JU_CSE: A Conditional Random Field (CRF) Based Approach to Aspect Based Sentiment Analysis

Braja Gopal Patra, Soumik Mandal, Dipankar Das and Sivaji Bandyopadhyay
Department of Computer Science & Engineering,
Jadavpur University, Kolkata, India
`brajagopal.cse@gmail.com, mandal.soumik@gmail.com,`
`dipankar.dipnil2005@gmail.com, sivaji_cse_ju@yahoo.com`

Abstract

The fast upswing of online reviews and their sentiments on the Web became very useful information to the people. Thus, the opinion/sentiment mining has been adopted as a subject of increasingly research interest in the recent years. Being a participant in the Shared Task Challenge, we have developed a Conditional Random Field based system to accomplish the Aspect Based Sentiment Analysis task. The *aspect term* in a sentence is defined as the target entity. The present system identifies *aspect term*, *aspect categories* and their sentiments from the *Laptop* and *Restaurants* review datasets provided by the organizers.

1 Introduction

In recent times, the research activities in the areas of Opinion Mining/Sentiment Analysis in natural language texts and other media are gaining ground under the umbrella of subjectivity analysis and affect computing[1]. The reason may be the huge amount of available text data in Social Web in the forms of news, reviews, blogs, chat and twitter etc. Majority of research efforts are being carried out for the identification of *positive* or *negative* polarity from the textual contents like sentence, paragraph, or text span regardless of the entities (e.g., *laptops*, *restaurants*) and their aspects (e.g., *battery*, *screen*; *food*, *service*).

Aspect is a multinomial distribution over words that represent a more specific topic in reviews (Jo and Oh, 2011). For example, in case of *Laptop* reviews, "*touchpad*" is considered an aspect. Similarly, given a predefined entity, an *aspect term* describes a specific aspect of that entity (e.g., for the entity "*restaurant*", "*wine*" can be an *aspect term*). *Aspect term* can be appeared as a single word (e.g., "*menu*") or multiple words ("*side dish*").

It is observed that for a particular entity, one or more number of *aspect terms* can be grouped into a single category (e.g., *aspect terms* "*drinks*", "*main course*" belongs to the same category, "*food*").

The main goal of the Aspect Based Sentiment Analysis (ABSA) (Pontiki et al., 2014) task is to identify the *aspect terms* and their categories from the given target entities as well as to identify the sentiments expressed towards each of the *aspect terms*. The datasets provided by the shared task organizers consist of customer reviews with human-annotations.

We have participated in all of the four tasks. A combination of Conditional Random Field (CRF) based machine learning algorithm and rule based techniques has been adopted for identifying the *aspect term*, *aspect category* and their sentiments. We have used several features like Part of Speech (POS), Stanford dependency relations[2], WordNet information, and sentiment lexicon (SentiWordNet[3]) to accomplish these tasks.

The rest of the paper is organized in the following manner. Section 2 provides the details of previous works. Section 3 provides an elaborative description of the data used in the task. Features used in these experiments are described in Section 4. The detailed setup of experimentation and analysis of the results are described in Sec-

[1]http://www.saaip.org/

[2] http://nlp.stanford.edu/software/lex-parser.shtml
[3] http://sentiwordnet.isti.cnr.it/

tion 5. Finally, conclusions and future directions are presented.

2 Related Work

It has been observed that most of the previous works on aspect detection were based on information extraction, to find the most frequent noun phrases (Hu and Liu, 2004). This approach is generally useful in finding aspects which are strongly associated with a single noun. But, one principal disadvantage of this approach is that it cannot detect the *aspect terms* which are of low frequency and noun phrases (e.g., different names of *dishes* like *Biryani, Dosa* and *Uttapam* etc. for the *aspect category*, *"food"*). The proposed work of such problem involves semantic hierarchy, rule-based or combination of both (Popescu and Etzioni 2005). More recent approaches of aspect detection are based on topic modelling, that use Latent Dirichlet Allocation (LDA) (Brody and Elhadad, 2010). But, the standard Latent Dirichlet Allocation (LDA) is not exactly suitable for the task of aspect detection due to their inherent nature of capturing global topics in the data, rather than finding local aspects related to the predefined entity. This approach was further modified in Sentence-LDA (SLDA) and Aspect and Sentiment Unification Model (ASUM) (Jo and Oh, 2011). Similarly, the identification of focussed text spans for opinion topics and targets were identified in (Das and Bandyopadhyay, 2010).

Snyder and Barzilay (2007) addressed the problem of identifying categories for multiple related *aspect terms* appeared in the text. For instance, in a restaurant review, such categories may include *food*, *ambience* and *service* etc. In our task, we call them as *aspect* or *review categories*. The authors implemented the Good Grief decoding algorithm on a corpus collected on restaurant review[4], which outperforms over the famous PRank algorithm (Crammer and Singer, 2001).

Ganu et al., (2009) have classified the restaurant reviews collected from *City search New York*[5] into six categories namely *Food, Service, Price, Ambience, Anecdotes, and Miscellaneous*. Sentiment associated with each category has also been identified and both the experiments were carried out using *Support Vector Machine* classifiers. Finally, they implemented the regression based model containing MATLAB regression function (*mvregress*) to give rating (1 to 5) to each review.

To determine the sentiment or polarity of the *aspect term* and *aspect category*, we need a prior sentiment annotated lexicon. Several works have been conducted on building emotional corpora in different English languages such as SentiWord-Net (Baccianella et al., 2010), WordNet Affect (Strapparava and Valitutti, 2004) (Patra et al., 2013) etc. Among all these publicly available sentiment lexicons, SentiWordNet is one of the well-known and widely used ones (number of citations is higher than other resources[6]) that has been utilized in several applications such as sentiment analysis, opinion mining and emotion analysis.

Several works have been performed on the automated opinion detection or polarity identification from reviews (Yu and Hatzivassiloglou, 2003; Hu and Liu, 2004). Yu and Hatzivassiloglou (2003) has focused on characterizing opinions and facts in a generic manner, without examining who the opinion holder is or what the opinion is about. Then, they have identified the polarity or sentiment of the fact using Naive Bayes classifier. Hu and Liu, (2004) has summarized the customer review and then identified the sentiment of that review. They have achieved promising accuracy in case of identifying polarity of the reviews.

3 Data

The sentences collected from the customer reviews of *Restaurants* and *Laptops* are used in these tasks. The training data of *Restaurant* reviews contains 3041 English sentences annotated with *aspect terms* and *aspect categories* along with their *polarity*. The training data of *Laptop* reviews contains 3045 sentences annotated with *aspect terms* along with their *polarity*. The test data contains 800 sentences from each of the review sets.

An example extracted from the corpus is as follows:

But the <u>staff</u> was so horrible to us.

Here, "*staff*" is the *aspect term* and its polarity is "*negative*". The *aspect category* is "*service*" and *polarity* of the *aspect category* is also "*negative*".

---

[4] http://people.csail.mit.edu/bsnyder/naacl07/
[5] http://www.citysearch.com/guide/newyork-ny-metro

[6] http://citeseerx.ist.psu.edu/index

4 Feature Analysis

In general, the feature selection always plays an important role in any machine learning framework and depends upon the data set used for the experiments. Based on a preliminary investigation of the dataset, we have identified some of the following features. Different combinations of the features have also been used to get the best results from the classification task.

Parts-of-Speech (POS): the *aspect terms* are basically represented by the noun phrases. On the other hand, the POS tag plays an important role in *aspect term* identification (Hu and Liu, 2004; Brody and Elhadad, 2010). Thus, we have used the *Stanford CoreNLP*[7] tool to parse each of the review sentences to find out the part-of-speech tag of each word and included them as a feature in all of our experiments.

POS Frequency: We have observed that the *aspect terms* surrounded by a noun or adjective are also denoted as *aspect terms*. Therefore, we have utilized this information in our system. For example, in the phrase "*external_JJ mouse_NN*". Here the word "*mouse*" is an object and *aspect term*. The word "*external*" is also tagged as *aspect term*.

Before be verb: We have observed that the nouns occur before the "*be*" verbs denote the *aspect terms* in most of the cases. e.g. "*The hard disk is noisy*". Here "*hark disk*" is an *aspect term* and is followed by the "*be*" verb "is".

Inanimate words: In case of the *Restaurant and Laptop* reviews, we observed that many of the inanimate nouns occur as *aspect terms*. We have used the hyponym tree of *RiTa.WordNet*[8] to identify the inanimate words. For example, in the following sentence, the words *food*, *kitchen* and *menu* are inanimate nouns occurred as *aspect terms*.

"The *food* is uniformly exceptional, with a very capable *kitchen* which will proudly whip up whatever you feel like eating, whether it's on the *menu* or not."

Dependency Relation for finding Object: We have identified the object based dependency relations from parsed sentences, as we have observed that the words occupied in such relations are represented as *aspect terms* in many cases. "*dobj*", "*obj*" and "*xobj*" are considered as the probable candidate relations for identifying the *aspect terms*. Here, the *Stanford Parser*[9] has been used to get the dependency relations.

Ontology Information (Liu, 2012)***:*** We have counted the *aspect terms* in the training data. The *aspect terms* occurred more than five times in the corpus are considered during our experiments. At first, we have tested this ontology information on the development set and observed that the *aspect terms* with frequency five or more also give better results in the test set.

Sentiment Words: We have used the sentiment words as a feature for the sentiment identification tasks (Liu, 2012; Brody and Elhadad, 2010). Words are identified as *positive*, *negative* or *neutral* using SentiWordNet[10].

WordNet Information: The *RiTa.WordNet* package has been used to extract different properties of the words.

For *aspect category* identification, we have matched the hypernym tree of each word with the four categories (*service, price, food*, and *ambience*). If the hypernym tree does not contain any of such words, we check the next level hypernym tree of the words derived from hypernym of previous word. We have checked up to the second degree hypernym tree. We also searched hypernym tree of the synset of each word.

Number of Sentence: It has been found that many reviews contain more than one sentence. Therefore, we have included the number of sentence as a feature based on the output of *Stanford Parser*. We have split the output of *Stanford Parser* by the mark, "*(S*".

In case of our experiments, the stop words are excluded. Total of 329 stop words was prepared manually.

5 Experimentation and Result Analysis

We have used the CRF++ 0.58[11], an open source tool for implementing the machine learning framework for our experiments. CRF is well known for sequence labeling tasks (Lafferty et al., 2001). Similarly, in the present task, the *aspect terms* use the context information and are represented in sequences. Many of the *aspect terms* are multiword expressions such as "*hard disk*". We have created different templates for different subtasks to capture all the relations between different sequence related features.

---

a. Classification of Aspect Term

Features used in case of identifying aspect terms are *POS, POS Frequency, Before be verb, Inanimate word, objects of the sentence, ontology information*. We have used several rules to identify these features. Then, we have used the CRF++ to identify the *aspect terms*. Some post processing techniques are also used in order to get better accuracy. The present system identifies only single word *aspect terms*. But it is found in the training data that many *aspect terms* consist of multiple words. Therefore, if there is a stop word in between two system identified aspect words, the stop word is also considered as a part of the *aspect term*. We have joined the aspect words along with the stop words to form a single but multiword *aspect terms*.

Precisions, Recalls and F-scores are recorded for our system in Table 1. The maximum F-scores achieved in the aspect term identification task for *Laptop* and *Restaurant* are 0.7455012 and 0.84012544, respectively. Our system performs better on *Restaurant* reviews than *Laptop* reviews.

	Laptop	Restaurant
Precision	0.4938838	0.6481481
Recall	0.7442396	0.8184855
F-score	0.59375	0.72342515

Table 1: JU_CSE system result for *aspect term* identification.

b. Classification of Aspect Category

Features used in this experiment are *POS, Dependency relations for object* and a few semantic relations of *WordNet*. In this subtask, we have also used *aspect term* knowledge as a feature. We identified the POS of the words using *Stanford CoreNLP* tool and used the words which are not listed in our stop-word list. The objects are identified from the dependency relations. The hpernym trees of these words are searched up to second degree to find four *aspect categories* (*service, price, food*, and *ambience*). If we don't find these four categories in the hypernym tree, we increase the frequency of *anecdotes/ miscellaneous* category. Frequency counts of these matched words are listed as a feature. The accuracy of the system for *aspect categories* in the *Restaurant* reviews are shown in Table 2.

Maximum F-score achieved in this *aspect category* identification is 0.8857715. The main problem faced in this task was to assign the *anecdotes/ miscellaneous* category to the respective reviews. There are many cases in which the *anecdotes/miscellaneous* categories occurred with other categories. In these cases, our system fails to identify the *anecdotes/miscellaneous* category.

Restaurant		
Precision	**Recall**	**F-score**
0.7307317	0.68029064	0.7046096

Table 2: JU_CSE system result for *aspect category* identification.

We have also observed that every review has at least one category. If any word of the review does not belong to any of the four categories, we assign these reviews with *anecdotes/ miscellaneous* category at the time of post processing.

c. Classification of Sentiment of Aspect term and category

Features used in these experiments are *POS, Positive, Negative* and *Neutral* words and number of sentences. Some reviews with multiple sentences contain different sentiments associated with different *aspect terms*. This observation also leads to *conflict* sentiment. Therefore, we have also included the *aspect term* and *aspect category* information during sentiment identification. The accuracy of the system is given in the Table 3.

Accuracy →	Aspect Term Sentiment	Aspect Category Sentiment
Laptop	0.5321101	NaN
Restaurant	0.65547705	0.6409756

Table 3: JU_CSE system result for *aspect term* and category sentiment identification.

Our system performs moderate in case of sentiment identification. Mainly, the system was biased towards the *positive* tags. It is found that the number of *positive* tags in the training data was more as compared to others. We have observed that a *conflict* tag occurs when an *aspect term* was present as both *positive* and *negative*. As the present system identifies the sentiment based on word level only, it was unable to detect the conflict tags. The feature, *number of sentences* fails to identify the *conflict* tags. Therefore, we need to find more suitable features for our system to improve the accuracy.

6 Conclusion

In this paper, we have presented a CRF based system for identifying the *aspect terms*, *aspect categories* and their sentiments. We believe that this problem will become increasingly important for common people. This task will not only be useful to common shoppers, but also crucial to product manufacturers and restaurateurs.

Overall accuracies of our system were moderate. In future, we will include more suitable features to improve accuracy of our system. We also intend to explore different machine learning algorithms for these tasks in future.

Reference

Benjamin Snyder and Regina Barzilay. 2007. Multiple Aspect Ranking Using the Good Grief Algorithm. In Proceedings of the *Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT 2007)*, pp. 300-307.

Bing Liu. 2012. Sentiment Analysis and Opinion Mining. *Synthesis Lectures on Human Language Technologies* 5, no. 1 (2012): 1-167.

Braja G. Patra, Hiroya Takamura, Dipankar Das, Manabu Okumura, and Sivaji Bandyopadhyay. 2013. Construction of Emotional Lexicon Using Potts Model. In Proceedings of the *6th International Joint Conference on Natural Language Processing (IJCNLP-2013)*, Nagoya, Japan, pp. 674–679.

Carlo Strapparava, and Alessandro Valitutti. 2004. WordNet Affect: an Affective Extension of Word-Net. In *LREC*, vol. 4, pp. 1083-1086.

Dipankar Das and Sivaji Bandyopadhyay. 2010. Extracting emotion topics from blog sentences: use of voting from multi-engine supervised classifiers. In Proceedings of the *2nd international workshop on Search and mining user-generated contents*, pp. 119-126.

Ganu Gayatree, Noemie Elhadad, and Amelie Marian. 2009. Beyond the stars: Improving rating predictions using review text content. In Proceedings of the *12th International Workshop on the Web and Databases*, Providence, Rhode Island.

Hong Yu and Vasileios Hatzivassiloglou. 2003. Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In Proceedings of the *Conference on Empirical Methods in Natural Language Processing (EMNLP-2013)*, pp. 129-136.

Koby Crammer and Yoram Singer. 2001. Pranking with ranking. In *NIPS*, vol. 14, pp. 641-647.

John Lafferty, Andrew McCallum, Fernando C.N. Pereira. 2001. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In Proceedings of the *18th International Conference on Machine Learning (ICML 2001)*, pp. 282-289.

Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In Proceedings of the *10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 168-177.

Ana-Maria Popescu and Oren Etzioni. 2005. Extracting product features and opinions from reviews. In proceedings of the *Human Language Technology Conference: Conference on Empirical Methods in Natural Language Processing (HLT-EMNLP)*. Morristown, NJ, USA, pp. 339–346.

Samaneh Moghaddam and Martin Ester. 2010. Opinion digger: an unsupervised opinion miner from unstructured product reviews. In Proceedings of the *19th ACM international conference on Information and knowledge management*, pp. 1825-1828.

Samuel Brody and Noemie Elhadad. 2010. An unsupervised aspect-sentiment model for online reviews. In Proceedings of the *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL)*.

Soo-Min Kim and Eduard Hovy. 2006. Extracting opinions, opinion holders, and topics expressed in online news media text. In Proceedings of the *Workshop on Sentiment and Subjectivity in Text,* pp. 1-8.

Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. 2010. SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. In *LREC*, vol. 10, pp. 2200-2204.

Yohan Jo and Alice H. Oh. 2011. Aspect and sentiment unification model for online review analysis. In Proceedings of the *fourth ACM international conference on Web search and data mining*.

Zhongwu Zhai, Bing Liu, Hua Xu, and Peifa Jia. 2011. Clustering product features for opinion mining. In Proceedings of the *fourth ACM international conference on Web search and data mining*, pp. 347-354.

Maria Pontiki, Dimitrios Galanis, John Pavlopoulos, Haris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. 2014. SemEval-2014 Task 4: Aspect Based Sentiment Analysis. In Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), Dublin, Ireland.