ISSN No. 0976-5697

Volume 5, No. 1, Jan-Feb 2014

International Journal of Advanced Research in Computer Science

RESEARCH ARTICLE

Available Online at www.ijarcs.info

Sentiment Analysis using Linguistic Structures - (Adv-Adj-Noun)

Aruna A S*, Praveen K Wilson
P.G Scholar* Assistant Professor
Department of Computer Science and Engg*, Department of IT
College of Engineering Perumon Perumon, Kollam-691601
aruna.arathi@gmail.com, praveenkwilson@gmail.com

Abstract: Nowadays individuals as well as institutions are paying increasing attention to sentiment analysis. Companies are interested in what bloggers are saying about their products. Politicians are interested in how different news media are portraying them. Governments are interested in how foreign news media are representing their actions. An automatic method is thus needed that is capable of processing and analyzing the information. In my research proposing for the first time, an Adverb-adjective- noun combinations (AAN) based sentiment analysis technique deploying linguistic analysis of adverbs of degree, adjective and abstract noun. Here define a set of general axioms (based on a classification of adverbs of degree into five categories, classification of abstract noun in two categories) for opinion analysis. This has been a significant advancement from the previous research on this domain. There is currently no automated domain-independent sentiment classification tool, with high accuracy that does not need a manually-annotated corpus. Such a tool is needed for opinion search, recommendation, summarization and mining of the increasingly web opinionated content.

Keywords: Sentiment analysis, Adverb-adjective-noun combination, adverbs of degree, abstract noun, domain-independent sentiment classification tool

I. INTRODUCTION

A. Overview:

The capability to study facts(data) about each living as well as non-living entity and derive conclusions (information) from those facts and then store them for future use and reference(knowledge), is an art which no other species has been gifted. This skill has been enriched over the time. With the advent of the internet, communicating across the globe has virtually been reduced to our palm. So, it is of utmost importance, to judicially use our vocabulary and grammar, to get the true feeling and sentiment across to the intended person/(s).Sentiment classification techniques can be used to analyze the opinion and sentiment information in the Internet. This is called Opinion Mining, Sentiment Extraction/Analysis, or Review mining. It is the area of research that attempts to identify the opinion/sentiment that a person expresses towards a product or an issue. It is a broad area of computational linguistics, natural language processing, text mining and machine learning. Sentiment analysis is concerned with analysis of direction-based text, i.e. text containing opinions and emotions.

The main objective of the research is to propose a linguistic approach to sentiment analysis where we ascribe a number from -1 for maximally negative opinion to +1 for maximally positive opinion to denote the strength of sentiment on a given topic in a sentence or document based on the score as assigned to the applicable adverb-adjective-noun combinations found in sentences.

B. Sentiment Classification:

Generally, the opinion expressed in a review document could either be a direct opinion or comparative opinion. Direct sentiment expressions on some target objects such as products, events, topics, persons. E.g.: "The picture quality of this camera is great." Comparison opinion expresses the similarities or differences of more than one object usually stating an ordering or preference. E.g.: "car x is cheaper than car y." Different types of comparatives are Non equal Gradable (less than), Educative (same), Superlative (longest). Opinion mining is carried at either sentence level, document level or feature level.

Sentence level opinion mining is performed by two tasks subjective or Objective.

Objective: I bought an iPhone a few days ago.

Subjective: It is such a nice phone.

For subjective sentences or clauses, classify positive or negative. Positive: It is such a nice phone. Negative: The phone has poor reception.

In document level, a document (e.g., a review) is classified based on the overall sentiment expressed by opinion holder.

Classes: Positive or negative

Assumption: each document focuses on a single object and contains opinions from a single opinion holder.

E.g., thumbs-up or thumbs-down, star ratings (2 stars, 3 stars…)

Opinions can also be made based on features as shown in example. "I bought an iPhone a few days ago. It was such a nice phone. The touch screen was really cool. The voice quality was clear too. Although the battery life was not long, that is ok for me. However, my mother was mad with me as I did not tell her before I bought the phone. She also thought the phone was too expensive, and wanted me to return it to the shop. …" Each feature of the product is classified and overall sentiment is judged.

C. Sentiment Detection:

Subjectivity classification [1] Subjectivity in natural language refers to aspects of language used to express opinions and evaluations. Subjectivity classification is stated as follows: Let S = {s1, . . . , sn} be a set of sentences in document D. The problem of subjectivity classification is to distinguish sentences used to present opinions and other

forms of subjectivity (subjective sentences set Ss) from sentences used to objectively present factual information (objective sentences set So),[2] where Ss U So = S. This task is especially relevant for news reporting and Internet forums, in which opinions of various agents are expressed.

Sentiment classification includes two kinds of classification forms, i.e., binary sentiment classification and multi-class sentiment classification. Given a document set D = {d1, . . . , dn}, and a pre-defined categories set C = {positive, negative}, binary sentiment classification is to classify each di in D, with a label expressed in C. If we set C* = {strong positive, positive, neutral, negative, strong negative} and classify each di in D with a label in C*,[3] the problem changes to multi-class sentiment classification. Most prior work on learning to identify sentiment has focused on the binary distinction of positive vs. negative. But it is often helpful to have more information than this binary distinction provides, especially if one is ranking items by recommendation or comparing several reviewers' opinions

II. RELATED WORKS

Two types of approaches have been used in sentiment classification studies: Machine learning and Semantic orientation. The machine learning approach involves text classification techniques. This approach treats the sentiment classification problem as a topic-based text classification problem. Any text classification algorithm can be employed, e.g., naïve Bayes, SVM, etc. Semantic orientation approach: The semantic orientation approach performs classification based on positive and negative sentiment words and phrases contained in each evaluation text .It does not require prior training in order to mine the data [4]. Two types of techniques have been used in previous sentiment classification research using the semantic orientation approach: (1) corpus-based techniques, and (2) dictionary-based techniques. Corpus-based techniques try to find co-occurrence patterns of words to determine their sentiments. Dictionary-based techniques use synonyms, antonyms and hierarchies in WordNet (or other lexicons with sentiment information) to determine word sentiments.

Each of the two approaches has its own pros and cons. The machine learning approach tends to be more accurate than the semantic orientation approach. However, a machine learning model is tuned to the training corpus; therefore, training is needed if it is applied elsewhere. In contrast, the semantic orientation approach has better generality. But its classification accuracy is often not as high as that of the machine learning approach.

An adverb-adjective combination based sentiment analysis technique [5] that uses a linguistic analysis of adverbs of degree defining a set of general axioms (based on a classification of adverbs of degree into five categories) that all adverb scoring techniques must satisfy. They used adverbs and adjectives for sentiment analysis. An adjective verb adverb based method [6] to quantify the intensity of sentiment on a continuous, real-valued scale, using careful methods to score combinations of adverbs, verbs and adjectives. Based on the score of the combination determining the expressed sentiment. Soo-Min Kim and Eduard Hovy present a system that, given a topic, automatically finds the people who hold opinions about that topic and the sentiment of each opinion [7].The system contains a module for determining word sentiment and another for combining sentiments within a sentence. Here describe an opinion as a quadruple [Topic, Holder, Claim, Sentiment] in which the Holder believes a Claim about the Topic, and in many cases associates a Sentiment, such as good or bad, with the belief. Mohamed Yassine and Hazem Hajj proposed a new framework for characterizing emotional interactions in social networks, and then using these characteristics to distinguish friends from acquaintances [8].They used facebook users as a casestudy.

The techniques used are k-mean clustering and SentiWordNet. An ontology based combination approach [9] for feature level sentiment classification. Here used the supervised learning techniques for classification of the sentiments in the software reviews. This paper proposed the combination of using Natural Language Processing techniques, ontology and support vector machine for classifying the software reviews are positive, negative or neutral.

III. PROPOSED SOLUTION

A. System Architecture:

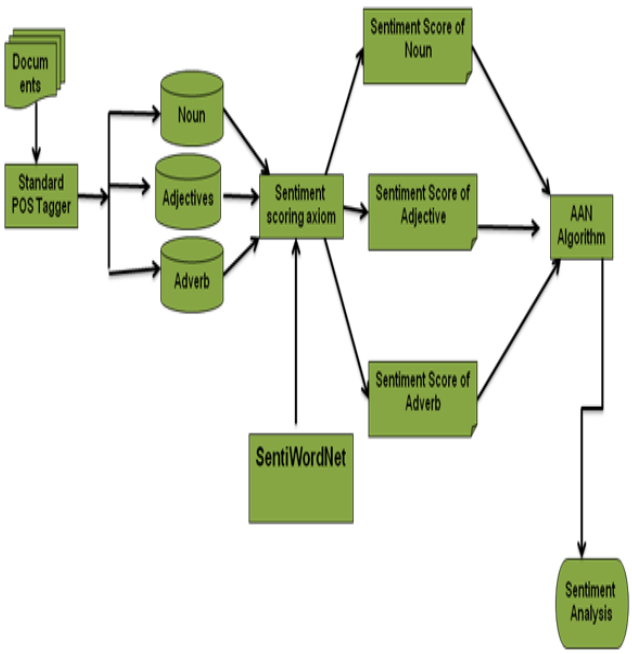

Figure.1 System architecture

B. Scoring adverb, adjective and noun:

Scoring of adverbs, adjectives and nouns are done based on SentiWordNet scores. In SentiWordNet words are arranged as synsets and each synset is assigned three sentiment scores positivity, objectivity and negativity.

a. Adverb Scoring Axioms:

In this paper, we only focus on adverbs of degree [10] such as extremely,absolutely, hardly, precisely, really - such adverbs tell us about the intensity with which something happens. Our approach introduces methods of scoring these adverbs of degree between 0 and 1 depending upon its intensity of modifying adjective. Assigning a score of one to an adverb indicate that it completely affirms an adjective,

whereas a zero score signifies that the adverb has no impact on an adjective.

Adverbs of degree are classified as follows:

a) Adverbs of affirmation: such as absolutely, certainly, exactly, totally, and so on.
b) Adverbs of doubt: such as possibly, probably, roughly, apparently, seemingly and so on.
c) Strong intensifying adverbs: such as astronomically, exceedingly, extremely, immensely, and so on.
d) Weak intensifying adverbs: such as barely, weakly, slightly, and so on.
e) Negations: such as not, never–these somewhat different than the preceding four categories as they usually negate sentiments.

After scoring adverbs, categorizing adverbs into one of the groups (Strong_affirmation, Weak_doubt, and Neutral) using k-means clustering. Scores getting from SentiWordNet for adverbs are in a scale of -1 to +1. So converting from -1 to +1 scale to [0, 1] scale using Min-Max normalization.

b. Adjective Scoring Axioms:

We used a way of scoring adjective on a scale of -1 (maximally negative) to + 1 (maximally positive).

Example: SentiWordnet provide two Senses for adjective
BEAUTIFUL.
Result from SentiWordNet:
Sense1: P: 0.75 0: 0.25 N: 0
Sense2: P: 0.625 0: 0.375 N: 0
Score (Beautiful, sense1) = (0.75 - 0) = 0.75
Score (Beautiful, sense2) = (0.625-0) = 0.625

Final score of adjective is obtained by dividing the difference between positive scores and negative scores by total number of senses of that word in adjective domain.

c. Noun Scoring Axioms:

We considering here only abstract class of noun are useful in opinion analysis .And also including noun with adverb-adjective combination will provide us much better result in opinion analysis. We classify abstract noun into these two sets:

Positive Abstract Noun such as joy, victory, faith and so on positively reinforces an opinion.

Negative Abstract Noun such as destruction, sorrow, stupidity, pain, failure, hatred, sadness and so on negatively reinforce an opinion.

Scoring of Noun will be either + 1 or -1 depending upon SentiWordNet positive and negative polarity decision.

$$\text{Scoring(Noun)} = \frac{(\sum \text{P-score}) - (\sum \text{N-score})}{n}$$

P-Score -- Postive score of a word obtained from SentiWordnet
N-Score -- Negtive score of a word obtained from SentiWordNet.
n is the no of Sense obtained for that word in Noun domain of SentiWordNet.
Final score will be -1 if Scoring (Noun) <0 and + 1 if Scoring (Noun) >0.

C. AAN Algorithm:

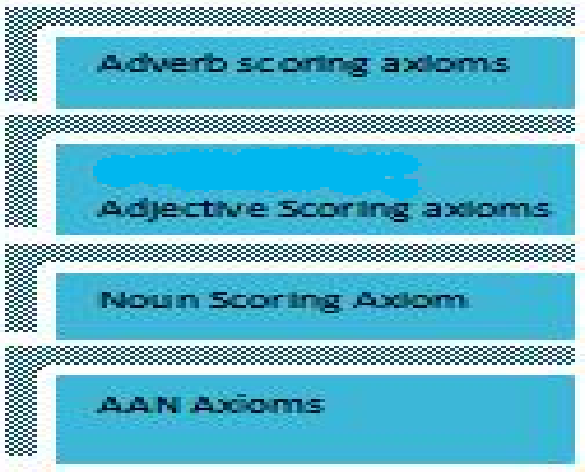

Figure.2 Steps of AAN algorithm

Here for getting final sentiment score (number between +1 and -1) applying Unary AAN algorithm and Binary AAN algorithm.

a. Unary AAN Algorithm :

Let AFF, DOUBT, WEAK, STRONG and MIN respectively be the sets of adverbs of affirmation, adverbs of doubt, adverbs of weak intensity, adverbs of strong intensity and minimizers. Suppose fSense is any unary AAN scoring function that takes as input, one adverb, one adjective, and one noun and returns a number between -1 and +1. AAN scoring function fSense should satisfy the following axioms:

a) fadv € AFF U STRONG & If sc(adj) < 0,then
fSense(adv, adj, noun) = (sc(adj) - (1- sc(adj)) x sc(adv)) *sc(noun)

If adv € AFF U STRONG & If sc (adj)>0, then
fSense(adv, adj ,noun) = (sc(adj) + (1- sc(adj)) x sc(adv)) *sc(noun)

b) If adv € WEAK U DOUBT & sc (adj) > 0, then:
fSense(adv, adj ,noun) = (sc(adj) - (1- sc(adj)) x sc(adv))*sc(noun)

If adv € WEAK U DOUBT & sc (adj) < 0, then:
fSense(adv, adj ,noun) = (sc(adj) + (1- sc(adj)) x sc(adv)) *sc(noun)

c) If adv € MIN then:
fSense (adv, adj , noun) = - sc(adj) *sc(noun)

To calculate the sentiment of the document considering two cases. 1. If Standard Deviation of fSense scores for all the sentences is above a certain threshold value. Then, Median of fSense scores for all these sentences is taken as final Sentiment scoring of that blog or article. 2. If Standard Deviation of fSense scores for all these sentences is below that threshold value. Then, Arithmetic Mean of fSense scores for all these sentences is taken as final Sentiment scoring of that blog or article.

For including social acronyms and words with intentional misspelling we have implemented a dictionary containing a set of social acronyms and words with intentional misspelling. It is impossible to include all possible social acronyms. So we have collected most common social acronyms in reviews and its expansions from the websites .If the social acronyms and misspelled words specified in the dictionary are present in the review, then replacing it with the exact word and find the

sentiwordnet score for it from sentiwordnet as by using unary and binary AAN. We also put an option to edit the social acronyms and misspelled words.

b. Binary AAN Algorithm:

Binary AAN algorithm is applied if the review containing combinations like "possibly less expensive".ie <adv1, adv2, adj, noun>[11].We assign a score to a binary AAN <advl . adv2 > <adj . noun> as follows. First, we compute the score fSense (adv2 , adj , noun). This gives us a score denoting the intensity of the unary AAN <adv2.adj. noun> which we denote AAC1. We then apply fSense to (adv 1,AAC1 ,1) and return that value as the answer.

c. Ternary AAN Algorithm:

Ternary AAN algorithm is applied if the review containing combinations like "very very very expensive".ie <adv1, adv2, adv3, adj, noun>.We assign a score this as follows. First, we compute the score fSense (adv3 , adj , noun). This gives us a score denoting the intensity of the unary AAN <adv3.adj. noun> which we denote AAC1. We then apply fSense to (adv2,AAC1 ,1) which we denote as AAC2.Then apply fSense to (adv3,AAC2,1) and return that value as the answer.

IV. RESULTS AND DISCUSSION

The main purpose of our research is to propose an automatic domain independent sentiment analysis tool with high accuracy and that does not require a manually annotated corpus. Our research addressed an effective solution for the Sentiment detection Using ANN. As we are using AAN combination that helps to achieve better results. As first step we are extracting AAN combinations from the review using POS tagging, which is to be analysed. Then assigning score for adverbs, adjectives and nouns based on SentiWordNet and applying the equation. After scoring, assigning a score for the AAN combination. Based on this score determining the expressed sentiment of the given sentence or document. Here performing multiclass sentiment classification as Strong Positive, Positive, Weak Positive, Neutral, Weak Negative, Negative and Strong Negative.

V. CONCLUSION AND FUTURE WORK

Sentiment analysis is an important current research area. Sentiment detection has a wide variety of applications in information systems, including classifying reviews, distinguishing synonyms and antonyms, extending the capabilities of search engines, summarizing reviews, tracking opinions in online discussions, and analyzing survey responses. In my project addressed an effective solution for the Sentiment detection using ANN. As this has taken the Adverb-Adjective-Noun Combination along with different domain combinations, it helps to achieve better results. Instead of aggregating the scores of both adverbs and adjectives using simple scoring functions, our work propose an axiomatic treatment of AAN combinations based on the linguistic categories of adverbs de fined.

One main future work proposed here is to make our work more accurate. That is including sentiment scores of words of foreign languages transliterated to English and emoticons which are common in a review. And another future study we like to propose is to increase the accuracy in analyzing reviews containing opinions from different opinion holders and on different topic. Finally, multilingual sentiment analysis is another major topic for future research.

VI. REFERENCES

[1]. "A survey on sentiment detection of reviews" Huifeng Tang, Songbo Tan, Xueqi Cheng Information Security Center, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080, PR C1hina

[2]. H. Yu and V. Hatzivassiloglou, towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences, In Proceedings of EMNLP-03, 2003

[3]. T. Wilson and J. Wiebe and R. Hwa, Just how mad are you? Finding strong and weak opinion clauses, AAAl-04, 2004.

[4]. B. Pang and L. Lee and S. Vaithyanathan, Thumbs up? Sentiment Classification Using Machine Learning Techniques, 2002.

[5]. F.Benamara et al., "Sentiment Analysis: Adverbs and Adjectives Are Better than Adverbs Alone," Proc. 2007 in Conf. Welogs and Social Media

[6]. AVA: Adjective-Verb-Adverb Combinations for Sentiment Analysis V.S. Subrahmanian and Diego Reforgiato, *SentiMetrix and University of Maryland*

[7]. S.O Kim and E. Hovy, Determining the Sentiment of Opinions, Coling04, 2004.

[8]. Mohamed Yassine Hazem Hajj," A Framework for Emotion Mining from Text in Online Social Networks", Department of Electrical and Computer Engineering American University of Beirut Beirut, Lebanon

[9]. Khin Phyu Phyu Shein, Thi Thi Soe Nyunt, Sentiment Classification based on Ontology and SVM Classifier, 2010 Second International Conference on Communication Software and Networks.

[10]. A. Lobeck, Discovering Grammar. An Introduction to English Sentence Structure, New York/Oxford: Oxford University Press, 2000.

[11]. J.K. Sing, Souvik Sarkar, Tapas Kr. Mitra Development of a Novel Algorithm for Sentiment Analysis Based on Adverb-Adjective-Noun Combinations