

Mining and Summarizing Customer Reviews

Authors: Minqing Hu and Bing Liu

Conference: Science of Knowledge Discovery and Data Mining(SigKDD)

Keywords:

Year: 2004

Report:

1. Abstract:

Merchants selling products on the Web often ask their customers to review the products that they have purchased and the associated services. As e-commerce is becoming more and more popular, the number of customer reviews that a product receives grows rapidly. For a popular product, the number of reviews can be in hundreds or even thousands. This makes it difficult for a potential customer to read them to make an informed decision on whether to purchase the product. It also makes it difficult for the manufacturer of the product to keep track and to manage customer opinions. For the manufacturer, there are additional difficulties because many merchant sites may sell the same product and the manufacturer normally produces many kinds of products. In this research, we aim to mine and to summarize all the customer reviews of a product. This summarization task is different from traditional text summarization because we only mine the features of the product on which the customers have expressed their opinions and whether the opinions are positive or negative. We do not summarize the reviews by selecting a subset or rewrite some of the original sentences from the reviews to capture the main points as in the classic text summarization. Our task is performed in three steps: mining product features that have been commented on by customers; identifying opinion sentences in each review and deciding whether each opinion sentence is positive or negative; summarizing the results. This paper proposes several novel techniques to perform these tasks. Our experimental results using reviews of a number of products sold online demonstrate the effectiveness of the techniques.

Introduction

- The paper's main intention is to summarize product review automatically, by collecting the data from overall customer reviews.
- Data mining and NLP are two base proceedings, which they have used to create their algorithm.
- The article concentrates on different fields such as the text summarization, semantic orientation and the text summarization.
- The work is very essential in the rapidly growing e-commerce and the increase in diversity of goods sold. Each product has to be reviewed separately with the corresponding specific characteristics, which would be impossible by manual entries.

Previous Works

- Former to the work, the studies had been done on the binary classification with thumbs up and thumbs down approach to the products.
- Many of the works also focused in summarizing the review of the products on the whole from the customer inputs. Even mining on the product reviews had not been taken into account.
- The researchers also not focused on the data summarizing with less data and the individual specification reviews of the product. On the contrast, the authors concentrated on extracting the subjective reviews of the product's data on different features of the product.
- Moreover the former works had been doing template instantiation by defining the templates and domains of the classes to which they had to classify. So, these works are limited to the templates. While the authors tried to achieve the independence of the templates.

Content

They have mainly classified their problem into three individual methods as

1. Mining product features of all the reviews given by customers.
2. Identifying the opinions are whether positive or negative.
3. Summarizing the reviews into results.

1. Mining Product Features

- In this process, they mainly take all the customer reviews of a product into the review database and stores there.
- There, in the review database they collect the frequent words of all the reviews to generate the semantic orientation. This database is maintained separately for every product.

2. Generating Semantic Orientation

- After all the reviews are collected to the database, by use of NLP, they search for the adjectives in the sentences, which are named as opinion words in the paper. They also collect the nearby nouns of the adjectives and tag them individually.
- After tagging all the frequent features are taken into consideration. In the paper more than 1% repeated words are considered as frequent words. *Generally WordNet database is used for these type of processes, but there they avoided it because it needs much amount of data for training.
- After collecting the frequent words, a seed adjective is set and all the synonyms are allied to that seed adjective.
- On the other hand antonym seeds are also set and their synonyms are also attached as said before.
- The adjectives which are not found in the seed values or their synonyms, then they are deleted as infrequent words.
- After collecting all the adjectives, based on positive and negative, they are awarded +1 point for positive case and -1 for negative cases.

- Based on these points the algorithm confirms whether the given reviews are positive or negative on the whole. This process is called semantic orientation.

3. Summarizing the opinions

- After stating whether the reviews are positive or negative for a seed word, they now summarize the whole product review into those adjectives and

corresponding nouns for all the frequent words taken.

- Now, they can also collect the main features of the product and output its overall review with this procedure.

Results

They have evaluated the summarization based on three aspects.

- Effectiveness of Summarization of opinion
- Effectiveness of feature extraction
- Accuracy of orientation Prediction of summarization

It is obvious that they have got the better results in the experiment evaluation. The maximum average precision of 0.8 and minimum average precision of 0.56 are resulted after the calculation of accuracy of orientation prediction, which is very satisfactory.

Discussion and Thoughts

The algorithms and requirements are explained very clearly, which makes even a good paper for the beginners.

- Figures were very well designed to illustrate the procedures and steps taken at different times.
- When it comes to the limitations, they have not look into the opinion sentences with pronoun resolutions.
- They have not even handled the nouns and verbs resolutions. The need of highlighting strong opinion is also needed, which they have stated that they are making it for their future prospects.

Conclusion

- On the whole paper appeases all the content required to understand their algorithms and experiments. Their views are well defined in the process. *The results are satisfying and the procedure also seems to be reasonable. However, they were pushing out their limitations by overcoming them in their future prospects.