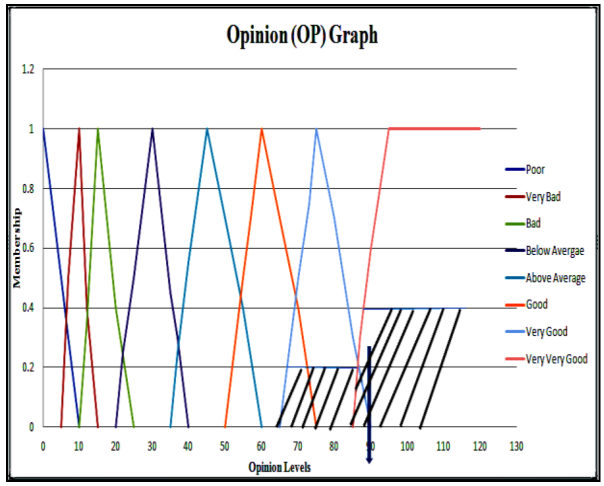

Figure 3: Opinion (OP) Graph (Representing the Fuzzy Domain Knowledge)

7.0 OPINION CLASSIFICATION RESULTS

Since this study focused on improving opinion mining in terms of classification and opinion scoring using SentiWordNet, AVAN, Opinion Degree and Fuzzy Logic, these four major elements have been selected to be the classification features to classify reviews using a suitable classifier. In order to classify a review, the literature shows several available classifiers. However, the most popular algorithms are SVM, SMO, the k- Nearest Neighbor Classifier, Naïve Bayesian Classifier, Decision Tree Classifier. To find out the most suitable machine language algorithm for the selected review rating, a 10-fold cross validation approach was used to train several Weka classifiers. Every fold of the sample training set is divided into training data and testing data. Here, a hybrid classification is carried out. Results of four best performance classifiers are shown in Table 5. These classifiers are: Sequential minimal optimization (SMO), a Library for Support Vector Machine (LibSVM), a logistic regression model (Logistic) and a tree J48. Moreover, these classifiers are commonly used in literature for opinion classification and sentiment analysis [16, 17, 18].

Table 5: Performance of Selected Classifiers

Classifier / Performance	SMO	Lib SVM	Logistic	Trees J48
Accuracy	92%	69%	92%	88%
Precision	0.923	0.665	0.921	0.884
Recall	0.92	0.69	0.92	0.88
F-Measure	0.921	0.623	0.92	0.82

Table 6: Class-wise performance using all selected classifiers

Classifier / Performance		SMO	LibSVM	Logistic	Trees J48
Accuracy		92	69	92	88
Excellent	Precision	0.947	1	0.95	0.947
	Recall	0.9	0.45	0.95	0.9
	F-Measure	0.923	0.621	0.95	0.923
Good	Precision	0.792	0.431	0.826	0.75
	Recall	0.864	1	0.864	0.818
	F-Measure	0.826	0.603	0.844	0.783
Fair	Precision	0.889	0	0.889	0.789
	Recall	0.842	0	0.842	0.789
	F-Measure	0.865	0	0.865	0.789
Poor	Precision	1	0.947	1	0.947
	Recall	1	0.947	0.947	0.947
	F-Measure	1	0.947	0.973	0.947
Very Poor	Precision	1	0.952	0.952	1
	Recall	1	1	1	0.95
	F-Measure	1	0.976	0.976	0.974

Table 7 : Performance of Classification Features Using All Selected Classifiers

Classifier	Overall Accuracy	SentiWordNet	AVAN	Opinion Degree	Fuzzy Logic
SMO	92	62%	62%	58%	60%
LibSVM	69	71%	69%	69%	69%
Logistic	92	92%	92%	93%	93%
Tree J48	88	88%	88%	88%	88%

Experiment Setup

100 opinion sentences were randomly selected out of 500 reviews of passenger reviews for an airline. These 100 instances were manually annotated as “Excellent”, “Good”, “Fair”, “Poor” or “Very Poor.” Such annotation is done in order to use this data for training using supervised learning approach.

7.1 RESULTS AND ANALYSIS

The proposed approach herewith is to classify an airline on-board services into a predefined set of Excellent, good, fair, poor and very poor using the best four performance classifiers. The results of the best four performance classifiers are shown here. Figure 4 and Table 6 show Accuracy, Precision, Recall and F-score for each classification feature using the four selected classifiers and the 5-class task (“Excellent”, “Good”, “Fair”, “Poor” ,“Very Poor.”) – Here all the classification features are used.

In view of the above, both SMO and Logistic have shown very good accuracy of 92%. However, when both of these classifiers were analyzed on individual analysis, Logistic has shown no variability and keeps on giving the same results. This shows that the Logistic classifier does not seem to perform brilliantly well with text analysis. On the other hand, SMO has been found from literature that it scaled very well with text. On this basis, it has been decided to select SMO as the classifier for comparing and analyzing results in this study [16, 17, 18].

Table 6 shows class-wise performance using all selected classifiers. This covers accuracy, precision, recall and F-measure. After examining the confusion matrix of Weka, it is

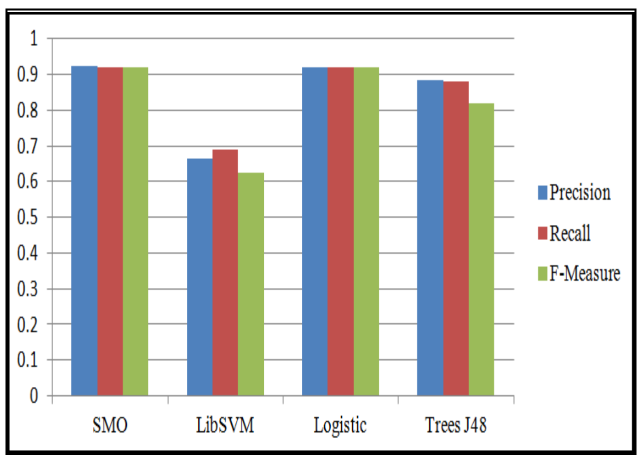

Figure 4: Classifiers’ Precision, Recall and F-Measure

noticed that most classification errors come from the “Fair” class. Moreover, Table 7 provides the accuracy for each classification feature per classifier.

Table 8 compares different proposed classification features using SMO classifier. As can be seen both SentiWordNet and AVAN performs at par and achieve an accuracy of 62%. Opinion Degree and Fuzzy logic have shown little less accuracy when compared to SentiWordNet and AVAN. However, AVAN, Opinion Degree and Fuzzy logic combined

Table 8: Comparison Among Selected Classification Features Using SMO Classifier

Classification Feature(s)	Accuracy	Precision	Recall	F-Measure
SentiWordNet (1)	62%	0.663	0.62	0.536
AVAN (2)	62%	0.663	0.62	0.536
Opinion Degree (3)	58%	0.443	0.58	0.476
Fuzzy Logic (4)	60%	0.46	0.6	0.497
(1) + (2)	87%	0.91	0.87	0.861
(1) + (3)	86%	0.906	0.86	0.847
(1) + (4)	86%	0.906	0.86	0.847
(1)+(2)+(3)+(4)	92%	0.922	0.92	0.92
(2)+(3)+(4)	92%	0.922	0.92	0.92
(2) + (3)	86%	0.906	0.86	0.847
(2) + (4)	86%	0.906	0.86	0.847
(3) + (4)	85%	0.902	0.85	0.833

Table 9: Performance Comparison (Jorge's Vs. SentiWordNet-AVAN-Fuzzy)

Classifiers	Acc.	Excellent		Good		Fair		Poor		Very Poor	
		Pr.	Re.	Pr.	Re.	Pr.	Re.	Pr.	Re.	Pr.	Re.
Work proposed by Jorge et al. (2011)											
Logistic	46.9	52.6	65	39.9	30.5	38.3	38.5	41.7	40	58.5	60.5
LibSVM	35.3	52.3	68	33.1	20.5	37.4	36.5	37.3	40.5	59.8	61
FT	43.7	49	59.5	27.6	18.5	37.6	37	39.7	35.5	55.1	68
SentiWordNet, AVAN, Opinion Degree and Fuzzy logic approach											
SMO	92	94.7	90	79.2	86.4	88.9	84.2	100	100	100	100
Logistic	92	95	95	82.6	86.4	88.9	84.2	100	94.7	95.2	100
LibSVM	69	100	45	43.1	100	0	0	94.7	94.7	95.2	100
Trees J48	88	94.7	90	75	81.8	78.9	78.9	94.7	94.7	100	95

has shown much better results when compared to SentiWordNet with an increase of 32% in accuracy.

This gives a clear indication that SentiWordNet can be much more efficient by adding one or more of the proposed classification features i.e. AVAN, Opinion Degree and Fuzzy logic. This is clearly shown in Figure 5.

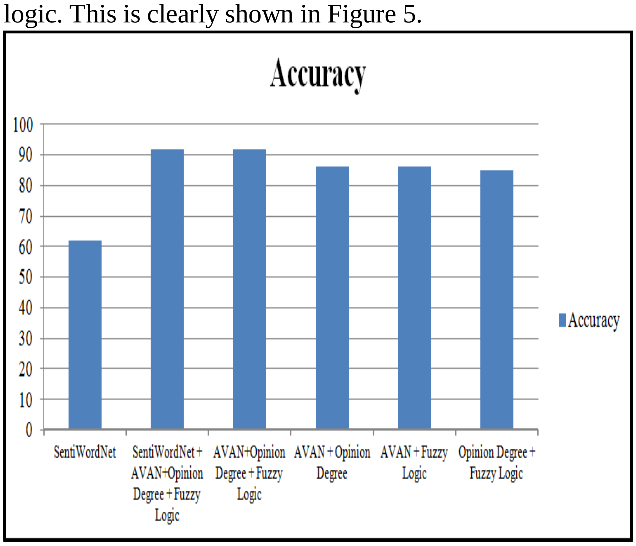

Figure 5 –Accuracy of Different Combination of Classification Features Using SMO

7.2 Benchmarking

This work has been compared and benchmarked with the work done by Jorge [19]. This is due to the following reasons:

- Authors used hotel dataset which is similar to onboard services of an airline passenger reviews
- Work is on reviews classification. Authors focused on measuring the polarity and strength of opinions. This is similar to the task set in this paper.
- They classify reviews using 5-class task "Excellent", "Good", "Fair", "Poor", "Very Poor." The paper uses a class which is similar to what is proposed here.

Jorge's objective was to measure the polarity and strength of opinions using over 1000 hotel reviews from booking.com. They have used a 5-class prediction model. This model classifies hotel reviews into "Excellent", "Good", "Fair", "Poor", "Very Poor." They have used the following classifiers: Logistic, LibSVM and Functional Tree (FT) classifiers to classify reviews using the following classifications features:

- Most Common Feature (MCF): Here the sentence is related to the feature with which it has more WordNet concepts in common.
- All Common Features (ACF): Here the sentence is linked to every feature with some concept in common.
- Most Salient Feature (MSF): Here the salience of the concepts in the sentence (that are also found in the feature cluster) are added to compute a score. Then, the sentence is linked to the highest score feature.

Jorge focused on measuring the polarity and strength of reviews by following these steps. First, those features that are key to customers are identified when evaluating a certain type of product. Second, those sentences that have opinions are located. The polarity and opinion strength of opinions are computed. Finally, a single score for each feature is computed. This is done based on the polarity of subjective sentences. At this stage a Vector of Feature Intensities is built. This vector represents the review and it is used as an input to a machine learning classifier that predicts a rating for the opinion.

They compared their results with previous approaches proposed by Carrillo de Albornoz and Pang [3, 20] and they found their approach is significantly better. Table 9 compares their results with the approach proposed in this study using

SentiWordNet, AVAN, Opinion Degree and Fuzzy logic. As can be seen from Table 8, the improvements are very significantly high as this study used different classification features which appeared to be better and more effective compared to what Jorge has proposed [19].

8.0 CONCLUSION

This paper emphasized the importance of finding more effective ways to enhance the classifications and polarity of reviews. SentiWordnet is an important resource which can be used as basis for such classification and polarity assignment. To further enhance the extracted scores of opinion words from SentiWordNet, it is important to properly measure degrees of opinions which can be expressed at different levels. Moreover, since opinions are fuzzy in nature, this paper proposed to use fuzzy logic approach to enhance scoring of an opinion. All such processing of opinion classifications and scores should be done for all opinion word types covering adjectives, verbs, adverbs and nouns (AVAN). In view of such important elements, this study identifies SentiWordNet, AVAN, Opinion Degree and Fuzzy Logic as classification features to classify customer reviews in a 5-class prediction task (Excellent, Good, Fair, Poor and Very Poor). The Results show an accuracy of 92% using SMO classifier for these features and this outperform previous work as shown in this paper. Moreover, AVAN, Opinion degree and Fuzzy Logic combined outperformed SentiWordNet alone by 30% accuracy.

REFERENCES

[1] Zabin, J. and Jefferies, A. "Social media monitoring and analysis: Generating consumer insights from online conversation," Aberdeen Group Benchmark Report. 37(9), (2008) .

[2] Esuli, A. and Sebastiani, F., “Determining Term Subjectivity and Term Orientation for Opinion Mining,” Proceedings of the 2006 EACL, (2006).

[3] Pang, B., Lee, L., Vaithyanathan, S., “Thumbs up? Sentiment classification using Machine Learning techniques,” In: Proc. of CoRR, (2002)

[4] Pang, B. and Lee, L., “Opinion mining and sentiment analysis. Foundations and trends in information retrieval,” 2(1-2), 1-135, (2008).

[5] Abbasi, A., “Affect intensity analysis of dark web forums,” Proceedings of the 2007 Intelligence and Security Informatics, IEEE, 282-288, (2007).

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

[7] Kim, S.M., Hovy, E., “Automatic identification of pro and con reasons in online reviews,” In: Proc. of COLING/ACL, pp. 483 – 490, (2006)

[8] Martineau, J., Finin, T., “Delta TF-IDF: An improved feature space for sentiment analysis,” In Proc. of 3rd AAAI Conference on Weblogs and Social Media, (2009).

[9] Brooke, J., “A semantic approach to automated text sentiment analysis,” PhD thesis, Simon Fraser University (2009)

[10] Mikalai Tsytsarau, Themis Palpanas, “Survey on mining subjective data on the web,” Data Mining Knowledge Discovery, Springer 2012, pp.478-514, (2012).

[11] Chien-Liang Liu, Wen-Hoar Hsaio, Chia-Hoang Lee, Gen-Chi Lu, and Emery Jou., “Movie Rating and Review Summarization in Mobile Environment,” IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews, Vol. 42, No. 3, May 2012, pp.397-406, (2012).

[12] Alvaro Ortigosa, José M. Martín, Rosa M. Carro, “Sentiment analysis in Facebook and its application to e-learning,” Computers in Human Behavior Journal Elsevier (2013).

[13] Kar, A. and Mandal, D. P., “Finding opinion strength using fuzzy logic on web reviews,” International Journal of Engineering and Industries. 2(1), 37-43. Li, C.-h. (2009). Sentence Factorization for Opinion Feature Mining. Proceedings of the 2009 Computational Aspects of Social Networks, 2009.

[14] Nadali, S., Murad, M. A. A. and Kadir, R. A., “Sentiment classification of customer reviews based on fuzzy logic,” Proceedings of the 2010 Information Technology (ITSim), 2010 International Symposium in: IEEE, 1037-1044, (2010).

[15] Rodrigues, L. M. and Dimuro, G. P., “Measuring the Quality of Internet Shopping: An Experiment Using Fuzzy Logic,” Proceedings of the 2011 Theoretical Computer Science (WEIT), 2011 Workshop-School on: IEEE, 60-66, (2011).

[16] Richa, S., Shweta, N. and Rekha, J., “Supervised Opinion Mining Techniques: A Survey,” International Journal of Information and Computation Technology. ISSN 0974-2239 Volume 3, Number 8, pp. 737-742. http://www. irphouse.com /ijict.htm, (2013)

[17] John, C., “Fast Training of Support Vector Machines using Sequential Minimal Optimization,” Platt. Microsoft Research. WA 98052, USA, jplatt@microsoft.com. http://www.research.microsoft.com/_jplatt, (2000)

[18] Jayashri, K. and Mayura, K., “Machine Learning Algorithms for Opinion Mining and Sentiment Classification,” International Journal of Scientific and Research Publications, Volume 3, Issue 6, June 2013 1 ISSN 2250-3153 www.ijsrp.org, (2013).

[19] Jorge, C. A., Laura, P., Pablo G. and Alberto D., “A Joint Model of Feature Mining and Sentiment Analysis for Product Review Rating,” ECIR, LNCS 6611, pp. 55–66, Springer-Verlag Berlin Heidelberg, (2011).

[20] Carrillo de Albornoz, J., Plaza, L., Gerv´as, P., “A hybrid approach to emotional sentence polarity and intensity classification,” In: Proc. of 14th CoNLL, pp. 153 – 161 (2010)

]