International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:9, No:3, 2015

Feature-Based Summarizing and Ranking from Customer Reviews

Dim En Nyaung, Thin Lai Lai Thein

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 is the task of opinion summarization. Therefore, 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. 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, Opinion Summarization, Sentiment Analysis, Text Mining.

I. 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 has the best service in this Country”. The purpose system addresses 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. 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 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.

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 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. And the ranking is also based on the grades given by previous users; it is useful information for both user and administrator.

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 their features by using probabilistic based model.

The remaining paper is structured as follows. Section II presents a brief review of the existing opinion mining systems. Section III presents the dependency relation for sentence

Dim En Nyaung is with the University of Computer Studies, Yangon, Myanmar (corresponding author e-mail: dimennyaung@ gmail.com).

Thin Lai Lai Thein is with the Department of Information Science, University of Computer Studies, Yangon, Myanmar (e-mail: tllthein@ucsy.edu.mm).

structure. Section IV represents about the Senti-WordNet Lexicon and section V presents architecture and functional detail of the proposed system. Section VI explains how to identify the polarity and score for opinion phrase. Section VII explains the summarizing and ranking. The experimental setup and evaluation results are presented in section VIII. Finally, section IX 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 as in [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 in [11] or machine learning classification algorithms are used in [12] and [15], or manually assigned values are used to classify as in [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 are included 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, [7] and [5] show how to extract 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 focused more on adjacent adjectives that modify feature nouns or noun phrases, than other opinion words/phrases in [8]. Some researchers considered that a product feature and its opinion words/phrases usually co-occur within a certain distance in the text as described in [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. [4] applied complicated linguistic analysis to identify associations between entities (i.e. features, topics) and opinions at finer granularity within sentences. 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. Dependency Relation for feature-opinion mining

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 in [7]. The dependency of words will form a dependency tree. From this tree, we attempt to capture the relation product feature and opinion using dependency relations between them. The syntactic structure of a sentence consists of dependencies shown in Fig. 1. In this figure (façade, impressive), (lobby, nice) and (lobby, newly renovated) are pairs of extracted by dependency relation.

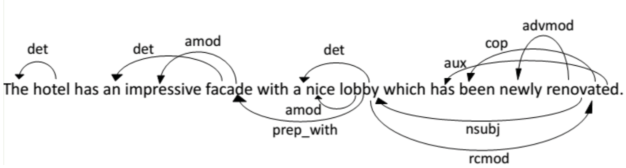

Fig. 1 The syntactic structure of a sentence consists of dependencies

IV. Senti-WordNet

SentiWordNet (SWN) is an extension of WordNet that was developed by [13], 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. Batista and Ribeiro [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 of 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 subjective senses tended to outperform the senses that are known to be most frequent.

V. Proposed System

This section presents the architecture and functional detail of the proposed system to identify feature-opinion pairs for summarizing. Fig. 2 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.

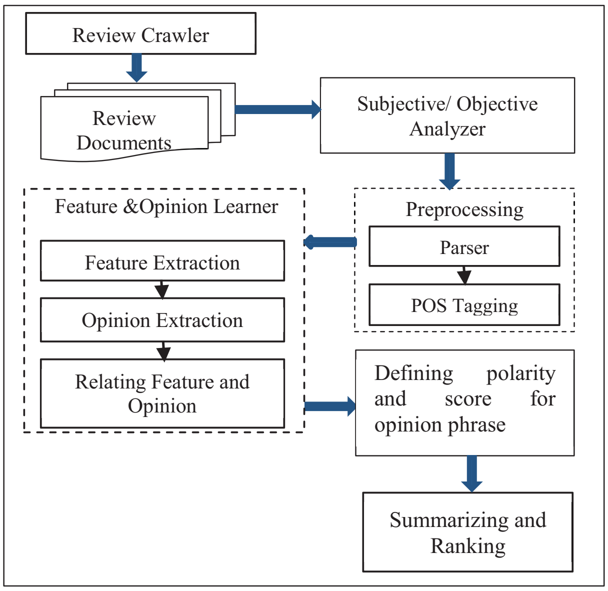

Fig. 2 Proposed System Architecture

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. 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 opinions. So, we have to classify the objectivity by exploiting the fact that only interested in subjective sentences that have expressed opinions on.

B. 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 in [3], so we used naive Bayes as our learning algorithm for classification. 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.

C. 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.

D. 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. The extracted pairs from dependency relation can contain which are relevant or not. Therefore, maximum entropy model is used to predict opinion-relevant product feature or not by choosing the class with the highest conditional probability p according to this model.

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 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. Prediction the 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:

$$cl\ x \ \rightarrow \arg\max p(y/x) \qquad (1)$$

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

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

$$\mathrm{Z(x)} = \ {}_{y} \alpha_i^{f(x,y)} \qquad (3)$$

where y 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 αi corresponds to one feature fi and can be interpreted as a weight for that feature.

We use the Generalized Iterative Scaling (GIS) algorithm 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 following:

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

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. 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 (1).

VI. Identifying Polarity and Scores for Opinion Phrase

To identify the polarity for the extracted opinion phrase, SWN is used for 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.

To determine the polarity of opinion phrase, use some rules such as:

Negation Negative → Positive // "not bad"
Negation Positive → Negative //"not good"
Negation Neutral → Negative//"not appropriable"
Neutral Positive → Positive //"more beautiful"
Neutral Negative → Negative//"almost double"
Positive Positive → Positive //"very good"

A. Handling the Context Dependent Opinion Word

To handle the context-dependent opinion word, the system uses the intra-sentence conjunction rule and inter-sentence conjunction rule.

- Intra-Sentence Conjunction Rule

The idea is tries to see whether another opinion word in the sentence (or clauses) is positive (or negative) which means that a sentence only expresses one opinion orientation unless there is a "but" word. e.g. The rooms are well decorated and

International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:9, No:3, 2015

airy. <well- decorated> and <airy> are opinion words for this sentence. Although well-decorated has the positive meaning, airy depends on the context to determine the polarity. Therefore the above rule is applied for this sentence.

- Inter-Sentence Conjunction Rule

If the above rule could not decide the opinion orientation, use the context of previous or next sentence to decide. That is the extension of intra-sentence conjunction rule to neighboring sentences.

The idea is that people usually express the same opinion (positive or negative) across sentences unless there is an indication of opinion change using words such as "but" and "however". E.g. The staff was very prompt. The food was very good. To determine the polarity of the word "prompt", we need to look the neighbor sentence.

VII. Summarizing and Ranking

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 noun phrase perfect to generate the complete summarization. And then we evaluate the relative importance of each feature according to related opinion score and use that measure to rank features. Equation (8) shows the standard deviation for positive and negative scores of product feature. Let P_i and N_i be the number of positive and negative opinions on feature f_i, i Є {1, … , n}. $PScore_i$ and $NScore_i$ will be the min-max normalized values of P_i and N_i. The standard deviation of these Score using:

$$STDScorei = \sqrt{\frac{1}{2}\left(\left(PScorei - \frac{PScorei+NScorei}{2}\right)^2 + \left(NScorei - \frac{PScorei+NScorei}{2}\right)^2 \right)} \quad (5)$$

VIII. 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.

$$\pi = \frac{TP}{TP+FP} \quad (6)$$

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

$$\rho = \frac{TP}{TP+FN} \quad (7)$$

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

$$F1 = \frac{2\rho\pi}{\rho+\pi} \quad (8)$$

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. It also satisfies the zero probabilities by smoothing the Naïve Bayes estimates by adding one Laplace. We can compare the results by simple Naïve Bayes. A binary classification consists of two classes subjective and objective. After the list of English stop words have been eliminated, the class of each unigram from the input sentence is estimated. The class which has the higher probability of the sentence is set as the class of this sentence. From classification results, true positive TP (number of correct subjective/objective unigrams the system identifies as correct), false positive FP (number of incorrect subjective/objective unigrams the system falsely identifies as correct), and false negatives FN (number of correct subjective/objective unigrams the system fails to identify as correct) are obtained. The data set are from the comment written by the user on TripAdvisor website. We used the 3000 dataset for subjectivity analysis. The dataset consists of 1500 subjective sentences and 1500 objective sentences. The results tested are shown in Table I.

TABLE I
Classifier's Performance Using IR Metrics

	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 5000 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. It is the fact that 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:

An obvious problem of any automatic methodology 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. The recall value is lower than precision which 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 described 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 II, 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

TABLE II
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	75.65	82.77	78.45

IX. 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.
[5] 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.
[6] 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.
[7] M. Hu and B. Liu, "Mining and Summarizing Customer Reviews", KDD'04, Seattle, Washington, USA, August 22–25, 2004.
[8] 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.
[9] B.Liu, "Web Data Mining", Springer, 2008.
[10] X. Ding, B. Liu, P.S. Yu, "A Holistic Lexicon-Based Approach to Opinion Mining", WSDM'08, Palo Alto, California, USA, February, 2008.
[11] 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.
[12] 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".
[13] 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.
[14] 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.
[15] K. Fragos, Y. Maistros and C. Skourlas, "A Weighted Maximum Entropy Language Model for Text Classification".

Dim En Nyaung was born in Yangon, Myanmar in 1986. She has done B.C.Sc (Computer Science) and B.C.Sc (Honors) (Computer Science) from the University of Computer Studies, Yangon (UCSY), Myanmar. She also got her M.C.Sc (Computer Science) from UCSY, Myanmar in 2010. All of the degrees are related with computer science. Now she is pursuing her Ph.D from UCSY, Myanmar in the area of Natural Language Processing. Her area of interest includes web mining, text mining, database administration, and information retrieval.

She is a Tutor at the University of Computer Studies, Yangon, Myanmar. She is now doing research concern with text mining and information retrieval.