International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)
Web Site: www.ijettcs.org Email: editor@ijettcs.org, editorijettcs@gmail.com
Volume 05, Issue 01, January - February 2016 ISSN 2278-6856

Knowledge Based Approach for Concept Level Sentiment Analysis for Online Reviews

Tharushika Silva[1], MGNAS Fernando [2]

[1]University of Colombo School of Computing,
35, Reid Avenue, Colombo 7, Sri Lanka
thrush.991@gmail.com

[2]University of Colombo School of Computing,
35, Reid Avenue, Colombo 7, Sri Lanka
nas@ucsc.cmb.ac.lk

Abstract: ***An Evolution of internet application and usage, World Wide Web (WWW) offers a great contribution to the global society. One of the major application domains is sharing of the knowledge of users' through a large repository of valuable opinions in the form of texts from different people on numerous products and services at the WWW. It is not a straight forward process to access these opinions because the available number is vast, and understanding what is actually meant has always been problematic considering the complexity of human languages. The main objective of the study is to provide accurate, clear cut information with a numerical justification quantifier for decision making activities based on a vast number of review comments when users are buying a product or obtaining service via WWW. This study simulates the cognitive ability of the human brain, up to some extent by identifying concepts and process with the aid of a large semantic knowledge-base that we create through natural language processing techniques. In this study, concept extraction from natural language text is achieved in two steps, part-of-speech based extraction and dependency based extraction. The experiment shows promising state of the art results, proving we have achieved successfully using human knowledge with the domain knowledge to analyze sentiments.***

Keywords: Online Reviews, Sentiment Analysis, www, Knowledgebase.

1. INTRODUCTION

When Online reviews are useful information for decision making activities on different aspects when buying products or obtaining a service. All over the people in world can write reviews and share opinions on the WWW. The opinion-rich reviews will not only be limited to help clients to make better judgements ,but they are also useful information for manufacturers of products and the service providers to take correct decisions at the right time in their decision making processes in their respective development activities. If there are several review comments available on the WWW for a particular product or service, it is very difficult to take a correct decision based on available comments due to the non-availability of reliable numerically weighted quality factors since most of them provide only positive and negative quantifiers.

The usage of review comments and opinions for a particular product or services has increased due to the enormous internet usage and ease of remote accessibility, as the number of product/services reviews grows. A survey performed by Faves.com on online shoppers has shown that 70% consult reviews or ratings before purchasing 62% rely on the popularity of information based on users' votes or ratings. Therefore, it is difficult to analyze the existing online information manually as well as the existing negative and positive level analysis techniques to take an accurate decision. Therefore, it is evident that it is vital and desirable to develop an efficient and effective sentiment analysis technique and that it should be capable of summarizing the sentiments of consumer reviews automatically to enhance the decision making process of organizations and to take correct decisions in consumer activities from customers' point of view.

Further, it was revealed that there is a lack of researchers to address how to analyze the user reviews rather than depend on negative and positive quantifiers.

The aim of the study is to provide a great and reliable support for decision making activities in connection with buying products, obtaining services, manufacturing products and providing services using the sentimental analysis using a variety of tools and methodologies. For example, in restaurant reviews, there are comments on aspects such as the overall restaurant, food, service, price and location. Before deciding to go to a particular restaurant, consumers often seek opinions from other users by reading their reviews and it will as well as be useful to service providers too.

The available feedback reviews are purely unstructured text with both subjective and non-subjective information. The extraction of knowledge from such a large amount of unstructured information is an expert task. Further, the rapid development of technology and the busy life style of the human beings needs accurate information for their day-day activities to survive in modern society. Therefore, to

extract knowledge for accurate judgment or decision making activities, the following techniques are greatly supportive: Mining for opinions and sentiments from natural language, involves a deep understanding of most of the explicit and implicit reviews, regular and irregular, syntactical and semantic rules proper of a language. The existing approaches mainly rely on parts of text in which opinions and sentiments such as polarity terms, affect words and their co-occurrence frequencies are explicitly expressed. However, opinions and sentiments are often conveyed implicitly through latent semantics, which make purely syntactical approaches ineffective.

In addition to the above, Concept-Level Sentiment Analysis can help semantic analysis of text through the use of web ontologies or semantic networks, which allow the aggregation of conceptual and affective information associated with natural language opinions. Also construction of comprehensive and common sense knowledge bases is the key to feature spotting and polarity detection, respectively.

This research study is mainly focused on how to identify the quality factors that have been expressed by consumers in a review and the level of satisfaction for each quality factor. Further, the ultimate goal of the study is to provide a weighted quality indicator as a numerical figure for different qualities based on un-structured comments in order to take correct decisions at the right time and at the most convenient place.

2. LITERATURE REVIEW

Set Research related to opinion mining that include subjective genre classification, sentiment classification, text summarization and terminology finding have been done in the past. Genre classification classifies text into different groups [1] and some techniques have the ability to detect opinion in a document [2]. But they have been unable to detect the semantic orientation (whether the opinion expressed in the document is positive or negative) of the opinions [1].

Hatzivassiloglou and Wiebe [3] has concluded that the presence and type of adjectives in a review sentence are the reason behind a sentence being subjective or objective. Aspect extraction from opinionated text was first studied by Hu and Liu [1] who also introduced the distinction between explicit and implicit aspects. However, there the authors only dealt with explicit aspects by adopting a set of rules based on statistical observations. For example, a sentence like "The food was good but it was very expensive" contains two aspects or opinion targets, namely general comment on the food and the price of the food. These are implicit in the word itself [4]. There are many types of implicit aspect range of enterprise systems and information management systems. The reason behind this effect is that semantic description behaviors and services based on ontology results are in better coordination of software agents in a multi agent system.

Expressions: Adjectives and adverbs are perhaps the most common types because most adjectives describe some specific attributes or properties of entities, e.g., expensive describes "price," and beautiful describes "appearance." Implicit aspects can be verbs too. In general, implicit aspect expressions can be very complex, e.g., "This camera will not easily fit in a pocket." "Fit in a pocket" indicates the size aspect. Although explicit aspect extraction has been studied extensively, limited research has been done on mapping implicit aspects to their explicit aspects [5].

As to the implicit aspects, the OPINE extraction system developed by Popescu and Etzioni [6] was the first that leveraged on the extraction of this type of aspects to improve polarity classification. However, their system is not described in detail and is not publicly available. Su [7] proposed a clustering method to map implicit concepts (which were assumed to be sentiment words) to their corresponding explicit aspects. The method exploits the mutual reinforcement relationship between an explicit aspect and a sentiment word forming a co- occurring pair in a sentence. Hai [8] proposed a two-phase co-occurrence association rule mining approach to match implicit aspects (which were also assumed to be sentiment words) with explicit aspects. Zeng and Li [9] proposed a rule-based method to extract explicit aspects and mapped implicit features by using a set of sentiment words and by clustering explicit feature-word pairs [4].

It is revealed that there are some approaches that rely on semantic lexicons, stressing the need for methods to create and maintain high quality lexical information per category. Therefore, they [10][14] have tried to analyze these sentiments based on the meaning of the sentence or the phrase.

Zhou [10] have proposed ontology supported polarity mining. With domain specific information, opinion/polarity mining can be enhanced by using ontology. Further, they stated that ontology has an intense effect on broad statements.

Binali, et al.[11] have explained the process of opinion mining by using a framework which has been logically derived by analyzing critically the existing research in opinion mining. The first step in opinion mining is item extraction. Item extraction is to know about that item for which an opinion can be extracted e.g. Camera, Mobile, MP3 player etc. It provides a general opinion about the item, whether it is good or bad. A negative opinion about any product does not mean that every aspect of the product is disliked. It is also important to provide a justification to the reason about the feature that make a thing good or bad. The second step, feature extraction which helps in providing a sound ground to the subjective opinion is important.

Yaakub et, al.[12] have also proposed an architecture that uses a multidimensional model to integrate customer's characteristics and their comments about products. Also, the ontology proposed by Yaakub et, al.[12] covers features and the characteristics of the mobile phone in general as well as in other technical terms.

The study [13] has presented a method of ontology-based sentiment clustering to cluster and analyze movie reviews. In their study, they have proposed the domain ontology to extract the related generic category, in order to find the class of the movie based on the theme (Comedy, Action, Thriller, Tragedy, Horror, Sci-Fi, Family) the domain ontology and the related adjectives are analyzed and applied to the Fuzzy C Mean (FCM) clustering process where the attributes with a fuzzy score are calculated and used as semi supervised learning for clustering techniques.

According to Anni et, 2014 approach relies on a core ontology of the task, augmented by a workbench for bootstrapping, expanding and maintaining semantic assets that are useful for a number of text analytic tasks. The workbench has the ability to start from classes and instances defined in an ontology and expand their corresponding lexical realizations according to target corpora. In their paper they have presented results from applying the resulting semantic asset to enhance information extraction techniques for concept-level sentiment analysis.

3. THEORETICAL BACKGROUND

This study uses three main theoretical bases in the sentimental concept analysis that can be used to identify quality in the stage of factor extraction.

3.1 Tokenizing the words

In lexical analysis, tokenization is the process of breaking up a stream of text into words, phrases, symbols, or other meaningful elements called tokens [15]. The list of tokens becomes input for further processing such as parsing or text mining. Tokenization is useful both in linguistics (where it is a form of text segmentation), and in computer science, where it forms part of lexical analysis [15].

3.2 POS Tagging

In corpus linguistics, part-of-speech tagging (POS tagging or POST), also called grammatical tagging or word-category disambiguation, is the process of marking up a word in a text (corpus) as corresponding to a particular part of speech, based on both its definition, as well as its context(i.e. relationship with adjacent and related words in a phrase, sentence, or paragraph). Among the linguistic tools available Natural Language Toolkit, provide great support for tokenizing as well as POS tagging [15].

3.3 Dependency Parsing

Typed dependencies and phrase structures are different ways of representing the structure of sentences [16], while a phrase structure parse represents nesting of multi-word constituents, a dependency parse represents dependencies between individual word tokens. A typed dependency parse additionally labels dependencies with grammatical relations, such as subject or dependent object. These dependencies are triplets consisting of a relation-type and two arguments known as governor and dependent. For this research Stanford dependency parser will be used [16].

4. DESIGN

The proposed design approach consist of three main components such as creating a knowledge base, generating rules and identifying feature and their priorities.

Quality factor extraction is performed according to the guidance given in the Figure 6.1. The proposed system takes a sentence of a review comments and extracts emotional words and their respective nouns using the

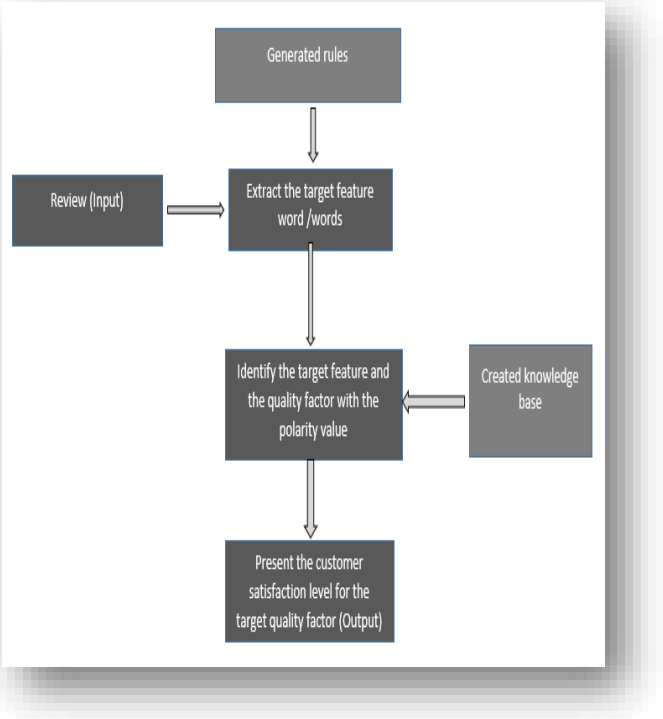


Figure 1 Overview of the design approach

generated rules as described in Figure 6.1 and 6.2. Further, knowledgebase will be used to identify the quality factors and the level of satisfaction. The major functionalities of the three component are described in detail as follows.

4.1 Creating a Knowledge Base

The key activity of the knowledgebase is to identify the features that have been expressed by the commenters in review comments. In the knowledgebase, each adjective is under a particular feature. For example, the adjective "good" belongs to the feature category of GENERAL, if the adjective is "tasty" it belongs to the category of “QUALITY”. The higher level view of the knowledgebase is shown in figure 6.2. The target features are identified using the actual reviews which are extracted from social media sites. There can be different quality factors in restaurant reviews, such as FOOD, DRINKS, SERVICE, AMBIENCE, and LOCATION. For example a comment

International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)
Web Site: www.ijettcs.org Email: editor@ijettcs.org, editorijettcs@gmail.com
Volume 05, Issue 01, January - February 2016 ISSN 2278-6856

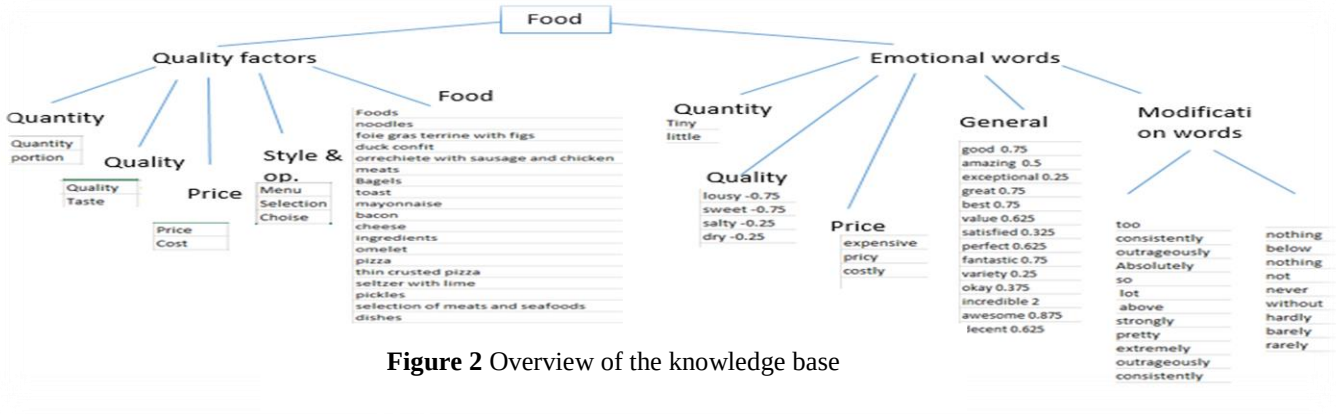


Figure 2 Overview of the knowledge base

as "great food at a reasonable price" the features that would be extracted are "GENERAL" and "PRICE". Therefore, each adjective word is categorized into a feature. This is the component in the proposed solution that includes domain knowledge in order to analyze sentiments. This study reveals that food concepts provide the following features as described in Figure 6.2: GENERAL, QUALITY, PRICE, QUANTITY, STYLE and OPTIONS. As a way of enhancing the knowledge in the knowledge base, the synonyms and antonyms are also included in the knowledge base. In addition to the above, the modification words for the adjective are also included in the knowledge base. For example such words are: "very", "not" and etc. when analyzing and calculating the polarity value of each feature, the effect of modification words will be considered.

The following shows the structure of the table which is used in the database of the study.

Quality_Factor_Type (QFT): The following example shows the main quality factors and their attributes as Quality, Quantity and Price of the food item.

QFT_ID (primary key)	Quality_Factor

Quality_Factor_Nouns: this includes the nouns that will be used to identify the quality factors if the identified category of the emotional words is 'GENERAL'.

Quality_ Words	QFT_ID foreign key refereeing QFT_ID in Quality_Factor_Type)

Emotional Words: this includes emotional words with their polarity values that were taken from SentiWordNet and their respective quality factors

Words	Polarity_Values	QFT_ID(foreign key refering QFT_ID in Quality_Factor_Type)

Modification Words: this includes the modification words and also whether they are enhancing or sentiment shifter.

Modification_Words	Mod_Word_Type

With the designed knowledgebase, inputs should be preprocessed. Each review comment will go through the steps of tokenizing, POS tagging and dependency parsing. For tokenizing, POS tagging Natural and for taking dependency parsing of the review Language Tool Kit (NLTK) and Stanford NLP tool have been used respectively. Subsequently, extracting the quality factor were processed after this stage. For that rule, generation has been carried out.

4.2 Generate rules

There are very simple sentences to complex sentences that exists in review comments. As such appropriate rules need to be generated to extract emotional words, their noun factors along with their modification words.

For example, the following rules elaborate how to extract the features from the review comments.

Rule 1- Simple sentence with adjective

Step 1: *if there is Noun phrase(NP) with Adjective phrase(ADJP)*
extract adjective (JJ)
Step 2*: search in knowledge base for the adjective (ADJP) word*
Step 3 *Return the count (priority value) with feature*

Rule 1 can be used for most simple type of review comments as described below:

The sentence with NP and an ADJP, Rule 1 can be applied and outcome of the analysis can be obtained. Hence if there is a noun phrase with an adjective, the adjective will be extracted and it will be searched in the knowledge base to get the feature and its appropriate polarity value.

Example review: "the bacon has been fantastic"

Rule 2: Complex sentence with adjectives and modifications

Step1 *: if there is NP with one or more than one RB(Modification) or PP(preposition phrase) with ADJP*
extract all RB and JJ
search in knowledge base for the adjective word(ADJP) store the value
search in knowledge base for the modification word(RB)
Step 2*: for each modification word*
extract the type of modification(enhancing/shifters) store the type
Step 3: *if all are the same type*
Step 3.1*:if the type is Enhancers*
*polarity value= pv of adj+ type of pol of adj (pv of mod*no of mod words)*
Step 3.2*: if the type is shifters*
*polarity value= pv of adj-type of pol of adj (pv of mod*no of mod words)*
Step 4*: if they are of a different type*
Step 4.1*:if the pol type of adj is positive*
take half of the value of the sentiment shifters and work as same for same type of mod words
Step4.2: *if the pol type of adj is negative take half of the value of the sentiment enhancer and work as same for same type of mod words*
Step 5: *Return the count with feature*

Rule 2 is an enhancement of the first rule. That is, adjective will be presented with modification words. Initial task of the system is the need to identify the value and type of the adjective. Then, for each modification word, identify the type. If all the modification words are sentiment enhancers, the overall polarity value of the adjective will be calculated. It is carried out using the addition of polarity value of adjective and type of the adjective getting the product of type of the adjective into the polarity value of the modification word and the number of modification words. If the modification words are sentiment shifters instead of adding it will be deducted from the polarity value of the adjective. Example review: "the food is always consistently outrageously good".
If the modification types are different and if the polarity type of the adjective is positive then from the value of sentiment enhancers only half will be taken. If the polarity type of the adjective is negative, half of the value of the sentiment shifters will be dropped. Example review: "the food is not very good".

5. IMPLEMENTATION

The implementation task explains the procedures adopted in implementing the concept-level opinion mining system proposed in Section 6 and describes the various tools and technologies used in the study. This section describes, how one can implement the proposed design approach.

Python 2.7 (stable version of Python) was used in this task with a vast number of in-built functions, vast availability in XML/HTML parsing libraries such as xml, and more features which are needed for supported scientific calculations and for providing an efficient multi-dimensional container of generic data with packages such as NumPy and SciPy for Concept Level Sentiment Analysis.
Implementation process consist of the three major tasks and is described in detail as follows.

5.1 Design and Implementation of the concept level knowledgebase

As discussed in section 3, there are different quality factors that can be identified from the restaurant domain. When constructing the knowledgebase, deep knowledge is extracted and embedded in the knowledge based with integrating the domain knowledge which are included in well recognized knowledge bases as described in the following sources.

Annotated data set

This dataset was obtained from SemEval-2015 ABSA Train Data. Here emotional words for different quality factors are annotated by humans. Therefore, it is easier to categorize emotional words and insert them into the knowledge base.

WordNet

WordNet has been used to add synonyms and antonyms for collected words from the reviews. WordNet made a high contribution to the expansion of the knowledgebase (http://wordnet.princeton.edu).

DBpedia

Some reviews contain comments about different food item. Hence, in that case the target factor is a food item. Therefore, knowledge is needed to identify that these different types of food types belong to the category of FOOD. For example, there can be the same general emotional word targeted on two different quality factors such as "Sandwich is good" and "Service is good". As such in order to identify the target quality factor the emotional word is not adequate. In other words, there should be a way to identify that Sandwich is a 'Food' and service is the quality factor 'Service'. For this a list of different types of food has been collected using DBpedia and it has made a great contribution to enhance the knowledgebase.

SentiWordNet

In order to calculate the polarity value of each of the emotional words, there are certain different mechanisms used by various researchers. Important mechanisms are explored in the literature review. Among the above explored mechanisms, SentiWordNet was used to get higher quantitative in this study. Further, it described numerical measure for words [17]. Therefore, with these collected knowledge it is required to explain the structure

of the knowledge base in detail.

Mysql 5.5.4

In this study, the database has been created using Mysql 5.5.25 with embedded advanced database techniques. Further, tables include motional words, quality factors, target features and emotional words. After developing the knowledge base, the main step can be carried out. That is, extracting the target factor from the review comments. It is explained below in detail.

5.2 Extracting Quality Factors

Initially, the POS tagged sentence adjectives will be extracted. After extracting the adjectives, the knowledge base Rule 1 will be applied as explained in the design. For each adjective, the knowledge base has been searched and the quality factor type and the polarity value located.

If the return category is GENERAL then the noun relevant to the adjective has been extracted. In order to do this, for each adjective extracted, relevant dependencies have been taken in the dependencies that have been stored in the previous preprocessing stage. From them the dependencies that have the pattern of noun with the adjective is considered and noun has been extracted. Then the noun is used to find the target quality factor in the knowledge base.

If the sentence contain modification words as explained in the design Rule 1.1 has been applied. In order to do that in the same way for each dependency adjective, modification combined dependencies will be taken. Then for the adjective based on the modification types polarity value has been calculated. Therefore, with the implementation proposed solution which is explained in the following can be evaluated.

6 Experiments

The evaluation of the proposed Concept Level Sentiment Analyzer done by testing the accuracy of the classification for various input datasets. The evaluation process involves two (02) major tasks such as, evaluation of the extracted quality factor and the level of satisfaction. In the level of satisfaction, positivity and negativity is evaluated using the above mentioned methodologies. Further, outcome of the justified random sample techniques embedded sample fields survey was used to evaluate the numerical value.

6.1 Evaluation of Quality Factors

As re-evaluation of the outcomes, Precision, Recall, F-Measure and Accuracy for understanding and measuring the relevance and the correctness of the sentient classification were used. Further, Recall measures were used to determine the completeness of quality of the results while Precision reflects the exactness of quality of the results. F- Measure is used to determine the harmonic mean of the precision and the recall of a test is used to measure accuracy.

A Trained data set from SemEval-2015 ABSA Data were used for the evaluation. In this study, newly generated quality factors have been added to the data set and modified accordingly. In each step of the testing process, sentences in the reviews are analyzed and if new quality factors have been found then they have been annotated. Basically the modified data set contains annotated quality factors such as QUALITY, QUANTITY, GENERAL, PRICE, and STYLE and OPTIONS. This data set is used to calculate the Recall and Precision.

6.2 Evaluation of the Level of Satisfaction

In order to evaluate the numerical values which were taken using the algorithm (Rule 1 and Rule 2), a questionnaire was designed and information collected from customers in different selected restaurants. The information in the questionnaire was enriched to evaluate the different patterns of review sentences including the complex review sentences. In this study, we have used group administered survey methodology to collect the information [18].

The outcomes from the evaluation is obtained and are shown in the table below.

Table 1: Results of the Evaluation

Precision	Recall	F-measure	Accuracy
92.2 %	78.7%	84.91 %	74.6%

Further, level of satisfaction is measured and evaluation success level obtained as 17% of strongly agreed, 66% of agreed, 15% moderate and 2% for disagree

7 Findings and Discussion

In principle, the accuracy of a sentiment analysis system is how well it agrees with human judgments. This is usually measured by precision and recall. However, according to research, human rates typically agree 79% of the time. Thus, a 70% accurate program is doing nearly as well as humans. So when compared with the results that have been achieved using the proposed solution it has an accuracy of 74.6% which is higher than 70%. Therefore, we can conclude that the proposed solution to determine the quality factor is acceptable when evaluating the review comment, and as such, quality factor can be used to evaluate the customer reviews is more accurate than the existing methodologies. In this study, 92.2% precision was obtained from the proposed approach. This implies that a majority of words are mapped with the knowledgebase including emotional words and a noun that is misspelt.

From this study, recall value obtained is 78.7% and benchmark value is 88.15. The experiment value is low compared to the benchmark. This is because missing words might be high. One reason is this experimental prototype knowledgebase is not comprehensively and fully updated. Therefore, in real implementation, expansion of the knowledgebase with more tokens is essential. Further, more tokens from reviews can be added to the knowledgebase and hence recall value can be increased. Another reason is that the ability to extract an emotional word. Most of the emotional words are adjectives. As such, if the NLTK POS tagging section could not identify a particular emotional

word as an adjective, the proposed system will not concentrate on extracting the emotional words and research opportunities are available to extract the emotional words in future studies. As we have observed, when POS tagging is carried out using Stanford Parser and using NLTK there are some differences. In most of the cases Stanford parser gives the proper POS tag, and as such we have used it in our implementation as mentioned, but there were situations where POS tag were incorrect. Therefore, we would be able to enhance the results of Recall if we do this manually, but it is hard to do it for every sentence in the data set. Hence, we identified that this opens the research area for NLP to find more accurate POS tags, so that solutions that use POS tagging will be more accurate.

In the discussing about the results on the level of satisfaction, we have achieved a considerable rate of answers that agree with the polarity values that have been obtained using our algorithm in the rules set.

8 Conclusion

Nowadays, Web 2.0 offers a great means to share knowledge, including opinions which may be useful to various kinds of people. Most of the existing approaches are based on word-level analysis of texts and are able to detect only explicit expressions of sentiment. Concept-level opinion mining focuses on going beyond a mere word level analysis of text and provide a more semantic analysis of text through the use of web ontologies or semantic networks empowering novel approaches to sentiment analysis, in potentially any domain.

This study has explained all the different stages of development, starting with the construction of linguistic resources, designing and implementing the proposed methodology and evaluation of the study. In order to use knowledge when classifying a quality factor a knowledge base is used which includes domain knowledge on a large scale. Various tools has been used for extracting emotional words and quality factors. The main difference of proposed solution with available research work related to this is the inclusion of domain knowledge to identify quality factors (both implicit and explicit) and the ability present a level of satisfaction with numerical figures using SentiWordNet.

9 Future Work

The current system is capable of sentiment classification only in the English language. Construction of multilingual lexical resources including sentiment seed networks and concept knowledge-bases cannot be trivial. Another possible future implementation could be, enforcing a text quality filter for the review in question, in order to reduce the number of incorrect classifications. We have considered the polarity detection as a binary classification problem in this study, hence neutral opinionated sentences are not detected. We left subjectivity detection (i.e., detecting whether a sentence contain an opinion or not) considering it as a subject for future work.

References

[1] Hu,M. & Liu,B., 2004. Mining and summarizing customer reviews in Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, NewYork,USA, pp.168-177

[2] Kessler, B. Nunberg, G. & Schutze, H., 1997. Auto matic Detection of Text Genre in Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics, Stroudsburg, PA, USA, pp.32-38

[3] Hatzivassiloglou, V. & Wiebe., 2000. Effects of Adjective Orientation and Gradability on Sentence Subjectivity in International Conference on Computational Linguistics

[4] S. Poria, E. Cambria, A. Gelbukh, and C. Gui, "A Rule-Based Approach to Aspect Extraction from Product Reviews."

[5] B. Liu, "Sentiment Analysis and Opinion Mining," no. May, 2012

[6] Popescu, Ana-Maria and Oren Etzioni. Extracting product features and opinions from reviews. in Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP-2005). 2005

[7] Qi Su, Xinying Xu, Honglei Guo, Zhili Guo, XianWu, Xiaoxun Zhang, Bin Swen, and Zhong Su. 2008. Hidden sentiment association in chinese web opinion mining. In Proceedings of International Conference on World Wide Web (WWW-2008), pages 959–968

[8] Hai Zhen,Kuiyu Chang, and Jung-jae Kim. 2011. Implicit feature identification via co-occurrence association rule mining. In Computational Linguistics and Intelligent Text Processing. 12th International Conference, CICLing 2011, Tokyo, Japan, February 20–26, 2011. Proceedings, Part I, volume 6608 of Lecture Notes in Computer Science, pages 393–404.

[9] Lingwei Zeng and Fang Li. 2013. A classification-based approach for implicit feature identification. In Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data. 12th China National Conference, CCL 2013 and First International Symposium, NLP-NABD 2013, Suzhou, China, October 10–12, 2013, Proceedings, volume 8202 of Lecture Notes in Computer Science, pages 190–202.

[10] Zhou,L. & Chaovali,P.,2008. Ontology-Supported Polarity Mining, Journal of The American Society for Information Science and Technology, 59 (1), pp. 98-110

[11] Haji, B.Vidyasagar, P. & Chen,W., 2009. A state of the art opinion mining and its application domains, in Yousef Ibrahim (ed), International Conference on Industrial Technology (ICIT 2009), pp. 1-6.

[12] Yaakub,R. M, Li. & Feng,Y., 2011.Integration of Opinion into Customer Analysis Model in proceedings of Eighth IEEE International Conference on e-Business Engineering, pp. 90-95

[13] C. Sivagami and S. C. Punitha, "Ontology Based Sentiment Clustering Of Movie Review," vol. 2, no. 4, pp. 375–378, 2013
[14] Anni Coden, Dan Gruhl, Neal Lewis, Pablo N. Mendes et, al "Semantic Lexicon Expansion for Concept-Based Aspect-Aware Sentiment Analysis", 2014
[15] N. Madnani, "Getting Started on Natural Language Processing with Python," pp. 1–16, 2013.
[16] http://nlp.stanford.edu/software/stanford-dependencies.shtml
[17] A. Esuli and F. Sebastiani, "SentiWordNet : A High-Coverage Lexical Resource," pp. 1–26, 2006.
[18] http://www.socialresearchmethods.net/kb/survtype.php

AUTHOR



Ashani Tharushika Silva is an ungraduated in University of Colombo School of Computing from 2011 to 2015. She has achieved Advanced Diploma of Management Accounting in Charted Institute of Management Accounting. She has research interests of Natural Language Processing, Text Mining, and Artificial Intelligence.



Dr. M G N A S Fernando is a Senior Lecturer in the Department of Information Systems Engineering, University of Colombo School of Computing, Sri Lanka. He has taught Courses on Data structures and Algorithms, Operating Systems, Data Communication & Networking, and Data Mining. Previously, he was Consultant/ Chief Information officer in the Ministry of Higher Education. Prior to his Ministry of Higher education Appointment, he was Assistant Network Manager, University of Colombo, Systems Analyst/Programmer at the University Grants Commission, Sri Lanka; he has over with over 27 years of experience Computing and ICT field. Dr. Fernando graduated from the University of Colombo in 1983 with a B.Sc. in Applied Science. He earned an M.Sc in Computer Science and PhD subsequently. His research areas are Data Mining, Algorithms and ICT and Education especially in e-learning technologies He has published several papers and he was a Students counselor, Students coordinator, Senior Treasurer, MSc coordinator of the University of Colombo school of Computing.