International Conference on Information Engineering, Management and Security
2015 [ICIEMS 2015]

ISBN	978-81-929742-7-9		**VOL**	01
Website	www.iciems.in		**eMail**	iciems@asdf.res.in
Received	10 - July - 2015		**Accepted**	31- July - 2015
Article ID	ICIEMS001		**eAID**	ICIEMS.2015.001

INTEGRATING ASPECTS BASED ON OPINION MINING FOR PRODUCT REVIEWS

Suganya S[1], Sureka K[1], Vishnupriya P[1]
[1]V.S.B.Engineering College

Abstract: *It is a common practice that merchants selling products on the Web ask their customers to review the products and associated services. As e-commerce is becoming more and more popular, the number of customer reviews that a product receives grows rapidly. For a popular product, the number of reviews can be in hundreds. This makes it difficult for a potential customer to read them in order to make a decision on whether to buy the product. We aim to summarize all the customer reviews of a product. This summarization task is different from traditional text summarization because we are only interested in the specific features of the product that customers have opinions on and also whether the opinions are positive or negative. We do not summarize the reviews by selecting or rewriting a subset of the original sentences from the reviews to capture their main points as in the classic text summarization. We only focus on mining opinion/product features that the reviewers have commented on. A number of techniques are presented to mine such features. The proposed system is used to decisive the customer reviews for multiple product reviews then find the aspect from the review and classify the review whether they wrote positive or negative. We only focus on mining opinion/product features that the reviewers have commented on and compare the more product and rank the product based on the reviews automatically.*

Keywords: Aspect –based, Opinion mining, Product reviews

INTRODUCTION

With the inception of the Web 2.0 and the explosive growth of social networks, enterprises and individuals are increasingly using the content in these media to make better decisions. For instance, customers check opinions and experiences published by other customers on different Web platforms when they planning to buy the products through online. On the other hand, for organizations, the vast amount of information available publicly on the Web could make polls, focus groups and some similar techniques an unnecessary requirement in market research. However, due to the amount of available opinionated text, users are often overwhelmed with information when trying to analyze Web opinions. So far, many authors have tacked the problem of human limitation to process big amounts of information and extract consensus opinions from a large number of sources relying on data-mining-based tools. Considering a similar problem, this work is an effort to create a tool that offers a set of summarization methods and help users digest in an easy manner the vast availability of opinions.

Three main components of Opinion Mining are:
1. Opinion Holder: Person that expresses the opinion is opinion holder.
2. Opinion Object: Object on which opinion is given.
3. Opinion Orientation: Determine whether the opinion about an object is positive, negative or neutral. The core of our system is a novel extension of aspect-based opinion mining methodology, which was developed by us for online shopping of products. The core of our system is concerned with the fact that

Cite this article as: Suganya S, Sureka K, Vishnupriya P. "INTEGRATING ASPECTS BASED ON OPINION MINING FOR PRODUCT REVIEWS." *International Conference on Information Engineering, Management and Security (2015)*: 01-06. Print.

Users refer differently to different kinds of generic products when writing reviews on the Web. For instance, when a person writes a movie review, he probably comments not only movie elements, but also movie-related people. The contributions of this paper are mainly three. First, to the best of our knowledge existing approaches do not address the special issues. So we developed a model for aspect-based opinion mining that specially considers these features. Secondly, as a result of the analysis of the domain, we created a special datasets that help representing the features of the mentioned domain. The rest of this paper is structured in the following manner.

2. Related work

opinion mining or sentiment analysis comprises an area of NLP, computational linguistics and text mining, and refers to a set of techniques that deals with data about opinions and tries to obtain valuable information from them. The aspect-based approach is very popular and many authors have developed their own perspectives and models. Other related approaches are unsupervised topic-based document modeling techniques, which model an input document as a mixture of topics. In this context, our work lies on a radically different paradigm, as the former consists in identifying the aspects reviewed in a piece of text based on a bag-of-words model of the document, rather than extracting individual feature mentions and their related opinions. Therefore, our work is not directly comparable to these kinds of works. Our work acknowledges the differences between domains that is discussed in the paper, and proposing a general model that works for all the domains. Also, our system does not require any training datasets and only a small amount of human support. Finally, one last related topic is the set of so-called concept-level sentiment analysis approaches. These approaches focus on a 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. Our approach is different from all these applications since it is aspect-based and analyzes opinions at the sentence level.

3. Background

In this section, we proposed our approach in general terms. The opinions are 5-tuples composed of the following parts.

• An entity: Proposed to denote the opinion objective other words. An entity can contain a set of components and attributes and, similarly, each entity component can have its own subcomponents and attributes.
• An aspect: Because it is difficult to study an entity at an arbitrary hierarchy level, this hierarchy is simplified to one or two levels, denoting as aspect every component or attribute of the entity.
• The Sentiment orientation, considering that opinions express a positive or negative sentiment about what they evaluate.
• The Opinion holder, which corresponds to the user that gives the opinion.
• Time: Time and date when the opinion was given. In this manner, opinions are considered to be a positive or negative view, attitude, emotion or appraisal about an entity or an aspect of that entity from an opinion holder in a specific time. The following concepts are also introduced:
• Entity expression: Corresponds to the actual word or phrase written by the user to denote or indicate an entity. As a result, entities
• are then generalizations of every entity expression used in the analyzed documents, or a particular realization of an entity expression.
• Aspect expression: As for an entity expression, the aspect expression is the actual word or phrase written by the user to denote or indicate an aspect. Thus, aspects are also general concepts that comprise every aspect expression.

3.1. Aspect identification

This stage aims to find and extract important topics in the text that will then be used to summarize. In their proposal, part-of- speech (POS) tagging and syntax tree parsing (or chunking) are used to find nouns and noun phrases or NPs. Then, using frequent item set mining, the most frequent nouns and NPs are extracted. The extracted sets of nouns and NPs are then filtered using special linguistic rules. These rules ensure that the terms inside those aspects that are composed of more than one word are likely to represent real objects together and also eliminate redundant aspects. They also extract non-frequent aspects using an approach by finding nouns or NPs that appear near to opinion words with high

frequency. This approach does not extract adjectives or any other kind of non-object aspects.

3.2. Sentiment prediction

The next phase is sentiment prediction, to determine the sentiment orientation on each aspect. This method relies on a sentiment word dictionary that contains a list of positive and negative words (called opinion words) that are used to match terms in the opinionated text. Also, since other special words might also change the orientation, special linguistic rules are proposed. Among others, these rules consider negations words ‘‘no’’ or ‘‘not’’ and also some common negation patterns. However, despite how simple these rules might appear, it is important to handle them with care, because not all occurrences of such rules or word apparitions will always have the same meaning.

Cite this article as: Suganya S, Sureka K, Vishnupriya P. “INTEGRATING ASPECTS BASED ON OPINION MINING FOR PRODUCT REVIEWS.” *International Conference on Information Engineering, Management and Security (2015)*: 01-06. Print.

3.3. Summary generation

The last step is summary generation, to present processed results in a simple manner. In this context, defined opinion quintuples are a good source of information for generating quantitative summaries. In this case, each bar above or below the x-axis can be displayed in two scales: (1) the actual number of positive or negative opinions normalized with the maximal number of opinions on any feature of any product and (2) the percent of positive or negative opinions, showing the comparison in terms of percentages of positive and negative reviews.

4. Proposed extension

Our extension, considers the same set of structured steps mentioned in Section. Here, we discuss issues on each one of the three steps and explain our own approach in the context of product reviews.

4.1. Aspect expression extraction

The aspects do not directly appear in a text but they exist in the manner of aspect expressions. Accordingly, when trying to apply opinion model to extract opinions from real data, concepts can be somewhat confusing or unclear. It is also unclear how aspects that appear more than once in a document are managed. Having noticed these issues, a model to build opinion tuples from an opinionated document has been developed here. We will not extract implicit nor not-frequent aspect expressions.

4.2. Determination of the opinion orientation

Taking the work as inspiration, a set of rules to determine the sentence orientation was developed, always considering opinion words as a basis.

4.2.1. Word orientation rules

In first place, we need to determine the orientation of each word in a sentence. In order to do so, we propose Algorithm 1.The algorithm applies a set of linguistic rules, which are explained below:

Algorithm 1.Word orientation

1: if word is in opinion- words then
2: mark (word)
3: Orientation Apply Opinion Word Rule (marked word)
4: else
5: if word is in neutral_words Then
6: mark (word)
7: orientation -0
8: end if
9: end if
10: if word is near a too_word then
11: orientation Apply Too Rules(orientation)
12: end if
13: if word is near a negation_word then
14: orientation Apply Negation Rules (orientation)
15: end if
16: return orientation

Word rules: Positive opinion words will intrinsically have a score of 1, denoting a normalized positive orientation, while negative ones will have associated a score of 1. Every noun and adjective in each sentence that is not an opinion word will have an intrinsic score of 0 and will be called neutral word.
Negation rules: A negation word or phrase usually reverses the opinion expressed in a sentence. Consequently, opinion words or Neutral words that are affected by negations need to be specially treated.
Too rules: Sentences where words ‘‘too’’, ‘‘excessively’’ or ‘‘overly’’ appear, are also handled specially. When an opinion word or a neutral word appears near one of the mentioned terms, denoted too words, its orientation will always be Negative (score=-1).

4.2.2. Aspect orientation rules

Having mentioned rules that help in determining each word orientation in a sentence, it is now explained how all these orientations should be combined to determine the final orientation of a sentence on a particular aspect. Our proposal is summarized in Algorithm 2 and it only considers words marked as opinion words or neutral words, which we call marked words, as they are the only ones that will provide the orientation for each sentence. The detailed process is explained below.

Algorithm 2: Opinion orientation

1: if but_word is in sentence then
2: orientation Opinion
Orientation (aspect,marked_words,but_clause)

Cite this article as: Suganya S, Sureka K, Vishnupriya P. "INTEGRATING ASPECTS BASED ON OPINION MINING FOR PRODUCT REVIEWS." *International Conference on Information Engineering, Management and Security (2015)*: 01-06. Print.

```
if orientation – 0 then
return orientation
else
orientation Opinion Orientation(aspect,marked_words,not but_clause)
if orientation – 0 then
return À1 Â orientation
else
return 0
end if
end if
else
for all aspect_position in aspect do
for all aspect_word in aspect_position do
orientation += suborientation
end for
final_orientation += orientation
end for
if final_orientation > 0 then
return 1
else
if final_orientation < 0 then
return À1
else
return 0
end if
end if
end if
```

4.3 Summarization

This proposal seems fairly, simple and effective for summarizing opinions. However, it lacks a robust way of measuring the importance of each evaluated aspect. Here, we attempt to measure the importance of each aspect simultaneously using the amount of positive and negative opinions of it. We also use that measure to rank aspects. We calculate the standard deviation of these scores using:

$$AVScorei = \frac{Scorei + NScorei}{2}$$

We propose that aspect-based summaries should include bar charts and a table that shows the actual values of PScorei, NScorei and Relative Importance for each aspect expression.

5. System architecture

Two different tasks need to be performed, aspect extraction and orientation determination, for which two submodules are included:

- **Aspect extraction sub-module:** in charge of applying the aspect extraction algorithm to a set of POS-tagged sentences.
- **Orientation determination sub-module:** This sub-module applies the algorithms presented in Section 4 to determine the orientation of an opinion on a given aspect. It also extracts the set of adjectives that appeared near each aspect. Results include the following features:
- **Aspect-based summaries:** Bar charts, in which each bar measures the number of positive and negative mentions of each attribute or component of one product. Bars are initially sorted according to Relative Importance.
- **Adjective bubble charts:** Nearby adjectives in all sentences where an aspect appears are shown in a bubble chart. The size of each bubble counts the times that each adjective is used to describe the aspect.
- **Original opinions**: A list of all original sentences is also displayed in an ad-hoc manner, separating them into positive or negative.

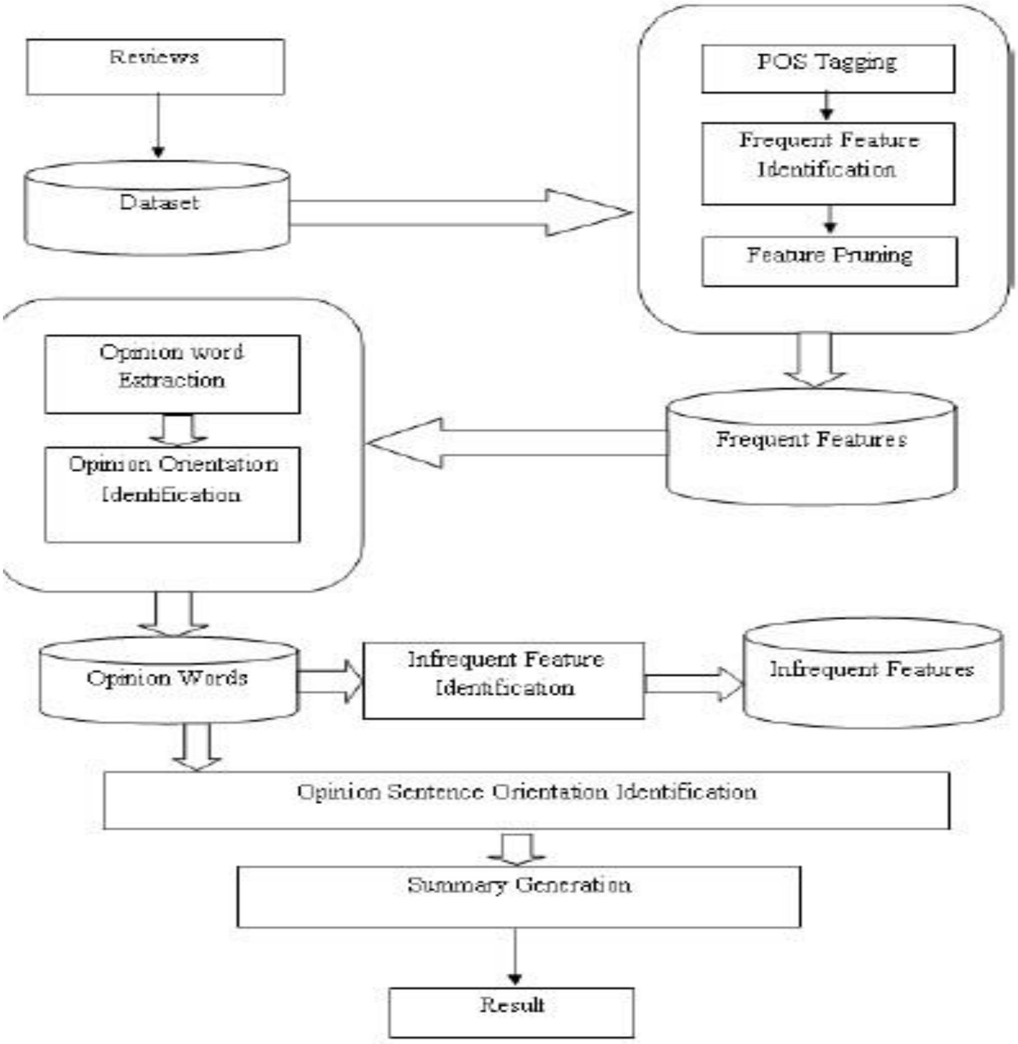

Fig.1: General Design of our system.

7. Conclusions and future work

In this study, we present a generic design of a tourism opinion mining system that aims to be useful in many industries. The core of our system is an extension of aspect-based opinion mining technique. On the one hand, the non-tailored algorithm for aspect expressions extraction, based on frequent nouns and NPs appearing in reviews, achieved a poor performance. This result shows that, in fact, multiple expressions are used to denote the same attribute or component in online product reviews. Therefore, not only the most frequent words need to be considered when extracting aspect expressions in order to achieve a better recall for this task. Our design and models for aspectbased opinion can be used in many possible applications in the online shopping domain. Benefits that may arise entail both merchants and customers.

7.1. Future work

For future work, the primary objective should be to improve Recall on the task of aspect expression extraction, finding infrequent and implicit aspect expressions. On the other hand, we have seen that tourism product reviews contain an important number of sentences that have no opinions. These sentences need to be filtered since they introduce noise to the opinion mining process. This also includes the problem of analyzing context and domaindependent opinions. New methods to determine subjectivity or sentiment orientation need to be tested on the tourism domain in order to improve the performance of these tasks. Future work should also tackle the problem of transforming aspect expressions into aspects. This is a difficult problem yet a crucial feature for any system like ours, because presenting aspect expressions to users implies redundancy and makes the analysis more complex. Here, the objective is to build or use ontologies, hierarchies or clusters of aspect expressions to make the system become easier to navigate and more intuitive for users. Finally, another extension of this work implies working with tourism products reviews written in different languages. Some of the NLP tasks that are used by our system, including sentence and word tokenizers, are generally machine leaning algorithms that need to be properly trained in order to generate good results. The vast availability of data in English to train these models contrasts with a relative scarcity for other languages. Therefore, there is an immense room for future work on this area.

Cite this article as: Suganya S, Sureka K, Vishnupriya P. "INTEGRATING ASPECTS BASED ON OPINION MINING FOR PRODUCT REVIEWS." *International Conference on Information Engineering, Management and Security (2015)*: 01-06. Print.

References

1. *Archak, N., Ghose, A., &Ipeirotis, P. (2007). Show me the money!:*
2. *Deriving the pricing power of product features by mining consumer reviews. In Proceedings of the 13th ACM SIGKDD international conference on knowledge discovery and data mining (pp. 56–65). ACM.*
3. *Bollegala, D., Matsuo, Y., Ishizuka, & M. (2007). An integrated approach to measuring semantic similarity between words using information available on the web. In HLT-NAACL (pp.340–347).*
4. *Cadilhac, A., Benamara, F., &Aussenac-Gilles, N. (2010). Ontolexical resources for feature based opinion mining: A case-study. In 23rd International conference on computational linguistics (p.77).*
5. *Cambria, E. (2013). An introduction to concept-level sentiment analysis. In F. Castro,*
   a. *Gelbukh, & M. González (Eds.), Advances in soft computing and its applications. Lecture notes in computer science (Vol. 8266,pp. 478–483). Berlin Heidelberg: Springer.*
6. *Cambria, E., Poria, S., Gelbukh, A., &Kwok, K. (2014). A common-sense based api for concept-level sentiment analysis. Making Sense of Microposts, 1(1), 1–2.*
7. *Cruz, F. L., Troyano, J. A., Enríquez, F., Ortega, F. J., & Vallejo, C. G. (2013). 'Long autonomy or long delay?' The importance of domain in opinion mining. Expert Systems with Applications, 40(8), 3174–3184. Decker, R., &Trusov, M. (2010). Estimating aggregate consumer preferences from online product reviews. International Journal of Research in Marketing, 27(4), 293–307. Ding, X., Liu, B., & Yu, P.(2008). A holistic lexicon-based approach to opinion mining. In Proceedings of the international conference on Web search and web data mining (pp. 231–240). ACM. Dueñas-Fenández, R., Velásquez, J. D., &LH' uillier, G. (2014). Detecting trends on the Q3 web: A multidisciplinary approach. Information Fusion 0 (0), in press. Fromkin, V., Rodman, R., & Hyams, N. (2010). An introduction to language. Wadsworth Publishing Company. Havasi, C., Cambria, E., Schuller, B., Liu, B., & Wang, H. (2013a). Knowledge-based approaches to conceptlevel sentiment analysis. Intelligent Systems, 28(2), 12–14. Havasi, C., Cambria, E., Schuller, B., Liu, B., &Wang, H. (2013b). Statistical approaches to concept-level sentiment analysis. Intelligent Systems, 28(3), 6–9. Hu, M., & Liu, B. (2004a). Mining and summarizing customer reviews. In Proceedings of the tenth ACM SIGKDD international conference on knowledge discovery and data mining (pp. 168–177). ACM. Hu, M., & Liu, B. (2004b). Mining opinion features in customer reviews. In Proceedings of the national conference on artificial intelligence (pp. 755–760). Menlo Park, CA; Cambridge, MA; London: AAAI Press. MIT Press; 1999. Hu, M., &Liu, B. (2006). Opinion extraction and summarization on the web. Proceedings of the national conference on artificial intelligence (Vol. 21, pp. 1621). Menlo Park, CA; Cambridge, MA; London: AAAI Press. MIT Press; 1999.Kim, H., Ganesan, K., Sondhi, P., & Zhai, C. (2011). Comprehensive review of opinion summarization.*

Cite this article as: Suganya S, Sureka K, Vishnupriya P. "INTEGRATING ASPECTS BASED ON OPINION MINING FOR PRODUCT REVIEWS." *International Conference on Information Engineering, Management and Security (2015)*: 01-06. Print.