Abstract—The increasing proliferation of mobile devices has resulted in people being able to express their opinions and sentiments more easily. To obtain people's preferences and evaluations about products and services, sentiments expressed over Websites and social network services need to be analyzed. However, due to the diversity of the media and the short lengths of messages, the sentiment data is fragmented and distributed. Nowadays, products are equipped with numerous functions, so product features are diverse and complex. In order to solve the sentiment data fragmentation problem and the complexity of the feature, we propose an ontology-based sentiment analysis system for mobile devices.

Keywords: Ontology, Sentiment analysis, Mobile device

1 Introduction

Social network services such as Facebook and Twitter and blogs have expanded due to the proliferation of mobile devices. Personal opinions about products can be easily posted and shared over the Web and on social networks, and people can easily obtain other people’s opinion by conducting Web and social network searches. With these data, there is no need to enter each item individually. However, this data is unstructured; this type of data has also been described as fragmentary because of the constraints of microblogging. Because reviews are scattered, it is difficult to obtain a comprehensive understanding of the overall opinion of everyone. An overall analysis of the review data created by professional power bloggers is hard to find as the data are subdivided according to each feature. In order to solve these problems, decision making by collecting and analyzing the opinions on the Web and on social networks is carried out. The methods used are called sentiment analysis methods.

Sentiment analysis can be divided into document-level, sentence-level, word-level, and feature-level analyses. Document-level analysis determines the polarity of the document according to the frequency and location of a specified term. Sentence-based analysis focuses on sentence-level adjectives. Word-based polarity analysis makes word lists that have polarity and extends them with WordNet. Finally, feature-based sentiment analysis finds frequent nouns and noun phrases in order to extract frequent features; it analyzes sentiments for these extracted features [1].

Initial studies considered preferences with respect to the overall target product itself. However, people's opinions and sentiments are complex. Thus, it was found that detailed classification and evaluation were needed of the features of the products. However, the problem is that detailed information on the features is inadequate for learning with a reasonable dataset. As an alternative, structured data can be used to obtain such knowledge by leveraging domain ontology [2]. It can conjugate the ontology mapping because ontologies have structural information that facilitates such an activity. Ontology structure can embody the kind of feature using a hierarchical domain with less data. By means of these ontology-based sentiment analyses, users can collect public opinions and preferences in detail for a feature.

Three aspects of these sentiment analysis methods make them suitable for mobile device environments. First, data generated on mobile devices can be personally gathered and in various forms of information than on a PC for public use. Fragmented data can be integrated into one role based on the user. Second, sentiments containing personal information need to be kept private. It is more suitable to keep personal information on a mobile device than on the Web. Private information has to be handled inside a mobile device to minimize access by external processes. Third, to analyze a user in the new media, a certain amount of data is needed. These cold start problems can be resolved by utilizing unused existing data pertaining to users on mobile devices.

Nowadays, products include various functions, resulting in diverse and complex features. Because previous approaches simply constructed feature lists from public opinions, a separate process is needed to evaluate the products according to their own priorities [3]. Domain-specific fine-grained features can be difficult for users to understand because the term itself may be strange to understand. There are differences depending on the level of the individual and the public’s preferred level. Positive rating details vary with the characteristics of the public, and even the results may vary with personal preference. In order to solve these two problems, a system that reflects personal standards and that supports a hierarchical architecture is required for understanding fine-grained features easily.
In this paper, we propose a personal sentiment analysis system for mobile phones using feature-level ontology. Our proposed system analyzes the differences between personal preferences and the public’s preferences. The former is extracted from mobile phones, whereas the latter is extracted from the Web and social networks. Personal preference is determined by the weights of different features and is provided to readers as a reference when a user writes a review. The system is configured on camera lens domain. We chose one product for feasibility tests and applied the results obtained to other products to determine whether the selected feature is appropriate.

2 Related Work

The uses of ontology-based text mining can be divided into four kinds: utilization of text classification in ontology evaluation, use of a domain-specific ontology in mapping, enhancement of the results of text classification, and use of domain ontology for classification. In particular, ontology-based sentiment analysis has been used for feature recognition using domain ontology to reduce the number of sentiment analyses [4] and secondary data mining techniques to reduce the number of rules in the rule-based classification [5]. There are dictionaries, such as SentiWordNet, that calculate polarity. Using these dictionaries, the polarity of each term can be obtained [6]. Sentiment Ontology Tree (SOT) study for product reviews includes the adding of sentiment value to the ontology-based feature tree [7]. Recently, it has been found that fragmentary comments of about 140 characters in social media such as Twitter need an efficient technique to extract meaningful sentiment information. In recent years, some studies have been conducted to extract sentiment in mobile devices because of the prevalence of smartphones. However, those analyses were limited to one media like SMS and Twitter [8].

3 Proposed System

Content-based analysis methods that use previous purchases and collaborative filtering analysis using the data of neighbors with similar inclinations can be used to generate recommendations. A general analytical model deals with one medium only. However, accurate analysis is difficult because of the fragmentary nature of microblogs. Our proposed model conducts analyses using data on the previous representation of individuals in a variety of media stored on smartphones. Personal opinion is expressed in various ways, but can be collected from a variety of media and mobile devices. Thus, this proposed system can integrate continuous information and opinions about one topic. The system comprises a Collective Sentiment Ontology Tree (CSOT) configuration module, data storage modules, a Personal Sentiment Ontology Tree (PSOT) configuration module, and an individual character extraction module. These modules are described below:

- **CSOT configuration module**: The system uses data about a specific topic, from social media such as Twitter Search and Web page history. In the preprocessing step, Tokenization, POS Tagging, and NER are carried out on the information gathered. The module extracts features using domain ontology and constructs an ontology tree based on the extracted features. It also uses SentiWordNet to calculate the polarity of the sentence that contains the feature, and makes instances of the CSOT ontology tree that include weighted polarity of terms.

- **Data storage modules**: These data modules are separated according to the three characteristics of the storage media of a mobile device. There is an internal saving module (which deals with messages, kakaotalk, history, etc.), an external saving module (which deals with information from Facebook, Twitter, Webpages, etc.), and a streaming module (which deals with calls, TV data, streaming information, etc.). Because the internal saving module uses local data storage, it easily handles the information. The external saving module is based on requests using parameters such as keywords; it stores the response data from servers. The streaming module cannot store all of the streaming data; therefore, it stores sentiment with features using SOT.

- **PSOT configuration module**: Either one big SOT or one SOT for each application can be stored. Because some applications can be removed, individual SOT structures are better to maintain. This technique also has an advantage over CSOT when using the same media. When a target topic has been determined, it merges all the individual SOTs constructed by each application into one SOT. It can also be extended using an external domain ontology, in which case the SOT can obtain more domain information and have a more hierarchical structure. The preprocessing step is also responsible for Tokenization and POS Tagging, and NER on the information gathered. However, it uses a feature-based ontology tree created previously in the CSOT configuration module because the amount of personal data is relatively small.

- **Individual character extraction module**: Personal polarity can be measured by comparing PSOT with CSOT. One is the personal information generated by internal applications such as browser bookmarks,
messenger, and SMS on mobile devices. The other is public information located on the Web and in social network data. It divides the data into major factors and minor factors depending on the degree of deviation by comparing the instance polarity value of CSOT and PSOT about the extracted features. The deviation in personal and public sentiment analysis of the target product is considered per individual character. Thus, it is easy to take advantage of the new product because it uses already existing data on mobile device.

Users can know their own feature preferences and sentiments by personal polarity. Instead of the ambiguous recommendation for a product, users can check the interest or dislike feature ratings. Thus, recommendations can be made that reflect the characteristics of individuals based on analysis of the existing information. A hierarchical ontology can identify the feature preferences with less personal data using the higher-level feature preferences. Locally stored ontology on mobile devices can be reused in a similar domain, without compromising privacy because the operation is done internally. When people write reviews that include their personal polarity, they can be helpful to other people. Instead of sharing a simple expression such as “like,” a review can provide additional information about the writer. The information gathered is “what kind of people like it.”

4 Use case

In order to demonstrate the feasibility of our proposed system, we chose camera lenses as the target domain. The major features of the lenses are different and depend on the purpose of usage; in addition, optical-related features are difficult to understand. There are various emotional preferences and expressions associated with pictures.

For our experiments, we chose Canon EF 200 mm USM lens as our target product and used the personal data of people who are interested in these products. We found 447 sentences containing the name of the target product on 20,354 history pages. The sentiment word was contained in 139 sentences among 447 sentences. From those sentences, 46 features were extracted. Using the 46 features, CSOT was constructed. We applied 139 sentiments on the instance of history CSOT. Using Daum twitter search (http://www.daum.net), we searched 139,000 tweets.
associated with the target product. Based on our constructed CSOT, the previously decided 46 features were used for the extraction of featured sentence. A total of 336 tweets associated with the 46 features were selected. After analyzing the selected messages, 87 sentiment words were extracted. We made an instance of CSOT and applied the 87 sentiments on the instance of Twitter CSOT.

In the mobile device environment, we extracted 132 personal datasets associated with the target camera lenses from Twitter (http://www.twitter.com), SMS, bookmark, Facebook (http://www.facebook.com), kakaotalk (messenger application; http://www.kakao.com/talk), tistory (blog; http://www.tistory.com), and slrclub (DSLR review site; http://www.slrclub.com). There were 21 sentiment sentences. Those sentiment sentences included 17 features. Based on the 17 features, a PSOT was constructed and the 21 sentiments applied on the instance of PSOT.

According to the importance of the evaluation, we divided the elements into major elements and minor elements. Major elements signify opposite polarities between instances of PSOT and CSOT. A minor element signifies a low difference in polarity. “Strength” and “stabilization” features were extracted from the results as major elements by comparing the value of the polarity between PSOT and history CSOT. The “stylish” feature was extracted as a minor factor between PSOT and history CSOT.

In the Twitter CSOT and PSOT case, “Image stabilization” had a different polarity value, so it was extracted as a major element. From these comparisons of PSOT, history CSOT, and Twitter CSOT, “Strength” and “Image stabilization” were regarded as major elements and “Stylish” as a minor element of the personal weighted feature. When we constructed PSOT, history CSOT, and Twitter CSOT about Canon EF 24 mm and 85 mm USM lens in the same way, there was no conflict between them and the previously extracted major and minor elements.

5 Conclusions

In this study, a personal sentiment ontology tree was constructed using data extracted from individual applications on a mobile device according to temporal relevance. PSOT, constructed from data on a mobile device, was compared with CSOT constructed from data on the Web and social networks. A personalized sentiment feature was extracted between them. Our proposed system was able to extract major and minor elements of personal features in a feasibility.
test conducted using camera lenses. Personalized polarity is a factor that can increase the accuracy of recommendation systems. The polarity information of a writer can help other people to consider his reviews. In order to demonstrate the performance of the system, we are currently developing an automated system for large-scale testing, and we plan to analyze the temporal relevance of topics in each media.

6 Acknowledgement
The corresponding author is Doo-Kwon Baik.

7 References


