An Insight into Task of Opinion Mining

Nidhi Mishra[1,*] and C.K. Jha[2]

[1] Buddha Institute of Technology Gorakhpur, UP, India
[2] Banasthali Vidyapith , Rajasthan, India
{mishra.nidhi7384,ckjha1}@gmail.com

Abstract. Recent years have brought the burst of popularity of community websites across the internet of opinionated text on web. Users express their views and opinions regarding products and services. These opinions are subjective information which represents user's sentiments, feelings or appraisal related to the same. People use such opinion rich sources to formalize knowledge and analyze it for further reuse. This leads to emergence of new field opinion mining which differs from traditional fact based information mining which are generally done by current search engines. With introduction of Blog track in TREC 2006, a considerable work has been done in this field which comprises of opinion mining at sentence level, passage or document level and feature level. This paper presents an insight into task of opinion mining. We find that task of opinion mining is directly related to degree of formalism of language used in data sources.

Keywords: Opinion Mining, Information retrieval, Sentiments, Blogs, Polarity.

1 Introduction

The Task of opinion mining [1] is involved with recognizing, classifying the opinionated text and determining the user's sentiments expressed in the text. The aim of information gathering is to find the demands and opinions of people. Users leave their comments and reviews through debate and personal notes on a variety of products and their services on various commercial websites such as online review sites and personal blogs. With increase in popularity of such opinion-dominated resources, there is growing need to mine the opinion linked contents to search out and identify the sentiments of others. There are two types of textual information in the entire world: facts and opinions. The facts are the objective expressions which describe about entities and their properties whereas the opinion are the subjective expression which describes people's opinions, feelings, emotions and attitudes towards entities and their properties. For e.g., fact: don-2 is the high budget movie in bollywood, and opinions: don-2 is the best action movie and in this shah rukh khan look was awesome. The opinion mining [1] [2] is generally associated with information retrieval (IR). In IR the algorithms function on factual data, whereas in opinion mining the algorithms function on subjective information. Hence opinion

* Corresponding author.

V.V. Das and P.M. El-Kafrawy (Eds.): SPIT 2012, LNICST 117, pp. 185–190, 2014.

mining is difficult task as compared to IR. The research in the field of opinion mining aims at classifying the opinion of a given text at the document level, sentence level, or feature level as positive, negative, or neutral [3] [4]. The applications of opinion mining are in (1) businesses and organizations: it includes product and service; market intelligence etc. and spends a group of money to track consumer sentiments. (2) Individuals: paying attention in other's sentiments when purchasing a product or use a service. (3) Ads placements: business and organization place ads in the user-generated content on web. Spam filtering refers to detection and removal of fake opinions that mislead the users by giving unworthy positive or negative opinions to some objects in order to sponsor or spoil the objects reputations. It is also a research issue in healthy opinion mining.

Rest of the paper is organized as follows Section 2 deals with brief review and discussion about task of opinion mining. Finally we conclude our discussion in Section 3.

2 Brief Review and Discussion about Task of Opinion Mining

In order to give more insight into the problem, in the subsequent subsections we describe various attempts to classify and formalize different opinion types. Hu and Liu [4] put most impact on their work and said that the components of an opinion are: **Opinion holder:** it is the person or organization that holds or gives a specific opinion on an object. **Object:** it is entity on which an opinion is expressed by user or company. **Opinion:** it is a view, attitude, or appraisal on an object done by an opinion holder. An **object** "O" is an entity which can be a product, person, event, organization, or topic, "O" is generally represented as a hierarchy consisting of components, its sub-components, and so on. Each node represents a component and has set of attributes of the component. O is the root node. An opinion can be determined on any node or its attribute. Hu and Liu, use the features in representing both components and their attributes. They present the review of a model as follows:

1. An object O is represented with a set of features, $F = \{f1, f2 \ldots fn\}$.
2. Each feature fi in F can be expressed with a finite set of words or phrases Wi, which are **s**ynonyms. That is to say: we have a set of corresponding synonym sets $W = \{W1, W2, \ldots, Wn\}$ for the features.
3. An opinion holder j comments on a subset of the features $Sj \subseteq F$ of object O.
4. For each feature $fk \in Sj$ that j comments on, he/she, chooses a word or phrase from Wk to describe the feature, and expresses a positive, negative or neutral opinion on fk.

With regard to opinion there could be two types of opinions: Direct opinion and indirect opinion

- Direct Opinions: sentiment expressions on some identifiable objects, e.g. products, services etc.
    - E.g., "the picture quality of the mobile camera is good"
- Comparisons: relations establishing similarities or differences in between different objects.
    - E.g., "car x is cheaper than car y."

Table 1. Presents insight into opinion mining at different levels

Classification of Opinion mining at different levels	**Assumptions made at different levels**	**Tasks associated with different levels**
1. Opinion Mining at Sentence level.	1. A sentence contains only one opinion posted by single opinion holder; this could not be true in many cases e.g. there could be multiple opinions in compound and complex sentences. 2. Secondly the sentence boundary is defined in the given document	Task 1: identifying the given sentence as subjective or opinionated Classes: objective and subjective (opinionated) Task 2: opinion classification of the given sentence. Classes: positive, negative and neutral.
2. Opinion Mining at Document level.	1. Each document focuses on a single object and contains opinion posted by a single opinion holder. 2. Not applicable for blog and forum post as there could be multiple opinions on multiple objects in such sources.	Task 1: opinion classification of reviews Classes: positive, negative, and neutral
3. Opinion Mining at Feature level.	1. The data source focuses on features of a single object posted by single opinion holder. 2. Not applicable for blog and forum post as there could be multiple opinions on multiple objects in such sources.	Task 1: Identify and extract object features that have been commented on by an opinion holder (e.g., a reviewer). Task 2: Determine whether the opinions on the features are positive, negative or neutral. Task 3: Group feature synonyms. Produce a feature-based opinion summary of multiple reviews.

2.1 Document Level Opinion Mining

Document level opinion mining is about classifying the overall opinion presented by the authors in the entire document text as positive, negative or neutral about a certain object [5] [6].The work done by Turney [5] on review classification presents an approach based on distance measure of adjectives found in text from preselected words with known polarity i.e. excellent or poor. The author presents a three step algorithm which processes the documents without user care. First step, the adjectives are extracted along with a word that provides contextual information. Second step, the semantic orientation is captured. This is done by measuring the distance from words of known polarity. The mutual dependence between two words is found by analysis of hit count with AltaVista search engine for documents that contain two words in certain proximity of each other. Third step, the algorithm counts the average semantic orientation for all word pairs and classifies a review as recommended or not.

In contrast, Pang et al. [3] present a work based on classic topic classification techniques. The proposed approach aims to test whether a selected group of machine learning algorithms can produce good result when opinion mining is perceived as document level, with two topics: positive and negative. Authors present the results using nave bayes, maximum entropy and support vector machine algorithms and the performed tests shown the good results as comparable to other ranging from 71 to 85% depending on the method and test data sets.

Apart from the document-level opinion mining, the next sub-section discusses the classification at the sentence-level, which classify each sentence as a subjective or objective sentence and determine the positive or negative opinion.

2.2 Sentence Level Opinion Mining

The sentence level opinion mining is an action that is associated with two tasks [7] [8] [9]. First task is to identify whether the given sentence is subjective (opinionated) or objective. The second task is to find opinion of a subjective sentence as positive, negative or neutral. Riloff and Wiebe [10] do the task of identifying subjective sentences through a method called bootstrap approach which uses high precision classifiers to extract a number of subjective sentences. Authors achieve around 90% accuracy during their tests. Yu and Hatzivassiloglou [13] discuss both sentence classification (subjective/objective) and orientation (positive/negative/neutral). For the sentence classification, author's present three different algorithms: (1) sentence similarity detection, (2) naïve Bayens classification and (3) multiple naïve Bayens classification. In the second step for opinion orientation as positive, negative and neutral, authors use a technique similar to the one used by Turney [5] for document level opinion mining. The technique takes the average of log-likelihood ratio (LLR) scores of seed words in sentence and use thresholds to decide whether it is positive, negative and neutral. Wilson et al. [17] pointed out that not only a single sentence may contain multiple opinions, but they also have both subjective and factual clauses. It is useful to pinpoint such clauses. It is also important to identify the strength of opinions.

Like the document-level opinion mining, the sentence-level opinion mining does not consider about object features that have been commented in a sentence. For this the feature level opinion mining is discuss in the next sub-section.

2.3 Feature Level Opinion Mining

The feature level of opinion mining is to not only determine the subjectivity and opinion of an object but also what author liked or disliked about the object [12] [14]. The feature level opinion mining is associated with following tasks: the commented object features are extracting, determine the opinion (positive, negative and neutral) of the object and then group the feature synonyms and produce the result. Hu and Liu do customer review analysis [19] through opinion mining based on feature frequency, in which the most frequent features is accepted by processing many reviews that are taken during summary generation. In opposite to Hu and Liu, Popescu and Etzioni [11], improved the frequency based approach by introducing the part-of relationship and remove the frequent occurring of noun phrases that may not be features.

2.4 Tools and Techniques

- **Sentence Delimiter [18]:** The given document is segmented into individual sentences by the help of sentence delimiter like question mark (?), dot or full stop (.). Sometimes identification of full stop in the sentence does not mark the end of sentence such as date ***12.1.2012,*** hence the rule based pattern matching could be used to identify sentence boundary.
- **Part of Speech Tagger [16]:** The part-of-speech tagger tool is used to assigns parts of speech to each word of a sentence such as noun, verb, adjective, etc. for example – happy plays hockey, here happy is noun not adjective hence this sentence is not opinionated.
- **Named Entity Recognition [16]:** It is also known as entity identification that's seeks to locate and classify atomic elements in text into predefined categories such as the names of persons, organizations locations, expression of times, percentages, names of object, its parts etc. It is used for noun phrase only.
- **Sentiword Net [15]:** It is a lexical resource in which each Word Net synsets is associated to three numerical scores Obj(s), Pos(s) and Neg(s), describing how objective, positive, and negative the terms contained in the synset are.

3 Conclusions

This paper presents an insight into task of opinion mining. Opinion mining aims at recognizing, classifying and determining opinion polarity of the opinionated text. Most of the approaches presume that the given text is opinionated. There are cases where objective sentence could be opinionated e.g., opinion drawn about a person in news articles column. Hence just identifying subjective sentence to have opinionated content is not enough. Sentence boundaries are not clearly defined on such blog sites. We find that the language used whether it is formal or informal in the text have direct impact on opinion mining as informal text are difficult to be processed i.e., syntactical, semantically or higher level analysis is tough. The task of opinion mining becomes difficult if there are opinion spams in text. Detecting fake opinions is a research issue. Some of the opinions are time sensitive hence identifying latest opinions on the subject is another research issue. We observe that opinion mining at three level i.e., document level, sentence level and feature level have got positive and negatives associated with them which are already discussed in section 2. The performances of Natural language processing tools also contribute towards the effective opinion mining.

References

1. Liu, B.: Opinion Mining and Summarization. In: World Wide Web Conference, Beijing, China (2008)
2. Liu, B.: Web Data Mining, Exploring Hyperlinks. Contents and Usage data (2007)
3. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up? Sentiment Classification using Machine Learning Techniques. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 79–86 (2002)

4. Dave, K., Lawrence, S., Pennock, D.M.: Mining the peanut gallery: Opinion extraction and segment classification of product reviews. In: Proceedings of WWW, pp. 519–528 (2003)
5. Turney, P.: Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In: Proceeding of Association for Computational Linguistics, pp. 417–424 (2002)
6. Das, St., Chen, M.: Yahoo! for Amazon: Extracting market sentiment from stock message boards. In: Proceedings of the Asia Pacific Finance Association Annual Conference (2001)
7. Hatzivassiloglou, V., Wiebe, J.: Effects of adjective orientation and gradability on sentence subjectivity. In: Proceedings of the International Conference on Computational Linguistics (2000)
8. Riloff, E., Patwardhan, S., Wiebe, J.: Feature subsumption for opinion analysis. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing (2006)
9. Riloff, E., Wiebe, J.: Learning extraction patterns for subjective expressions. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing (2003)
10. Wiebe, J., Bruce, R.F., O'Hara, T.P.: Development and use of a gold standard data set for subjectivity classifications. In: Proceedings of the Association for Computational Linguistics, pp. 246–253 (1999)
11. Popescu, Etzioni.: Extracting product features and opinions from reviews. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing (2005)
12. Wilson, T., Wiebe, J., Hoffmann, P.: Recognizing contextual polarity in phrase-level sentiment analysis. In: Proceedings of the Human Language Technology Conference and the Conference on Empirical Methods in Natural Language Processing, pp. 347–354 (2005)
13. Yu, H., Hatzivassiloglou, V.: Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing (2003)
14. Bethard, S., Yu, H., Thornton, A., Hatzivassiloglou, V., Jurafsky, D.: Automatic extraction of opinion propositions and their holders. In: Proceedings of the Association for the Advancement of Artificial Intelligence Spring Symposium on Exploring Attitude and Affect in Text (2004)
15. Cardie, C., Wiebe, J., Wilson, T., Litman, D.: Combining low-level and summary representations of opinions for multi-perspective question answering. In: Proceedings of the Association for the Advancement of Artificial Intelligence Spring Symposium on New Directions in Question Answering, pp. 20–27 (2003)
16. Jin, W., Hay Ho, H., Srihari, R.: Opinion Miner: A Novel Machine Learning System for Web Opinion Mining and Extraction. In: Proceeding of International Conference on Knowledge Discovery and Data Mining, Paris, France (2009)
17. Wilson, T., Wiebe, J., Hwa, R.: Just how mad are you? Finding strong and weak opinion clauses. In: The Association for the Advancement of Artificial Intelligence, pp. 761–769 (2004)
18. Dey, L., Mirajul Haque, S.K.: Studying the effects of noisy text on text mining applications. In: Proceedings of the Third Workshop on Analytics for Noisy Unstructured Text Data, Barcelona, Spain (2009)
19. Hu, M., Liu, B.: Mining and Summarizing Customer Reviews. In: Proceedings of the Tenth International Conference on Knowledge Discovery and Data Mining, Seattle, WA, USA (2004)