Improved Feature-based Summarizing and Mining from Hotel Customer Reviews

Dim En Nyaung, Thin Lai LaiThein
University of Computer Studies, Yangon
dimennyaung@gmail.com, tllthein@ucsy.edu.mm

Abstract

Due to the rapid increase of Internet, web opinion sources dynamically emerge which is useful for both potential customers and product manufacturers for prediction and decision purposes. These are the user generated contents written in natural languages and are unstructured-free-texts scheme. Therefore, opinion mining techniques become popular to automatically process customer reviews for extracting product features and user opinions expressed over them. Since customer reviews may contain both opinionated and factual sentences, a supervised machine learning technique applies for subjectivity classification to improve the mining performance. In this paper, we dedicate our work to the main subtask of opinion summarization. The task of product feature and opinion extraction is critical to opinion summarization, because its effectiveness significantly affects the identification of semantic relationships. The polarity and numeric score of all the features are determined by Senti-WordNet Lexicon how intense the opinion is for both positive and negative features. The problem of opinion summarization refers how to relate the opinion words with respect to a certain feature. Probabilistic based model of supervised learning will improve the result that is more flexible and effective.

Keywords: *Opinion Mining, Summarizing, SentiWordNet, Text Mining, Sentiment Analysis.*

1. Introduction

Tourism is a dynamic and growing industry, with the Internet offering a multitude of new ways of conducting tourism business and promoting tourism destinations. Most users of tourism services book their travels on the Web. However, current information technologies are hardly capable of making full use of the potential of the Web for tourism business. Traditional search engines do not provide users efficient means to access the information they require, retrieving vast numbers of web pages in response to queries expressed in keywords. Instead, users often want specific and brief answers to complex questions like "Which hotel have the best service in this Country". The purpose system address to provide the information facility from users interested in tourism services and retrieves answers by summarizing from the Web.

Therefore, opinion mining is a research subtopic of data mining aiming to automatically obtain such useful knowledge. It has been widely used in real-world applications such as e-commerce, business-intelligence, information monitoring and public polls. Opinion mining seeks to determine the sentiment, attitude or opinion of an author expressed in texts with respect to a certain topic.

On the web, there are increasing numbers of review web sites, where users post their comments on a product (e.g. hotel and restaurant) and provide their positive or negative evaluation. These websites are important resources providing advice to new users and helping them with their travel plans. Among them TripAdvisor is nowadays important tool for travelers when deciding which hotel to stay in, and which restaurant and tourist attractions to visit. The contents on such travel websites is user-generated, thus giving access to the opinions of many individuals. Automatic analysis of sentiment expressed in such customer reviews has a lot of potential for applications in the tourism domain.

In this study, the overall problem we address is the analysis of customer reviews with respect to specific features of a tourism product. Our eventual goal is to generate a feature-based summarization on a product based on this analysis. When contributing opinions to the travel websites, users typically select feature for a number of facets (cleanliness, location, etc.). Customer-based services such as hotel are an area where multiple factors may impact customer sentiment. For instance noise, nearby construction, weather, even customer expectations.

Since the reviews on webpages are written in natural language and are unstructured-free-texts scheme,the task of manually scanning through large amounts of review one by one is computational burden and is not practically implemented with respect to businesses and customer perspectives.By summarizing the comments, customers and administrators would know the features which are generally liked and disliked by the customer. So customer can get valuable facts which hotels should stay according to their desire and administrator can know directly the strength and weakness of theirs so that necessary improvement can be done in those areas.

The specific problem for opinion mining is how to associate descriptions of different product features with sentiment expressions found in a review. This paper presents a method for identification of extraction patterns that relate the types of expressions. This system evaluation demonstrates the perfect extraction in features and opinion and relating the feature with opinion phrase. Opinion summarization is the task of producing a sentiment summary, which consists of sentences from reviews that capture the author's opinion. The summarization task is interested in features or objects on which customers have opinions.

Particularly, the main contributions in this work are:

• NLP and dynamic programming techniques to identify the features and sentiment words in reviews and determine their sentiment orientations.

• To relate product features and opinion phrase which reflects the inherent quality of products in terms of user interest which describe in review by using probabilistic based model.

The remaining paper is structured as follows. Section 2 presents a brief review of the existing opinion mining systems. Section 3 presents architecture and functional detail of the proposed system. The experimental setup and evaluation results are presented in section 4. Finally, section 5 concludes the paper with possible enhancements to the proposed system.

2. Related Work

Previous work has attempted to perform opinion mining at three different levels – the document level, the sentence level and the feature level [1]. At the document level, whole documents are classified into either "positive" or "negative" according to the overall sentiment expressed in the text. To predict the polarity of the opinion expressed in documents, sentiment words such as "excellent", "poor", "enjoy", and "dislike", are used as input into statistical [11] or machine learning classification algorithms[12, 15], or manually assigned values are used to classify [10]. However, the assumption does not always hold and not all sentences in a product review express subjective opinions. Instead, many sentences present factual information. To deal with this fact, the sentence- or clause-level sentiment classification is performed, which consists of two subtasks – distinguishing subjective from objective sentences and determining the polarity (positive, negative) of each subjective sentence. The representative studies on subjectivity sentence classification include in [2] and [3].

A product review usually contains comments on different aspects or features of a product, e.g. picture quality and battery life for a camera, or opinions of different subjects on a topic, e.g. persons or organizations. The document-level and sentence-level sentiment classification can determine the overall sentiment in a document or sentence but is unable to indicate which specific features of an object are evaluated positively and which negatively. The third variety of opinion mining techniques is intended to reveal the opinions expressed towards individual features. This problem involves two subtasks – extracting different features of a product and associating each feature with its corresponding opinions. To address the first sub problem, Somprasertsri et al. [7] extracted nouns and noun phrases as candidate feature terms based on patterns of part-of-speech tags and selected feature terms using likelihood-ratio test. Hu et al. [8] used association rule mining to find infrequent features by exploiting the fact that they are only interested in features that the users have expressed opinions on.

To associate features and their corresponding opinions, Hu and Liu [8] focused more on adjacent adjectives that modify feature nouns or noun phrases, than other opinion words/phrases. Some researchers considered that a product feature and its opinion words/phrases usually co-occur within a certain distance in the text [9].

However, the simple statistics-based approaches (e.g. co-occurrence) are not sufficient in some situations, for example, if more than one feature or topic is mentioned in a sentence. T. Ahmad et al. applied complicated linguistic analysis to identify associations between entities (i.e. features, topics) and opinions at finer granularity within sentences [4]. They focused on analyzing the grammatical structure of sentences and representing it using a formal expression e.g. <feature, modifier, opinion> and derived associations from the expression.

2.1 Senti-WordNet

Senti-WordNet (SWN) is an extension of WordNet that was developed by Esuli and Sebastiani [14], which is intended to augment the information in WordNet with information about the sentiment of the words in WordNet. Our research uses the information provided by sentiment in some detail, so we will describe it here. Each synset in SWN has a positive sentiment score, a negative sentiment score and an objectivity score. When these three scores are summed they equal one, so they give an indication of the relative strength of the positivity, negativity and objectivity of each synset. Esuliand Sebastiani [14] obtained these values by using several semi-supervised ternary classifiers, all of which were capable of determining whether a word was positive, negative, or objective. If all the classifiers agreed on a classification then the maximum value was assigned for the associated score, otherwise the values for the positive, negative and objective scores were proportional to the number of classifiers that assigned the word to each class.

The drawback in using SWN is that it requires word sense disambiguation to find the correct sense of a word and its associated scores. Whilst there has been significant research into this problem, we decided that it was out of scope to use any sophisticated word sense disambiguation for this research, so we simply took the highest positive and negative values that we could find for each word. This is based on the assumption that in a subjective document it is reasonably likely that the most subjective sense of a word is being used. Preliminary testing confirmed that using the most subjectivesenses tended to outperform thesenses that are known to be most frequent.

3. Proposed Opinion Mining System

This section presents the architecture and functional detail of the proposed system to identify feature-opinion pairs for summarizing. Figure 1 presents the complete architecture of the proposed system, which consists of five different functional components - review crawler, subjective/objective analyzer, preprocessing, feature and opinion learner, and summarization. Further details about these modules are presented in the following sub-sections.

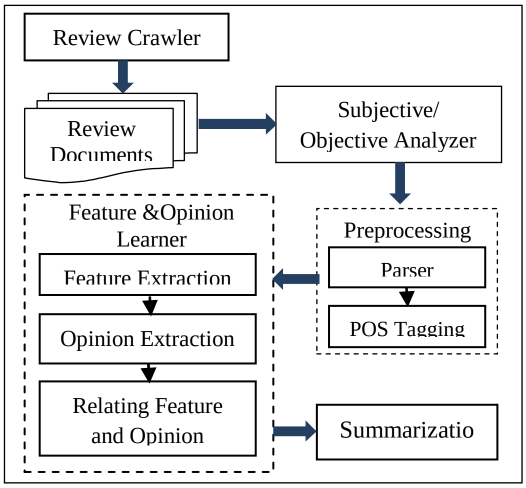

Figure 1. The Proposed System Architecture

3.1 Review Crawler

For a target review site, the review crawler crawl the related web pages and retrieves review comments only. The filtered review comments will be proceeding for more processing steps. It has been found that noisy reviews that are introduced either without any purpose or to increase/ decrease the popularity of the product may cause problem while extracting real features and opinion. So, we have to eliminate the objectivity by exploiting the fact that we are only interested in subjective sentences that have expressed opinions on.

3.2 Subjective/Objective Analyzer

Subjective sentences express the reviewer's sentiment about the product and objective sentences do not have any direct or support of that sentiment. Therefore, the idea of filtering out objective sentences can increase the system performance in terms of efficiency and accuracy. Naïve Bayes performs well for text mining [3] so we used naive Bayes as our learning algorithm. Whereas simple naive Bayes would model a document as the presence and absence of particular words, multinomial Naive Bayes explicitly models the word counts and adjusts the underlying calculations to deal with in. The proposed method for subjectivity/objectivity determination works in two phases – training and classification. For training phase, manually labeled sentences are used as trained data, which is later used to identify subjective unigrams for new dataset. In the second phase, the classification is centered on the probability of unigrams from test dataset using the training data.

3.3 Preprocessing

In the preprocessing, only subjective sentences are submitted to a pipeline for Parts-Of-Speech (POS) tags. POS tagging is used for sentence splitting and to assign lexical categories to the words in text. Maxent tagger from Stanford NLP is used for tagging the sentence. There are 36 tags in Maxent tagger. The system is used 20 tags among 36 tags of Maxent tagger to get the features which express the sentiment and also the opinion words which related to those words. As observed in [1], noun phrases generally correspond to product features, adjectives refer to opinions and adverbs are generally used as modifiers to represent the degree of expressiveness of opinions. In the system, POS-based filtering pattern will extract texts for further processing.

3.4 Feature and Opinion Learner

To learn the feature and opinion pairs, the parser from previous step are analyzed and generate all possible information components from them.

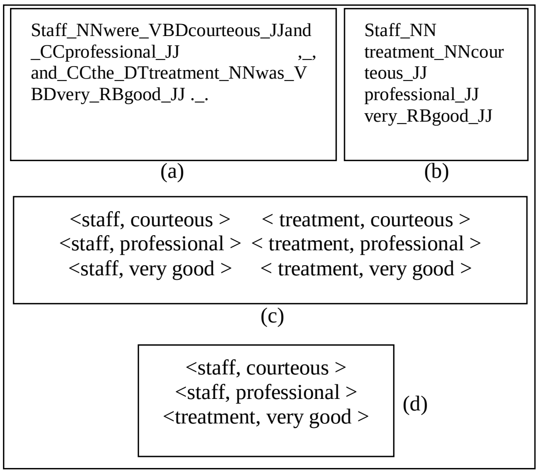

Figure 2. (a) A POS-tagged sentence, (b) extracted feature and opinion from pattern, (c) extracted information components, and (d) feature and opinion pairs

3.4.1 Extracting Features

In general, most product features indicating words are nouns or noun phrases. To summarize the reviews completely, feature extraction phase plays the critical role. Therefore, to recognize the all of the features in simple and complex sentence, defining the pattern is the effective way. As a result, the system can extract the features a though reviews are not written in grammatical structure. The linguistic filtering patterns to identify a noun phrase are the following:

- NN
- NN NN, JJ NN
- NN IN NN, NN JJ NN
- NN IN DT NN, NN NN IN NN

where NN, JJ, DT, and IN are the POS tags for noun, adjective, determiner, and preposition respectively defined by the Maxent Tagger. In this paper, we use the domain ontology to get the domain related features and to define the synonym set for features. Product feature candidates are identified by POS tags and only the features which are stored in the domain ontology are valid.

3.4.2 Extracting Opinion Phrases

Feature related opinion words are extracted in this phase. Since, to know how intense the features for the customer is also important for the decision making process, we need to extract the negation and modifier words. The linguistic filtering pattern to identify the opinion phrases is the following:

1. Single Verb, e.g., “satisfy”, “like”
2. Single Adjective, e.g., “great”, “good”
3. One or more adjective, e.g., “nice little”, “not good”
4. One or more adverb and an adjective, e.g., “very good”, “not so bad”

3.4.3 Identifying Polarity and Scores for Opinion Phrase

To identify the polarity for the extracted opinion phrase, SWN is used by applying the rule based for each opinion word. To combine the polarity of each modifier, adjective and verb expressed in the SWN, we use predefined rules that can also satisfy the some of the negation. But it has some weakness for negation sentences. However, it can find most of the polarity of the sentences. For each feature, the score of the related opinion phrase are examined and ranked based on the score value that is calculated from SWN lexicon. In the previous work, they used classification technique to define the polarity of the word. So using the rule-based method is more efficient, simplify and accurate.

3.5 Predictingthe Relation of Product Feature and Opinion Pair

Previous work in text classification has been done using maximum entropy modeling with binary-valued features or counts of feature words. In this work, we present a method to apply Maximum Entropy modeling for prediction of feature and opinion pairs in a different way it has been used so far, using weights for both to emphasize the importance of each one of them in the relation of feature and opinion pairing task. Maximum entropy model is used to predict which feature word should be related with the opinion word with maximum probability. This task can be reformulated as a classification problem, in which the task is to observe linguistic class y∈Y. We can implement classifier cl: X→Y with a conditional probability model by simply choosing the class y with the highest conditional some linguistic context x∈X and predict the correct probability p in the context x: X→Y with a conditional probability model by simply choosing the class y with the highest conditional some linguistic context x∈X and predict the correct probability p in the context x:choosing the class y with the highest conditional some linguistic context x∈X and predict the correct probability p in the context x: X →Y with a conditional probability model by simply choosing the class y with the highest conditional some linguistic context x∈X and predict the correct probability p in the context x:

$$cl(x) \rightarrow \arg\max_{y} p(y/x) \quad (1)$$

The conditional probability of p(y/x) can be defined as the following:

$$p(y/x) = \frac{1}{Z(x)} \prod_{i=1}^{k} \alpha_i^{f(x,y)} (2)$$

$$Z(x) = \sum_{y} \alpha_i^{f(x,y)} (3)$$

wherey refers to the outcome, x is the history (or context), k is the number of features and Z(x) is a normalization factor. Each parameter α_i corresponds to one feature f_i and can be interpreted as a weight for that feature.

We use the Generalized Iterative Scaling (GIS) algorithm [17] to estimate parameters or weights of the selected features. Under the maximum entropy framework, the probability for a class y and object x depends solely on the features that are active for the pair (x, y), where a feature is defined here as a function f: X x Y → {0, 1} that maps a pair (x, y) to either 0 or 1. The feature is defined as follows:

$$f_{cp,y'} = \begin{cases} 1 & \text{if } y=y' \text{ and } cp(x)=\text{true} \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

where cp(x) is contextual predication that returns true or false, corresponding to the presence or absence of useful information in some context, or history x Є X. For example, to predict which the class of product feature-opinion candidate belongs.

Let us consider example “Staff were courteous and professional, and the treatment was very good.” as shown in Figure 2(a). We can extract several product feature opinion candidates such as “staff, courteous”, “staff, professional”, “staff, very good”, “treatment, courteous”, “treatment, professional”, and “treatment, very good”. Each such pair becomes a pair candidate. For effective relation extraction, we identified the valid product features by using product ontology. The maximum entropy model is used to predict opinion-relevant product feature. Firstly, for each pair, we compute several features automatically. We denote the features employed for learning as learning features, discriminative from the product features we discussed above. We will simply choose the class with the highest conditional probability p according to Equation 1.

3.6 Summarization

The summarization task is different from traditional text summarization because we only mine the features of the product on which the customers have expressed their opinions. We do not summarize the reviews by selecting a subset or rewrite some of the original sentences from the reviews to capture the main points as in the classic text summarization. After all the previous steps, we are ready to generate the final feature-based review summary. Complete summary generation consists of the following steps:

- For each subjective sentence, feature and related opinion phrase are put into positive and negative categories according to the opinion sentences’ orientations.
- The pattern extracting the noun phrase perfect to generate the complete summarization.

4. Experimental Results

In this section, we present the experimental details of the proposed opinion mining system.To evaluate the method, standard IR performance measures.For evaluation of the experimental results, we calculate the true positive TP (number of correctly the system identifies as correct), the false positive FP (number of incorrectly the system falsely identifies as correct), true negative TN(number of incorrectly the system identifies as incorrect), and the false negatives FN (number of correctly the system fails to identify as correct)to measure the effectiveness of our approach. By using these values we calculate the following performance measures:

- Precision (π): the ratio of true positives among all retrieved instances.

$$\pi = \frac{TP}{TP + FP} \quad (5)$$

- Recall (ρ): the ratio of true positives among all positive instances.

$$\rho = \frac{TP}{TP + FN} \quad (6)$$

- F1-measure (F1): the harmonic mean of recall and precision.

$$F1 = \frac{2\rho\pi}{\rho + \pi} \quad (7)$$

There are four types of experiments: the evaluation of the subjective/objective analyzer, the evaluation of the feature extraction, the evaluation of the opinion word extraction and the evaluation of feature and opinion learner.

Evaluation of subjectivity/objectivity analyzer:

The accuracy of classification is high because multinomial Naïve Bayes performs well in text mining.Italso satisfies the zero probabilities by smoothing the Naïve Bayes estimates by adding oneLaplace. We can compare the results by simple Naïve Bayes. A binary classificationconsists of two classes subjective andobjective. After the list of English stop word have been eliminated, the class of each unigram from the input sentence is estimated.The class which has thehigher probability of the sentence is set as the class of this sentence. From classification results, true positive TP(number of correct subjective/objective unigrams thesystem identifies as correct), false positive FP (number ofincorrect subjective/objective unigrams the system falselyidentifies as correct), and false negatives FN (number ofcorrect subjective/objective unigrams the system fails toidentify as correct) are obtained. The data set are from the comment written by the user on TripAdvisor web page. We used the 3000 dataset for subjectivity analysis. The dataset consists of 1500subjective sentences and 1500 objective sentences. The results tested are shown in Table 1.

Table1. Classifier's performance using IR metrics on testing dataset

	Subjective Class	Objective Class
Precision (%)	77.98%	87.34%
Recall (%)	93.63%	62.44%
F-score (%)	85%	72.82%

Evaluation of the feature extraction step:

Since the proposed system use the domain ontology, the precision of this task can be very good because most of the extracted features are relevant. However recall is not as good as a precision because the set of ontology labels cannot totally cover the terms of hotel domain. For feature and opinion extraction step, we use 1000 review sentences as testing data from TripAdvisor webpage.

Evaluation of the opinion extraction step:

Since most of the reviewers do not follow the grammatical rules while writing reviews the proposed system can miss some opinion words. As a result the errors come from the syntactic parser and incomplete of lexicon.Implicit opinion expressions and typo can also make not to good the precision value. Therefore some of extraction rules that extract expression of recommendations do not perform very well which imply a loss of precision.

Evaluation of the relation between feature and opinion pair:

Since terminology and complex proper names are stored in ontology, an obvious problem of any automatic method for concept extraction is to provide objective performance evaluation. Therefore manual evaluation has been performedto judge the overall performance of the proposed system. From the extraction results, Table 2 summarizes the performance measure values for this step. The recall value is lower than precision indicating that certain correct feature-opinion pairs could not be recognized by the system correctly because of the fact is already mentioned in the previous evaluation step.However, the ontology is useful thanks to its list of properties between concepts which allows recognizing some opinions expressed about implicit features.Therefore, almost all identified feature-concept pairs are correct. Our results can compare with other adjacent based and pattern method which describe in [8] and [6] because they are the opinion summarization most relevant to our work and they have evaluated their performance on product review datasets. According to the results showed in Table 2, we conclude that the proposed approach is more flexible and effective than the adjacent based approach and opinion pattern based approach.

According to the feature and opinion pairs, we can get the summarization in detail and complete form. The following shows an example summary for the feature "service" of a hotel.

Feature: **Service**

Positive

- Medical care service → good
- Airport shuttle→very convenience
- Room Service → impeccable

Negative

- Currency exchange→ did not perform
- Check out service→slow

Table2. Prediction feature and opinion pairs using IR metrics on different method

	Precision (%)	Recall (%)	F-score (%)
Adjacent Based	68.65	57.93	62.69
Pattern Based	59.65	59.95	59.72
Our Approach	72.65	78.77	75.45

5. Conclusion

In this paper we have proposed summarization for each feature from user generated contents of hotel domain. We focused on extracting relations between product features and opinions. We have proposed a novel way to capture the actual relations of product features in sentences regardless the distance from them to opinions. Experimental results show the effectiveness of the proposed approaches. However, the system doesn't handle comparative sentences which require further training and classification. As part of our future work, we would like to understand the reasons behind the unsatisfactory performance on the comparative sentence.

References

[1]. B. Liu, "Sentiment Analysis and Subjectivity. Handbook of Natural Language Processing", Second Edition, (editors: N. Indurkhya and F. J. Damerau), 2010.

[2]. A. Kamal, "Subjectivity Classification using Machine Learning Techniques for Mining Feature-Opinion Pairs from Web Opinion Sources", New Delhi – 110025, India.

[3]. J. Wiebe and E. Riloff, "Creating Subjective and Objective Sentence Classifiers from Unannotated Texts".

[4]. T. Ahmad and M.N Doja, "Rule Based System for Enhancing Recall for Feature Mining from Short Sentences in Customer Review Documents", International Journal on Computer Science and Engineering (IJCSE), ISSN : 0975-3397, Vol. 4 No. 06, June 2012.

[6]. S.S. Htay and K.T. Lynn, "Extracting Product Features and Opinion Words Using Pattern Knowledge in Customer Reviews", The Scientific World Journal, Volume 2013 (2013), Article ID 394758, September 2013.

[7]. G. Somprasertsri and P. Lalitrojwong, "Mining Feature-Opinion in Online Customer Reviews for Opinion Summarization", Journal of Universal Computer Science, vol. 16, no. 6 (2010), 938-955, March 2010.

[8]. M. Hu and B. Liu, "Mining and Summarizing Customer Reviews", KDD'04, Seattle, Washington, USA, August 22–25, 2004.

[9]. F. Wogenstein, J. Drescher, D. Reinel, S. Rill and J. Scheidt, "Evaluation of an Algorithm for Aspect-Based Opinion Mining Using a Lexicon-Based Approach", WISDOM '13, Chicago, USA, August 2013.

[10] B.Liu, "Web Data Mining", Springer, 2008.

[11] X. Ding, B. Liu, P.S. Yu, "A Holistic Lexicon-Based Approach to Opinion Mining", WSDM'08, Palo Alto, California, USA, February, 2008.

[12] J.I. Sheeba and Dr.K. Vivekanandan, "Improved Sentiment Classification From Meeting Transcripts", IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 5, September 2012.

[13] S. Rill, S. Adolph, J. Drescher, N. Korfiatis, D. Reinel, J. Scheidt, O. Sch¨ utz, F. Wogenstein, R. V. Zicari, "A Phrase-Based Opinion List for the German Language".

[14] A. Esuli and F. Sebastiani.SentiWordNet: a publicly available lexical resource for opinion mining. In Proc. of LREC 2006 - 5th Conf. on Language Resources and Evaluation, Volume 6, 2006.

[15] F. Batista, R. Ribeiro, "Sentiment Analysis and Topic Classification based on Binary Maximum Entropy Classifiers", ProcesamientodelLenguaje Natural, Revista nº 50 marzo de 2013, pp 77-84.

[16] K. Fragos, Y. Maistros and C. Skourlas, "A Weighted Maximum Entropy Language Model for Text Classification".

[17] G. Liang and B. Yu, "Maximum Entropy Models: Convergence Rates and Application in Dynamic System Monitoring".