

Feature-Based Opinion Mining Using Ontological Resources

Dim En Nyaung
University of Computer Studies, Yangon
dimennyaung@gmail.com

Abstract

One of the important types of information on the Web is the opinions expressed in the user generated content, e.g., customer reviews of products, forum posts, and blogs. Customer reviews of products are focused in this paper. Mining opinion data that reside in web is a way to track opinions of people on specific product. Opinion mining is a recent subdiscipline of computational linguistics which is concerned not with the topic a document is about, but with the opinion it expresses. To aid the extraction of opinions from text, recent work has tackled the issue of determining the orientation of subjective terms contained in text, i.e. deciding whether a term that carries opinionated content has a positive or a negative connotation. In this paper the task of deciding whether a given term has a positive connotation, or a negative connotation by using feature-based opinion mining with ontology where opinions expressed towards each feature of an object or a product are extracted and summarized. In this context, the goal is to study the role of domain ontology to structure and extract object features as well as to produce a comprehensive summary.

1. Introduction

With the dramatic growth of web's popularity, the number of freely available online reviews is increasing at a high speed. Merchants selling products on the Web often ask their customers to review the products that they have purchased and the associated services. As e-commerce is becoming more and more popular, the number of customer reviews that a product receives grows rapidly [3].

Therefore, opinion mining is a growing research area both in natural language processing and information retrieval communities as it aims at finding subjective information, which may be more relevant to users than factual information in many applications. Companies, politicians, as well as customers need powerful tools to track opinions, sentiments, judgments and beliefs that people may express in blogs, reviews, audios and videos data regarding a product/ service/ person/ organization/ etc., [1]. A significant number of websites, blogs and forums allow customers to post reviews for

various products or services (e.g., amazon.com). Such reviews are valuable resources to help the potential customers make their purchase decisions. In the past few years, mining the opinions expressed in web reviews attracts extensive researches [2, 10]. Based on a collection of customer reviews, the task of opinion mining is to extract customers' opinions and predict the sentiment orientation. The aim is not to compute the general orientation of a document or a sentence, since a positive sentiment towards an object does not imply a positive sentiment towards all the aspects of this object [11], as in: The picture quality is good, but the battery life is short.

In feature-based opinion mining, the task goes to the sentence level, i.e., what aspects of an object that people liked or disliked. The object could be a product, a service, a topic, an individual, an organization, etc. For example, in a product review, this task identifies product features that have been commented on by reviewers and determines whether the comments are positive or negative. In the sentence, "the battery life of this camera is too short," the comment is on the "battery life" and the opinion is negative. As defined in [3], a feature can be a "part-of" of a topic (such as the screen of a camera) or a property of the "part-of" of the topic (such as the size of the screen).

This paper is to study the role of domain ontology in feature-based opinion mining. The context is to study how domain ontology can be used to:

- *structure features*: an ontology is more suitable than a simple hierarchy where features are grouped using only the "is-a" relation [4, 5]
- *extract explicit and implicit features from texts*: the lexical component as well as the set of properties of the ontology can help to extract, for each feature, the set of the associated opinion expressions.
- *produce a discourse based summary of the review*: the ontology can guide the process of identifying the most relevant discourse relations that may hold between elementary discourse unit.

2. Related Work

Since the research on opinion retrieval is relatively new, the state-of-the-art opinion retrieval techniques on the Association for Computing

Machinery (ACM) portal and Google Scholar are identified in early 2011 [6]. Identified techniques were then classified under text classification approach, lexicon-based approach, probabilistic approach, and other emerging approaches.

Most of the current opinion mining work mostly focuses on mining product review data [1], because of the wide availability of review data and their relatively obvious sentiment orientations such as good, bad and so on. The opinion words are extracted using the resulting frequent features, and semantic orientations of the opinion words are identified with the help of WordNet [8]. WordNet can be interpreted and used as a lexical ontology. The orientation of each opinion sentence is identified and a final summary is produced. POS tagging is the part-of-speech tagging [3] from natural language processing, which helps us to find opinion features. Then produce a structured summary that informs about positive or negative statements for product features.

Su et al [7] proposed two opinion mining approaches, namely feature-based approach and similarity-based approach. The feature-based approach incorporates computational features at punctuation-, word-, collocation-, phrase-, sentence-, paragraph- and document-level in a coarse fine multi-pass classification framework. The similarity-based approach estimates the similarity between the example sentences and testing sentence and identifies the similar example sentence testing sentence pair.

Hu et al [3] proposed the idea of opinion mining and summarization. It uses a lexicon-based method to determine whether the opinion expressed on a product feature is positive or negative.

In this paper, ontological resources for feature-based opinion mining are proposed.

3. An Ontology Based Opinion Mining

Ontologies have been widely used in a variety of natural language applications. Ontologies describing similar domain information varied significantly in syntax and semantics depending on the nature of the ontology language used. Hence, ontologies written in different languages needed to be modified and refined in order to get useful ontological data. The important for NLP systems is not only to get an accurate opinion in texts but also to go beyond explicit features and to propose a fine-grained analysis of opinions expressed towards each feature. The works using ontology aim at organizing features using a model of representation: ontology. The use of ontology would have several advantages in the domain of opinion mining to:

Structure features: Ontologies are tools that provide a lot of semantic information. They help to define concepts, relationships and entities that describe a domain with unlimited number of terms. This set of terms can be a significant and valuable lexical resource for extracting explicit and implicit features.

Extract features: Ontologies provide structure for these features through their concept hierarchy but also their ability to define many relations linking these concepts. This is also a valuable resource for structuring the knowledge obtained during feature extraction task. In addition, the relations between concepts and lexical information can be used to extract implicit features.

4. Featured-Based Opinion Mining System Using Ontology

The feature-based opinion mining system needs three basic components: a lexical resource L of opinion expressions, a lexical ontology O where each concept and each property is associated to a set of labels that correspond to their linguistic realizations and a review R . Having, for an object/product, the set of its associated features $F=\{f_1, \dots, f_n\}$, research in feature-based opinion mining focus on extracting the set F from reviews for each feature f_i of F , extract the set of its associated opinion expressions $OE=\{OE_1, \dots, OE_j\}$. Once the set of couples (f_i, OE) were extracted, a summary of the review is generally produced.

The extracted features correspond exclusively to terms contained in the ontology. The feature extraction phase is guided by domain ontology, build manually or semi-automatically, which is then enriched by an automatic process of extraction/clustering of terms which corresponds to new feature identification [9]. Same features are grouped together using semantic similarity measures. New features are added to their ontology concepts using a corpus based method where sentences contains a combination of conjunction word and already recognized concept are extracted. This process is repeated iteratively until no new concepts are found. Ontologies have also been used to support polarity mining. For example, ontology that is manually built for camera reviews and then incorporated it into the polarity classification task which significantly improves performance over standard baseline.

A review R is composed of a set of elementary discourse units (EDU). An EDU is a clause containing at least one elementary opinion unit (EOU) or a sequence of clauses that expressing an opinion. An EOU is an explicit opinion expression composed of a noun, an adjective or a verb with its

possible modifiers (actually negation and adverb) as described in the lexicon L. The lexicon contains a list of opinion terms where each lexical entry is of the form: [POS, opinion category, polarity, strength] where POS is the part of speech tagging of the term, opinion category can be a judgment or a sentiment, polarity and strength corresponds to the opinion orientation (positive, negative and neutral) and the opinion strength (a score between 0 and 2). For example, the term for ‘good’: [Adj, judgment, +, 1].

Ontology: For the camera domain, a pre-existent ontology is used as a basis coupled with additional information that required information is gathered from several web sites. It is adapted to this application by manually reorganize, add and delete concepts in order to describe important camera features. Our domain ontology actually contains 202 concepts, 31 object properties and 503 labels.

In a review R, an opinion holder h comments on a subset S of the features of an object/product using some opinion expressions. For example, in the following product review, EDUs are between square brackets, EOUs are between embraces whereas object features are underlined.

[I bought the camera yesterday]_a. [Even if the camera is {excellent} in picture quality]_b, [the design and the size are {very basic}]_c [which is {upsetting} in this brand]_d.

The figure below gives an overview of the proposed system.

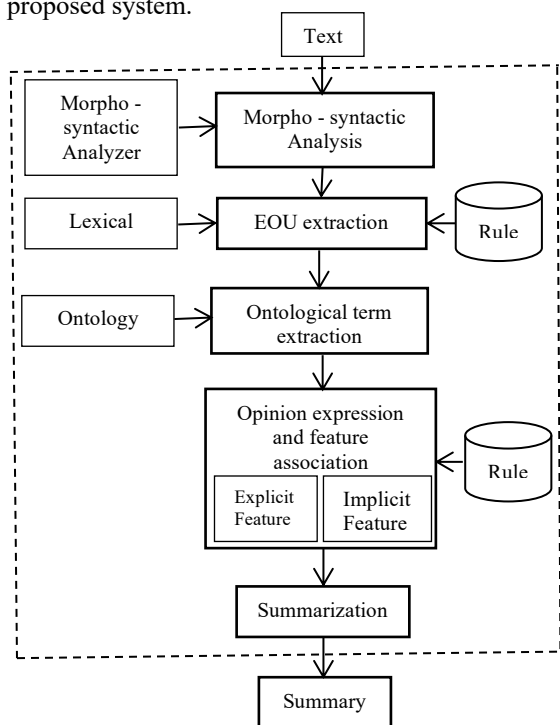


Figure 1. Overview of the proposed system

First, each review R is parsed using the Morpho-syntactic analyzer. To perform part-of-speech (POS) tagging as many words can have multiple POS tags depending on their usages. The part-of-speech of a word is a linguistic category that is defined by its syntactic or morphological behavior. It determines the right part-of-speech tagging each word belongs to: is it a verb, an adjective, a noun, preposition, etc., and the set of dependency relations. The review is then segmented in EDUs using the discourse parser. The segmented conjoined phrases can be separated into clauses. The segmented are connected to each other using a small subset of “veridical” discourse relations, namely:

- *Contrast (a, b)* implies that (a) and (b) are both true but there is some defeasible implication of one that is contradicted by the other. Possible markers can be although, but.
- *Result (a, b)* indicated by markers like so, as a result, indicates that the EDU_b is a consequence or result of the EDU_a.
- Continuation (a, b) corresponds to a series of speeches in which there are no time constraints and where segments form part of a larger thematic.
- Elaboration (a, b) describes global information that was stated previously with more specific information.

For each EDU, the system:

1. Extracts EOUs using a rule based approach
2. Extracts features that correspond to the process of term extraction using the domain ontology
3. Associates, for each feature within an EDU, the set of opinion expressions
4. Produces a summary based on the information.

4.1 Extracting Elementary Opinion Units

EOU composed of one and only one opinion word (a noun, an adjective or a verb) possibly associated with some modifiers like negation words and adverbs. EOU is the smallest opinion unit within an EDU. For example, “really not good” is an EOU. An EOU can also be simply an adverb as in less productive. Adverbs are also used to update the opinion lexicon, as in less productive where the opinion word productive is added.

4.2 Extracting Features

This step aims at extracting for the review all the labels of the ontology. This step also involves extracting explicit and implicit features from the customer reviews. Since each concept and its associated lexical realizations correspond to explicit features, the lexical component of the

ontology in the review can be projected in order to get, for each EDU, the set of features F. The linking features to opinion expressions can partially solve that the lexical ontology does not cover all the linguistic realizations of concepts and properties in a given domain.

To extract implicit features, ontology properties are used. For example, the property “good at” links “camera” and “picture quality” concepts.

4.3 Associating opinions expressions to extracted features

In this step, the extracted opinion expressions in step 1 have to be linked to the features extracted in step 2 i.e. to associate to each EDU_i the set of couples (f_i, OE_i). During this step, the following cases can be distinguished:

Case 1: Known features and known opinion words
For example, if the lexicon contains the words really, good and excellent and the ontology contains the terms used camera and picture quality as a linguistic realization of the concepts camera and picture quality, then this step allows the extraction from the EDU “really good camera with excellent picture quality” the couples (camera, really good) and (picture quality, excellent).

Case 2: Known features and unknown opinion expressions, the opinion lexicon can be automatically updated with the retrieved opinion word.

Case 3: Unknown features and known opinion expressions, the domain ontology can be updated property or by adding a new concept or a new by adding a new feature to an existing concept or property in the right place to the ontology. However, since a user may express an opinion on different objects within a review, need to be done carefully. So, the ontology should be manually updated to avoid error.

Case 4: Opinion expressions alone, in order to retrieve the associated concept, the ontology properties in the ontology are used. This kind of EDU expresses the implicit features.

Case 5: Features alone, an EDU with features alone can also be an indicator of the presence of an implicit opinion expression.

4.4 Production of the Summary

After all the previous steps, generating the feature-based review summary is the final stage,

which is straightforward and consists of the following steps:

- For each association opinion expression to extracted features are put into positive and negative categories according to the opinion expression. A count is computed to show how many reviews give positive/negative opinions to the feature.
- All features are ranked according to the frequency of their appearances in the reviews. Any types of rankings can be used. For example, ranking features according to the number of reviews that express positive or negative opinions.

The following example shows the summary for the feature “picture quality” of a camera.

Feature: picture quality

Positive: 8

- Overall this is a good camera with really good picture clarity.
- The pictures are absolutely amazing - the camera captures the minutest of details.
- The pictures are sharp – the resolution of the camera is excellent.

...

Negative: 3

- The pictures come out hazy if your hands shake even for a moment during the entire process of taking a picture.
- This camera is not easy to carry and use, the design is very basic.
- The picture quality is poor. Pictures produced by this camera were blurry and in a shade of orange when focusing on a display rack about over 20 feet away from the object.

5. Experimental Evaluation

There are three types of experiment: the evaluation of the extraction of elementary opinion units, the evaluation of the features extraction step and finally, the evaluation of the link between the retrieved opinion expressions and the retrieved object features.

Evaluation of the EOU extraction step:

The system missed some EOU for two reasons. The first one is due to missed opinion words in the lexicon and to implicit opinion expressions. The second reason is the errors that come from the syntactic parser because of typos and dependency link errors.

Table1. Evaluation of the EOU Extraction

Precision	0,5486
Recall	0,6535
F-measure	0,7976

Evaluation of the features extraction step:

Since the corpus is in the camera domain, the precision of this task is 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 do not totally cover the terms of the corpus.

<Start> This (camera) is {excellent} in (picture quality) and it is also {easy} to use. It would be {nicer} to see {little} (zoom) sign on the side <Stop>
--

Figure 2. Result of EOU (embraces) and ontological terms (parentheses) extraction

Evaluation of the link between EOU and features:

The system is able to extract opinion expressions which do not contain words present in the lexicon. It is the case with “sharp” which has been correctly associated to “lens” and “image” even if the word “sharp” was not in the lexicon.

Table2. Evaluation of the link between EOU and features

Precision	0,5692
Recall	0,5733
F-measure	0,5712

6. Conclusion

In this paper a technique for mining and summarizing product by using ontology is proposed. The objective is to provide a feature-based summary of a large number of customer reviews of a product sold online. This problem will become increasingly important as more people are buying and expressing their opinions on the Web. Summarizing the reviews is not only useful to common customers, but also crucial to product manufacturers. The proposed techniques are very promising in performing their tasks because the use of the ontology allows improving the feature extraction and the association between an opinion expressions and object features. Moreover, the ontology is useful thanks to its list of properties between concepts which allows recognizing some opinions expressed about implicit features.

References

- [1] Y. Fang, L. Si, N. Somasundaram and Z. Yu, “Mining Contrastive Opinions on Political Texts using Cross-Perspective Topic Model”. WSDM’12, Seattle, Washington, USA, February 8–12, 2012.
- [2] Q. Su, X. Xu, H. Guo, Z. Guo, XianWu, X. Zhang, B. Swen and Z. Su, “Hidden Sentiment Association in

Chinese Web Opinion Mining”. WWW 2008, Beijing, China, April 21–25, 2008.

- [3] M. Hu and B. Liu, “Mining and Summarizing Customer Reviews”. KDD’04, Seattle, Washington, USA, August 22–25, 2004.

- [4] Carenini, Giuseppe, Raymond T. Ng and Ed Zwart, “Extracting Knowledge from Evaluative Text”. In Proc of the 3rd international conference on Knowledge capture, 2005.

- [5] BlairGoldensohn, Sasha, Kerry Hannan, Ryan McDonald, Tyler Neylon, George A. Reis and Jeff Reynar, “Building a Sentiment Summarizer for Local Service Reviews”. WWW2008 Workshop : Natural Language Processing Challenges in the Information Explosion Era (NLPIX 2008), 2008.

- [6] SYLVESTER O. ORIMAYE, SAADAT M. ALHASHMI and SIEW EU-GENE, “Performance and Trends in Recent Opinion RetrievalTechniques”. The Knowledge Engineering Review, Vol. 00:0, 1-28, Cambridge University Press, 2012.

- [7] R. Xu and C. Kit, “Incorporating Feature-based and Similarity-based Opinion Mining – CTL in NTCIR-8 MOAT”. Proceedings of NTCIR-8 Workshop Meeting, Tokyo, Japan, June 15–18, 2010.

- [8] A. Esuli and F. Sebastiani, “Determining Term Subjectivity and Term Orientation for Opinion Mining”. Padova, Italy.

- [9] X. Ding , B. Liu, Philip S. Yu, “A Holistic Lexicon-Based Approach to Opinion Mining”. WSDM’08, Palo Alto, California, USA, February 11-12, 2008.

- [10] A. Bakliwal, P. Arora, A. Patil, V. Varma, “Towards Enhanced Opinion Classification using NLP Techniques”. Proceedings of the Workshop on Sentiment Analysis where AI meets Psychology (SAAIP), IJCNLP 2011, Chiang Mai, Thailand, November 13, 2011, pages 101–107.

- [11] Alexandru Lucian Gînscă, “Fine-Grained Opinion Mining as a RelationClassification Problem”. Imperial College Computing Student Workshop (ICCSW’12), 2012