

# Aspect Based Classification Model for Social Reviews

Jibran Mir

Computer Science Dept  
Shaheed Zulfiqar Ali Bhutto  
Institute of Science and Technology  
Islamabad, Pakistan  
jibmir@gmail.com

Azhar Mahmood

Computer Science Dept  
Shaheed Zulfiqar Ali Bhutto  
Institute of Science and Technology  
Islamabad, Pakistan  
azharmahmood8@hotmail.com

Shaheen Khatoon

College of Computer Science and  
Information Technology  
King Faisal University  
Al Ahsa, Saudi Arabia  
sha.xin@live.com

**Abstract**—Aspect based opinion mining investigates deeply, the emotions related to one's aspects. Aspects and opinion word identification is the core task of aspect based opinion mining. In previous studies aspect based opinion mining have been applied on service or product domain. Moreover, product reviews are short and simple whereas, social reviews are long and complex. However, this study introduces an efficient model for social reviews which classifies aspects and opinion words related to social domain. The main contributions of this paper are auto tagging and data training phase, feature set definition and dictionary usage. Proposed model results are compared with CR model and Naïve Bayes classifier on same dataset having accuracy 98.17% and precision 96.01%, while recall and F1 are 96.00% and 96.01% respectively. The experimental results show that the proposed model performs better than the CR model and Naïve Bayes classifier.

**Keywords**—POS; Chunking; Word Case; Feature Set; Dictionary; NER; IOB tagging

## I. INTRODUCTION

Opinions are the key elements that influence a person's decision making capability. It is natural for people to consult with friends or family whenever they make a decision, and so, to have access to diverse opinions. Similarly, to gain better business insight, organizations conduct surveys, opinion polls and focus groups discussions to better understand their customer opinions about a product or service. In the same way, an individual customer, before buying a product tries to find his family's or friend's opinion about it [1]. Hence, for marketing, public relations, and political campaign, gathering public and customer opinions has become a profitable business. Recently it has been made possible for everyone to have internet access and therefore express his/her opinion about something in a matter of one click on social networks or other sites [2]. As a result, one can find huge amounts of user experience about usage of a product and/or service online. Now, instead of consulting friends or family, people prefer to read reviews online in order to make a purchase decision about a product or service. Consequently, companies can utilize these reviews to understand their customer response about the product usage instead of conducting traditional surveys to collect customer

feedback. However, due to the huge amount of user generated contents it is almost impossible for companies to go through each review and respond accordingly. Therefore, there is a need to extract useful information from this huge amount of data. Given the size and growth rate of this kind of data, this could only be possible by using an automated system. The opinions expressed in social networking sites or product review sites have already assisted in reshaping the political and businesses structure. It was social networking sites like Facebook that had changed the political thinking of Arab people in 2011 [1].

Consequently, sentiment analysis is a new trend in information science. The sentiment analysis is a process of classifying and identifying the opinions related to a topic. In addition to this, sentiment analysis is conducted at three main levels such as: document level, sentence level and aspect based level [3]. The document level determines the overall sentiment of product as a whole, whereas, sentence level sentiment analysis evaluates whether each sentence expressed positive, negative or neutral opinion. In contrast to this, aspect level sentiment analysis investigates opinions on bases of some specific aspect. For instance: "the price of the canon is very reasonable." Here "canon" is a product, "price" is an aspect and "very reasonable" is a sentiment word. Therefore, aspect based opinion mining investigates opinions about specific features of a product or service. However, identifying features from a sentence is another challenge. Therefore, this research work will focus on aspect based opinion mining. There are different types of aspect based opinion mining such as regular and comparative, explicit and implicit, and multi-aspect sentiment analysis. Each type comes with huge challenges due to natural language processing issues. Analysis of human text is so complex, that none of the aspect based opinion mining techniques can accommodate all the types mentioned above. That is the reason why aspect based opinion mining researchers worked in different directions. Some work on domain adaptability and some work only on sentiment analysis, however, most of them work on aspect identifications. There are some challenges in aspect based opinion mining:

The first challenge is the identification of implicit feature. Implicit features are those that are not used in a sentence

directly instead they are implied in a sentence. For instance: "this mobile is too expensive" so here author refers expensive to the aspect "price". The second challenge is multi-aspect sentence detection, where a sentence holds one explicit and one implicit aspects or both explicit or both implicit. For instance: "this phone's battery life is great but it is too expensive", this sentence is a multi-aspect sentence which contains one explicit and one implicit aspect. The third challenge is the detection of a comparative sentence, the example of this sentence is, "the Canon camera is better than the Nikon camera". Another challenge is domain and language adaptability. If an aspect based opinion mining is designed for Chinese language it should be equally applicable to other languages. Similarly, if an aspect based opinion mining algorithm is designed for product domain it should be easily applicable to other domains such as hotel accommodation. Since, aspect based opinion mining has been applied in business domain, therefore, this study initiated an aspect based opinion mining model that will classify aspects and features of social reviews. In addition to this, the proposed model is able to classify explicit aspects efficiently and can identify and classify aspects from complex social reviews. The rest of the paper is organized as follows: Section II presents the literature review of different aspect based opinion mining algorithms. Section III presents the proposed model design. Section IV shows the experimental results and Section V offers the conclusion.

## II. LITERATURE REVIEW

Authors in [4] proposed a feature based opinion mining using ontologies. Moreover, this research, introduced a vector based sentiment analysis method. Domain of the dataset is very important, if we know the domain of dataset in advance that will be helpful to find the domain aspects and sentiment words. However, there is no vibrant way to find dataset domain before sentiment analysis. The main contributions of this research are divided in to four parts: NLP, ontology based aspect detection, polarity assignment and opinion mining. This research successfully addresses the semantic relation in aspect identification process. Furthermore, they developed a mathematical solution for sentiment analysis. There is no discussion of implicit aspect identification method neither there is any method for multi-aspect sentence and comparative sentence detection. Domain independent method has been developed for electronic products in [5]. The crucial focus of this study is to develop a system of electronic product domain independent and explicit/implicit aspect identification from online customer's reviews. There are a few limitations to this research, there is no method defined that can handle multi-aspect sentences. It is domain independent only for camera, mobile phone and DVD player. It is not language adaptable. It does not group to gather semantically related aspect terms [6] and utilizes supervised approach with a balanced dataset[7]. Aspect identification and classification is a basic step in sentiment analysis, moreover, to make a domain adaptable, aspect method classification is important. Therefore, authors in [8] introduced an approach which will classify aspects with respect to domain. The main purpose of this research is to develop a model that will establish an association between product feature and domains. This will assist to build a domain

independent solution. There is no discussion of implicit feature identification. In addition to this, there is no mechanism defined to handle multi-aspect and comparative sentences nor it is language adaptable. The strengths of this model are: aspect identification and sentiment analysis. One weakness of this approach is that the polarity assignment is done by global lexicon, however the global lexicon is unable to define polarity of polysemous words [9].

Today, in aspect based opinion mining, most of the researchers carried out their work in explicit and implicit aspect detection. Authors in [10] introduced a semi supervised model for aspect identification. They have introduced two statistical methods such as "Seeded Aspect and Sentiment Model" and "Maximum Entropy SAS Model" in order to discover explicit/implicit features. This model is not automatic as it requires users to provide some prior knowledge and user may be not aware of that domain knowledge [11]. General aspects can be found by this method, however, it is unable to find domain specific aspects [12]. According to authors in [13] in information retrieval, although facts such as finding the relevant information are based upon precision and recall play an important role, opinions also play a crucial role to know the sentiments about the searched item. Therefore, they established a search engine that not only brings out facts about the searched items but also mines opinions about them. The limitations of this study are that there is no discussion about implicit feature identification, no discussion about multi-aspect sentence detection and that it is neither domain nor language adaptable. The weakness of this study is that by using the TF-TFID algorithm sentiment analysis cannot be evaluated properly [14]. As mentioned above there are many challenges in aspect based opinion mining, however, authors in [15] initiated a method for implicit aspect identification for reviews in Chinese only. There is no discussion of multi-aspect sentence detection. It is not domain or language adaptable. This method is not tested on bigger corpora. It should be adaptable for different domains and languages [16].

Aspect based opinion mining is widely used in business intelligence, where the interest is to discover the customer's opinion about the product. Authors in [17] described a model that finds the product's weaknesses and its competitive features found online in Chinese reviews. It is domain specific and single lingo supportive. Authors manually discriminated positive and negative opinion words. The method proposed in this paper is not able to properly discover customer satisfaction [18]. So far, aspect based opinion mining has been used in product domain, however, authors in [19] transformed the aspect-based opinion mining technique presented in [20] to be applicable on tourism domain. Most of the work has been conducted on physical items, however, for intangible services there is no opinion mining system. There is a difference in the reviews of physical items and service products, for instance hotels. The reviews of physical products are usually short and easy to handle whereas, service domain reviews are verbose and difficult to handle. This technique is domain specific and single lingo supportive. Neither there is a discussion of implicit nor multi-aspect identification. There is no mechanism defined to handle comparative sentences. The model described in [19] is unable to identify confusing and ambiguous terms in tourism

reviews [21]. Moreover, hardly 35% of the explicit aspects are detected by this model [22].

Aspect identification is an important step in aspect based opinion mining because it is a fundamental step in fine grain sentiment analysis. Most of the researches have been conducted in explicit aspect identification. However, they ignored the detection of implicit aspect. Therefore, authors in [23] focused on the identification of implicit aspect in Chinese reviews by using hybrid association rule mining. There is no discussion of comparative sentences, the model is not language adaptable while some of the features are extracted manually. This model used the most occurring explicit features as implicit feature indicators, however, it ignored the features that occur less often, even if they are most important [24]. Aspect based sentiment analysis is an emerging science and different scientists have adopted different ways to identify implicit and explicit aspects. This requires extensive prior knowledge of the dataset domain, however, some studies like [24] stated that this issue can be solved without prior knowledge, whereas previous studies require extensive prior knowledge. The main contribution is that they introduced an algorithm that discriminates the sentences that have implicit feature and the ones that don't have. They defined a threshold for the value of an aspect. There is no discussion of multi-aspect sentence detection or of comparative sentences. It is neither language nor domain adaptable. The removal of miss-spelling and the reduction of similar implicit features have been done by manual clustering. A detailed comparison of model based methods and statistical based methods have been given in [25]. It must be noted that the research is biased towards model based methods. Hence, they have implemented and tested CRF model based technique. The research results are quite satisfying compared with other model based and statistical based methods. Their main contribution was the classification of feature words, opinion words and intensifiers and the feature set definition. Disadvantages of this research are that this model will not give fine results when applied to other domains, there is no detail discussion about implicit aspect identification and comparative sentences. The biggest disadvantage of this method is that the dataset training is very complicated and required a great need of care.

### III. PROPOSED MODEL

Most of the previous studies developed different aspect based opinion mining models for product and service domain [1-8]. In this study, aspect based opinion mining model for social reviews is proposed. It consists of five main phases where information flow is top down similar to water fall approach as shown in Figure 1. Social reviews are more complex and it is difficult to extract aspects from them. That is the reason model structure is more complex and divided into five phases namely Pre-processing, Auto Tagging, Training and Classifier, Testing and Dictionary Phase. The selection of classifier is very important in any machine learning problem. This research used Conditional Random Fields (CRF) [26] which requires well-trained dataset and features set as input. The model's performance is highly depended on these two inputs. The core reason for using this classifier is to solve the NER (Name Entity Recognition) problem. CRF is proven good

in detecting NER in plain text [26]. Since, the dataset is consisted of social movie reviews which contain names of movies, actors, directors and writers, therefore, CRF is the right choice for this NER problem.

#### A. Pre-processing

The dataset for movie social reviews has been crawled from social websites and each review has been saved in a separate text file. Approximately, 2000 reviews have been crawled for 2000 different movies. These movie reviews are recent and written by different writers. The movie reviews are complex and detailed in contrast with the product reviews that are usually short and simple. For instance "this phone is light weight and cheaper in price". It can be observed from this example that this is a simple and short sentence. The review is talking about aspects, such as "weight and price", of a particular phone. In addition to this, "light" and "cheaper" are opinion words. However, the social review example is: "The Mayan Empire grew from about the year 400 to 900. At their height they became a people very advanced in science. Mayan notation for numbers made arithmetic easier for Mayan children than our numbers make it for our children. The Dresden Codex shows that they may not have understood exactly what eclipses were, but they knew when they were coming". It can be observed that this is complex and detailed comparing to common product review. Therefore, this requires a great deal of effort to find feature words and opinion words from social reviews.

#### B. Auto Tagging and Dataset Training

Supervised machine learning methods are effective but they require a well-defined example dataset for training, moreover, preparing dataset for training is usually a manual and tedious task [26]. Therefore, an automated process has been developed to prepare the dataset. It involves five subtasks such as tokenizing, POS tagging, chunking, word case and IOB tagging (Inside Outside Beginning). The NLP [27] tasks such as tokenizing, POS tagging and chunking are all performed by using OpenNLP software. In first step each review is tokenized into a list of tokens and saved into a text file for further processing. Next each token Part of Speech (POS) tag is assigned something which is required for the next step to identify named entity. In next step POS tagged tokens are used to detect entity using chunking. In next step word case is assigned to each chunk. For instance: if the token starts with a capital letter then the word will be tagged as TC (Title Case), and if all the letters of the token word are capitals, then the word will be tagged with UC (Upper Case). The other will be tagged as LC (Lower Case). The reason for using the 'word case tag' is to identify the movies and person names, since it is observed that people used title case for person name and upper case for movie name. The last column of Figure 2 presents the IOB tagging. IOB label is given based upon POS and word case information for given token. For instance, for the token "Cornell" its corresponding POS is "NNP" and if its Word Case is "TC" therefore, IOB tag should be "B-PERS". Twenty one (21) such patterns are derived as shown in Table I. As a result of tagging each token is annotated with Token name, POS tagger, chunk, word case and IOB tagging. Moreover,

Figure 2 is an excerpt of the trained file. The annotation of dataset with these five columns is the major contribution of this work. The same process will be repeated for test file.

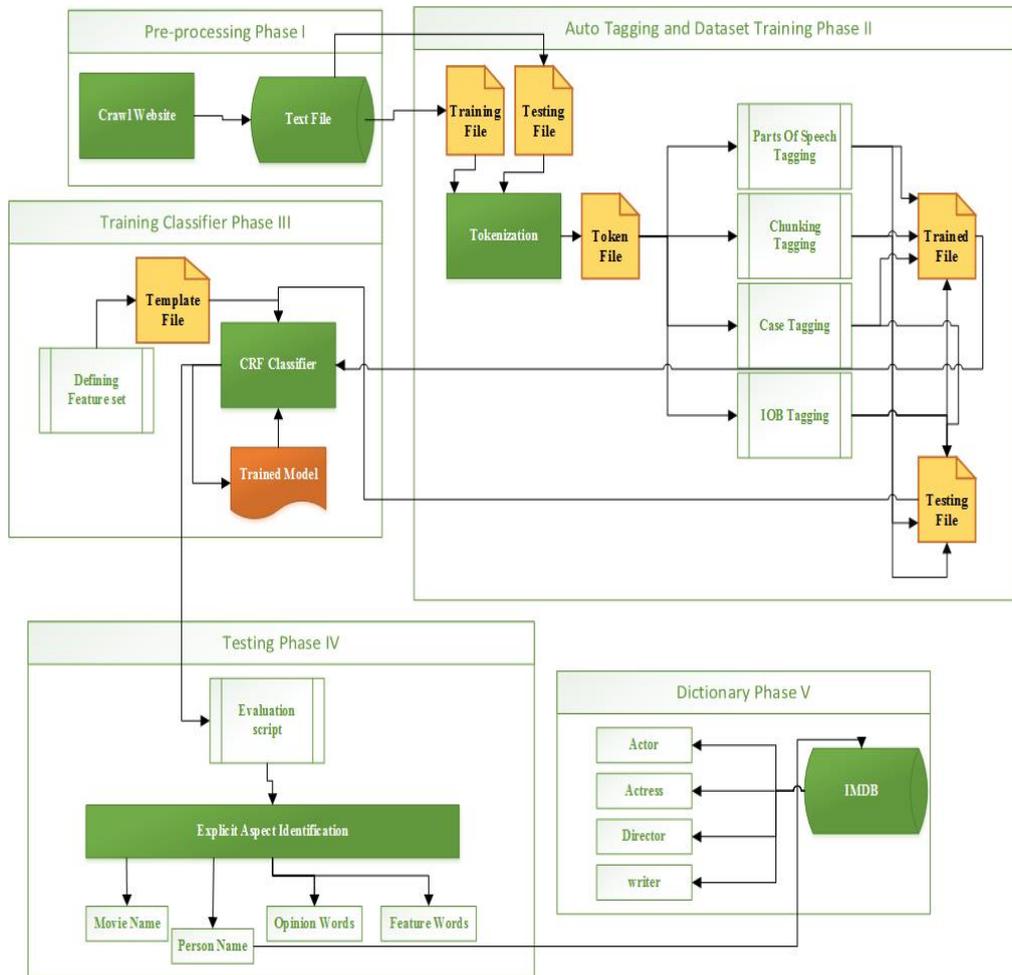


Fig. 1. The Proposed model of aspect identification and classification for social reviews

```

Cornell NNP I-NP TC B-PERS
Wilde NNP I-NP TC I-PERS
classic NN I-NP LC O
THE DT B-NP UC B-MOVIE
NAKED NNP I-NP UC I-MOVIE
PREY JJ I-NP UC I-MOVIE
. . O LC O
    
```

Fig. 2. Example of IOB tagging

```

Cornell NNP I-NP TC B-PERS B-PERS
Wilde NNP I-NP TC I-PERS I-PERS
classic NN I-NP LC O O
THE DT B-NP UC B-MOVIE B-MOVIE
NAKED NNP I-NP UC I-MOVIE I-MOVIE
    
```

Fig. 3. Example of output file

### C. Training Classifier

CRF++ CLI software [28] has been used to train and test the proposed model. This classifier takes two inputs, the template file and the trained file and outputs the model file for training purposes as shown in Figure 1. Trained file will be the

output of previous phase, however, template file will be developed according to the trained file. Template file contains a set of rules regarding the training of the classifier. In this way, classifier looks at trained file by following the feature set or set of rules defined in template file. Table II shows an excerpt of the template file. Forty feature sets have been written in template file. The Model file is a binary file and will be used for testing purposes. In order to perform testing CRF++ software takes two inputs, the model file and the test file. After performing the testing, CRF++ output would be a six column text file as shown in Figure 3. Therefore, the last column in Figure 3 is labeled by CRF classifier and the previous column is label by IOB tagging subtask in Phase II. In this way these two columns can be compared in order to calculate accuracy.

### D. Testing

The output text file produced by Training Classifier phase, will be used for evaluation purposes. The output text file will be given to Perl script and it will calculate precision, recall and f1 measures for movie name, person name, opinion words and

feature words. To know the proposed model accuracy, performance evaluation is crucial.

TABLE I. A LIST OF PATTERNS DISCOVERED FROM DATASET

Sr.No	Pattern	Entity Name
1	IF WC[0] = UC	Movie
2	IF POS[0] = JJ OR POS[0] = NNP OR POS[0] = NN AND WC[0] = TC	Person
3	IF POS[0] = POS AND POS[-1] = NNP OR POS[-2] = NNP AND WC[0] = TC	Person
4	IF POS[0] = JJ OR POS[0] = NNP OR POS[0] = NN AND WC[0]=TC AND WD[-1] = by AND CK[-2] = I-VB OR CK[-1] = B-VB	Person
5	IF POS[0] = JJ OR POS[0] = NNP OR POS[0] = NN AND WC[0]=TC AND WD[-1] = by AND POS[-2] = RB AND CK[-3] = B-VP AND CK[-3] = I-VP	Person
6	IF POS[0] = JJ OR POS[0] = NNP OR POS[0] = NN AND WC[0]=TC AND WD[-1] = by AND POS[-2] = NN	Person
7	IF POS[0] = JJ OR POS[0] = NNP OR POS[0] = NN AND WC[0]=TC AND POS[-1] = JJ OR POS[-1] = NN AND POS[-2] = DT	Person
8	IF CK[0] = B-VP AND NOT CH[-3] = B-NP AND WC[0] = TC	Person
9	IF POS[0] = NN AND POS[-1] = NNP AND WC[0] = TC	Person
10	IF POS[0] = CC AND POS[-1] = NNP AND POS[-2] = NNP AND WC[-1] = TC AND WC[-2] = TC	Person
11	IF POS[0] = CC AND POS[+1] = NNP AND POS[+2] = NNP AND WC[+1] = TC AND WC[+2] = TC	Person
12	IF WD[0] = , AND POS[+1] = NNP AND POS[+2] = NNP AND WC[+1] = TC AND WC[+2] = TC	Person
13	IF WD[0] = , AND POS[-1] = NNP AND POS[-2] = NNP AND WC[-1] = TC AND WC[-2] = TC	Person
14	IF WD[0] = , AND POS[+1] = NNP AND WC[+1] = TC	Person
15	IF WD[0] = , AND POS[-1] = NNP AND WC[-1] = TC	Person
16	IF CK[0] = B-NP AND WC[0] = LC OR WD[-1] = . AND POS[0] = DT AND POS[+1] = NN AND WC[+1] = LC	Feature Word
17	IF POS[+1] = JJ AND WC[+1] = LC	Feature Word
18	IF POS[0] = NN	Feature Word
19	IF POS[0] = DT AND CK[0] = B-NP AND WC[0] = LC AND POS[+1] = RBS AND WC[+1] = LC	Opinion Word
20	IF CK[0] = B-AD AND WC[0] = LC	Opinion Word
21	IF CK[0] = B-VP AND WC[0] = LC	Opinion Word

TABLE II. A LIST OF FEATURE SET DEFINITION

Sr.No	Feature Set	Word Token
1	U00:%x[-2,0]	The
2	U01:%x[-1,0]	movie
3	U02:%x[0,0]	seems
4	U03:%x[1,0]	amazing

### E. Dictionary Phase

At this level, model will classify all the explicit aspects but it will not be able to discriminate a person as actor, actress, director or writer. The reason for this is the availability of the less information about a person's gender and job title in the review text. For that reason a dictionary has been used to identify gender and job title.

## IV. EXPERIMENTAL RESULTS

The dataset consists of 2000 movie reviews crawled from internet movie database (imdb) official website [imdb.com/reviews/index](http://imdb.com/reviews/index). Moreover, the proposed model, CR model and Naïve Bayes have been implemented on the movie reviews dataset. This dataset has been annotated with IOB tagging scheme. There are basically 9 different tags been used. B-MOVIE I-MOVIE, B-PERS I-PERS, B-OPINION I-OPINION, B-FEATURE I-FEATURE and O. They stand for movie name, person name, opinion word and feature word. Here "B-" stands for beginning of entity name, "I-" stands for continuity of the entity name and "O" shows it doesn't belong to any entity name. An example of tagging scheme is the movie name "V FOR VENDATTA" which will be annotated like "V B-MOVIE FOR I-MOVIE VENDATTA I-MOVIE". For experimental purposes 119 movie reviews have been taken for training and 51 movie reviews have been taken for testing. Therefore, the ratio is 70% and 30% for training and testing. At this stage, 170 review comments have been used for training and the overall accuracy is 97.48% of the proposed model. Now, there is no need to train the classifier by using IOB tagging subtask from Phase II, moreover, there is no need to further define any new feature set. The CRF classifier accomplished this job on its own. In other words, the last column in Figure 2 will be labeled by CRF classifier and not by Phase II, this whole process is called Self-tagging. Manually, quality check has been done at every 100 self-tagging reviews. If the CRF classifier self-tagged 100 reviews correctly and there is no error then these 100 reviews are added to the already trained dataset. In this way, 700 reviews have been trained and the rest of the 1300 reviews was kept for testing. The overall accuracy of 2000 reviews of the proposed model was 98.17%. Tables III - V show the classified aspects for the proposed model, CR model and Naive Bayes. The proposed and CR model use the CRF classifier for aspect classification. The results show that the proposed model outperforms the CR model and Naive Bayes classifier. The reason why CR model is not performing well on this social review dataset is because it was not designed for the social domain. It doesn't provide a well-defined method for dataset training, no method is defined for the name entity recognition problem and finally there is no feature set definition in it. That is why, the feature word and

opinion word precision, recall and f1 is pretty good, while movie name and person name precision and recall is not quite satisfactory. Similarly, the results show that the Naive Bayes classifier is not efficient regarding NER problem. The incorrectly identified aspects are 6,491 only by the proposed model, whereas, the CR model incorrectly identified aspects are 37,388. Finally, the Naive Bayes incorrectly classified aspects are 71,121 as shown Figure 4. Table VI depicts the overall precision, recall, F1 and accuracy for the proposed model, CR model and Naive Bayes classifier. The proposed model's overall accuracy is 98.17% which is way better than the CR model and Naive Bayes classifier. Figure 5 shows the graphical representation of the overall accuracy.

TABLE III. 2000 REVIEWS RESULTS FOR PROPOSED MODEL

Aspects	Precision	Recall	F1
Feature Word	96.12%	95.74%	95.93%
Movie Name	96.15%	96.19%	96.17%
Opinion Word	97.34%	97.53%	97.43%
Person Name	91.08%	91.36%	91.22%

TABLE IV. 2000 REVIEWS RESULTS FOR CR MODEL

Aspects	Precision	Recall	F1
Feature Word	82.72%	89.20%	85.84%
Movie Name	54.11%	19.42%	28.58%
Opinion Word	76.46%	81.93%	79.10%
Person Name	73.83%	81.94%	77.67%

TABLE V. 2000 REVIEWS FOR NAIVE BAYES CLASSIFIER

Aspects	Precision	Recall	F1
Feature Word	43.5%	79%	55.6%
Movie Name	74.5%	45.8%	56.6%
Opinion Word	56.5%	65.2%	59.8%
Person Name	57.37%	65.47%	57.8%

TABLE VI. COMPARISON OF RESULTS FOR THE PROPOSED MODEL, CR MODEL AND NAIVE BAYES CLASSIFIER

	Precision	Recall	F1	Accuracy
PM	96.01%	96.00%	96.01%	98.17%
CR	77.95%	81.35%	79.62%	91.25%
NB	57.37%	65.47%	57.8%	68.9%

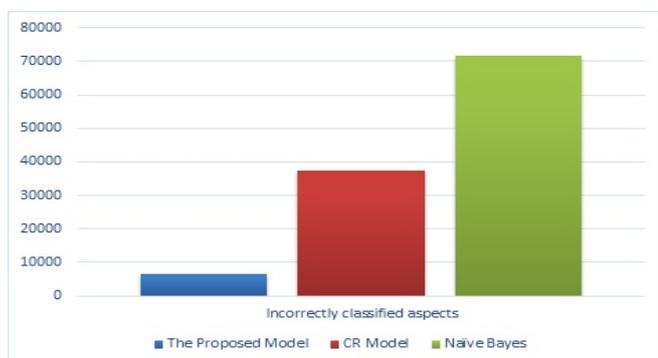


Fig. 4. Number of incorrectly identified aspects for proposed model, CR model and Naive Bayes

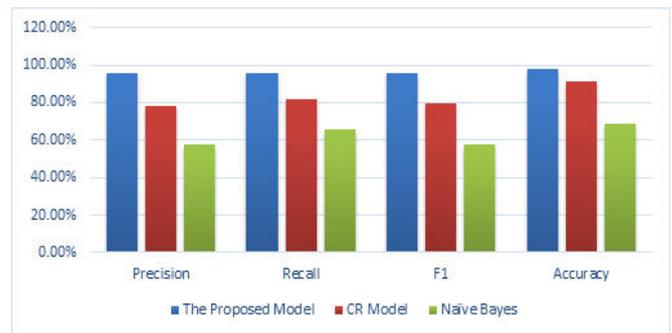


Fig. 5. A comparison of the proposed model, CRF model and Naive Bayes classifier

### V. CONCLUSION

Nowadays more people are engaged in generating online data. With the availability of plenty of data, the need of a mechanism that will extract useful information automatically emerged. This has opened the doors for aspect based opinion mining. This research has implemented an aspect based opinion mining method for identifying aspects from a social movie reviews dataset. The main contributions of this research were data training (phase II), feature set definition (phase III) and dictionary (phase V). The overall accuracy of our proposed method is 98.17% and precision, recall and f1 respectively are: 96.01%, 96.00% and 96.01%. The experimental results show that the proposed model performs much better than the CR model and the Naive Bayes Classifier. Future work involves implementation of a model that will identify implicit aspects and calculate aspect wise sentiment analysis. Moreover, we will avoid the dictionary usage and find out the patterns for deep aspect classification. There are some other issues which are more challenging and tedious, for instance, comparative sentences, specific writing style of a person, number of times an entity reemerges in a dataset etc. These challenges will be the focus of the future research.

### REFERENCES

- [1] B. Liu, Sentiment analysis and opinion mining, Synthesis lectures on human language technologies Vol. 5, Morgan & Claypool, 2012
- [2] J. Mir, M. Usman, "An effective model for aspect based opinion mining for social reviews," Tenth International Conference on Digital Information Management, pp. 49-56, 2015
- [3] T. Chinsha, S. Joseph, "A syntactic approach for aspect based opinion mining," IEEE International Conference on Semantic Computing, pp. 24-31, 2015
- [4] I. Penalver-Martinez, F. Garcia-Sanchez, R. Valencia-Garcia, M. A. Rodriguez-Garcia, V. Moreno, A. Fraga, J. L. Sanchez-Cervantes, "Feature-based opinion mining through ontologies", Expert Systems with Applications, Vol. 41, No. 13, pp. 5995-6008, 2014
- [5] A. Bagheri, M. Saraee, F. De Jong, "Care more about customers: unsupervised domain-independent aspect detection for sentiment analysis of customer reviews", Knowledge-Based Systems, Vol. 52, pp. 201-213, 2013
- [6] A. Bagheri, M. Saraee, F. De Jong, "ADM-LDA: An aspect detection model based on topic modelling using the structure of review sentences", Journal of Information Science, Vol. 40, No. 5, pp. 621-636, 2014
- [7] F. Tian, F. Wu, K.-M. Chao, Q. Zheng, N. Shah, T. Lan, J. Yue, "A topic sentence-based instance transfer method for imbalanced sentiment

- classification of Chinese product reviews”, *Electronic Commerce Research and Applications*, Vol. 16, pp. 66-76, 2015
- [8] C. Quan, F. Ren, “Unsupervised product feature extraction for feature-oriented opinion determination”, *Information Sciences*, Vol. 272, pp. 16-28, 2014
- [9] M. Zimmermann, E. Ntoutsis, M. Spiliopoulou, “Extracting opinionated (sub) features from a stream of product reviews using accumulated novelty and internal re-organization”, *Information Sciences*, Vol. 329, pp. 876-899, 2016
- [10] A. Mukherjee, B. Liu, “Aspect extraction through semi-supervised modeling”, *50th Annual Meeting of the Association for Computational Linguistics: Long Papers*, Vol. 1, pp. 339-348, 2012
- [11] Z. Chen, B. Liu, “Mining topics in documents: standing on the shoulders of big data”, *20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1116-1125, 2014
- [12] L. Zhang, B. Liu, “Aspect and entity extraction for opinion mining”, in *Data Mining and Knowledge Discovery for Big Data*, pp. 1-40, Springer, 2014
- [13] M. Eirinaki, S. Pisal, J. Singh, “Feature-based opinion mining and ranking”, *Journal of Computer and System Sciences*, Vol. 78, No. 4, pp. 1175-1184, 2012
- [14] L. Lizhen, S. Wei, W. Hanshi, L. Chuchu, L. Jingli, “A novel feature-based method for sentiment analysis of Chinese product reviews”, *Communications, China*, Vol. 11, No. 3, pp. 154-164, 2014
- [15] H. Xu, F. Zhang, W. Wang, “Implicit feature identification in Chinese reviews using explicit topic mining model”, *Knowledge-Based Systems*, Vol. 76, pp. 166-175, 2015
- [16] K. Ravi and V. Ravi, “A survey on opinion mining and sentiment analysis: tasks, approaches and applications”, *Knowledge-Based Systems*, Vol. 89, pp. 14-46, 2015
- [17] W. Zhang, H. Xu, W. Wan, “Weakness Finder: Find product weakness from Chinese reviews by using aspects based sentiment analysis”, *Expert Systems with Applications*, Vol. 39, No. 11, pp. 10283-10291, 2012
- [18] D. Kang, Y. Park, “Review-based measurement of customer satisfaction in mobile service: Sentiment analysis and VIKOR approach”, *Expert Systems with Applications*, Vol. 41, No. 4, Part 1, pp. 1041-1050, 2014
- [19] E. Marrese-Taylor, J. D. Velasquez, F. Bravo-Marquez, “A novel deterministic approach for aspect-based opinion mining in tourism products reviews”, *Expert Systems with Applications*, Vol. 41, No. 17, pp. 7764-7775, 2014
- [20] B. Liu, *Web data mining: exploring hyperlinks, contents, and usage data*, Springer Science & Business Media, 2007
- [21] M. Afzaal, M. Usman, “A novel framework for aspect-based opinion classification for tourist places”, *Tenth International Conference on Digital Information Management*, pp. 1-9, 2015
- [22] S. Y. Ganeshbhai, B. K. Shah, “Feature based opinion mining: A survey”, *IEEE International Advance Computing Conference*, pp. 919-923, 2015
- [23] W. Wang, H. Xu, W. Wan, “Implicit feature identification via hybrid association rule mining”, *Expert Systems with Applications*, Vol. 40, No. 9, pp. 3518-3531, 2013
- [24] K. Schouten and F. Frasincar, “Finding Implicit Features in Consumer Reviews for Sentiment Analysis”, in *Web Engineering: Springer*, 2014, pp. 130-144.
- [25] L. Chen, L. Qi, F. Wang, “Comparison of feature-level learning methods for mining online consumer reviews”, *Expert Systems with Applications*, Vol. 39, No. 10, pp. 9588-9601, 2012
- [26] C. Sutton A. McCallum, *An introduction to conditional random fields*, Now Publishers, 2012
- [27] J. Baldridge, “The opennlp project”, url: <https://opennlp.apache.org>, (accessed 2 February 2012), 2005
- [28] T. Kudo, “CRF++: Yet another CRF toolkit”, Software available at <http://crfpp.sourceforge.net>, 2005