

An Opinion Analysis Model for Implicit Aspect Expressions based on Semantic Ontology

Lirong Qiu

*School of Information Engineering, Minzu University of China
Beijing, China
E-mail: qiu_lirong@126.com*

Abstract

Analyzing user opinion and building features model of entities is very important for sentiment analysis and opinion mining, which is the fundamental work for social network analysis. Opinions of entity can be classified into two types: explicit aspect expressions and implicit aspect impressions, and implicit aspect expressions might have some relationships which can be denoted by domain specific knowledge, existing previous works have not addressed in much detail. This paper defines an opinion mining model for implicit aspect expression based on semantic ontology, and proposes semantic similarity estimation approaches of entity by the domain ontology.

Keywords: *aspect-based opinion, opinion analysis, semantic ontology, sentiment analysis*

1. Introduction

With the development of reviews, blogs and such social media on the Web, individuals and organizations are increasingly using and are easily affected by the opinions in these media for decision making [1]. How to process the large amount of information and to produce consistent results of people's opinion is an active research area in natural language processing and Web mining. Opinion mining or sentiment analysis is the computational study of people's opinions, appraisals, attitudes, and emotions toward entities such as products, services, organizations, individuals, events and their different aspects [2]. Opinion mining or sentiment analysis system is needed, which can overcome mental and physical limitations of human beings, such as avoiding subjective biases.

Opinions can be expressed about many entities, such as products, services, organizations, individuals, events and so on. An opinion is a positive or negative view about an entity from an opinion sources, which is often stated a person or organization that holds the opinion. Each entity has many aspects or features. And people often have positive opinions on some aspects of the entity and may have negative opinions on other aspects of the same entity. For example, in the sentence "*The food is good in that restaurant, but the service is very poor.*" The entity is the "*restaurant*", and the feature are "*food*" and "*service*". The opinion of the entity is quite different according the different features or aspects.

Feature-based opinion mining is to extract and summarize people's opinions expressed on entities and aspects of entities. Specially, the opinion of the entity can be classified into two types: explicit aspect expressions and implicit aspect impressions. For example, "*image quality*" in "*The image quality of this camera is very good.*" is an explicit aspect expression, and "*heavy*" is an implicit aspect expression in "*This camera is too heavy*", which implies the aspect "*weight*". Many implicit aspect expressions imply some specific comments of the entity, e.g., expensive (price), and large (size), which can be quite complex.

When we look into the details of the examples of implicit aspect expressions, we find

that there are some intrinsic properties that existing previous works have not addressed in much detail. First of all, implicit aspect expressions might have some relationships which can be denoted by domain specific knowledge. For example, for a cell phone, comments on voice quality are usually mentioned. However, a sentence like “*iphone is lousy*”, also refers to voice quality of the cell phone “*iphone*”. Secondly, vocabularies used in the same kind of entities tend to be highly overlapping, especially the synonyms of the same attributes. Thirdly, to identify the relation between opinion word and aspects can be a word alignment task with the help of domain knowledge.

We argue that when performing opinion analysis, more attention is needed to distinguish between attributes that are mentioned with and without explicit aspects. In this paper, we study the problem of the aspect-based opinion analysis, especially the implicit opinions, based on semantic ontology. By semantic ontology based opinion analysis, we aim to fulfill two tasks: 1) identify some common opinion words in target domain and 2) the associations between aspects and opinion words are measured based on domain ontology.

The remainder of the paper is organized as follows. In Section 2, we provide an overview of related work on sentiment analysis. Section 3 presents our work on aspect-based opinion analysis with semantic ontology approach. The empirical analysis and the results are presented in Section 4, followed by the conclusions, discussions, and future work in Section 5.

2. Related Work

Aspect-based opinion mining aims to obtain detailed information about opinions, which composed of two core tasks: aspect extraction and entity extraction [2], which are ongoing research problems, with the vast majority of existing work done in the NLP and Web mining domain.

More recently, there have been some important works concerning feature-based opinion mining. Qiu et al. utilized domain-independent and unsupervised approach to identify the relationships between aspects and opinion words to perform aspect extraction in the paper [8]. For aspect extraction, existing solutions also comprises sequence learning algorithms such as Hidden Markov Models and Conditional Random Field with help of manual labeling effort [9,10].

Semantic rule based sentiment analysis technology essentially belongs to unsupervised learning algorithm, Turney proposed its typical algorithm SO-PMI in 2002 [16]. The algorithm only select “excellent” and “poor” as the standard word to judge positive or negative property of the comments. Calculate the distance between the comment word and above two sentimental standard words based on point wise mutual information, thus complete the sentiment classification procedure.

SO-PMI algorithm uses point wise mutual information to measure the distance between comment word and seed word. The fundamental principle is, if the PMI value of comment word and seed word gets bigger, the probability of comment word and seed word appearing in the same time is higher, thus both words tend to have same sentiment polarity.

There are some existing works that tried to find implicit aspects indicated by adjectives [4,7]. In [5], a dictionary-based approach was proposed by studying the problem of mining attribute nouns of opinion adjectives.

In [3], sentiment analysis on product reviews are studied by a method named hierarchical learning with sentiment ontology tree.

There are also some works tried to improve feature-based opinion mining by employing ontologies in the selection of features, for example, paper [11] aimed to improve feature-based opinion mining by employing ontologies in the selection of features and to provide a new method for sentimental analysis based on vector analysis.

And paper [12,13] mainly focus on how to calculate the polarity by taking into consideration the features of an entity. Paper [14] describes a sentiment analysis system based on ontology and Paper [15] utilize ontologies at the feature selection stage and provide a new vector analysis-based method for sentiment analysis by using SentiWordNet.

Given the great deal of similar work in opinion mining and a variety of methods have developed for knowledge management based on ontology, but less of them are focus on implicit opinion analysis.

Our work mainly focuses on obtaining the implicit aspect of the entity based on ontologies to improve the results of aspect extraction obtained in those previous works.

3. Implicit Opinion Analysis Model based on Semantic Ontology

From the perspective of part of speech, noun are usually used to represent the entity or its attributes, while adjectives and adverbs are typically used to express the user's emotional views, and this means that evaluation of words are formed from adjective or adverb words. If we can have access to all the emotional polarity of evaluation word, it's easy to figure out the author's emotional tendencies. For example, in the following comment "The iPhone mobile phone looks beautiful, trendy, personality and fashion", we can see that this author use some positive word to express his point of view like "beautiful, trendy, personality and fashion", so it is easy to determine the emotional bias of the author is positive.

This method is based on an emotional dictionary with words that have been marked with emotional tendency, and we call this method as the method based on emotional dictionary.

Methods based on emotional dictionary can preliminary judgment text emotional tendencies, but this kind of method is not applicable to all cases. This is because none of emotional dictionaries can contain all of the evaluation words, and part of emotional words have different emotional polarity in different contexts.

An ontology is a formal naming and definition of the types, properties, and interrelationships of the entities that really or fundamentally exist for a particular domain [14]. The implicit opinion analysis model based on semantic ontology method is made out of the following considerations:

First, the entities of opinions are mostly noun phrases and the entity phrases may have some other names that can be defined as semantic items in ontologies, which can bear or express opinions.

Second, the ontology should be directly related to an opinion expression, comprised both the entities and aspects.

Third, the implicit opinion may be lack of the exact expression of aspects and only have the entity expressions and opinions.

An opinion is connected with some information, such as opinion holder, object entity, entity aspect, opinion orientation, and the time when opinion is expressed by the holder. Opinion mining technique based on semantic rules which are based on the semantic ontologies, can calculate the distance between evaluation words and the seed words in the entities to achieve the goal of opinion analysis.

3.1. Model Formalization

Based on those considerations, the model should have the ability to discover all opinions based on ontology given a collection of opinionated documents. Inspired by the previous work cited in [2], we gave the formally definitions of opinion analysis model as follows:

Definition (Opinion): Opinion is correlated with a quintuple (E,A,Op,H,O) , and $(e_i,a_{ij},oo_{ij},h_k,o_i)$ is an explicit definition, where $e_i \in E$ is the name of an entity, a_{ij} is an

aspect of e_i , Op is the set of expressions of opinion orientation, $h_k \in H$ is the opinion holder and o_i is the domain ontology related to the domain of the opinion entity.

Given a collection of opinionated documents D , the objective of opinion mining is to discover all opinions based on ontology, and record all the quintuples $(e_i, a_{ij}, oo_{ij}, h_k, o_i)$ in D .

The main task of our work is consists of two sub-tasks: (1) identify some common opinion words as seeds in the ontology; (2) a bootstrapping method is employed to expand the seeds.

An entity can be a product, service, person, event, organization, or topic and also has some expressions and other names that refer to the same entity. For example, the brand *Motorola* (entity name) can be expressed in several ways, e.g., “*Moto*”, “*Mot*” and “*Motorola*” itself. As cited in paper [14], in the smartphone ontology, the “interface” attribute could also be expressed as “monitor”, “screen” or “display”, which are synonymous words.

For example, the aspect ontology of smartphone entity is shown in Figure 1. The domain ontology should define the alias and the semantic relation between the words.

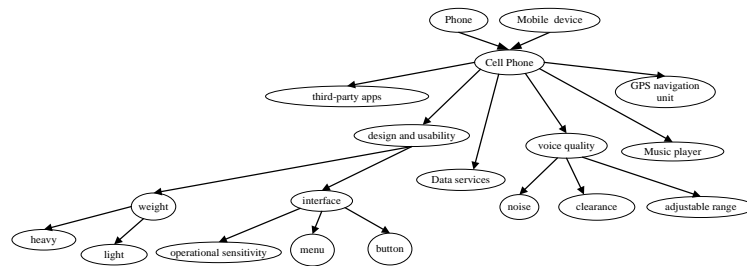


Figure 1. The Aspect Ontology of Cell Phone Entity

Definition (Entity): Entity set $E = \{e_1, e_2, \dots, e_m\}$, and $\forall e_i \in E$ can be expressed with any one of the expressions $\{e_{i1}, e_{i2}, \dots, e_{in}\}$, which is defined in ontology O_e .

Definition (Aspect): To the same kind of entity set E , the set of the aspects is finite $A_i = \{A_{i1}, A_{i2}, \dots, A_{in}\}$, and each aspect $a_{ij} \in A_{ii}$ of the entity also can be expressed by any different expressions $AE_{ij} = \{ae_{ij1}, ae_{ij2}, \dots, ae_{ijm}\}$, which is defined in ontology O_a .

Definition (Ontology): Ontology O is domain ontology set of different types for opinion analysis $O = \{O_e, O_a, Op, O_h\}$. The Ontology O is a combination of an upper and some domain ontologies.

This paper does not indent to discuss how to create the domain ontology, so we assume there are some entity ontologies and can be integrated by means of ontology mapping and ontology learning technologies.

For appending the sets of synonyms and hyponyms to the ontology, the ontology learning approach is adopted, such as OntoGen, which is a semi-automatic, data-driven ontology editor.

3.2. Semantic Labeling Based On Ontology

$D = \{d_1, d_2, \dots, d_n\}$ is a weighted documents set, and d_i ($1 \leq i \leq n$) is the i th record of the opinion document. $E = \{e_1, e_2, \dots, e_m\}$ is the entity items set, and e_j ($1 \leq j \leq m$) denotes the entity set and $w[d_i][e_j]$ ($1 \leq i \leq n, 1 \leq j \leq m$) is the weight of e_j in d_i .

Given the entity e_j , we need to select the ontology o_i , the likelihood of entity e_j co-occurring with the ontology o_i is given by:

$$p(e_j, o_i) = p(e_j | o_i) p(o_i)$$

$$p(c_k | o_i) = \sum_{p_i: domain(o_i)} sim(c_k, p_i)$$

The likelihood of a concept c_k occurring with the ontology and o_i and entity e_j is given by:

$$p(c_k|o_i, e_j) = \frac{p(e_j|c_k)p(c_k|o_i)}{\sum_{l=1}^K p(e_j|c_l)p(c_l|o_i)}$$

And the likelihood of a concept c_k occurring with the entity e_j is given by:

$$p(e_j|c_k) = \frac{\sum_{i=1}^N n(o_i, e_j)p(c_k|o_i, e_j)}{\sum_{m=1}^M \sum_{i=1}^N n(o_i, e_m)p(c_k|o_i, e_m)}$$

$$p(e_j|o_i) = \sum_{k=1}^K p(c_k|o_i)p(e_j|c_k)$$

The occurring value of concept c_k and candidate entity e_j is given by:

$$occ(c_k, e_j) = \sum_{c_k \in o_i} n(o_i, c_k) \log \sum_{l=1}^K p(c_k|o_l)p(e_j|c_k)$$

Given a concept c_k defined in an ontology o_i , a set of weighted candidate entities e_j can be selected by defining a threshold constraints.

For a given set of candidate entity e_j , the aggregate likelihood candidate implicit aspect is given by:

$$entity\text{-}aspect(a_{ij}, e_j) = \sum_{o_i} \sum_{e_j} (occ(c_k, e_j) \times \gamma)$$

$$\gamma = p(c_k|o_i, e_j)$$

The *entity-aspect* weights of a_{ij} and e_j are normalized to values in [0,1]. So far we have presented a way of generating a number of weighted aspects a_{ij} for an entity e_j .

4. Use Case Scenario

4.1. Baseline Scenario

Describe a baseline scenario that better illustrates the usability of the proposed approach. For demonstration purposed, Let us suppose an example of the evaluation on iPhone from Weixin (like twitter in Chinese). The smartphone ontology which is available at <http://www.productontology.org/id/Smartphone>, has been populated using the ontoGen cited in [14].

For example:

User A: (1) I bought *iPhone 6 plus* yesterday. (2) It has *fashion* and *trendy appearance*, *large size* and *high-resolution screen*, *long battery life* and *high quality photography*, I like it very much. (3) But my friend B thought it had *high price*, even though it was beautiful. (November 20th, 2014)

In the example, sentence (1) is an objective descriptive sentence, which shows the fact of buying the phone without any sentimental tendency. Sentence (2) is a positive comment on iPhone from user A. Sentence (3) is a negative comment from user B. The commenting time is the time that the users hold the corresponding comments.

From the traditional opinion mining technology, the two opinions will be extracted as follows:

The opinion 1: <iPhone, positive, user A, November 20th, 2014>

The Opinion 2: <iPhone, negative, user B, November 20th, 2014>

It is worth noting that the entity words are usually used to express the several aspect of the entity and belongs to different domain in different context condition. In fact, in the example, there are some aspects in the opinion. Especially user B, he has a positive attitude to the appearance of the entity, although has a low evaluation of the price.

4.2. Evaluation

As already explained, the basic idea behind the proposed approach is to take advantage of ontology for providing more elaborate implicit aspects of the entity. The proposed methodology is divided into two phases: (1) find the ontology regarding the given entity, and (2) extract the implicit aspects of the entity, based on the concepts and properties included in the ontology.

The method need to be input both a corpus of opinions and the domain ontology.

Step1. Extract the comment word set E from the document to be analyzed. The extraction adopts the algorithm of part of speech tagging, mainly chooses noun and adjective as entity aspects and opinion words.

Given the example above, the words in this dictionary are labeled in italics.

Step 2. Select “*iphone*” and corresponding ontology as input and analyze semantic relations based on pointwise mutual information. The aspect of opinions are abstracted based on ontology technology, described in section 3.

Opinion 1 with different aspects: <*iPhone, appearance, fashion and trendy*>, <*iPhone, screen, large size*>, <*iPhone, battery life, long*>, <*iPhone, high quality, photography*>.

Opinion 2 with different aspects: <*iPhone, appearance, beautiful*>, <*iPhone, price, high*>.

Step3. Calculate the average semantic orientation of sentences based on SO-PMI algorithm.

5. Conclusion and Future Work

Opinions of entity can be classified into two types: explicit aspect expressions and implicit aspect impressions, and implicit aspect expressions might have some relationships which can be denoted by domain specific knowledge, existing previous works have not addressed in much detail.

This paper addressed the problem of detecting and semantically typing the implicit aspect expression based on semantic ontology, and proposed a semantic similarity estimation approach of entities by the domain ontology.

Directions for future work include trying to increase recall by identifying relationships between aspects and entities that defined in ontology, and identifying multiple aspect expressions of the same entity.

Aspect expressions are usually nouns and noun phrases, but can also be verbs. Future work will also include more precise and accurate definitions as well as a complete description of the algorithms of the model.

Acknowledgements

Our work is supported by Beijing Natural Science Foundation (4153062) , the 985 special funds in School of information engineering, Minzu university of China and the Program for New Century Excellent Talents in University (NCET-12-0579).

References

- [1] B. Pang and L. Lee, “Opinion mining and sentiment analysis”, *Foundations and Trends in Information Retrieval*, vol. 2, no 1, (2008), pp. 1-135.
- [2] L. Zhang and B. Liu, “Aspect and Entity Extraction for Opinion Mining”, *Data Mining and Knowledge Discovery for Big Data*, vol. 1, (2014), pp. 1-40.
- [3] W. Wei and J. A. Gulla, “Sentiment learning on product reviews via sentiment ontology tree”, In *Proceedings of Annual Meeting of the Association for Computational Linguistics ACL* (2010).
- [4] Z. Hai, K. Chang and J. Kim, “Implicit feature identification via co-occurrence association rule mining”, *Computational Linguistics and Intelligent Text Processing*, (2011), pp. 393-404.
- [5] G. Fei, B. Liu, M. Hsu, M. Castellanos and R. Ghosh, “A dictionary-based approach to identifying aspects implied by adjectives for opinion mining”, In *Proceedings of International Conference on Computational Linguistics*, (2012).
- [6] Y. Choi, C. Cardie, E. Riloff and S. Patwardhan., “Identifying Sources of Opinions with Conditional Random Fields and Extraction Patterns”, In *proceedings of the Human Language Technology Conference and the Conference on Empirical Methods in Natural Language Processing*, (2005).

- [7] Q. Su, X. Xu, H. Guo, Z. Guo, X. Wu, X. Zhang, B. Swen and Z. Su, “Hidden sentiment association in Chinese web opinion mining”, Proceedings of International Conference on World Wide Web (2008).
- [8] G. Qiu, B. Liu, J. Bu and C. Chen, “Opinion word expansion and target extraction through double propagation”, Computational Linguistics, (2011). [11] P. Isidro, V. Rafael and G. Francisco, “Ontology-Guided Approach to Feature-Based Opinion Mining”, Natural Language Processing and Information Systems Lecture Notes in Computer Science, vol. 6716 (2011), pp. 193-200.
- [9] W. Jin and H. Ho, “A novel lexicalized HMM-based learning framework for web opinion mining”, In proceedings of International Conference on Machine Learning, (2009).
- [10] N. Jakob and I. Gurevych, “Extracting opinion targets in a single and cross-domain setting with conditional random fields”, In proceedings of Conference on Empirical Methods in Natural Language Processing, (2010).
- [12] A. Balahur and A. Montoyo, “Semantic approaches to fine and coarse-grained feature-based opinion mining”, In NLDB LNCS, vol. 5723, (2009), pp. 142-153.
- [13] Z. Lili and L. Chunping, “Ontology based opinion mining for movie reviews”, In KSEM LNCS, vol. 5914, (2009), pp. 204-214.
- [14] E. Kontopoulos, C. Berberidis, T. Dergiades and N. Bassiliades, “Ontology-based sentiment analysis of twitter posts”, Expert Systems with Applications, vol. 40, no.10, (2013), pp. 4065-4074.
- [15] I. Peñalver-Martinez, F. Garcia-Sanchez and V. Garcia, “Feature-based opinion mining through ontologies”, Expert Systems with Applications, vol. 41, no. 13, (2014), pp. 5995-6008.

Author



Lirong Qiu, she received her Ph.D. in Computer Sciences (2007) from Chinese Academy of Science. Now she is an associate professor of computer sciences at Information Engineering Department, Minzu University of China. Her current research interests include different aspects of natural language processing, artificial intelligence and distributed systems.

