Building Ontology for Different Emotional Contexts and Multilingual Environment in Opinion Mining

Wan Tao, Tao Liu, Wei Yu, Gan Huang
School of Computer and Information Science, Anhui Polytechnic University
Key Laboratory of Computer Application Technology, Anhui Polytechnic University
Wuhu, China
{Taowan, Liutao}@ahpu.edu.cn

Abstract—**With the explosive growth of various social media applications, individuals and organizations are increasingly using their contents (e.g., reviews, forum discussions, blogs, micro-blogs, comments, and postings in social network sites) for decision making. These contents are typical big data. Opinion mining or sentiment analysis focuses on how to extract emotional semantics from these big data to help users to get a better decision. That is not an easy task because it faces many problems, such as different context may make the meaning of the same word change variously, at the same time multilingual environment restricts the full use of the analysis results. Ontology provides knowledge about specific domains that are understandable by both the computers and developers. Building ontology is mainly a useful first step in providing and formalizing the semantics of information representation. We proposed an ontology DEMLOnto based on six basic emotions to help users to share existed information. The ontology DEMLOnto would help in identifying the opinion features associated with the contextual environment, which may change along with applications. We built the ontology according to ontology engineering. It was developed on the platform Protégé by using OWL.**

Keywords—Ontology; opinion mining; social media; emotional context; multilingual environment; OWL

I. INTRODUCTION

Today large amount of information is supplied by the social media users. They share information about products, services and their experiences on the social networks. The information may contain users' sentiments, such as joy, sad, like, or dislike (Sentiment and emotion mean similar, in the following, we usually use emotion in a more specific situation). Users' sentiment orientation and emotion of the topic or event can not only provide decision-making basis in business but also provide support for government's public opinion monitoring. But it is very hard to integrate users' comments and feedback into normal application system. Currently, opinion mining is increasingly important than ever before, especially in customer preference analysis and prediction. Most opinion mining attempt to identify the polarity of sentiment in three categories: positive, negative or neutral. But how to identify the polarity of sentiment is a difficult task. The same word in different contexts may convey different emotions. For example, the word 'high' in sentence 'My salary is high' represents the positive emotion but in 'The price is high' may represent the negative emotion. Otherwise,

multilingual applications have produced many problems, such as how to map a word in Chinese to a similar word in English.

In this paper, we construct ontology *DEMLOnto* to help to solve the above problems. This ontology would help in identifying the opinion features associated with the different contextual environment that contains positive or negative sentiments. Also some methods are proposed to deal with multilingual environment. The remaining portion of the paper is organized as follows: Section II firstly introduce the special environment of social media, then describes emotion and opinion mining, provides the background research used for why and how to integrate the ontology into opinion mining; Section III provides the definition of our ontology *DEMLOnto*, and focuses on how to design our ontology on emotions and multi-languages these two aspects; Section IV provides constructs of the ontology in OWL(a Web Ontology Language), focuses on discussion about how to design ontology, explain how to get new describing construct, provides the building process on the Protégé, and introduce how to use *DEMLOnto* to extract information from the social media; and Section V briefs on conclusions and future work.

II. BACKGROUND RESEARCH

A. Social Media Environment

Social media supports the interaction among people in which they create, share, discuss, or exchange ideas in virtual communities and networks. Sentiment expression in micro-blog posts often reflects user's specific individuality due to different language habit, personal character, opinion bias and so on [1].

Social media brings us with typical big data. Firstly, it produces large amount of data. For instance, daily online QQ users count for 160 million. Their storage is 300G. Secondly, data update fast. Twitter posts about 14300 per second. Thirdly, social media data is extremely various. This kind of variety is not only limited into its data type which is structured or mainly unstructured, but also means the following: many new words produced by media users (e.g. "童鞋" means "同学", which is "classmate" in English); kinds of contents inserted by supplying simple "Like" or "Dislike" tools, star-rating systems, tag-based annotation and navigation, and so forth. Lastly, there is uncertainty in social media data. While using the same word, people may deliver different sentiment

orientations depending on the underlying context, just as our example in Section 1. Our work mainly has relationship with the last two points. We found that vocabulary is the basis of sentiment analysis. When we do sentiment analysis, first of all, we should have a reserve of lexical knowledge, so that the massive sentiment analysis does not require relearning subjective point of view, then the efficiency of sentiment analysis is improved greatly. Ontology knowledge base is a good vocabulary knowledge base, which can help to understand concepts and the relationship among all the attributes in the concepts.

B. How to do Opinion Mining

Opinion mining, also called sentiment analysis, is the field of study that analyzes people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes[2]. The task of determining positive and negative orientations of the information present in textual form is considered as a fundamental issue in opinion mining.

In general, opinion mining has been investigated mainly at three levels: document level, sentence level, entity and aspect level. It can be used for three varied objectives: polarity identification, subjectivity or objectivity identification and feature/aspect based analysis[2]. In this paper we focuses on opinions which express or imply positive or negative sentiments then make a summary at different levels.

Opinion mining consists of the following tasks[3]:*1)* entity extraction and categorization which is often done by ETL technology; *2)* aspect extraction and categorization which mainly focus on opinion features; *3)* opinion holder extraction and categorization; *4)* time extraction and standardization; *5)* aspect sentiment classification which often determine the actual feeling about the mining object by assign a numeric sentiment rating to the aspect; *6)* outputting the analysis results in tuples.

C. Opinion and Emotion Description

According to [3], an opinion can be defined as a quintuple

$$(e_i, a_{ij}, s_{ijkl}, h_k, t_l),$$

where e_i is the name of an entity, a_{ij} is an aspect of e_i, s_{ijkl} is the emotional 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 emotional s_{ijkl} is positive, negative, or neutral, or expressed with different strength/intensity levels, e.g., 1 to 5 stars as used by most review sits on the Web. That is, the opinion s_{ijkl} must be given by opinion holder h_k about aspect a_{ij} of entity e_i at time t_l. Any mismatch is an error. At last, the feature pairs and emotional words can help us make decision.

Emotions are our subjective feelings and thoughts. Based on [4], people have six primary emotions, i.e., love, joy, surprise, anger, sadness, and fear, which can be sub-divided into many secondary and tertiary emotions. Each emotion can also have different intensities. The strength of an emotion or opinion is typically linked to the intensity of certain emotions, e.g., joy and anger.

We think if we can integrate features (i.e. emotions, aspects, etc) into the ontology, we will share them more comfortable. In Section 3 we will focus on emotional context, polarity strength and multi-language these aspects by integrating them into ontology.

D. Ontology Description

Ontology has been defined as the specialization of the conceptualization by [5]. The main aim of ontology is to provide knowledge about specific domains that are understandable by both the computers and developers. Ontology plays important roles in sharing sources and defining terms precisely for future uses such as meta-data. It also helps to interpret a text review at a finger granularity with shared meanings and provides a sound semantic ground of machine understandable description of digital content.

In former researches, reference [6] has proposed a feature ontology that uses a multidimensional model to integrate customer's characteristics and their comments about products. This approach first identifies the entities and then emotions present in the customers reviews related to mobiles are transformed into an attribute table by using a 7 point polarity system (-3 to 3). The limitation of their approach is that the ontology proposed by them is too general and is lack of reasoning ability among multiple products. In [7], a whole framework for building domain ontology of video is proposed. The semantic link network model is used to mine and organize video resources based on their associations. A semantic-based video organizing platform is provided for searching videos.

However, we found that even previous researches have focused emotion analysis at different levels but domain knowledge, context relationship, and multilingual application were not considered together during those researches. We think if ontology can be integrated into the 6 tasks of opinion mining (introduced in Section 2), it will make the mining results to be structured, make the results easy to be utilized and shared.

III. THE ONTOLOGY OF EMOTIONS AND MULTI-LANGUAGE

A. How to Deal with Emotions

We have developed *DEMLOnto* ontology ascended from [8,9] with some changes. Our ontology is designed on six basic emotions: *love, joy, anger, sad, fear, surprise*. They are structured in a taxonomy that covers from six basic emotions concepts to the most specific emotional words as data objects or instances. Each of the emotional data is related with the seven emotional strength levels by means of data ranges, as well as three polarity values to stand for positive, negative, or neutral.

Definition 1: Ontology is defined as

$$O = \{ C, A^C, R, A^R, H, I, X \},$$

where C is a set of concepts; A^C is a set of multiply properties which one is belong to a concept; R is a set of relationships; A^R is a set of multiply relationships which one is belong to a property; H is a set of hierarchy relationships among concepts; I is a set of instances; X is a set of axioms.

In Fig. 1, we explain how concept Joy and Anger are constructed into the emotional ontology. In particular, when the application environment is not the same, we will make record of the different emotion-feature pairs into the ontology base, such as (price, i-Phone, highness, -1), (Tom's salary, incoming, highness, +2).
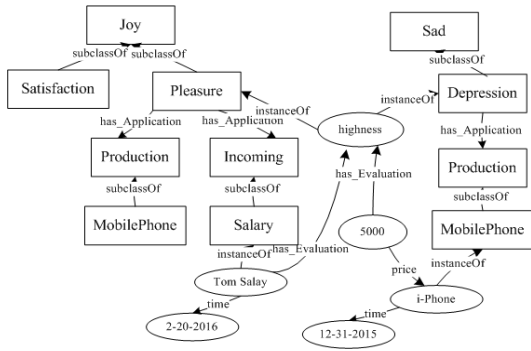

Fig. 1 Emotional Ontology(Fragment)

Fig.2 is a fragment of our ontology *DEMLOnto*'s definition. In definition, we use many properties to explain the feature of an emotion. For instance, polarity of an emotion which value is [1, 0, -1] respectively represents the emotion is positive, negative, or neutral; power represents the degree of positive or negative. We use (-3, 3), there are totally 7 numbers to stand for the different degree of an emotion. Why do we design two properties to measure an emotion? We do it just for that different applications maybe require different granularity in analysis process.

$O_{em}= \{C_{em}, A_{em}^{C}, R_{em}, A_{em}^{R}, H_{em}, I_{em}^{C}, X_{em} \}$ *where*
$C_{em}= \{$ *Joy, Anger, Sad, Fear, Surprise, Love* $\}$
$A_{em}^{C}=\{ A_{em}^{C}$ *(Joy)*, A_{em}^{C} *(Anger),…* $\}$
A_{em}^{C} *(Joy)*$=\{$ *polarity(integer),power(integer)*$\}$
R_{em}^{C} *(Joy)* $=\{$ *subclassOf, instanceOf ,is_Similarwith, has_Application, has_Apptime, has_Evaluation…,*$\}$
A_{em}^{R} *(Joy,is_Similarwith)*$=\{inChinese(text), inEnglish(text), inSpanish(text)…\}$
$H_{em}=$*(Is-a, Superclass, Subclass…)*
$I_{em}^{pleasure}=\{highness, goodness, fineness…\}$
$X_{em}=\{polarity\_rules, power\_rules…\}$
…

Fig. 2 Ontology DEMLOnto Definition（Fragment）

B. How to Deal with Multi-Languages

In order to deal with multiply languages in some application, we extended the original ontology with some changes from [8].

The ontology has three root classes: *Emotion, Word, Application. Emotion* is the root for all the emotional concepts. *Word* is the root for the emotion-denoting words, i.e. the words which each language provides for denoting emotions, easily to be extended. Now we supposed it originally had two subclasses: *EnglishWord* and *ChineseWord*, All original words can get by information extraction from the Web (just as task 1 of opinion mining). *Application* includes objects which the emotions come from.

And in Fig.2 we desigened *is_Similarwith* relationship to help us dynamically classify the new words which are

provided for denoting emotions into *Word* concept. The ontology's structure will be listed in Fig. 4 which is worked out by Protégéin Section 5.

IV. BUILDING THE ONTOLOGY IN OWL

A. OWL Overview

OWL is a recommendation of the W3C from 2004. In 2009, OWL 2 is proposed as an extension and revision of the OWL. OWL 2 also has three increasingly-expressive sublanguages: OWL Lite, OWL DL, and OWL Full. Just as the former OWL. OWL Lite is simple but lack of logical expression ability; OWL Full is powerful for representing complex statements but not useful for reasoning with them due to their computational properties. OWL DL is the subset of OWL designed for applications that need the maximum expressiveness without losing computational completeness and decidability. It is based on Description Logics, a particular fragment of first order logic, in which concepts, roles, individuals and axioms that relate them are defined.

In our work, we apply OWL DL to construct our ontology.

B. Building Process of DEMLOnto

According to [10], building an ontology includes:*1)* defining classes in the ontology; *2)* arranging the classes in a taxonomic (subclass–superclass) hierarchy; *3)* defining properties and describing allowed values for these properties; *4)* filling in the values for properties for instances.

At first, English emotional classes were extracted as the opinion as positive and negative based on the scales developed in previous works by [6], and Chinese classes were gained from [11]. All of these are read and evaluated by human beings.

We use Protégé 5.0 as the building platform. Protégé is mainly an ontology editor developed by Stanford University. Its reasoners can automatically classify the concepts into a hierarchy of emotional concepts.

In OWL, we can define new constructs and axioms to fit for our needs. For instance, in Section 3 we have designed *is_Simliarwith* to handle similar words in different language contexts. It helps us map the Chinese words to the similar English words, and then they can be classified into the correct emotional class by the reasoner. The results are shown in Fig. 3. We can see in the Fig. 3(a), "狂怒"(which means 'Fury') is a subclass of Chineseword. At the same time "狂怒" has property *is_Simliarwith* with "Fury" that is a subclass of Anger. As seen in Fig. 3(b), "狂怒" is automatically add into subclasses of Anger by the reasoner Pellet(a plug-in of Protégé). Otherwise, we also defined *has_application, has_apptime, has_evaluation*, etc properties in ObjectProperty of different concepts. The ontology structure is shown in Fig.4, which is displayed by OntoGraf (a plug-in of Protégé).

In summary, the ontology *DEMLOnto* can be supplementing accumulatively under the help of opinion extraction. Conversely, the built ontology can fix the mining results into common knowledge to help us eliminate ambiguity in opinion mining.
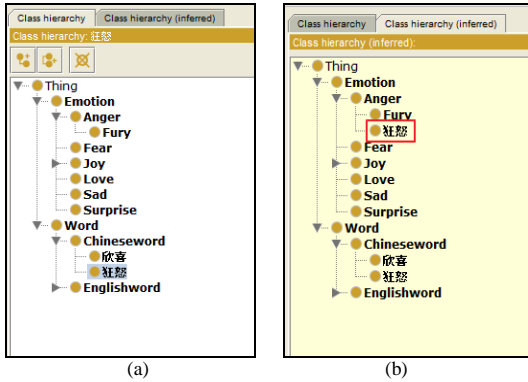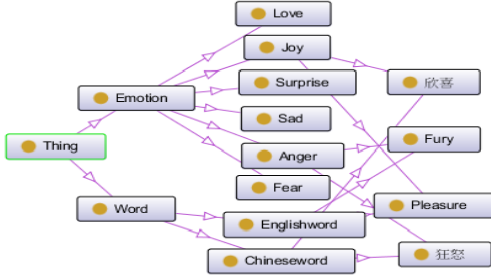
Fig. 3 Reasoning in Multi-Language

Fig. 4 DEMLOnto Structure (Fragment)

C. How to Use DEMLOnto

Ontology can help us to extract effective opinion, to make us mine more useful results. Now, we have used *DEMLOnto* to help us to extract opinion from micro-blog on domains of weather, book, and shopping. Under the guidance of the emotion ontology, we annotate entities at DOM level based on context distance and co-occurrence number. The annotation results were added to the entity record. And similarity of the ontology and entity is calculated by using (1).

$$Sim(C,O) = \frac{\max \sum_{j=1}^{n} \mathrm{Re}\, l(C, A_j)}{\sqrt{mn}} \qquad (1)$$

$$\mathrm{Re}\, l(C, A_j) = \alpha \times \left(\frac{Itemspan(C, A_j)}{\max\{Itemspan(C, A_j)\}} \right) + (1 - \alpha) \times \frac{Itemtimes(C, A_j)}{\sum_{j=1}^{n} Itemtimes(C, A_j)}$$

$$Itemspan(C, A_j) = \frac{\sum_{i=1}^{m} Span(c_i, A_j)}{m}$$

where O is ontology *DEMLOnto*; m is properties of an application entity in *DEMLOnto*; n is properties of an entity in web page being extracted; C represents an entity semantic values set; $Span(c_i, A_j)$ is context distance between semantic C_i and ontology attribute A_j in the same page; $Itemspan(C, A_j)$ is mean value of context distance between each semantic item and the attribute A_j in the same page of an entity C.

According to the Google ranking, we selected five top social media sites which mainly about books, weather and shopping to our research. If the computation output of (1) is large than 0.62(this value is determined by several tests), we judge that it is a good result. The evaluation index of information extraction is the accuracy rate of P, the recall rate

of R and F-measure. The experiment result is shown in TABLE I.

TABLE I. THE COMPARISON OF DIFFERENT DOMAIN

Domain	P(%)	R(%)	F(%)
Weather	84.8	84.4	84.6
Book	83.9	82.8	83.3
Shopping	82.4	81.7	82.1

V. CONCLUSION

Currently, our ontology is developed based on manually selected classes from some baseline datasets, and our research work is limited at extraction stage. Our next target is to make use of machine learning method to develop ontology from actual applications automatically and effectively. Also we will integrate ontology into opinion mining deeply. Protégé OLV is a plug-in published in 2015. It can help us searching for relevant vocabularies in the Linked Open Data catalogue, for easing the development of ontology by reusing existing vocabularies at low fine grained level. So, taking use of Protégé OLV may largely promote the efficiency of building and reusing of the ontology. In the next, we will focus on it.

Acknowledgment

References

[1] Wu F., Song Y., Huang Y., Microblog sentiment classification with contextual knowledge regularization. Twenty-Ninth AAAI Conference on Artificial Intelligence 2015.

[2] Dinakar S, Andhale P, Rege M. Sentiment analysis of social network content. Information Reuse and Integration (IRI), 2015 IEEE International Conference on IEEE, 2015:189-192.

[3] Liu B, Zhang L. A survey of opinion mining and sentiment analysis, Mining Text Data. Springer US, 2012:459-526.

[4] Parrott, W. Gerrod. Emotions in social psychology, Psychology Press, 2001.

[5] T. R. Gruber. A translation approach to portable Ontology Specifications. Knowledge Acquisition, 1993,5(2):199–220.

[6] Yaakub M R, Li Y, Zhang J. Integration of sentiment analysis into customer relational model: The Importance of Feature Ontology and Synonym. Procedia Technology, 2013, 11(1):495-501.

[7] Zheng Xu, Yunhai Liu, Lin Mei, et al. Semantic based representing and organizing surveillance big data using video structural description technology. The Journal of Systems and Software, 2015, 102:217-225.

[8] Baldoni M, Baroglio C, Rena P, et al. From tags to emotions: Ontology-driven sentiment analysis in the social semantic web. Intelligenza Artificiale, 2012, 6(1):41-54.

[9] Francisco V, Gervás P, Peinado F. Ontological reasoning to configure emotional voice synthesis. Proceedings of the 1st international conference on Web reasoning and rule systems. Springer-Verlag, 2007:88-102.

[10] Noy N F, Mcguinness D L. Ontology Development 101: A Guide to Creating Your First Ontology. And Stanford Medical Informatics, 2001.

[11] Xu Linhong, Lin Hongfei,et al. Constructing the affective lexicon Ontology, Journal of the China Society for Scientific and Technical Information. 2008, 27(2):180-185