Evaluation of Usefulness of Customer Reviews from Multiple Viewpoints

1910131 Sota Hayato

Recently, customer reviews about products and services become popular due to the rapid spread of online shopping. User's decision on choice of a product to buy is greatly influenced by customer reviews posted by other users who have already used that product. However, there exists both useful and non-useful customer reviews. When a huge amount of reviews are posted in online shopping web sites, it is rather costly and time-consuming to find useful reviews among them. Therefore, it is necessary to develop a technique to automatically evaluate the usefulness of reviews and show the results to users to help their purchase.

In previous work on estimating the usefulness of customer reviews, major methods are training a classifier using a sentence length and part-of-speech as features to determine whether a given review is useful or not. Another studies aim at evaluating the usefulness of customer reviews from a specific point of view, such as extracting comparative opinion and identifying an entity mentioned by a reviewer. However, different users may think what useful reviews are differently. It is insufficient to simply classify whether a review is useful or to evaluate the usefulness of a review from a single viewpoint in order to satisfy various users who have their own demands. Therefore, it is necessary to evaluate the usefulness of reviews on not a single viewpoint but multiple viewpoints.

The goal of this thesis is to develop a system that can not only classify if customer reviews are useful but also clarify what is useful in them or how useful for users they are. More precisely, we propose a system that evaluates customer reviews from multiple viewpoints and displays the results of the evaluation to users. We propose seven viewpoints for evaluating the usefulness of reviews. Our proposed system analyzes reviews in diversified ways by evaluating them from each viewpoint separately. The final system will be designed to provide many useful functions to users, such as to display a list of reviews in order of a score of the viewpoint specified by a user, or to display only reviews that are highly rated from the viewpoint. Our system makes it easier for users to find helpful customer reviews that meet personal preference or requirements of individual users.

Based on the findings of the previous studies on evaluating the usefulness of reviews as well as our past experiences and insight, seven viewpoints for evaluating the usefulness of reviews are proposed: "A reviewer shows reason for his/her opinion," "A reviewer explains a product in detail," "A reviewer compares a product with others," "A reviewer may (or may not) actually

use a product," "A reviewer shows reason for his/her rating," "A review is long" and "A review is easy to read." For a given set of reviews written about a certain product, the system evaluates the usefulness of them from each of the seven viewpoints separately. That is, the system consists of seven subsystems used for evaluation of each viewpoint. In this thesis, we focus on three of seven viewpoints and propose methods to automatically evaluate the usefulness of reviews from them.

To evaluate reviews from Viewpoint1(A reviewer shows reason for his/her opinion), we aim to detect sentences that contain evaluation of a product and reason of it. A rule-based method is designed by considering opinion words (such as "便利です"(*benri-desu*;convenient), "実用的です"(*jitsuyô-teki-desu*;practical)) and keywords (such as "ので"(*node*), "ため"(*tame*)) that are conjunctions indicating reason of something. Precisely, after dependency analysis of a given sentence, the system judges that the sentence includes an opinion to a product and its reason by checking either of the following requirements: (1) a chunk including a word in the form of *renyôkei*(the conjugation indicating that its head is a predicate) modifies another chunk including an opinion word, (2) a chunk including the above keyword (e.g. *node*, *tame*) modifies another predicative chunk. In the evaluation experiment, the test data was constructed by retrieving reviews from the online shopping web site and annotating them with the label indicating whether they express the reviewer's opinion and reason for it. The performance of the proposed system was evaluated on this dataset. The recall, precision, and F-measure were around 0.8, 0.45, and 0.6, respectively. It was found that the recall was relatively high, while the precision was low since the first requirement was often fulfilled in sentences not including reason for a reviewer's opinion.

To evaluate reviews from Viewpoint2(A reviewer explains a product in detail), we define "degree of explanation", a score that represents how detailedly a reviewer explains a product, and propose a method to calculate it. First, for each category of products such as "PC" and "book", keywords relevant to the category are obtained. Nouns and compound nouns are retrieved as the keywords from descriptions of products in the dataset of the e-commerce site "Rakuten Ichiba". In addition, significance of the keyword for the category is measured by using TF-IDF for each keyword. The keywords and their significant scores are stored in the lexicon. Next, for a given review, keywords are extracted by looking up the lexicon, and their significance scores are summed up. Finally, the degree of explanation is calculated by the weighted sum of the total of the significance scores and the length of the review. In the evaluation experiment, for a given pair of reviews written for the same product, the proposed system judged which explains the product

in more detail by comparing the degree of explanation of two reviews.

The accuracy of the proposed method was around 0.77, which was better than the baseline that simply selected a longer review.

To evaluate reviews from Viewpoint3(A reviewer compares a product with others), we propose a method to classify a review if it includes comparison among products. In this study, the method is designed to detect sentences that explicitly represent comparison. The following three types of rules are developed. The first one is a rule to check whether a sentence contains both a keyword indicating comparsion (such as "比べる"(*kuraberu*;compare) and "他のメーカー"(*hoka-no-mêkâ*;other maker) and an opinion word. The second one is a rule using keywords indicating that the reviewer bought a new product to replace old one, such as "買い替え"(*kaikae*;buy to replace). If such a keyword is found in the beginning of the review, it is regarded that the reviewer compare old and new products in the whole review. The third one is a rule to check whether an opinion word is the head of the conjunction "より"(*yori*;than) that often indicates comparison. Results of the experiment showed that the precision of the detection of reviews which contain comparison was sufficiently high for several rules. On the other hand, many rules could detect only a few reviews or no review including comparison, since the number of reviews with comparison was a quite few in the test data.