Estimation of User's Rating from Various Viewpoints on Online Review

ZHANG Bo (1510034)

School of Information Science,
Japan Advanced Institute of Science and Technology

February 10, 2017

Keywords: User's Rating, Viewpoint, Opinion Mining, Supervised Machine Learning.

In recent years, people can post their reviews and give ratings about a product or service in various websites such as online shopping sites. There exists a lot of useful information in the users' reviews for both consumers and enterprises. The consumers can choose their desirable product or service by reading the reviews written by the others or seeing the users' ratings. In addition, the enterprises can use the users' reviews to improve their own products or services, develop a new product/service and decide their management strategy.

However, it requires much labor to read a vast amount of review text. To alleviate this problem, methods called "reputation analysis" or "opinion mining" are widely studied. Opinion mining is a technique to analyze users' comments on a given target (product or service) to reveal the reputation of the target. It typically includes a process to judge whether a sentence expresses a positive or negative opinion for the target. One of the ways to present the results of the opinion mining to the users is to show the rating scores of the target. More precisely, a system predicts and shows the rating score that evaluates the target from a certain viewpoint. For example, screen, battery life and camera are examples of the viewpoints of a smartphone. If the users' ratings for each viewpoint are given, the consumer can quickly know the reputation of it. A goal of this thesis is to

propose a method to estimate the user's rating for the viewpoints from the given review texts.

The proposed method consists of two steps. In the first step, comments (sentences), which express the opinion from the viewpoint toward the target, are extracted from the review. Hereafter, we call them as "opinion sentence for the viewpoint". The review text is segmented into the sentences by splitting it at a position of punctuation such as a period, exclamation mark and question mark. Since punctuation can be expressed by different characters in Japanese texts, they are normalized by simple rules. Then, the sentences are analyzed by Japanese morphological analysis tool MeCab. Next, the opinion sentences for the viewpoint are extracted by checking if it includes a keyword of the viewpoint. For each viewpoint, a set of the keywords of the viewpoint is manually prepared. It consists of words that express the viewpoint itself, words selected from the most frequent 100 nouns in the review corpus, words excerpted from a thesaurus, and words extracted by pattern matching of a coordinate structure. In the second step, the users' ratings for the viewpoint are estimated. For each viewpoint, a model that can estimate the ratings is trained by L2 regularized logistic regression. The model accepts not all the sentences in the review but only the opinion sentences for the viewpoint, which are extracted by the first step, as an input. When no sentence is extracted by the first step, the proposed system does not try to estimate the rating and just output "unknown". Content words and sentiment words are used as features for machine learning. These features (words) are distinguished if they are followed by a negative expression. That is, four types of the features are used: "⟨cw⟩", "⟨cw⟩+NEG", "⟨sw⟩" and "⟨sw⟩+NEG", where ⟨cw⟩ and ⟨sw⟩ represent content word and sentiment word respectively. The sentiment words may be more intimately related to the user's rating than the ordinary content words. Therefore, the weights of ⟨sw⟩ and ⟨sw⟩+NEG are determined as 1.0, while the weights of ⟨cw⟩ and ⟨cw⟩+NEG are determined as 0.2. These weights were defined by intuition.

An experiment is conducted to evaluate our proposed method. Hotel reviews are crawled from a website "Rakuten Travel". In Rakuten Travel, the users can put their ratings from six viewpoints: service, location, room, facilities, bath and meal. The ratings of some viewpoints can be omitted.

In this experiment, the reviews with the ratings for all six viewpoints are used for the training and test data. The number of the reviews is 272,665.

First, the method to extract the opinion sentences for the viewpoint is evaluated. Five hundred reviews are manually annotated with the gold opinion sentences. Precision, recall and F-measure of opinion sentence extraction are calculated. The precision for all viewpoints except for "room" is higher than 90%. The recall of all viewpoints except for "facilities" is higher than 70%. The recall of the viewpoint "facilities" is relatively low due to a lack of the keywords. Since there are many facilities in the hotels, many keywords are related to this viewpoint. However, our method fails to prepare these keywords exhaustively.

Next, the method to estimate the ratings of the viewpoints is evaluated. Two baselines are compared with the proposed method. Baseline 1 is a method that does not extract the opinion sentences for the viewpoint and estimate the ratings with all sentences in the review. Baseline 2 is similar to the baseline 1, but the model is trained with the same amount of the training data as in the proposed method. The accuracy and root mean squared error (RMSE) are measured by 5-fold cross validation. The accuracy of the proposed method is 0.5208, which is 0.0417 and 0.0178 higher than the baseline 1 and 2 respectively. When RMSE is compared for each viewpoint, the proposed method outperforms the baselines with respect to "location", "meal" and "bath", but not with respect to "service", "room" and "facilities". When the accuracy is compared, the proposed method is better than the baselines with respect to all viewpoints except for "facilities". In addition, an error analysis is conducted for randomly chosen 100 reviews where the difference between the estimated and gold ratings is great. It is found that major causes of the errors are (1) the sentences that do not express the opinion are wrongly extracted as the opinion sentences, (2) the extracted opinion sentences are not actually related to the viewpoint, (3) a lack of the sentiment words in a sentiment lexicon and so on.

In future, the proposed method should be improved based on the results of the error analysis. It includes improvement of the extraction of the opinion sentences for the viewpoint, expansion of the sentiment lexicon, precise identification if the sentiment words express the opinion toward the given

target or the viewpoint using dependency parser. The parameter optimization of L2 regularized logistic regression is also required. In addition to the current four features, new features for the rating estimation should also be explored.