IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 18, Issue 2, Ver. I (Mar-Apr. 2016), PP 24-28
www.iosrjournals.org

Aspect Based Sentiment Analysis Survey

Naveen Kumar Laskari[1], Suresh Kumar Sanampudi[2]
[1](Department of IT, BVRIT Hyderabad, JNT University Hyderabad, India)
[2](Department of IT, JNTUH College of Engineering Jagitial, Karimnagar, India)

Abstract:*Sentiment analysis or Opinion mining is becoming an important task both from academics and commercial standpoint. In recent years text mining has become most promising area for research. There is an exponential growth with respect to World Wide Web, Mobile Technologies, Internet usage and business on electronic commerce applications. Because of which web opinion sources like online shopping portals, discussion forums, peer-to-peer networks, groups, blogs, micro blogs and social networking applications are extensively used to share the information, experience and opinions. In sentiment analysis, the opinion is evaluated to its positivity, negativity and neutrality with respect to the complete document or object. But this level of analysis does not provide the necessary detailed information for many applications. To obtain more fine grained analysis we need to go to Aspect Based Sentiment Analysis (ABSA). Aspect Based Sentiment analysis introduces a suite of problems which require deeper NLP capabilities and also produces a rich set of results.*
Keywords:*Sentiment Analysis, Aspects, Term, Opinion Mining and Polarity.*

I. Introduction

Sentiment Analysis (SA), so called Opinion Mining (OM) is the study through which analysis of people's opinions, sentiments, appraisals, evaluations, attitudes and emotions towards entities of various kinds expressed in Text. An entity can be a product of various types, services, organizations, individuals, issues, events, topics and their aspects. Opinion mining or sentiment analysis have spread to every possible domain, from health care, financials, consumer products, service, telecommunications and e-commerce to social events, political campaigns and elections.

The two expressions SA and OM are interchangeable, they express the mutual meaning. However, in view of some researchers SA and OM have different notions. Opinion Mining extracts and analyses people's opinion about an entity while Sentiment Analysis identifies the sentiment expressed in a text then analyses it. From Data Mining standpoint Sentiment Analysis or Opinion Mining can be considered as multi-step classification problem.In the first step, there are three main classification types like Document Level, Sentence Level and Aspect Level. In second step related to identification of polarity of the document or sentence or aspect as positive or negative or neutral. Document level sentiment analysis aims to classify an opinion document as positive or negative by considering the whole document as basic information unit. Sentence level sentiment analysis aims to determine whether a particular sentence expresses positive or negative opinion

Classifying opinionated texts at document level or at sentence level is useful in many applications; but it does not provide necessary details needed for many applications. A positive opinionated document about a particular entity does not mean that the author has complete positive opinion about all the features of the entity. Likewise, a negative opinionated document does not mean that the author does not at all like all the features of the entity. In a typical opinionated text the authors writes both positive and negative opinions with respect to entities and their attributes. Majority of current approaches, however attempt to identify the overall polarity of a document, sentence, paragraph or text irrespective of the entities involved. To obtain these hidden details, we need to go to Aspect Based Sentiment Analysis.

Aspect Based Sentiment Analysis aims to identify the aspects of entities being used in expressing sentiments. It is also used to determine the sentiment that expressed by author towards each aspect of the entity. Aspect Based Sentiment Analysis is much critical in mining and summarizing opinions from any kinds of datasets.In the last decade several Aspect Based Sentiment Analysis systems have been developed for wide variety of entities like Movie reviews, Travels, Digital cameras, Services, Computers and Restaurants.

In more detail, Aspect Based Sentiment Analysis systems receive a set of texts (product reviews, comments, forum discussions and messages from web 2.0 sources) discussing about a particular entity (Ex: Phone). The system attempts to find the main (frequently discussed features) Aspects ('Screen', 'Size', 'Price') of the entity and to find the sentiment expressed towards each aspect and their summary of polarity.

The rest of the paper is organized as follows. Section 2 provides the details of Data sets and required pre-processing to be done for Aspect Based Sentiment Analysis. Section 3 provides the details of Aspect Based sentiment Analysis Tasks, Section 4 gives the comparison of results from various approaches used by researchers and section 5 provides conclusion and future scope of the ABSA.

II. Data sets and pre-processing

To evaluate the performance of the Aspect Based Sentiment Analysis and Sentiment Analysis various train and test datasets are available for domains like restaurants, laptop reviews and hotels (which have been used as a part of SemEval and other conferences). Researchers have collected other datasets from tripadviser.com, amazon.com and other resources for analysis. Before implementing the actual task of ABSA, pre-processing of the data has to be done to get more insights of the data. Researchers have used techniques like removingpunctuations and stop words, normalization of data, tokenization using Stanford Parser, Parts-of-Speech tagging using Stanford PoS tagger and OpenNLPchunkerto get chunk level information.

III. Aspect Based Sentiment Analysis Tasks

The main goal of Aspect Based Sentiment Analysis is to identify the aspects of the given target entities and sentiment expressed for each aspect. To meet the objectives of Aspect Based Sentiment Analysis, the work can be done primarily in two phases. In phase-1 extraction of aspect terms and grouping aspect terms into aspect categories will be done. In phase-2 identification of polarity of the aspect terms and polarity of the aspect categories of each sentence will be done. These phases of activities are divided into four sub tasks namely: Aspect Term Extraction (ATE), Aspect Term Polarity (ATP), Aspect Category Detection (ACD) and Aspect Category Polarity (ACP)

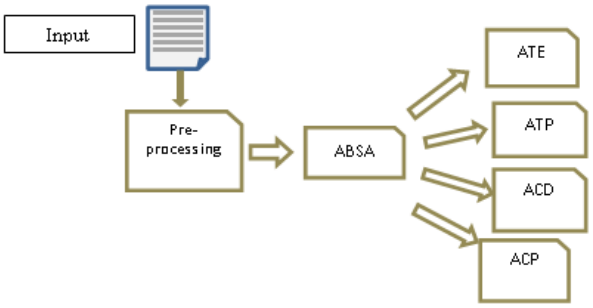

Fig.1 Aspect Based Sentiment Analysis Tasks

3.1 Aspect Term Extraction

Aspect Term Extraction (ATE) isto identify all the aspect terms present in each review sentence or comment. It recognizes aspects of the entity and more generally it can be seen as an information extraction task.Most of the researchers stated that an aspect can be expressed by a noun, verb, adverb and adjective.Out of which 60% - 70% of aspect terms are explicit nouns.The aspectterms could also consist of multiword entities suchas “battery backup” and “special chicken biryani” etc. Dealing with multiword entities and aspects is much critical than single word aspects.In the process of extracting aspect terms, researchers have used various features like Word N-grams, Bigrams, Name List, Head word, Word cluster, Casting, POS tagging, Parse dependencies, Relations and Punctuation marks. Methods like Conditional Random Fields (CRF), Support Vector Machines (SVM), Random trees and Random Forest have been used for extracting aspect terms.

Various methods have been proposed by researchers for Aspect Extraction Task like (i) FREQ baseline, (ii) H&L method, (iii) H&L+W2V method and (iv) FREQ+W2V method. All the methods are of unsupervised and are important to be used across domains with minimal changes. The FREQ baseline returns the most frequent (distinct) nouns and noun phrases of the reviews in each dataset (restaurants, hotels, and laptops), ordered by decreasing sentence frequency (how many sentences contain the noun or noun phrase). The method of Hu and Liu (2004), dubbed H&L, first extracts all the distinct nouns and noun phrases from the reviews of each dataset and considers them candidate distinct aspect terms. In addition to the above mentioned methods researchers have used LDA (Latent Dirichlet Allocation) based methods.

In [6] a co-occurrence based method for category detection and a dictionary based sentiment classification algorithm is used. Aspects can be identified by annotation process in [6]. However, by using the training set to count how often each word appears within an aspect, a simple probability can be computed representing the chance that this word is an aspect word or not. This probability is used to filter the set of noun phrases, such that only noun phrases remain that have at least one word for which the aspect probability having >= 0.05 and for those noun phrases, all leading words in the noun phrase with a probability below 0.05 are removed. This will remove words like determiners from the initial noun phrase, as those are not included in the aspect term.

In [7] author modelled Aspect extraction as sequential labelling task and extract features to be used for CRF training. Besides the common features used in traditional Named Entity Recognition (NER) systems, and also utilize extensive external resources to build various name lists and word clusters.In [8] supervised machine learning algorithm is used to extract the aspect term. An aspect can be expressed by a noun, adjective, verb or

adverb. But the recent research in (Liu, 2007) shows that 60-70% of the aspect terms are explicit nouns. The aspect terms could also consist of multiword entities.In [9] Aspect term can be extracted by casting it as a sequence tagging task where each token in a candidate sentence is labelled as either Beginning, Inside or Outside (BIO). Then employed conditional random fields (CRF) and a liner chain CRF used to estimate the conditional probability.

In [10] the authors employ a graph co-ranking approach. They model aspect terms and opinion words as graph nodes, and then they generate three different sub-graphs defining different types of relations between the nodes. Finally they rank the nodes using a combined random walk on the three sub graphs to obtain a list of reliable aspect term candidates.In [11] Conditional Random Fields (CRF) is used to extract aspect terms and used BIO model for representing aspect terms.

3.2 Aspect Term Polarity

For a given set of aspect terms within a sentence, the task is to determine the polarity of each aspect term: positive, negative, neutral or conflict (i.e., both positive and negative).In identification of Aspect term polarity various features like Word N-grams, Polarity of neighboring adjectives, Neighboring POS tags and Parse dependencies and relations have been extensively used by researchers.

In [6] sentiment of aspect can be computed by using sentiment value of each n-gram and distance between the n-gram and the aspect. In [7] the author built the aspect lexicon based on other information such as POS for polarity identification. In [8] Random Forest classification method has been developed and a new class called conflict has been introduced. In the classification process they have used features like local context, POS, chunk, prefix and suffix. In [9] aspect term polarity has been extracted by using various features like word N-grams, polarity of neighboring adjectives, neighboring POS tags and parse dependencies and relations.

In [10] the author developed a polarity lexicon reusing the generated Word2Vec model for the corresponding domain with the intuition that a polarity word in a domain should be more "similar" to a set of "very positive" words than to a set of "very negative" words, and vice versa. And employed the in-domain generated Word2Vec models because the polarity of words may vary between domains and wanted to capture the polarity for each particular domain.In [11] the words affecting the sentiment of the aspect term are assumed to be close in most of cases and thus used a context window of 10 words in both directions around the target aspect term.

3.3. Aspect Category Detection

Aspect Category Detection is to identify the major categories discussed in each sentence. Aspect categories are typically coarser than the aspect terms as defined in Aspect Term Extraction, and sometimes they do not necessarily occur as terms in the sentence.

Aspect category detection is based on a set of binary Maximum Entropy classifiers, one for each class. The final decision is simply assembled from decisions of individual classifiers.In [5] Aspect category classification is based on a set of one-vs-all binary classifiers, one classifier for each category found in the training set. For each sentence in the training set, they extracted features from all words in the sentence to create a training example.

In [6] the co-occurrence based algorithm is used for category detection. The central idea in this algorithm is a co-occurrence matrix that captures the frequency of the co-occurrences between words (i.e., the lemmas of the words) in the sentence and the annotated aspect category. This gives a mapping from words to aspect categories.In [9] aspect category detection is considered as multi label classification problem, i.e., given an instance, it should predict all labels that instance fit into.In [11] Aspect category detection is based on a set of binary Maximum Entropy classifiers, one for each class. The final decision is simply assembled from decisions of individual classifiers.

In [14] Aspect category detection task is done by identifying every entity E and attribute A pair E#A towards which an opinion is expressed. Authors in [14] have used supervised classification approach where they used C one-vs-all Random Forest classifier, for each of the C{entity, attribute} pair of aspect categories in the training data, with bag of words based approach.

3.4 Aspect Category Polarity

Aspect Category Polarity is the task which takes the information from the previous task (Aspect Category Detection) and determines the polarity (positive, negative, neutral or conflict) of each aspect category discussed in review sentence. In [6] sentiment of aspect category is computed by calculating the distance between n-gram and the corresponding aspect.In [9] aspect category polarity has been detected using just unigram and bigram features.

In [11] for aspect category polarity detection window size is not restricted, the complete sentence is considered, and maximum entropy classifier is used for distinguishing one category with other. In [14] for sentiment polarity classification, authors have extracted Bag of Words and WordnetSynset features from both train

and test data. Then implemented on variety of classifiers (like Stochastic Gradient Descent, SVM, Adaboost) multiple times and stored the confidence scores obtained from decision functions of each of these classifiers.

IV. Evaluation and Comparison of Results

The performance of the various approaches for the subtasks of Aspect Based Sentiment Analysis hasbeen evaluated. The performance can be evaluated through precision (P), recall (R) or F-score (F) depending on the subtask, which are defines as

$$P = \frac{TP}{TP+FP}, \qquad R = \frac{TP}{TP+FN} \qquad \text{and} \qquad F = \frac{2.P.R}{P+R}$$

Where TP (True Positive), TN (True Negative), FP (False Positive) and FN (False Negative) are the cases correctly classified or incorrectly classified, and F-score is the some average of precision and recall metrics.The following table will give summary information about the performance of the work done so far.

Table 1 Comparison of results

S. No	Paper title	Restaurant Dataset			Laptop Dataset		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score
1	Supervised Machine Learning System for Aspect Category Classification and Opinion Target Extraction	0.6637	0.6155	0.6268	0.6425	0.4483	0.5086
2	Commit-p1wp3: A co-occurrence based approach to aspect-level sentiment analysis	0.909	0.388	0.544	0.836	0.148	0.252
3	DLIREC: Aspect Term Extraction and Term Polarity Classification System	0.8625	0.8272	0.8401	0.8480	0.6713	0.7455
4	IITP: Supervised Machine Learning for Aspect based Sentiment Analysis	0.7344	0.6850	0.7088	0.6530	0.6490	0.6509
5	V3: Unsupervised Aspect Based Sentiment Analysis for SemEval-2015	0.694	0.694	0.45	0.683	0.683	0.25
6	UWB: Machine Learning Approach to Aspect-Based Sentiment Analysis	0.8270	0.7628	0.7936	0.7733	0.4954	0.6039
7	SIEL: Aspect Based Sentiment Analysis in Reviews	0.64	0.51	0.57	0.64	0.51	0.57

V. Conclusion and Future Scope

Most of the current active work is done part of the Aspect Based sentiment analysis by annotation process for specific to a particular domain or application like laptop reviews, restaurant reviews, movie reviews and travels. Researchers could not able to apply methods across all the datasets, because of not matching of aspect words in cross domain frame work has not been defined.

A framework can be designed for Aspect Based Sentiment Analysis through which sentiment analysis can be done for any domain datasets.So for most of the work carried out with respect to Sentiment Analysis, Opinion Mining and Aspect Based Sentiment Analysis with machine learning techniques only, moreover Deep Learning is becoming a prominent are for the research. Deep learning techniques can be implemented towards Sentiment Analysis and Aspect Based Sentiment Analysis.

References

Journal Papers

[1]. K Vivekanandan et.al International Journal of Computer Applications, "Aspect Based Sentiment Analysis:Survey", Volume 106-No.3

Books

[2]. Liu, Bing. "Sentiment analysis and opinion mining." Synthesis Lectures on Human Language Technologies 5.1 (2012): 1-167.

[3]. Pang, Bo, and Lillian Lee. "Opinion mining and sentiment analysis." Foundations and trends in information retrieval 2.1-2 (2008): 1-135.

Theses:

[4]. Pavlopoulos, Ioannis, and Ιωάννης Παυλόπουλος. "Aspect based sentiment analysis." (2014).

Proceedings Papers

[5]. Toh, Zhiqiang, and Jian Su. "NLANGP: Supervised Machine Learning System for Aspect Category Classification and Opinion Target Extraction."

[6]. Schouten, Kim, Flavius Frasincar, and Franciska de Jong. "Commit-p1wp3: A co-occurrence based approach to aspect-level sentiment analysis." Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). 2014.

[7]. Zhiqiang, Toh, and Wang Wenting. "DLIREC: Aspect Term Extraction and Term Polarity Classification System." (2014).

[8]. Gupta, De1epak Kumar, and Asif Ekbal. "IITP: Supervised Machine Learning for Aspect based Sentiment Analysis." SemEval 2014 (2014): 319.

[9]. Malhotra, Nishtha, et al. "SAP-RI: A Constrained and Supervised Approach for Aspect-Based Sentiment Analysis." SemEval 2014 (2014): 517.
[10]. Garcıa-Pablos, Aitor, MontseCuadros, and German Rigau. "V3: Unsupervised Aspect Based Sentiment Analysis for SemEval-2015 Task 12."
[11]. Brychcın, Tomáš, Michal Konkol, and Josef Steinberger. "UWB: Machine Learning Approach to Aspect-Based Sentiment Analysis." SemEval 2014 (2014): 817.
[12]. Pontiki, Maria, et al. "Semeval-2014 task 4: Aspect based sentiment analysis." Proceedings of the 8th international workshop on semantic evaluation (SemEval 2014). 2014.
[13]. Pontiki, Maria, et al. "Semeval-2015 task 12: Aspect based sentiment analysis." Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), Denver, Colorado. 2015.
[14]. Guha, Satarupa, Aditya Joshi, and Vasudeva Varma. "SIEL: Aspect Based Sentiment Analysis in Reviews." SemEval-2015 (2015): 759.