Performance evaluation of ABSA system for Restaurant Domain

Shyam A. Kanade
M. Tech. Scholar
NIIST Bhopal,
RGPV, MP, India

Abhishek Singh Chauhan
Asst. Professor
NIIST Bhopal,
RGPV, MP, India

Abstract

Today we are having abundant sources and quantity of information available online. This is result of advancements in the field of computing and communication technologies. Due to the availability of restaurant choices, the restaurant selection has influenced the decision making. So, certain websites are dedicated to the task which provides restaurant summaries in the form of ratings, some textual reviews. But still it is challenge to get impressive (densely informative) summary of review without reading all the reviews. To tackle this situation here aspect based sentiment analysis (ABSA) system is introduced. The main objective of ABSA system is to identify aspects and their polarity. Here the task of aspect term extraction was accomplished by dictionary based approach. Then classifiers (SVM and KNN) were used to identify the polarity and the source group of aspect terms. The classifiers were trained from the training dataset (provided for task 4 in SemEval). In this paper the performance evaluation is described for ABSA system for the restaurant domain.

Keywords: ABSA, dictionary based approach, SemEval, SVM and KNN.

I. Introduction

The field of opinion mining is outcome of tremendous efforts in the different aspects such as Increasing Computing Power, Statistical/Learning Algorithms and Improved Data Collection/management. It can be commented as Good machine learning is just the intelligent application of statistical processes. The reasons behind why data mining is becoming one of the hot research areas can be stated as, data is being produced, data is being warehoused, and the process of computation is now affordable, availability of commercial data mining software. It is also observed that the interest in the customer relationship management has become stronger now. Apart from that the amazing applications of the data mining cannot be ignored.

For any domain the customer satisfaction so important, because it is the driving force behind the success of every business [1, 2, 23]. Happy customers spend more money, help spread positive word of mouth and return to your domain. It is always said that, eat food timely to stay healthy. So such hotels and restaurants are always first choice of the customers who are living outside their home for either education purpose or for work. So, this paper deals with the restaurant domain. There are several search engines and websites available to deal with this issue. Generally search engine examines the index of query (entered by user) and then provides listings of web pages which match with it. Along with its list search engine also provides a short summary, and some details of text document. If we consider restaurant domain then for our queries search engine provides following results:

Search engine	Query	Results
Google [9]	Good restaurants in Mumbai	60,800,000
	Good restaurants in Pune	22,400,000
Bing [10]	Good restaurants in Mumbai	94,20,000
	Good restaurants in Pune	28,70,000

This scenario shows that the search engine may be the useful tool, but have some drawbacks as well. The above table illustrates how large volume of data is made available to the customer for his query. The result generated by search engines contains list of websites (like burp, tripadvisor, or websites of restaurants like KFC, etc.) along with some advisory websites which may sometimes offer rating and reviews of the restaurants. Most of this data is not useful. From this scenario we can understand the need of generation of aspect-wise summary rather than generation of overall result. A scenario for restaurants queries mentioned above generally produces no useful information greater than 30 to 50 per cents. It would be helpful for customer if query may yield a result like, "There are 240 hits out of which 80% are Thumbs Up and 40% are Thumbs Down". The solution for such problem can be provided by the Aspect Based Opinion Mining (ABSA).

II. Contributions to Data Mining

Data mining has large family composed of various algorithms. And the desire of improving the efficiency and accuracy of existing algorithm, researchers are working hard to expand the scope. So, new approaches are continuously being introduced. The complex application increases the gap between the user and application. The following Table.1 insights the comparison of the different approaches used in the field of opinion mining.

Table 1. Contribution of researchers

Year	Authors	Methodology	Description
2009	Plaban et al.[13]	RAKEL	Classifying news sentences into multiple emotion categories
2007	Strapparava, et al. [14]	CLaC	Unsupervised knowledge-based system
		UPAR7	Rule-based system
		SWAT	Supervised system using an unigram model
		CLaC-NB	Supervised corpus-based system
		SICS	Word-space model and a set of seed words
2011	Zhu, et al. [15]	Aspect-Based Opinion Polling	Does not require labeled training data thus flexible (for domains and language)
2012	Liu, et al. [16]	Feature selection criterion (Unigram)	Achieved accurate sentiment-classification and reduced system response time
2011	Xu, et al. [17]	Use Linguistic feature and SVM classifier	Achieved less complexity
2010	Ganesan, et al. [18]	Uses only POS tags.	Word Graph Data Was Used To Find Promising Paths That act as candidate summaries for individual input document
2014	Kim et al. [19]	A supervised method that directly	The proposed algorithm is much better in dealing

		maps between implicit features and words in a sentence	with this lack of data
2014	Caroline Brun, et al. [24]	In XRCE, robust parser was used to feed the classifier	Need performance improvement for neutral and conflict polarities.
2014	Svetlana Kiritchenko, et al [25]	NRC-Can., used 5 binary classifiers (automatically compiled polarity lexica for restaurant from YELP)	Identified aspect category and polarity successfully in both domains (restaurant and laptop) and proved that most successful features are those derived from automatically generated in-domain lexical resources.

With the emergence of the Internet as a social and interactive platform, an increasing share of public discourse and opinion making is taking place on the Web. Customer reviews represent a very prominent example where people share their opinions and experiences online. For companies and consumers such genuine customer voices represent extremely valuable information and they would like to have tools that automatically analyze and summarize this textual data. Following Table 2 shows the performance comparison among different approaches applied in [1, 2, 6, 8 and 12]

Table 2 Performance comparison of different approaches

Particulars	Approach and result		Dataset
Training Set	Classifiers	Accuracy	
WordNet	Naïve Bays Classifier	35.0	Text Dataset (SemEval)
	Support Vector Machine	27.0	
	Vector Space Model	34.8	
	Naïve Bays Classification with TF-IDF	47.8	
SentiWordNet	Aspect-Based Opinion Polling [4]	75.5	Chinese restaurant reviews collected (from [3])
	Feature selection criterion (Unigram)	86.5	Chinese movie reviews from Internet Blogs

From above table it can be concluded that there is scope for improvement in accuracy for the aspect based opinion mining task. This objective can be fulfilled by making the use of either a novel approach or combination of different approaches.

Before introducing the implementation of the aspect based opinion mining system the basics of opinion mining must be made clear to the readers. Hence this section deals with the different terms along with some basic concepts in the field of opinion mining [5, 11] in detail. The term **Data mining** is defined as the process of efficient discovery of non-obvious valuable patterns from a large collection of data or alternatively it can be defined as the practice of examining large pre-existing databases in order to generate new information.

The term **Opinion mining** deals with the study of opinions of people towards entities and their aspects. Here, the term **entity** refers to different topics, products, organizations, events and individuals. And the term **aspect** deals with the different attributes related to the entity. The sentiment or the opinion can be evaluated as either on binary scale or on ternary scale or on multivariate scale. In general, opinion text is classified into two major categories: Subjective and Objective. The **subjective** part deals with the opinion itself while the **objective** part deals with facts within the text. The binary scale evaluation of opinion mining produces an output as either **positive** or **negative**. For ternary scale evaluation these terms contains additional term, **neutral**. The multivariate scale evaluation of opinion mining produces an output as either **anger** or **sadness** or **disgust** or **joy** or **surprise** or **fear** etc. The data mining problem solving techniques deals with decision trees, nearest neighbor classification, neural network, rule induction and k-means clustering. This technique does not incorporate data warehousing, software agents and online analytical processing.

III. ABSA System for Restaurant Review

In this section we have discussed about different aspects of implementation of ABSA system. It includes dataset details, system overview and implementation.

a. Dataset Details

In this paper we have used "Trial dataset (For Restaurant Domain)", which is provided by SemEval [7, 21] (Link for dataset: http://alt.qcri.org/semeval2014/task4/). This database for restaurant is having 3044 English sentences appeared in the review. This data base includes annotations for aspect terms, aspect category-specific polarities, and aspect term polarities occurring in the sentences. From various databases it can be found that the restaurant review data set has clearly distinct statistical characteristics from other dataset. SemEval also provide dataset for laptops. From this dataset and various datasets we can comment as follows:

1. The implicit features have balanced frequency distribution in the restaurant reviews set,
2. All sentences have aspect category as an implicit function
3. This dataset contains restaurant reviews in sentences. These sentences are annotated using XML tags. In general this sentence is described in standard format which contains sentence ID No., sentence, Aspect terms, and Aspect categories as shown in the following example.

```
</sentence>
<sentence id="870">
<text>
In addition, the food is very good and the prices are reasonable.
</text>
<aspectTerms>
<aspectTerm term="food" polarity="positive" from="17" to="21"/>
<aspectTerm term="prices" polarity="positive" from="43" to="49"/>
</aspectTerms>
<aspectCategories>
<aspectCategory category="food" polarity="positive"/>
<aspectCategory category="price" polarity="positive"/>
```

SemEval dataset can be summarized as follows

- Sentences: 3,041
- Term tokens: 3,693
- Term types: 1,212
- Aspect terms: 5
- Aspect polarities: 3

b. Overview of the system

When the text input is applied the “Aspect and Polarity Identification” is done preliminarily. This stage plays an important role, for selection of desired restaurant. Then, “Get Recommendation” stage allows the user to select either particular Local Location i.e. “City” or to select the restaurant “Globally”. This both stages take information about Aspect, Polarity or emotion, and Restaurant from the database. Diagrammatic representation is shown in the Figure 1 shown below. The output is generated in the form of "Textual Recommendation, Ranking and Performance Graph".

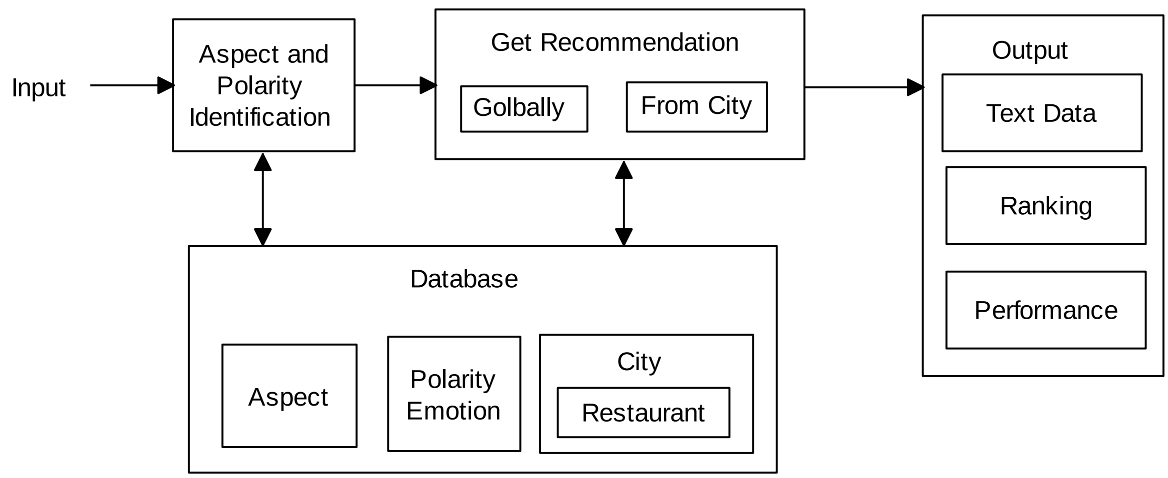

Figure 1 Overview of Aspect based opinion mining for restaurant review

c. Implementation

In this paper a sentiment analysis system is proposed to detect aspect terms, aspect categories and sentiment expressed towards aspect terms and categories in customer reviews. **Support Vector Machines (SVMs)** are a popular machine learning method for classification, regression, and other learning tasks. We trained a Support Vector Machine (for classification) on the training data provided. We have found that SVM is an effective text classification or categorization tool. It is robust method for large and small feature spaces. It is one of the state-of-the-art learning algorithms. For the task of aspect term identification and categorization, SVM classifiers were trained. Two state-of-the-art SVM classifiers, one to detect the sentiment of review and one to detect the sentiment term within a review. Here SVM was trained using library package (**LIWC**) for **aspect polarity identification**. In cross-validation over the training data, the kernel was implemented in library as well as a classifier. Here the cross-validation avoids the overlapping of the test sets. It is achieved by first dividing data into k-subsets. Then each subset in turn is used for testing and remainder for training. For aspect term identification SVM classifier was trained with the Semeval restaurant dataset. This dataset is having annotated terms in the form of review.

Basically aspect based sentiment analysis for restaurant reviews deals with identification of aspect and its polarity.

i. **Classification mechanism:** Here, the two different mechanisms are proposed for the classification purpose. The concept of classification is shown in figure below. The first one is Support Vector Machine approach and kNN classifier approach. The different aspects of this have been already discussed in the literature review chapter. The training of both SVM and kNN is done by the aspect related terms which are specified in the database.

ii. **Identification of the Aspect term:** This subtask can be visualized as SemEval Subtask of Aspect Term Extraction. The main objective of this task is to detect aspect terms in a sentence. For the aspect term identification along with occurrence of the same term in the sentence, a dictionary based approach is proposed. For this purpose a dictionary need to be prepared from the training data set and accordingly aspect terms present in the review sentences.

iii. **Identification of the Aspect Polarity:** This subtask can be visualized as SemEval Subtask of Aspect Term Polarity. The main objective of this task is to: to detect sentiment towards a given aspect term Here SVM classifier was used for identification of polarity of sentence regarding the aspect terms. Here SVM was trained using library package (LIWC [20]) for polarity identification as either positive or negative or neutral otherwise. The dictionaries generated for polarities contains word as shown in the following

Positive : Good, nice, excellent, tasty, amazing, awesome, fantastic, etc.

Negative : Dislike, disgusting, rude, unpleasant, etc

Neutral : Otherwise

As a preliminary data processing step, Term-Document Matrix was generated for training. This training matrix was generated for each sentence in the restaurant review. TDM shows the occurrences of the positive or negative terms, this is because we are interested in Polarity detection only. The condition of absence of keyword is denotes the third state as ‘Neutral’.

iv. **Identification of aspect source:** The main objective of this subtask is to: detect aspect categories discussed in a sentence. It can be visualized as SemEval subtask as Aspect Category Detection. The objective of the task is achieved by making the use of SVM classifier. Initially five dictionaries were generated (for the database provided by SemEval 2014 [7, 21]), where these dictionaries corresponds to aspect terms and aspect categories. Here also, the Term-Document Matrix was generated for training. But difference being that TDM now shows the occurrences or non occurrences of four aspect terms.

IV. Experimental Results and Performance Evaluation

The summery evaluation is one of the difficult tasks in the Natural Language Processing. One of the reasons behind this can be understood by the following example:

Table 3. Summery generated from Review of restaurant

Input	**Output**	
Query	**Polarity**	**Aspect**
I recommend the garlic shrimp, okra (bindi), and anything with lamb.	Positive	Food
The menu was impressive with selections ranging from a burger, to steak, to escargot.	Positive	Food
Reasonable prices.	Positive	Price
In addition, the food is very good and the prices are reasonable.	Positive	Price

The above Table 3 shows that even though the same output (in the format of "Polarity" followed by "Aspect") the input may be different. In other words it can be said that, generating an IDEAL SUMMERY for the document (sentence, or phrase) is almost impossible task. Apart from this issue, the different matrices in the summery with respect to different services or applications impose limitations to compare the two of them. Hence in this the method recommended and followed by many researchers is used for evaluation. The statistical approach is used to **evaluate the performance** of proposed method, which utilizes Accuracy, Precision, Recall, F1 (or simply F) measure [22].

We have compared the performance of the system for aspect and its polarity identification with two top teams along with the baseline provided in the SemEval 2014 (in Table 4). The sentiment analysis always makes frequent use of the notions of precision, recall and accuracy. Precision is measure of proportion of selected items that the system got right or it is the fraction of extracted aspect that is relevant to the problem. Recall is the proportion of target items that the system selected. F measure is the parameter that gives overall performance of the system. The most important parameter is accuracy which defines how well a binary classifier identifies the given condition. The performance of proposed algorithm is compared with different methods as shown in the Table 4.

Table 4 Performance Evaluation

Task/Method	**Aspect Term Extraction**	**Aspect Term Polarity**	**Aspect Category Detection**		**Aspect Category Polarity Identification**
Parameter	F1	Accuracy	F1	Accuracy	Accuracy
Baseline	47.15	64.28	63.89	------	65.65
XRCE	83.98	77.68	82.28	------	78.14
NRC-Can.	80.18	80.15	88.57	------	82.92
KNN	------	70.0114	------	61.8578	------
Proposed method	------	85.4680	------	92.0197	------

This table shows clearly improved performance in the terms of accuracy. We also improved the algorithm in terms of time. We further reduced execution time for the algorithm. The results also showed that SVM classifier improves the performance significantly. The use of document frequency terms plays crucial role in improvement in the accuracy.

V. Future scope

Dealing with the implicit sentences is also challenge for the researchers. The implicit sentences generally have the absence of aspect term (directly). In general there may be several sentences in the review having implicit functions. The example review is shown below

The excellent sushi was quite expensive!!!!!

This review indicates the concept of implicit sentence. Here, 'Food' is the aspect, having 'Positive' polarity, which is directly mentioned but there is absence of term related to 'Prices' which is indicating the 'Negative' aspect in the review.

Frankly speaking, the foods overall are good. I waited for a pair of chopsticks about 10 mins. After that, a waitress named Victoria gives me the chopsticks with a sentence, "Do you even wanna a fork and a knife?"

The review mentioned above does not have any negative emotions. But overall we can say that it has two aspects, the first one is 'Food' and second one is 'Service'. And these aspects are having the polarities of emotion as 'Positive' and 'Negative' respectively. This is an ideal output generated by human observation. But the algorithms in NLP need to be modified to tackle with such situation.

VI. Conclusion

This paper described different issues related with the ABSA system to generate structured summary from review or from query. Here restaurant domain was considered. In this paper preliminarily extracted the features from the reviews using inverse document frequency. Here the matrix was used to solve this issue. Then this preprocessed data was classified by the classifier. In this paper KNN, SVM classifiers were used. The performance was compared with baseline and two other methods. It can be concluded that to train or make the machine to have decision ability irrespective of domains is one of the major research challenge.

References

1. D. Talele, et al., "Extracting And Analyzing Sentiments Of The Crowd Using Naïve Bayes Classification", Asian Journal of Computer Science And Information Technology, ISSN 2249-5126, AJCSIT, 2013
2. G. Ganu, N. Elhadad, and A. Marian, "Beyond the stars: Improving rating predictions using review text content", In Proceedings of the 12th International Workshop on the Web and Databases, 2009.
3. www.DianPing.com.
4. N. Kobayashi, K. Inui, and Y. Matsumoto, "Extracting aspect-evaluation and aspect-of relations in opinion mining", Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, Prague, Czech Republic, pp. 1065–1074, 2007.

5. Bing Liu, "Handbook of Natural Language Processing", Editors: N. Indurkhya and F. J. Damerau, Second Edition, CRC Press, Taylor and Francis Group, Boca Raton, FL,2010, ISBN 978-1420085921
6. Miller, G.A. et al. "A lexical database for the English language", available at http://www.cogsci.princeton.edu/wn.
7. Maria Pontiki, Suresh Manandhar et.at. "SemEval-2014 task 4: Aspect based sentiment analysis", In Proceedings of the 8th International Workshop on Semantic Evaluation, SemEval '14, Dublin, Ireland, 2014
8. D. Ravishankar, K. Thirunarayan, and T. Immaneni, "A Modular Approach to Document Indexing and Semantic Search", In: Proceedings of the IASTED International Conference on Web Technologies, Applications and Services, pp. 165 -170, July 2005.
9. www.google.com
10. www.bing.com
11. Landauer, T., et al., "Handbook of Latent Semantic Analysis", Lawrence Erlbaum Associates, 2007.
12. Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, "Introduction to Information Retrieval", Cambridge University Press., 2008
13. P. M. Plaban Kumar Bhowmick, Anupam Basu, "Reader perspective emotion analysis in text through ensemble based multi-label classification framework", Computer and Information Science, 2:11, 2009
14. Carlo Strapparava, Rada Mihalcea, "SemEval-2007 Task 14: Affective Text", Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval-2007), pages 70–74, Prague, June 2007.
15. Jingbo Zhu, Huizhen Wang, Muhua Zhu, Benjamin K. Tsou, Matthew Ma., "Aspect-based Opinion Polling from Customer Reviews", IEEE Transactions on Affective Computing (TAC), 2(1), 37-49, 2011.
16. Chien-Liang Liu, Wen-Hoar Hsaio, Chia-Hoang Lee, Gen-Chi Lu, and Emery Jou, "Movie Rating and Review Summarization in Mobile Environment", IEEE Vol. 42, No. 3, May 2012.

17. Kaiquan Xu , Stephen Shaoyi Liao , Jiexun Li, Yuxia Song, "Mining comparative opinions from customer reviews for Competitive Intelligence", Conference on Decision Support Systems 50 (2011) 743–754.
18. Ganesan, K. A., C. X. Zhai, and J. Han, "Opinosis: A Graph Based Approach to Abstractive Summarization of Highly Redundant Opinions", Proceedings of the 23rd International Conference on Computational Linguistics (COLING), Beijing, China, pp. 340–348, 2010.
19. Kim Schouten and Flavius Frasincar, "Finding Implicit Features in Consumer Reviews for Sentiment Analysis", In Proceedings of the 14th International Conference on Web Engineering (ICWE 2014), pages 130–144. Springer, 2014.
20. http://www.liwc.net
21. http://alt.qcri.org/semeval2014/task4
22. Powers, David M W, "Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation" Journal of Machine Learning Technologies, Vol. 2 , Issue 1, page no 37–63, 2011.
23. Saif M. Mohammad, Xiaodan Zhu, "Sentiment Analysis of Social Media Texts", Tutorial at the 2014 Conference on Empirical Methods on Natural Language Processing, Doha, Qatar, October 2014
24. Caroline Brun, Diana Nicoleta Popa, et al., "XRCE: Hybrid Classification for Aspect-based Sentiment Analysis", In Proceedings of the International Workshop on Semantic Evaluation (SemEval). pp. 838, 2014.
25. Svetlana Kiritchenko, et al., "NRC-Canada-2014: Detecting Aspects and Sentiment in Customer Reviews", In Proceedings of the International Workshop on Semantic Evaluation (SemEval). pp. 437, 2014.