Implicit Aspect Identification Techniques for Mining Opinions: A Survey

Mily Lal
ISB&M School of Technology,Pune

Kavita Asnani
Padre Conceicao College of Engineering,Goa

ABSTRACT

Gathering information by finding out what other people think is always been a common behavior .It has become necessary to summarize the information obtained due to its growing availability and popularity in the form of online review sites and personal blogs . Aspect extraction is one major step for mining opinions. Extracting aspects still remains to be a challenging in problem in opinion mining domain. Most of the research works have only concentrated in extracting explicit aspects. Implicit aspects are also important because they relate to sellers, services and logistics. Without knowing it mined opinions can be of no use. This paper describes techniques and approaches that promises to enable implicit extraction for opinion seeking systems

General Terms

Opinion mining, Data mining, Sentiment Analysis

Keywords

Aspects extraction, implicit aspect, Opinions sentiments, sentiment lexicon, sentiment classification

1. INTRODUCTION

The number of customer reviews is increasing very rapidly with the rapid popularity and availability of e-commerce. It has become very difficult for users to read all these unstructured reviews and obtain useful information. All aspects of the transactions are rarely focused by the users. Therefore, it has become necessary to provide aspect-based opinion mining results to help users make Aspect extraction is one major step for mining opinions. Aspect includes both explicit and hidden or implicit aspects. Most of the research works have only concentrated in extracting explicit aspects only a few have attempted in extracting implicit features. The implicit aspects, which are implied by some words or phrases, are very important that they can convey the opinions and help in better understanding of customer reviews. [8][19][21].

2. SIGNIFICANCE OF ASPECTS

Aspect based opinion mining has attracted many works recently with various approaches in extracting aspects related to opinions expressed on online reviews. Aspect can be defined as the subject of the review such as the product like "cell phone" or it can be the products attribute like "display". An aspect that appears as noun or noun phrase in a sentence is known as the explicit aspect where as the aspect which is implied in the sentence is called implicit aspects.

For example in the sentence "Lovely display, I wish to buy one despite its slight heaviness", the word "display" is an explicit aspect and "weight" is an implicit aspect because "heavy" implies to its weight.

Aspect engineering is an extremely basic and essential task for Opinion Mining. Converting a piece of text to an aspect vector is the key step in any data driven approach to Opinion.

Without knowing aspects in a corpus, the mined opinions will be of little use.

Majority of the work has been done in finding explicit aspects from online customer reviews. Some had completely ignored implicit aspect identification [3][4]. Hu & Liu had partially addressed implicit aspect identification by applying the same methods used for explicit aspect identification [6]. However few works have tried considering implicit aspects during aspect identification phase. They have almost succeeded in improving the precision and recall measures [8][19][21].

3. OPINION MINING FRAMWORK

The main tasks of opinion mining are 1) identification of aspects 2) Selecting significant aspects 3) Classification of sentiments into positive and negative aspects 3) Summarization of the results.
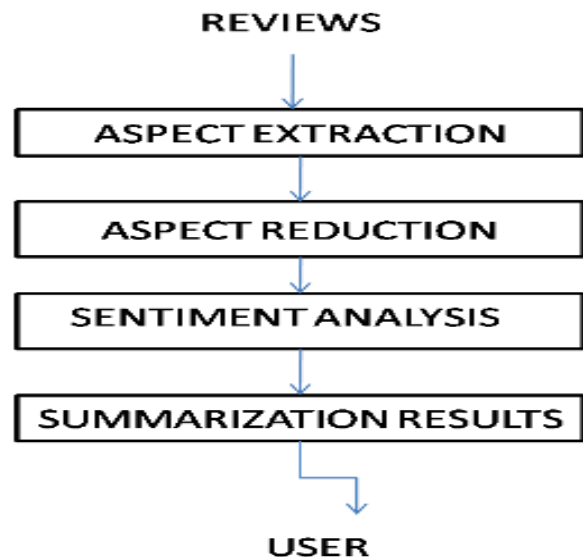

Fig 1: OPINION MINING FRAMEWORK

4. ASPECT EXTRACTION

This is the basic step in opinion mining. From a given text document, the target object's aspect commented by the author is extracted. It is different from the tradition aspect extraction in the area of machine learning.Aspects can be divided into explicit and implicit. If the aspect itself appears in the text, for example in the following sentences:"The battery life is long.", the aspect "Batter Life" appears in the sentence explicitly. However, in the sentence, "The camera is small enough to put in my pocket.", the aspect "size" is not directly mentioned but only implied by the word "small".The aspect extraction phase deals with i)aspect types which identifies the type of aspects used for opinion mining, ii)aspect selection which is used to select good aspects for opinion classification, iii)aspect weighting mechanism which weights each feature for good

recommendation and iv) reduction mechanisms which reduces aspects for optimizing the classification process.

4.1 Detecting Explicit Aspects

Explicit aspects are first identified and extracted as the first step in aspect extraction as they are explicitly specified in the reviews. This would indirectly help in the identification of implicit aspects that are hidden or implied in a review sentence. Mostly nouns and noun phrases are extracted as they represent potential words that are frequently talked about. This is possible by various techniques like POS (Parts Of Speech) tagger used for separating various part of speech tokens [15][17], Dependency parser which extracts aspects based on dependency relationship [5][17],Term co-occurrence (aspects which occur together and recording their frequency)[15][17]. Opinion words are most likely to be adjective, verbs and adverb phrases appearing in the review sentence [15] [17]. Aspects can be extracted using domain ontology [13]. Some papers have focused their work on predefined set of review aspects [6] [8].

4.2 Deriving Hidden Associations

The most likely aspect for an opinion word occurring in an implicit sentence can be obtained using the association between opinion phrases and explicit aspects.

Hu & Liu [6] had partially derived implicit aspects by applying the same methods used for explicit aspect extraction. Another approach used is clustering of opinion phrases. The clustered opinions are then labeled with property names. Association measures are calculated on the opinion cluster and explicit feature using different techniques like PMI (Pointwise Mutual Information), Likelihood ratio test (LRT), Co-occurance association rules (CoAR) [5][8][19].It has been observed that LRT and PMI did not perform satisfactorily in small set of documents whereas CoAR gave better results[19].

Implicit aspects can also be identified by clustering explicit features. It is then labeled by a representative term. Co-occurrence matrix can be used which records the frequent association of the explicit aspect and opinion phrases that occur together. [15][19] Association rules can be applied to select opinion and aspect pairs that frequently appear together [19][21]. The set of properties of the ontology can also help to extract, for each aspect, the set of the associated opinion expressions [13]. Corpus based approach was found to be relevant in finding context specific mappings of different aspects or attributes relating to aspects with opinions [9][22]. But this approach was found to be ineffective in many cases. One of the shortcomings was the size of the corpus which can limit the association between the aspect and opinion in a specific domain or context. All aspects and all opinions cannot be found in the same corpus. The short comings of this approach can be overcome by using dictionary based approach [23]. Opinion attributes can be derived from by the use of dictionary. For example for the opinion *"expensive"* can be related to "*high price*". The number of related opinions can be increased by the use of multiple dictionaries. Using dictionaries will not be limited to a specific context therefore larger coverage than the corpus. Evaluation of this approach showed better results than the corpus based approaches [23].

5. SENTIMENT ANALYSIS

Once the aspects and opinions are collected from the reviews they will be classified by using different methodologies to provide a good recommendation system. In document sentiment classification, supervised machine learning approaches are used for predicting the overall sentiment of the document [15][17].In sentence level sentiment classification, where the documents obtained from reviews is parsed into sentences [15][17]. Word or Phrase sentiment classification, the word level consolidation of sentiments has been done. It provides a fine grained sentiment classification [15][17].

6. SUMMARIZATION

Summarization leads to the final step of sentiment analysis. Opinions from a single customer will not give promising results. It is necessary to analyze opinions from a larger crowd. Summarization of opinions found on aspects and their entities can be of both quantitative and qualitative nature. Aspect based summary is one of the commonly used summarization methods [4] [5]. Opinions are summarized based on a particular aspect. The user can get the quantitative analysis about sentiments each aspect. It is also possible to drill down to get the actual opinions of the review sentences. Opinion summary with a timeline helps the user get the exact knowledge of opinions about a target and also helps in future analysis like figuring out what changes people's opinions [24]. Visualization of summaries and ranking of opinionated sentences to show the strongest opinion about an aspect gives a clear knowledge about the opinion statistics of the target.[4][5]

7. CONCLUSION

Sentiment detection has a wide variety of applications in e-commerce. It helps in classifying, summarizing reviews and in other real time applications.This paper focuses on the survey of techniques and approaches that promise to enable mining of implicit aspects from customer reviews. Challenges still exist in the area of implicit aspect identification. Finding accurate solutions is one of the main existing issues. Very few attempts have been made to extract implicit features. Implicit aspects play a significant role in determining sentiments from customer reviews. It has been observed that extracting implicit aspects in addition to explicit features has resulted in increase in sentiment analysis results.

8. REFERENCES

[1] Turney, p. (2002). Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews, In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, Philadelphia, Pennsylvania.

[2] Pang, B Lee, L., and Vaithyanathan, S. (2002). Thumbs up? Sentiment Classification Using Machine Learning Techniques, *In Proc. of EMNLP.*

[3] Hu, M. and Liu, B. 2004. Mining and summarizing customer reviews. International Conference on Knowledge Discovery and Data Mining (ICDM).

[4] Popescu, A. and Etzioni, O. (2005). Extracting product features and opinions from reviews, In Proceedings of the Human Language Technology Conference and the Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP).

[5] Ana-Maria Popescu and Oren Etzioni .OPINE: Extracting Product Features and Opinions from Reviews. Proceedings of HLT/EMNLP 2005 Demonstration Abstracts, pages 32–33,Vancouver, October 2005.

[6] Liu, B., Hu, M., And Cheng, J. 2005. Opinion observer: analyzing and comparing opinions on the web. In WWW

'05: Proceedings of the 14th international conference on World Wide Web. ACM, New York, NY, USA, 342–351.

[7] Ding, X., Liu, B. and Yu, P. S. (2008). A holistic lexicon-based approach to opinion mining, In Proceedings of the Conference on Web Search and Web Data Mining (WSDM).

[8] Qi Su,Kun Xiang,Houfeng Wang,Bin Sun and Shiwen Yu(2006).Using Pointwise Mutual Information to Identify Implicit Features in Customer Reviews.ICCPOL ,LNAI 4285 ,pp.22-30,Springer(2006).

[9] Almuhareb, A. 2006. *Attributes in Lexical Acquisition*. Ph.D. *Dissertation*, Department of Computer Science, University of Essex.

[10] Zhu J., H.Wang, M,Zhu and B.K.Tsou.2011.Aspect based opinion polling from customer reviews. IEEE Transactions on Affective Computing, 2(1):37-49.

[11] Qiu, G., B. Liu, J. Bu, and C. Chen. "Opinion word expansion and target extraction through double propagation." Computational Linguistics, 2011.

[12] Jin, W. and H. Ho. "OpinionMiner: a novel machine learning system for web opinion mining and extraction. "In Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2009), 2009b.

[13] Anaïs Cadilhac, Farah Benamara, Nathalie Aussenac-Gilles . "Ontolexical resources for feature based opinion mining : a case-study. Proceedings of the 6th Workshop on Ontologies and Lexical Resources "(Ontolex 2010), pages 77–86,Beijing, August 2010

[14] Pang B, Lee L. (2008) "Opinion Mining and Sentiment Analysis", *Foundations and Trends in Information Retrieval*, Vol. 2, Nos. 1-2, pp. 1-135, 2008.

[15] Nilesh M. Shelke, Shriniwas Deshpande, Vilas Thakre ,Survey of Techniques for Opinion Mining, International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 9, September 2013

[16] Yu Zhang, Weixiang Zhu Extracting Implicit Features in Online Customer Reviews for Opinion Mining, WWW '13 Companion Proceedings of the 22nd international conference on World Wide Web companion.

[17] G.Vinodhini* , RM.Chandrasekaran, Sentiment Analysis and Opinion Mining: A Survey, Volume 2, Issue 6,June 2012 International Journal of Advanced Research in Computer Science and Software Engineering

[18] Bing Liu. Sentiment Analysis and Opinion Mining, Morgan & Claypool Publishers, May 2012.

[19] Zhen Hai, Kuiyu Chang and Jung-jae Kim. Implicit Feature Identification via Co-occurrence Association rule Mining.Springer(2011)

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

[21] Wei Wang ,Hua Xu, Wei Wan "Implicit feature identification via hybrid association rule mining" Expert Systems with Applications: An International Journal Volume 40 Issue 9, July, 2013

[22] Hartung, M and A. Frank, 2010. A Structured Vector Space Model for hidden Attribute Meaning in Adjective- Noun Phrases. *Coling* 2010.

[23] *Geli Fei 1 Bing Liu ,"*A Dictionary- Based Approach to Identifying Aspects Implied by Adjectives for Opinion Mining". *Proceedings of COLING 2012: Posters*, pages 309–318,COLING 2012, Mumbai, December 2012

[24] KU, L.-W., LIANG, Y.-T., AND CHEN, H.-H. 2006. Opinion extraction, summarization and tracking in news and blog corpora. In AAAI Symposium on Computational Approaches to Analysing Weblogs (AAAI-CAAW). 100–107.