Summarization by Feature-Opinion Pairs from Web Opinion Sources

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

***Abstract* – With the rapidly growth Internet, web opinion sources is also dynamically developed. The valuable information useful for both customers and manufactures expressed over them. One of the important types of information on the Web are opinions expressed in the user generated content e.g., customer reviews of products, forum posts, and blogs. These are 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 opinions from each sentence. In this paper, we dedicate our work to the main subtask of summarization for the product. Therefore, product feature and opinion extraction is critical because its effectiveness significantly affects the identification of semantic relationships. The Probabilistic based model of supervised learning and dependency tree will improve the result that is more flexible and effective.**

***Keywords* - Opinion Mining, Summarization, Sentiment Analysis, Text Mining, dependency grammar**

I. INTRODUCTION

Nowadays, opinion mining or sentiment analysis is a research subtopic of data mining aiming to automatically obtain useful knowledge. It has been widely used in real-world applications such as e-commerce, business-intelligence, and 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 in making decision and marketing plan. 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. On the other hand, reviews on a product found on such websites can be used for the purposes of marketing research and customer relationship management by tourism businesses. 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 sentiment classification 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.

The specific problem we address is how to associate descriptions of different product features with sentiment expressions found in a review. We present a method for identification of extraction patterns that relate the types of expressions. Our evaluation demonstrates the perfect extraction in features and opinion and defining the polarity of opinion.

Particularly, the main contributions in this work are:

- NLP and dynamic programming techniques to identify the features and sentiment words in reviews.
- To define the polarity of the opinion word which reflects the inherent quality of products in terms of their features 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.

II. 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.

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.

III. PROPOSED SYSTEM

This section presents the architecture and functional detail of the proposed system to identify feature-opinion pairs with sentiment classification. Figure 2 presents the complete architecture of the proposed system, which consists of five different functional components - review crawler, preprocessing, feature extraction, opinion extraction, sentiment classification and summarization. Further details about these modules are presented in the following sub-sections.

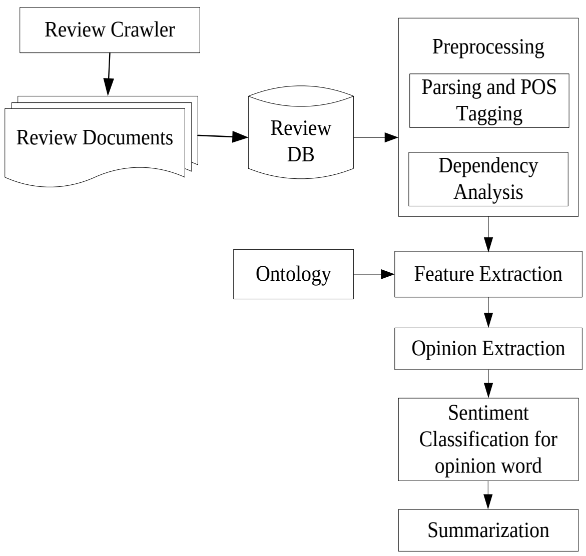

Fig 1. Proposed System Architecture

A. *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. To be crawl the review pages from TripAdvisor website; we use Crawler4j as web crawler. And the review documents are stored in the review database.

B. *Preprocessing*

In the preprocessing step, review 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 likewise adjectives and adverbs refer to opinions. In the system, POS-based filtering pattern will extract texts for further processing.

The relationship between product feature and opinion define by the dependency tree. Dependency grammars represent sentence structures as a set of dependency relationships. A dependency relationship is an asymmetric binary relationship between a word called head or governor, and another word called modifier or dependent [7]. The dependency of words will form a dependency tree. Then, attempt to capture the relating product feature and opinion using dependency relations between them. The syntactic structure of a sentence consists of dependencies shown in Figure 1. In this figure (façade, impressive), (lobby, nice) and (lobby, newly renovated) are pairs of extracted by dependency relation.

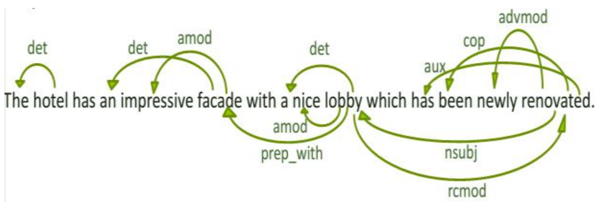

Fig 2. The syntactic structure of a sentence consists of dependencies

To measure the semantic similarity/distance between words and concepts, there are several ways that can be categorized as node based and edge based approaches, which correspond to the information content approach and the conceptual distance approach, respectively. The edge-based distance method is more intuitive, while the node-based information content approach is more theoretically sound. Both have inherent strength and weakness. In this paper, we use the edge-based method for better results.

1) *Edge-based Approach*

The edge based approach is a more natural and direct way of evaluating semantic similarity in a taxonomy. It estimates the distance (e.g. edge length) between nodes which correspond to the concepts/classes being compared. Given the multidimensional concept space, the conceptual distance can conveniently be measured by the geometric distance between the nodes representing the concepts. Obviously, the shorter the path from one node to the other, the more similar they are.

The weight between two nodes c1 and c2 is calculated as follows:

$$wt(c_1, c_2) = \frac{wt(c_1 \rightarrow_r c_2) + wt(c_2 \rightarrow_{r'} c_1)}{2d} \qquad (1)$$

given

$$wt(x \rightarrow_r y) = max_r - \frac{max_r - min_r}{n_r(x)} \qquad (2)$$

where $\rightarrow_r$ is a relation of type r, $\rightarrow_{r'}$ is its reverse, *d* is the depth of the deeper one of the two, *max* and *min* are the maximum and minimum weights possible for a specific relation type *r* respectively, and n_r (x) is the number of relations of type *r* leaving node *x*.

In determining the overall edge based similarity, most methods just simply sum up all the edge weights along the shortest path. To convert the distance measure to a similarity measure, one may simply subtract the path length from the maximum possible path length:

$$sim(w_1, w_{2)} = 2d_{max} - [min_{c_1 \in sen(w_1) c_2 \in sen(w_2)} len(c_1, c_2)] \quad (3)$$

where d_{max} is the maximum depth of the taxonomy, and the *len* function is the simple calculation of the shortest path length (*i.e.* weight = 1 for each edge).

C. 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. That will be used for further summarization.

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 almost entirely even 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.

2) Extracting Opinion word

Feature related opinion words are extracted in this phase. Let us consider example “Staff were courteous and professional, and the treatment was very good. We can extract several product feature opinion candidates such as “staff, courteous”, “staff, professional”, and “treatment, good”. Each such pair becomes a pair candidate. For effective relation extraction, we identified the valid product features by using product ontology.

D. Sentiment Classification for Opinion Word

To identify the polarity for the extracted opinion word, Naïve Bayes is a very simple probabilistic model that tends to work well on text classifications [19] and usually takes orders of magnitude less time to train when compared to models like support vector machines. We will show in this paper that a high degree of accuracy can be obtained using Naïve Bayes model, which is comparable to the current state of the art models in sentiment classification.

The Naïve Bayes model involves a simplifying conditional independence assumption. That is given a class (positive or negative), the words are conditionally independent of each other. This assumption does not affect the accuracy in text classification by much but makes really fast classification algorithms applicable for the problem.

In our case, the maximum likelihood probability of a word belonging to a particular class is given by the expression:

$$P(x_i|c) = \frac{Count\ of\ x_i\ in\ documents\ of\ class\ c}{Total\ no\ of\ words\ in\ documents\ of\ class\ c} \quad (4)$$

The frequency counts of the words are stored in hash tables during the training phase. According to the Bayes Rule, the probability of a particular document belonging to a class c_i is given by,

$$P(c_i|d) = \frac{P(d|c_i) * P(c_i)}{P(d)} \quad (5)$$

If we use the simplifying conditional independence assumption, that given a class (positive or negative), the words are conditionally independent of each other. Due to this simplifying assumption the model is termed as “naïve”.

$$P(c_i|d) = \frac{(\prod P(x_i|c_j)) * P(c_j)}{P(d)} \quad (6)$$

Here the x_i is the individual words of the document. The classifier outputs the class with the maximum posterior probability. We also remove duplicate words from the document, they don’t add any additional information; this type of naïve bayes algorithm is called Bernoulli Naïve Bayes. Including just the presence of a word instead of the count has been found to improve performance marginally, when there is a large number of training examples.

Naive Bayes classifiers due to their conditional independence assumptions are extremely fast to train and can scale over large datasets. They are also robust to noise and less prone to overfitting. Ease of implementation is also a major advantage of Naive Bayes.

D. Summarization

Summarization techniques become very useful in showing the general idea of the reviews. 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, which is straightforward and consists of the following steps:

- For each discovered feature, related opinion sentences are put into positive and negative categories according to the opinion sentences’ orientations.
- The noun phrase extracting with pattern make to perfect for generating the complete summarization.

The following shows an example summary for the feature “service” of a hotel. Note that the individual opinion sentences (and their corresponding reviews, which are not shown here) can be hidden using a hyperlink in order to enable the user to see a global view of the summary easily.

Feature: **service**

Positive: 8

- Good medical care service for the tourists.

- ATMs are available on either side of the hotel and the hotel takes all major credit cards.
- Check in was smooth and the staff was very friendly and helpful.
- Service is very good and charm.

….

Negative: 2

- The hotel does not yet accept credit cards and all payment is in USD.
- Service was impeccable.

IV. 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 feature extraction, the evaluation of the opinion word extraction and the evaluation of sentiment classification. The data set are from the comment written by the user on Myanmar hotels on TripAdvisor website.

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. 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 features are correct. For feature and opinion extraction step, we use 1000 review sentences.

TABLE I
EVALUATION OF FEATURE EXTRACTION

Precision	85%
Recall	72%
F-measure	78%

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. 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 sentiment classification for opinion word:
An obvious problem of any automatic method for concept extraction is to provide objective performance evaluation. Therefore manual evaluation has been performed to judge the overall performance of the proposed system. From the extraction results, Table II summarizes the performance measure values for this step. Our results can compare with other opinion lexicon and pattern based method which describe in [18] 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 II, we conclude that the proposed approach is more flexible and effective than the lexicon based approach and pattern based approach.

TABLE II
PREDICTION SENTIMENT CLASSIFICATION USING IR METRICS ON DIFFERENT METHOD

	Precision (%)	Recall (%)	F-score (%)
Opinion Lexicon	68.65	57.93	62.69
Pattern based	59.65	59.95	59.72
Our Approach	75.65	82.77	78.45

V. CONCLUSIONS

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.
[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", Procesamiento del Lenguaje 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".
[18] A. Hamouda and M. Rohaim, "Reviews Classification Using SentiWordNet Lexicon", The Online Journal on Computer Science and Information Technology (OJCSIT) Vol. (2) – No. (1), Reference Number: W11-0123, pages 120-123.
[19] B. Liu, Web Data Mining, chapter Opinion Mining, Springer, 2008.