Product Domain Ontology Construction: Based on Online Review Features and Opinion Quintuple

Jia XUE¹, Rui-xin HE² and Shu-fang CAO³

¹,³ Jiangsu University Information Department, No. 301, Xuefu Rd., Jingkou Dist., Zhenjiang 212013, Jiangsu Prov., China
² Baidu USA LLC, 1195 Bordeaux Dr, Sunnyvale, CA 94089, USA

Keywords: Online review, Domain ontology, Ontology construction.

Abstract. Based on the current literature research of online review and opinion mining demand, this paper focuses on ontology construction in ontology-based opinion mining procedure to improve the effect of opinion mining from evaluating online review. This study extracts domain concepts both from professional and user-generated contents. All elements in opinion quintuple are expressed in the ontology. The result shows that this new research method makes the improvement compared with the traditional process, which indicates the efficiency of the ontology created.

Introduction

Opinion mining from online reviews have enjoyed a huge burst of research activity since the end of the last century. As stated by Dave [1], ideally, an opinion mining tool would process a set of search results for a given item, generating a list of product attributes (quality, features, etc) and aggregating opinions about each of them (poor, mixed, good). Due to achieve this goal, some researchers tried to mine opinions based on theory of natural language process (NLP) and data mining technology. In the initial stage, the research process includes four main tasks: identify product features, identify opinions regarding product features, determine the polarity of opinions and rank opinions [2], which is called non-ontological method without using any ontology. According to the non-ontological method, product features are treated as mutually independent individuals, which may lead to confusion in sentiment computing stage. In fact, however, most concepts are essentially hierarchical and progressive. Thus, it is necessary to describe relations among product features during mining opinions. Recently, some researches have explored the way of opinion mining with ontology describing hierarchical relations among product features, whereas lacking attention on ontology construction.

An ontology is a formal and explicit specification with shared conceptualization [3]. In addition, an ontology works with defining concepts, relations, axioms and other elements. Products features are defined as concepts and their relationship relations. Concepts extraction is the first and basic work for creating an ontology, however, previous research shared one deficiency in this stage, which is discussed in relevant work.

This paper focuses on ontology construction in opinion mining and our contributions are as follows. Firstly, this paper analyses the features of online reviews by considering these reviews into product domain concepts extraction. By this way, this research makes a list of all necessary concepts in the product field. Secondly, this research shows that an opinion quintuple is considered during creating an ontology for a greater application of opinion into commercial operation.

Related Work

Ontology and Domain Ontology

Ontology initially developed in philosophy area, referring to explanation and illustration of objective existence. Robert [4] introduced the concept into information science area. He believes that an
ontology defines the basic terms and relations comprising the vocabulary of a topic area as well as the rules for combining terms and relations to define extensions to the vocabulary, while the acknowledged definition is then gave by Gruber [3]. Generally, common components of ontology include:

A. **Individuals.** Instances or objects.

B. **Classes.** Sets, collections, concepts, classes in programming, types of objects, or kinds of things.

C. **Attributes.** Aspects, properties, features, characteristics, or parameters that objects (and classes) can have.

D. **Relations.** Ways in which classes and individuals can be related to one another.

E. **Restrictions.** Formally stated descriptions of what must be true in order for some assertion to be accepted as input.

F. **Axioms.** Assertions (including rules) in a logical form that together comprise the overall theory that the ontology describes in its domain of application.

G. **Events.** The changing of attributes or relations.

A domain ontology (or domain-specific ontology) represents concepts which belong to a part of the world, such as biology or movies. Each domain ontology typically models domain specific definitions of terms.

**Ontology Construction**

Gruber [5] proposed a preliminary set of design criteria of ontology:

A. **Clarity.** An ontology should effectively communicate the intended meaning of defined terms. Definitions should be objective.

B. **Coherence.** An ontology should be coherent: that is, it should sanction inferences that are consistent with the definitions.

C. **Extendibility.** An ontology should be designed to anticipate the uses of the shared vocabulary.

D. **Minimal Encoding Bias.** The conceptualization should be specified at the knowledge level without depending on a particular symbol-level encoding.

E. **Minimal Ontological Commitment.** An ontology should require the minimal ontological commitment sufficient to support the intended knowledge sharing activities.

Based on these principles, several organizations come up with different methods to construct an ontology. Among these above methods, the Seven-Step method shows the highest applicability in commercial field. As indicates by its name, the Seven-Step method has seven steps: a. determine the fields of knowledge and scope of ontology; b. examine the reuse of existing ontology possibilities; c. list important terms in the ontology; d. define classes and class hierarchy; e. define class properties; f. custom properties sub-surface; g. create an instance.

When construct an ontology, the ontology description language should be chose at first. Ontology is commonly encoded using ontology languages. At present, the most powerful language is the Ontology Web Language (OWL) and it is recommended by the World Wide Web Consortium (W3C).

**Ontology-based Opinion Mining**

Researchers used to find solutions or discovery knowledge by data analysis or mathematical calculation. For example, literature [6-7] proposed theoretical formulations or algorithms to optimize systems. They play the role in many cases, but with the development of Web3.0, or language web, ontology plays the core role in its framework. It is necessary for corporations to merge their marketing information processing behavior into the new generation of web. With reasoning function of an ontology, computers can find new knowledge automatically.

Since the research of opinion mining usually based on specific product domain such as computers, movies and home appliances, ontology is called product domain ontology in opinion mining process, which is different from general ontology. Generally, ontology is used to improve the effect of product features extraction accuracy and precision. There are usually many synonyms such as system and operation in computer. The comment refer to the same concept, which is the operation system.
However, when calculating the similarity of words system and operation with HowNet word similarity, the figure is just 0.09. In contrast, an ontology has the function of declare synonyms relationships in a specific field. Besides, product features should not be regarded as irrelevant words, in fact, they are correlative. For example, when a potential consumer wants to figure out the appearance of a cell phone, consumers are concerned about the height, color and touch of the product. Thus, all these relevant features are expected when searching the information.

Several researches involve the construction of a product domain ontology. Yue XZ [8] firstly extracts laptop domain concepts from the website called 168.com. After that, the important step is to classify the brands and components of notebook computers. They involve a variety of product characteristics, but there is a lack of consistent relationship between product models and features. For example, not all laptop models have the feature of fingerprint identification. Feng SF and Wang SG [9] design a vehicle domain ontology construction process. The first step is to collect concepts from industrial professional websites. The second step is to create the ontology based on defined feature system and sentiment knowledge base. Their research proposes feature system should include two segments, which are product parts and attributes. Tang XB and Lan YT [10] show that concepts are extracted from the operating manual, but only some features are collected. However, Li JH et al [11] add review holder into their ontology, which is more suitable for opinion quintuple than the other studies. However, literature [9-12] did not consider time as one element of an opinion.

Although researchers above explored the concepts and relations a product domain ontology should involve, shortcomings are still obvious:

A. Lacking a sufficient combination of the profession and socialization of a domain. Indeed, on the one hand, product domain always refers to terminologies and technical parameters, those concepts only appear in specialized documents, such as corporate websites, product specification and so on. On the other hand, consumers usually publish comments with oral and personalized style. Previous researches are supposed to take these two sides into consideration simultaneously.

B. Incomplete product feature system. Only a few studies have noticed what features ontology should have. What these studies need to improve is to add product brands and components to the system. As for products on e-commerce websites, many elements have influence on consumers’ purchasing decision. For instance, the recipient on websites, the speed of transportation of products and the deliveryman are all one of the influencing factors. A product ontology should contain all these features vital of the e-commerce enterprise.

C. Ignoring time or holder. There are five members of a complete opinion quintuple, whereas most researches exclude time or holder.

Therefore, this paper propose a new method to construct a product domain ontology based on product features and opinion quintuple.

**Approach**

**Opinion Quintuple**

Before starting opinion mining, the definition of an opinion should be clarified. Several researchers indicated the different definition. Kim and Hovey [12] provided with the earliest definition. They described an opinion as a quadruple [Topic, Holder, Claim, Sentiment] in which the Holder believes a Claim about the Topic, and in many cases associates a Sentiment, such as good or bad, with the belief. For example, they considered the following mobile phone review:

_This Huawei phone is so beautiful!_

In the sentence above, a topic (appearance), a holder (the corresponding consumer ID), a claim (appearance is so beautiful), and the sentiment (so beautiful) are contained. Afterwards Liu Bing defined an opinion is a quintuple:

\[(e_i, a_{ij}, S_{ikl}, h_k, t_l),\]
where $e_i$ is the name of an entity, $a_{ij}$ is an aspect of $e_i$, $s_{ijkl}$ is the sentiment on aspect $a_{ij}$ of entity $e_i$, $h_k$ is the opinion holder, and $t_l$ is the time when the opinion is expressed by $h_k$. The sentiment $s_{ijkl}$ is positive, negative, or neutral, or expressed with different strength/intensity levels [13]. Compared with other definitions, it has the highest completeness. The other researchers seldom mention time or holder. However, in business reality where data is assert and consumer-oriented marking is pursued, nowadays time series analysis, consumer relationship management, precise marking, etc. are all devoted to find new knowledge between consumer related data, time and other types of data. There are good reasons to believe that the inclusion of time and consumer opinions will contribute to the following opinion-based data mining activities. Thus, the recommended method can describe those five elements of an opinion.

**Online Review Features**

Online users submit product reviews on shopping websites, BBS, social website and various kinds of APP. These reviews are a mixture of professional and oral contents. Product domain concepts can be collected from following approaches:

A. **Official contents.** For example, industrial regulations, related laws, books, journals, user mammals, corporate websites, thesaurus, etc., these above types of content ensure that the concepts collected are fully expressed, except in official settings.

B. **Folksonomy and social tags.** By this mean, concepts are selected from social tags gathered from websites with rich tags, such as douban.com and sina.blogs. In this way, according to the user's point of view, the product characteristics that consumers really care about can be distinguished. Nevertheless, it is always accompanied by a lot of redundant tags. In this paper, this approach is not recommended.

C. **Text mining.** This way based on a mass of text corpus using NLP technology. The advantage is concepts are collected from users’ perspective, meanwhile, they can express all the concepts that consumers concern about.

Considered that online review is always with terminologies and personalized content, this paper proposes that domain concepts should be collected from official resource as well as online review. In the next section, the proposed method is illustrated by an example of mobile phone construction.

**Concept Extraction**

This research extracts mobile phone domain concepts from online reviews as well as corporate websites. Before action, a user defined dictionary is needed in case that corporations usually create new vocabularies or phrases during the process of marketing. For Chinese documents, a dictionary is essential for a high accuracy in text segmentation period. This research collects 116 new vocabularies and phrases from bbs.cnmo.com/. Most of them are description of phone brands, models and parts. Besides, this research gathers 34 seller defined phrases from www.vmall.com. All of them describe the function of the smart phone.

Then, this study crawls 10 thousand phone reviews from www.jd.com, totally 86 thousand sentences with 970 thousand characters. With NLPIR2014, a Chinese language processing tool, 160 domain concepts could be collected. Afterwards, this paper extracts 140 professional concepts from product detail illustration documents. Next, after a series processing procedure including merging, removing duplication, washing, 84 mobile phone domain classes are generated.

**Class Hierarchy Levels**

Every item involved in product life cycle could be the factor effecting consumer’ purchasing. According to Michael Porter’s Value Chain Model, a firm's value chain forms a part of a larger stream of activities including primary activities and support activities. Primary activities start with inbound logistics, in which period a firm arranges the inbound movement of materials, parts, or finished inventory from suppliers to manufacturing or assembly plants, warehouses, or retail stores. Then it comes to operations period, followed with outbound logistics. Afterwards, marketing and sales are
also necessary. Eventually, it points out that service including all the activities required to keep the product and also service working is effective for the buyer when the products are sold out. At the same time, delivered is also considerable. Support activities includes infrastructure, technological development, human resources Management and procurement. For network products, from the perspective of consumers, all relevant factors can have an impact. Reasonably, this research reasonably classifies products into four aspects, each of which has its branches. Fig. 1 shows classes and their hierarchy.

Notice that there can be more than three levels for a specific type of product. This depends in part on the complexity of the product.

**Properties Definition**

There are at least two types of properties should be defined for the ontology. As for the object properties, there are 7 types as below:

A. **Sub-class Of.** This property describes the relationship between a father class and its son classes. For example, camera sub-class of part indicates that camera belongs to phone parts.

B. **Equivalent To.** This property describes the equivalence relationship among classes. For example, some properties may have different expressions with the same meaning.

C. **Disjoint Of.** This property describes the situation that two classes share no individuals.

D. **Part Of.** This property clarifies the relation between whole and part. Its inverse property is Part.

As for data properties, there are two types, describing person and product respectively. For example, a person has its ID, phone number, address and so on. A product has its price, height, weight, height, parts and sub-parts, color and so on. Especially, time is an indispensable index in the process of product classification.

**Phone Domain Ontology Execution**

So far, there are several available ontology building tools, among which Protégé is an outstanding one. Protégé is a free and open-source platform, which provides a growing user community with a suite of tools to construct domain models and knowledge-based applications of ontology. Previous work has proved the efficiency of Protégé on ontology creating. This paper uses Protégé5.0.0 to execute a phone domain ontology. Figure 2 shows the interface of this tool.
Figure 2. Protege5.0.0 interface.

First, this paper downloads a fit version of Protégé from https://protege.stanford.edu. Then, this research only need to create the ontology step by step in its default workspace. After that, the important process is to create classes, object properties and data properties in corresponding function models. At last, the results of the research output it into OWL form:

```
<Declaration> <Class IRI="#Battery"/> <Declaration> <Declaration> <Class IRI="#HandFeeling"/>
<Declaration> <Declaration> <Class IRI="#Screen"/> <Declaration> <DataProperty IRI="#Height"/>
<Declaration> <Class IRI="#Crust"/> <Declaration> <Declaration> <Class IRI="#System"/>
<Declaration> <Declaration> <Class IRI="#Entertainment"/> <Declaration> <Declaration> <Declaration> <Class IRI="#Appearance"/>
<Declaration> <Declaration> <Class IRI="#Specification"/> <Declaration> <Declaration> <Class IRI="#Quality"/>
<Declaration> <Declaration> <Class IRI="#USB"/> <Declaration> <Declaration> <Declaration> <Class IRI="#BisonFone"/>
<Declaration> <Declaration> <DataProperty IRI="#RAM"/> <Declaration> <Declaration> <Declaration> <Declaration> <Class IRI="#Storage"/>
<Declaration> <Declaration> <Declaration> <Class IRI="#Weight"/> <Declaration> <Declaration> <Declaration> <Declaration> <Declaration> <Class IRI="#PhoneFingerPrint"/>
```

Figure 3. Mobile phone domain ontology OWL expression.

With jena or other APIs, computers can understand the ontology and apply it into further action.

**Experiment for Contrast**

**Sentiment Dictionary Creating**

This research manages opinion mining with the support of phone sentiment dictionary, which is usually used in sentiment analysis. Domain sentiment dictionary is created from existing sentiment dictionaries. This paper adopts Hownet sentiment dictionary. It divides sentiment words into four categories, namely positive and negative description words, sentiment degree words, positive and negative emotion words and sensory words. It contains 8,936 emotional words and 219 intensity words. These words make it the most popular dictionary in the field of emotional analysis. However, a majority of sentiment words are irrelevant or slightly relevant with phone domain. These words are noneffective for sentiment analysis. Considering this factor, this paper suggests that users extract emotional words from online comments. Finally, this study generates an emotional dictionary in the smart phone domain, as shown in Table 1.
Table 1. Phone domain sentiment word dictionary.

| Sentiment words   | Degree words               |
|-------------------|----------------------------|
| Pos. 1 great, good, fast, awesome, …… | Low 0.5 a little, a bit, kidn of, slightly, …… |
| Neg. -1 worst, ugly, small, shit, bug, …… | Mid. 0.8 very, pretty, fairly, …… |
|                   | High 1.2 much too, extremely, excessive, …… |

According to the polarity of each emotional word, this paper scored each emotional word. Degree words are categorized into three levels and the assignment rises from 0.5, 0.8 to 1.2 with higher degree.

**Experiments and Results**

This section conducts two experiments to focus on the effect of the structure of ontology. This research collects over 10 thousand phone reviews from the most popular 3C product retailer - www.JD.com. After defining document processing and affective dictionary, the method of opinion mining in references is realized [14]. The pseudo code of the ontology-based method algorithm is as follows:

- **Step 1:** process the text, extract opinion phrases.
- **Step 2:** identify sentiment.
- **Step 3:** if the sentiment polarity is a, the sentiment intensity is b, generate a score record as +a·b of the corresponding topic.
- **Step 4:** add up all score records for every topic.
- **Step 5:** compute scores of all topics based on the topic hierarchy level.
- **Step 6:** rank topics by their importance and scores.
- **Step 7:** visualization of sentiment calculation result.

The results are as below. The indexes of accuracy, recall and F-value are usually used to evaluate the efficiency of opinion mining method.

Table 2. Accuracy recall and F value.

| Method         | Accuracy | Recall | F-value |
|----------------|----------|--------|---------|
| Non-ontology   | 63.54%   | 67.80% | 65.60%  |
| Ontology-based | 74.62%   | 82.47% | 78.35%  |

This above table shows that present method has a better accuracy, recall and F-value than the traditional way, which indicates the efficiency of the structure of the phone domain ontology in this paper.

**Summary and Future Work**

This paper proposes a new product domain ontology construction method based on online review features and opinion quintuple. The key point of this research innovation is that our ontology involves rich information, which can facilitate future knowledge discovery activities. Besides, in order to improve the effect of opinion mining in online comments, this research proposes the concept of extracting from professional knowledge and social documents. In addition, time is added to ontology as a data attribute, which is beneficial to facilitate time series comments and other mining activities.

This study hopes to expand this work in the following directions. To begin with, with the continuous change of domain knowledge, the method of exploring ontology in research would be constantly updated. Besides, in the context of the big data era, the research would explore how to use massive data to manage ontology construction in order to meet more needs of the big data era.
Acknowledgement

This work was financially supported by the Jiangsu Provience Social Science Foundation (16TQB009) and Jiangsu Provience University Philosophy and Social Science Foundation (2017SJB1892).

References

[1] Dave, K., Lawrence, S., & Pennock, D. M. (2003, May). Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In Proceedings of the 12th international conference on World Wide Web (pp. 519-528). ACM.

[2] Popescu, A. M., & Etzioni, O. (2007). Extracting product features and opinions from reviews. In Natural language processing and text mining (pp. 9-28). Springer, London.

[3] Studer, R., Benjamins, V. R., & Fensel, D. (1998). Knowledge engineering: principles and methods. Data & knowledge engineering, 25(1-2), 161-197.

[4] Neches, R. (1991). Enabling technology for knowledge sharing. Ai Magazine, 12(3), 36-56.

[5] Gruber T R. Toward principles for the design of ontologies used for knowledge sharing?[J]. International Journal of Human-Computer Studies, 1995, 43(5–6):907-928

[6] H. Runxin, G. Humberto. Numerical Synthesis of Pontryagin Optimal Control Minimizers Using Sampling-Based Methods[C]//IEEE 56th Annual Conference on Decision and Control (CDC). Melbourne, AUSTRALIA: IEEE CDC, 2017: 733-738.

[7] He, R., & Gonzalez, H. (2016, July). Zoned HVAC control via PDE-constrained optimization. In American Control Conference (ACC), 2016 (pp. 587-592). IEEE.

[8] Y. Xiaozheng. (2008). A Domain-Ontology-Based Opinion Mining System (Doctoral dissertation, Beijing: Beijing University of Posts and Telecommunications).

[9] F. Shufang and W. Suge. (2011). Research on Construction of Automobile Domain Ontology Knowledge for Opinion Mining. Computer Applications and Software, 28(5), 45-47.

[10] T. Xiaobo and Lan Yuting. (2016). Sentiment Analysis of Microblog Product Reviews Based on Feature Ontology. Library and Information Service, 60(16), 121-127.

[11] L. Jinhai, H. Youshi and M. Yunle. (2016). Hierarchical Mining of Online Reviews Based on Domain Ontology. Systems Engineering, 34(10), 39-47.

[12] Kim, S. M., & Hovy, E. (2004, August). Determining the sentiment of opinions. In Proceedings of the 20th international conference on Computational Linguistics (p. 1367). Association for Computational Linguistics.

[13] Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis lectures on human language technologies, 5(1), 1-167.

[14] Z. Junyong. (2014). Research On Ontology-based Internet Public Opinion Mining (Doctoral dissertation, Chongqing: Chongqing University).