

Title	Twitterデータを用いた商品価値モデルによるレビュー 検索手法
Author(s)	鄂, 香伶
Citation	
Issue Date	2021-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/17088
Rights	
Description	Supervisor: 由井 蘭 隆也, 先端科学技術研究科, 修士 (情報科学)

Abstract

Recent years, with the popularization of online shopping, there are a large amount of data on product evaluation on the Internet. Because these product reviews are expressed from the consumer's standpoint, these data can help users better understand the product's experience and make choices.

Although a large amount of evaluation data can help users obtain more detailed information, it has become more difficult for users to find useful information from such a large amount of data. As ordinary users, they often feel disorganized when facing the massive amount of information on the Internet. As a result, there have been some websites and platforms designed to provide users with information to post evaluations. However, while providing evaluation information, this type of website also displays product advertisements and purchase links.

This paper proposes an evaluation retrieval method that uses the commodity value model and implements an evaluation retrieval system based on this method. This evaluation retrieval method uses the special feature of 'different product categories have different evaluation points', and these evaluation points are extracted and combined with the evaluation performance to create a product value model. The evaluation retrieval system uses Twitter data and takes three types of cosmetic products as the research objects, namely foundation, lipstick and cheek blush.

The proposal method is mainly divided into two parts. The first part is the construction of the commodity value model, and the second part is the processing of sorting the retrieved evaluation information according to the needs of users. The specific operation of the first part is divided into the following steps. First, analyze the product characteristics of different product types from the evaluation information on the TF-IDF method network, and perform K-means clustering on these characteristics to obtain the evaluation theme. Then the collected 3000 evaluation information was manually tagged to make a data set for machine learning. Finally, use machine learning to train the data set to obtain the commodity value model.

The second part is achieved by automatically scoring the search results. The scoring standard is the evaluation criterion that users input when searching. The higher the score, the more consistent the piece of information is with the user's search expectations. And the search results are displayed on the front page in the order of the scores from high to low.

Finally, the retrieval system is constructed in the order of data storage layer, data access layer, business logic layer and presentation layer. The data storage layer contains database and log files. The data access layer contains database management and log management modules. The business logic layer contains commodity value models and information management modules. The Presentation layer includes front-end display pages and back-end systems.

Two evaluation experiments were carried out in this paper. The two experiments evaluated the proposal technique from two aspects. The purpose of the first evaluation experiment is to verify the search efficiency of the proposal method. @cosme and Twitter Keyword searches were selected as two comparison methods. The experiment

requires five experimenters to set a search target in advance and then use three search methods in any order to search. During the experiment, the time and the number of clicks for the experimenter to reach the retrieval goal were counted and analyzed. The purpose of the second evaluation experiment is to evaluate the automatic scoring process for search results. The experiment uses MAP (mean average precision) index to evaluate. In order to obtain the MAP value, the experiment randomly selected six search targets set by the experimenter in the first evaluation experiment and used the proposal method to search. The first ten content of the search result is manually judged whether the content is related to the search target, and expressed by 1/0, then calculate the AP value and take the average.

The results of the first evaluation experiment show that the proposal method improves user retrieval efficiency in terms of time and operation. Through the classification of the search target, it can be analyzed that the proposal method is especially suitable for the case of inputting product characteristics to search for the corresponding product. The results of the second evaluation experiment show that the automatic scoring process can sort the data according to user expectations. However, it also showed two problems with the proposal method. The first is the lack of distinction between advertising information and evaluation information, and the second is the lack of data for machine learning.

In future research, the proposed method should solve the problem of the distinction between advertising information and evaluation information and the lack of data for machine learning. In addition, because Twitter has the function of publishing pictures, the search results obtained by analyzing text information are limited. It is expected that the information volume of the retrieval system can be increased by adding the analysis function of the image to the retrieval system.