International Journal of Grid and Distributed Computing
Vol. 10, No. 1 (2017), pp.123-136
http://dx.doi.org/10.14257/ijgdc.2017.10.1.12

Research on a Personalized Recommendation Algorithm

Min Yang [1], Yong Ma[2] and Junlan Nie [3,*]

[1,2] Department of Computer Science and Technology, Bao Tou Medical College, Baotou 014040, china
[3]College of Information Science and Engineering, Yanshan University, Qinhuangdao 066004, china
[1]281929119@qq.com, [2]136627228@qq.com and [3]Niejll3@163.com
[3,]Corresponding author*

Abstract

Recommendation system is a new technology to recommend products for customers from huge amounts of products, which infers those objective users' preferences based on their personal information or online behavior. This paper studied the main personalized recommendation technology for current E-commerce. It proposed a hybrid recommendation algorithm based on opinion mining. This system combines web data mining technology, that is, takes advantage of user-generated-content by mining customers' online reviews. It is well known that online reviews can directly reflect customer's real emotion and expectation, so it's appropriate to extract a customer's latent interest and preference from his/her reviews, thus refine recommendation and improve accuracy. Meanwhile, an experiment was conducted and the result demonstrated that our system could generate a reliable and realistic recommendation.

Keywords*: Recommendation System, Personal Recommendation, E-commerce, Text Mining*

1. Introduction

In e-commerce environment, in the face of tremendous commodities and information provided by Internet merchants [1-6], people have plenty of choices, however, not possible to check and select commodities as in real environment, falling into the dilemma of difficult selection and waste of time and energy. Either traditional search engine like Baidu, Google or commodity search engine like Etao [6-10], it retrieves enormous information according to user input key words. Even so, the searched commodity information is universal and it's selected by user with lots of time and energy. Statistics show that before there is no recommendation system helping make decision, user needs check on average 11.7 articles before finding satisfactory merchandise. After introduction of recommendation mechanism, user needs check on average only 6.6 items, almost reducing workload by 50%. US ChoiseStream company found that after survey, 45% consumers prefer choosing websites which support recommendation function [11-19]. For high-end consumers, the percentage reaches up to 69%. If Internet merchants want to increase client loyalty and possession rate, it requires considering how to let user reduce online shopping time and purchase commodity to its heart desire [20-25].

Individualized recommendation system plays following functions in electronic business: (1) Attempt to convert commodity browser into purchaser;

(2) Increase cross selling to boost sales amount;

(3) Increase user adhesiveness to enhance user satisfaction and loyalty. On the e-commerce platform, it's not hard to find that websites which can provide personalized recommendation system can yield more throughput than common e-business websites.

ISSN: 2005-4262 IJGDC

The paper takes opinion mining technology, from the perspective of users' online review, to extract commodity features which are interesting to users as to predict their preference for some article [26-30]; then recommend to them items which are in exact accordance with their interest by providing intelligent shopping guide. To be specific, offer to them items which can meet their fuzzy requirements, try to understand user better, help raise website flow, and increase user's adhesiveness, as it has very strong and practical value [30-32].

The individualized recommendation based on online review opinion mining aims to recommend merchandise which accords the most to users' preference with the use of real user review data at Internet e-commerce website to find their interest or preference through analysis of their review information [33].

The Internet has dramatically changed the way people express their opinions. Many commercial sites such as Amazon, Taobao and so on, allow users to provide product evaluation. In addition, in the BBS, blog, micro-blog and other Internet applications can also see the user to express their views. These comments and opinions are often referred to as user generated content or user generated media. These texts can reveal a variety of valuable information, but because the text is unstructured data, makes it as easy to handle as structured data.

Opinion mining is a cross type technology of data mining, text mining, Natural Language Processing, and artificial intelligence. It also has strong practicability in the process of text data.

Opinion mining is based on the understanding of the subjectivity of the text in the web, and it is more active in the research of foreign countries. The main opinion classification, mining, feature based on the viewpoint of comparative sentences and correlation mining, view search and view the sub field of fraud

Opinion classification: the task of mining as a typical classification of the problem, that is, in accordance with the positive evaluation, negative evaluation, neutral evaluation and other emotional colors will be divided into different types of text

For example, a database of the evaluation of a product data, the user evaluation of the system to determine the views of the user is positive or negative. This classification is the evaluation of the overall level of the document, and cannot take into account the preferences of the specific content of the user.

Based on the characteristics of the point of view: the task began to be specific to the level of the statement to further find the details, that is, to determine the user like or not like the characteristics of the object or other key content. Object is the product, service, subject, individual and organization of a broad sense.

For example, given a section of the review, the need to identify specific features of the product is reviewed, as well as these comments are positive or negative. In the sentence "this camera's battery capacity is very big", the evaluation is aimed at "the battery capacity", contains the emotion is positive.

Comparative sentence and comparative study: comparison is a common way of assessment. In this assessment, an object directly and one or more of the alternatives are compared to the established criteria.

For example, the A camera's battery capacity is smaller than the B camera, need to find these comparative sentences and analyze the relationship between the expression of the sentence.

Point of view search and point of view fraud: a view of the search system allows users to search for any object on the point of view.

A fraudulent view is that some people who have made the point of view that they are not realistic or malicious to promote their own products or services, or to harm the reputation of their competitors. As the true identity of the network users are unknown, making fraud detection challenging.

Since the beginning of 2004, Liu Bing and other scholars in their series of studies, the identification of product features is as an important work of opinion mining, and in the mining of the characteristics of the basis of emotional analysis

This thought quickly attracted the attention of many scholars. How to improve the accuracy of product feature extraction, how to learn the features, how to accurately obtain the opinion words, has become a hot research direction in the field of opinion mining.

Opinion mining is the mining of the paragraph of the subjective sentence, so in the process of mining, to screen out the objectivity of the text does not contain emotion. User comments are divided into four elements: Topic, Holder, Claim, emotion, on the basis of this, the main steps of opinion mining are:

Identification of an opinion topic: identifying key topic terms and extracting ontology concepts in the field;

The identification of the Holder: the author and the other subjects involved in the comments;

Claim of choice: planning the general scope of the statement, the range or extent, and screen to get rid of the emotional and unrelated to the objective statements;

Sentiment analysis: determining the emotional tendency of the holders of the comments contained in the statement of opinion statement.

2. Objective of Individualized Recommendation

The objective of individualized recommendation is to boost sales on the basis of recommending items which can satisfy users' preference through collecting and analyze individual consumers' online behavior and purchasing record data, and withdraw their potential preferences. So far the common individual recommendation algorithm includes recommendation based on content and collaborative filtering recommendation. Content-based recommendation does to users items which match well with their preferences after acquiring their interests through recording their online browsing (like web pages they often click and what time they click) and purchasing logs. Collaborative filtering algorithm, as its name implies, firstly analyzes user interests, then utilizes collaborative thinking to recognize users who have similar preference with target user or item similar to recommended one, next combines those similarity information to screen out items which are possibly interesting to users to finally complete prediction. GroupLens research team developed collaborative filtering system based on user rating. By advantage of its tremendous database information, the system is used to recommend movie and news, e.g. Douban, which recommends movie and music.

Individual recommendation technology [34-35] is the kind of technique to make recommendations to relative users after analysis of contents and relationship based on extracted features of objects. At present, content-based recommendation technology and that based on collaborative filtering are facing some problems although they can initially meet requirements of recommendation [36-37].

The content-based recommendation technology is often restricted to the condition of acquiring recommended object's characteristics. To be specific, if a movie is recommended, it needs to have information such as film title, film type, director, casts, even contents or keywords. Comparatively, collaborative filtering technology doesn't require any descriptive information regarding object's feature nor is affected by such information whether it's correct or not; instead, it totally depends on user's scorings of objects [38-39]. Even though there's complete descriptive information regarding the object, some studies have demonstrated that the collaborative filtering approach reaches better recommendation effects than content-based solution.

To overcome the limit, the paper proposes an individual recommendation algorithm based on user online review, i.e. full use of user's review on item. Since a majority of users would make more comments on contents to their attentions or show stronger affections when they're reviewing items, we think product feature which is marked too much of comments on item in one category is the feature that current user concerns potentially. The idea of the algorithm: from reviews obtain feature of item of the kind interesting to user to search horizontally other users who have higher interest similarity with it; then compare vertically the item of the kind which gains the highest score among those product features as to recommend it to user.

3. Individualized Recommendation Model based on Online Review Mining

The major task of this paper is to offer personalized recommendation to user with relative recommendation system technology through analyzing user interest and the similarity between users after comparing product's similarity and good/bad points by decomposing user review data into granularity of product feature, with the use of such data about one product at one e-commerce website. The main thinking is: create one review analysis and user recommendation model based on product feature and analyze key features of each kind of product. The prerequisite of assumption of the method is: when user wants to buy one item and refers to other user comment, he cares more about article features which are more interesting to him. Likewise, when a user comments an item, he would make more and specific comments on features which he concerns much. Therefore, we assume of all comments made by a user for the same category, the product feature he comments the most is his concerned feature. In the meantime when a user's rating of concerned item is generally lower, he would make higher request. Different from other relative works on traditional recommendation system, this paper discusses mainly text data about product comment through relative model, rather than merely relying on user's overall rating data on product. A big contribution to do in that way is user's reviews or opinions can be specifically studied on the level of product feature and that system's analysis of product and user becomes more accurate and user interest can be meticulously defined.

After data pre-treatment, the system acquires structural user data, product feature data, product rating data, then statistically generates user vector, commodity collection and goods total score. In the next, we introduce the working flow of proposed individual recommendation model:

(1)Feature extraction: withdraw features of all commodities and obtain each user's overall rating of goods and average level of feature rating;

(2)User similarity calculation: with user review vector, calculate user's weight of product feature; then get users with higher similarity through user data analysis;

(3)Commodity similarity calculation: calculate commodity with higher similarity based on its feature data.

(4)Generate recommendation: from recommendation result by user with higher similarity, get that of commodity which is highly expected by target user; validate with commodity rating data.

3.1. Feature Extraction

During feature extraction, when we only have few commodity features, we can use Bootstrapping method to learn the text. Bootstrapping is a semi-supervised learning technique, with the main thinking: use a few man defined commodity feature words as feature's seed set, then according to learned pattern, withdraw relative glossaries from unlabeled text; labeled words are used as seed set which can join the next cycling process. Through specified iteration times, or till no new seed produces, the cycle ends and output extracted words and patterns. The merit is lower requirement for training set. With only a little sample information, it's possible to learn automatically patterns relevant with annotated information.

3.2. Calculation of Commodity Similarity

Commodity features derive not only from user review data and also from product's ontology structure, which in turn reflects details of one item. To get it simple, after extracting product features, we make only use of item features fetched from user reviews, which can help easily generate Item-Feature model. For item of one category, all its features are expressed with F vector. $F=\{f_1, f_2, ..., f_p\}$.

There are many methods for calculating item similarity, of which cosine similarity is adopted to compute Item-Feature matrix, i.e. use commodity' characteristic item to represent commodity. Matrix value can be score of one item acquiring from one of its feature item or weight of one feature item.

3.3. Calculation of User Similarity

The core of collaborative filtering method is calculation of similarity. Since the paper proposed is personalized recommendation based on opinion mining, which bases on product feature, so when similarity between users is calculated, we introduced User-Feature model, hereafter named UF model.

UF model records each user's rating information about items. User-feature matrix represents the matrix of comments by user i for feature k, marked as $UF_{m\times p}=\{UF_{ik}\in UF \mid 1<i<m, 1<k<p\}$. The value of UF_{ik} is decided by experimental requirement. If UF_{ik} records rating result by user for one feature k, UF matrix reflects user's visual evaluation of one item's specific feature; if UF_{ik} records weight of user i for feature k, UF matrix reflects the degree of its preference for the feature.

We explain by citing UF as weighting matrix of user for one feature to calculate user similarity with Pearson correlation coefficient in the formula:

$$sim(i,j)=\frac{\sum_{k=1}^{n}(UF_{i,k}-\overline{UF}_i)(UF_{j,k}-\overline{UF}_j)}{\sqrt{\sum_{p=1}^{n}(UF_{i,p}-\overline{UF}_i)^2}\sqrt{\sum_{q=1}^{n}(UF_{j,p}-\overline{UF}_j)^2}} \tag{1}$$

If we use given item feature to analyze, i.e. all users' item features of the same dimension, then user similarity can be computed by cosine similarity method in the formula:

$$sim(i,j)=\cos(i,j)=\frac{UF_i\times UF_j}{\|UF_i\|^2\times\|UF_j\|^2}=\frac{\sum_{k=1}^{n}UF_{ik}UF_{jk}}{\sqrt{\sum_{p=1}^{n}UF_{ip}^2}\sqrt{\sum_{q=1}^{n}UF_{jq}^2}} \tag{2}$$

3.4. Generate Recommendation

Use-item matrix is the commonest matrix used to stand for user i's rating of item j, put as $R_{m\times n}=\{R\in R_{ij}\,|\,1<i<m,1<j<n\}$ .User collection is $U_m=\{u_1,u_2,...,u_m\}$; item collection is $I_N=\{i_1,i_2,...,i_N\}$. Generally speaking, R matrix records rating numbers of a user for one item j.

$$R_{m\times n}=\begin{bmatrix} r_{11} & \cdots & r_{1n} \\ \\ r_{m1} & \cdots & r_{mn} \end{bmatrix}$$

With user similarity formula and R matrix, we can predict target user's possible scoring value for item it has not bought, with the equation:

$$p_{u,s}=\overline{r}_u+\frac{\sum_{u^{'}\in U^{'}} sim(u,u^{'})\times(r_{u^{'},s}-\overline{r}_{u^{'}})}{\sum_{u^{'}\in U^{'}} sim(u,u^{'})} \qquad (3)$$

4. Experiment Design and Discussion

4.1 Experimental Environment

Most experiments here are completed in Eclipse compilation environment with Java language. Since the experiments use both Chinese and English reviewing data, the author pre-processes Chinese data by applying Chinese morphological analysis system ICTCLAS during word segmentation and tagging, which is open source software furnished by CAS Institute of Computer Technology, with Hierarchical Hidden Markov Model HHMM as core of word segmentation technology, including Chinese segmentation module, part of speed tagging module, named entity recognition module and new word recognition module. The system has wide applicability and supports user custom dictionary and several coding schemes including traditional Chinese, GBK, UTF-8,UTF-7,Unicode. The newest version of ICTCLAS realizes the segmenting speed theoretically up to approximately 500KB/s by single machine and the precision reaches up to 98.45%. Moreover, it supports all around the application and development in all kinds of environments. It is shown in table1.

Table 1. ICTCLAS Word Segmentation Results Sample

Word segmentation	The screen is awesome, very clear, can download the software very much, the reaction is very good, the camera function. Personal feeling is good. The sound is clear, put the volume or not so awesome. Battery usage too Quick, boot slow, easy to report the wrong exit when playing the game. Small keyboard usage is very small, even if the use of keys Disk button is too small, the error rate of personal feeling to make 99.5%. Mobile phone built-in ROM is too small
Result of word segmentation	The screen /n is awesome /d /n, /wd /d /a /wd very clear, /v can download the /v software /ude1 /n

	Pretty much /a /ude1 /vi, /wd reaction /d also /d pretty good /a /ude1 /d, /wd camera /vn Function /q /n /n people feel /n is also a good /vshi /a /ude1 /d. /wj sound /n clear /a, /wd /f /v /n or /c /d is not the volume of /vshi /d /n is awesome. /wj battery /n enable With /v degree /qv too /d fast /a, /wd boot /vi slow /a, /wd play /v games /ng /n easy /a /n wrong /vd exit /v. /wj small /a keyboard /ude1 /n use /vn degree /qv very /a small /wd, /d /d on the /v to use the /v keyboard /ude1 /n /d button /n too small /a /y, /wd /v by /p Rate /q /v people /n feel /n make 99.5%/m /v. /wj mobile /n /f /v inside ROM/x /d too small /a

4.2. Experiment of Personalized Recommendation

4.2.1Personalized Recommendation Review Data. Review data used here is online hotel ratings, which are English data captured by TripAdvisor during March 14, 2015 to May 15, 2015, totaling 235,793 pieces. Those reviews were made by users after self experience in the commodity, including many emotional vocabularies which are complimentary or derogatory. Through initial artificial filtering, partial invalid comments are removed. The dataset used as experimental testing data is quite reasonable.

English text data and Chinese text data are treated in different way. We need to do following jobs:

(1) shift capital letters to small ones;

(2) eliminate punctuation, stop words, uncommon phrases (frequency <5);

(3) use porter stemmer to extract root.

Feature extraction algorithm

Input: Review data set $\{d_1, d_2, ..., d_{\|D\|}\}$, Characteristic seed word $\{T_1, T_2, ..., T_k\}$,Dictionary V, Selection threshold p, Iteration number I
Output: According to the characteristics of goods isolated comments
Setp1: Separate comments from sentences, $X = \{x_1, x_2, ..., x_M\}$
Setp2: In X, each sentence is matched with a set of seeds, and Record matching rate is Count (I)
Setp3: Mark a feature label for each sentence $a_i = \arg\max_i Count(i)$, if it has the same value, it will have more than one label for the sentence.
Setp4: Calculate the x^2 value of each word in V
Setp5: Sort the words in each feature according to the x^2 value, Top-p the words into the characteristics of the seeds of the word list T_i
Setp6: If the seed characteristic word list is no longer changed, or the number of cycles is I, to step 7, otherwise to step 2.
Setp7: Output sentences with marked feature words.

During feature extraction, we firstly define artificially a group of seed feature words; next use bootstrapping algorithm to add relative features. In the experiment, we choose threshold p and p=5, iteration I=10 times. Table2 lists out initially artificial defined feature words.

Table 2. Manually Defined Seed Feature Words

Features	Seed characteristic word
Value	value, price, quality, worth
Room	room, suite, view, bed
Location	location, traffic, minute, restaurant
Cleanliness	clean, dirty, maintain, smell
Check in/Front Desk	stuff, check, help, reservation
Service	service, food, breakfast, buffet
Business Service	business, center, computer, internet

If to avoid calculating missing value of vector as to get better calculation result, it needs each review must contain description of 7 kinds of features mentioned above. In that case, only 184 hotels and 780 pieces of reviews can meet that condition. From the whole quantity of reviews, the big number of reviews is adequate to manifest the average rating level of user group for hotels. Hence the experiment combines all reviews of the same one hotel to form holistic review (h-review) as to estimate the average rating value of each hotel. After those jobs, it remains about 1900 hotels and 108,891 pieces of review. Table lists the general statistical information of the experimental data.

Table 3. Corpus Statistics

Number of hotels	1900
Comment number	108,891
Average number of sentences per comment	8.45+0.2
Average number of words per feature	9,58+5.2

After word cutting, hotel's source data contains following metadata. After BootStrapping, we fetch out a few core fields which fit model calculation, each review formalized to a Vector. Review, Aspect, Vector respectively contains fields as below. It is shown in table3

Table 3. Data Format

Data format	Include field
Review	Author, Content, Date, Number of Reader, Number of Helpful Judgment, Overall rating, Value aspect rating, Rooms aspect rating, Location aspect rating, Cleanliness aspect rating, Check in/front desk aspect rating, Service aspect rating and Business Service aspect rating
Aspect	Author, Content, Date, Ratings (Overall, Value, Rooms, Location, Cleanliness, Check in/front desk, Service and Business Service) and Aspect Segments (Overall, Value, Rooms, Location, Cleanliness, Check in/front desk, Service and Business Service)
Vector	Hotel_ID, Overall_Rating, Value_Rating, Room_Rating, Location_Rating, Cleanliness_Rating,

	Check_in/front_desk_Rating, Service_Rating, Business_Service_Rating

Because the Review and Vector vectors are too complex, here only the Aspect vector is listed as an example. It is shown in Figure1.

```
<Author>JaneMichigan009
<Content>High Tech Room We had a 20 hour layover in Amsterdam and Citizen M was perfect. The
staff was wonderful. The room was like nothing we had ever stayed in before. The rooms are tiny
- like in New York City but it seems spacious and very cleaverly put together. Enjoyed every
minute of our stay. It is a short walk to airport and train station. FABULOUS!
<Date>Jan 3, 2009
<Rating>5      5      4      5      5      4      4      5
<Aspects>
3      29(minute):1     51(enjoy):1     270(fabulous):1
0
10     41(city):1       60(hour):1      74(perfect):1   111(spacious):1 338(tiny):1     361
(amsterdam):1   424(york):1     3836(tech):1    3917(layover):1 3944(citizen):1
5      7(walk):1        82(station):1   95(airport):1   144(short):1    182(train):1
0
2      0(staff):1       48(wonderful):1
0
0
```

Fig1. Aspect vector instance

Each hotel has a unique number ID, in the source data, all true score data are marked for 0-5 points, -1 points is empty data, there is no score in the web page. Finally, 263 features were selected, and 137653 effective users were selected.

4.2.2. Experimental results of individualized recommendation . After extraction of product features, annotate semantic contents of review corpora: product feature word, opinion word and degree word; obtain frequency of semantic contents and length of review text. To put experimental procedure more explicit, we firstly define following nouns then describe the process of the algorithm [40].

(1)User set: $U_m = \{u_1, u_2, ..., u_m\}$

(2)To be recommended (hotel): $I_N = \{i_1, i_2, ..., i_N\}$

(3)Product feature set: $F = \{f_1, f_2, ..., f_n\}$

(4)m users on the n merchandise (hotel) score: $R_{m\times n} = \{R_{ij}\}$

Personalized recommendation algorithm

Input: Review data: $\{d_1, d_2, ..., d_{\|D\|}\}$, Nearest neighbor number k, Recommended number p;
Output: p hotel is recommended by the target user u;
Setp1: Get the user set, the product set, the overall score.
Setp2: Extract the characteristics of the goods, to obtain the overall evaluation of each user to the evaluation and the evaluation of the overall evaluation of the number of signs;
Setp3: Generate User-feature matrix, using TF-IDF algorithm to calculate the preference matrix C;
Setp4: Based on matrix C, using Pearson correlation coefficient to calculate the similarity of users, to select the Top-k nearest neighbors of the target user;
Setp5: Generate the User-Item matrix, and generate the total score matrix R;
Setp6: Predict the target user of the product has not yet commented on the score, ranking top-p product will be recommended to the target user.

The experiment is made to examine the accuracy of predictive rating value of commodity waiting for recommendation. Since experimental data differs from common numeric type data used by traditional recommendation system, so the selection of evaluation indicators is largely limited. Here we use mean absolute error MAE to evaluate. MAE in essence measures the error between the predicted rating by the proposed model and user real rating, so it's quite suitable to confirm the effectiveness of the proposed algorithm. In the experiment, we note that the choice of similar user number k has effect on result of MAE. Here, we define k=5, 7, 10 separately to validate the result. Table4 lists comparison of partial prediction results and user's true score.

Table 4. The Recommended Results of the Predictive Score and the True Score

Hotel_ID	User	True score	K=5	K=7	K=10
618542	Blaize	5	4	4.5	4.5
	circuit	5	5	5	5
74806	Boeing	4	4	4	3.5
	Blaize	4	4	4	3.5
	Scott	3	4	4	4
256659	Harvsman	4	3.5	3.5	3.5
	Boeing	4	4	4	4
81444	Amy	4	4.5	4.5	4.5

It can be seen from the table4, when k=5, the average accuracy of the predictive score is the higher. From Table 5, we see when k=5, the method gets the best result. That is because in practical data, over big user group and commodity can inevitably lead to the problem of data sparsity. Only the users with highest similar interest are chosen, it has the ideal prediction result. Therefore, when data is sparse, the threshold for choosing the nearest neighbor should not be set too high.

Table 5. MAE in Different K Values

Nearest neighbor K	5	7	10
MAE	0.55	0.56	0.58

4.2.3 Online recommendation based on multi objective decision making.

1 Multi objective recomme nded review data

The data used in this paper is a review of the online mobile phone, data from the http://www.it168.com website. It is China's largest individual and enterprise IT products to buy, interactive web site, the data is relatively objective, comments are more professional

We downloaded from the site to download the 4 equivalent price of mobile phone reviews. Such as Nokia-N96, Nokia-N97, Nokia-C6, XT702 MOTO. For each cell phone, we randomly selected 100 comments.

The experimental selection of such data is motivated by the following considerations:

(1). The 4 phones have a large number of online reviews.

(2). The 4 phones are at the same price, but also has a consistent configuration.

(3).The average score of these 4 phones are similar, it is difficult for potential consumers to make a purchase decision.

2 Multi objective recommended experimental results

Table 6 lists 6 frequent features extracted from the mobile phone data, the second column lists the original frequencies that appear in online reviews. The third column lists the approximate words of these features, in a very common phenomenon in Chinese synonyms. We have screened out with the characteristics of synonyms together with the first column. The final column lists the final frequency of the commodity characteristics.

Table 6. Frequent Features of Mobile Phones

Features	Original frequency	Homoionym	Final total frequency
keyboard	150	Key key Button	218
Battery	207		208
screen	117	Shell	178
Effect	83	Sound effects / sound effects, sound quality / tone / volume	166
function	138		137
Price	112	Price, value	128
Appearance	80	Style, appearance Modelling Color	115
Memory	96		96
system	93		93
Software	89		89
Displayer	74		74
Speed	71		71
Design	68		68
Photograph	39		66
video	63		63
Camera	62		62

Table 7 lists the fuzzy decision matrix and feature weights in this experiment, which indicates the importance of the user's attention to the product characteristics in the comments, as the input data of PROMETHEE Fuzzy. Table 4 is listed as 7 models of the phone's fuzzy positive output flow, fuzzy negative output flow, net output stream. By calculating the expected value of the net output stream, it is easy to get the ranking results of the goods. MOTO-XT702＞Nokia-N97＞Nokia-N96＞Nokia-C6。

Table 7. Fuzzy Decision Matrix and Feature Weight

Features	Weight	Fuzzy decision matrix			
		Nokia-N97	Nokia-N96	Nokia-C6	MOTO XT702
keyboard	0.119	(0.311,0.501,0.682)	(0.191,0.379,0.575)	(0.292,0.486,0.682)	(0.194,0.379,0.575)
Battery	0.113	(0.124,0.310,0.509)	(0.094,0.258,0.458)	(0.137,0.318,0.518)	(0.178,0.371,0.570)
screen	0.098	(0.385,0.583,0.766)	(0.304,0.5,0.695)	(0.284,0.483,0.679)	(0.439,0.639,0.813)
Effect	0.091	(0.295,0.488,0.683)	(0.407,0.605,0.793)	(0.262,0.444,0.635)	(0.504,0.704,0.864)
function	0.075	(0.361,0.558,0.751)	(0.388,0.587,0.777)	(0.359,0.559,0.749)	(0.394,0.593,0.783)
Price	0.07	(0.133,0.311,0.510)	(0.133,0.314,0.514)	(0.212,0.401,0.601)	(0.229,0.422,0.619)
Appearance	0.63	(0.460,0.660,0.828))	(0.334,0.530,0.719)	(0.352,0.550,0.745)	(0.458,0.658,0.846)
Memory	0.052	(0.369,0.561,0.746)	(0.482,0.682,0.869)	(0.198,0.384,0.584)	)(0.140,0.340,0.540)

system	0.051	(0.409,0.609,0.793)	(0.273,0.465,0.657)	(0.317,0.515,0.714)	(0.388,0.588,0.775)
Software	0.049	(0.207,0.396,0.591)	(0.403,0.603,0.800)	(0.302,0.500,0.692)	(0.331,0.531,0.724)
Displayer	0.04	(0.289,0.485,0.677)	(0.253,0.447,0.647)	(0.231,0.417,0.616)	(0.445,0.645,0.82)
Speed	0.039	(0.324,0.512,0.698)	(0.373,0.568,0.758)	(0.38,0.58,0.76)	(0.547,0.747,0.913)
Design	0.037	(0.364,0.558,0.754)	(0.303,0.498,0.688))	(0.293,0.487,0.685)	(0.308,0.508,0.704)
Photograph	0.036	(0.312,0.512,0.799)	(0.3,0.5,0.695)	(0.338,0.533,0.725)	(0.203,0.386,0.586)
video	0.034	(0.305,0.505,0.7)	(0.462,0.662,0.846)	(0.154,0.328,0.522)	(0.348,0.548,0.744)
Camera	0.034	(0.31,0.509,0.709)	(0.315,0.515,0.703)	(0.338,0.538,0.728)	(0.189,0.389,0.584)

The method proposed in this paper has a practical effect on the online reviews of the users. From the experiments, we see that the final sorting results are consistent with the average ranking of mobile phones on the IT168 website.

Moreover, the fuzzy decision matrix of PROMETHEE fuzzy calculation, and the index weight can also provide the key commodity feature recognition, which can provide the depth information of the merchandise.

For example, the four mobile phone top-5 hot features are: "keyboard", "battery", "display", "effect", "function" ".

That is to say, the user is most concerned about the mobile phone feature is the 5 characteristics. And then we can also come to colleagues, MOTO-XT702 in the battery, sound, screen, appearance of these features are significantly better than the other three phones.

5. Conclusion

This paper presented a personalized recommendation model based on online review. Using real data, 2 kinds of empirical tests are carried out. Based on the opinion mining for personalized recommendation, the experimental results show that the proposed model can effectively predict the user's expectations, and can achieve good results. The results of the experiments are consistent with the real product ranking and description, and it also proves the effectiveness of multi objective decision making in opinion mining.

Acknowledgement

This work was supported by National 863 high technology research and development plan, Hebei province science and technology support project(1420311D).

References

[1] Z. Narisa and T. Jiafeng, “Strong. Based on the emotion lexis of online product reviews personalized recommendation method”, Zhengzhou University Journal (Natural Science Edition), vol. 2, **(2011)**, pp. 48-51.

[2] Z. Narisa and Jiafeng, “Online consumer reviews”, Fuzzy intelligent product recommendation system based on system engineering, vol. 11, **(2013)**, pp. 116-120.

[3] X.-Q. Tan and H. Shan, “A review of the research on music personalized recommendation system”, Modern library and information technology, vol. 9, **(2014)**, pp. 22-32.

[4] W. Wei, W. Hongwei and M. Yuan, “Collaborative filtering recommendation algorithm research: Considering online review of affective tendency”, System engineering theory and practice, vol. 12, **(2014)**, pp. 3238-3249.

[5] H. heYan, “Visual analysis of online social network”, Journal of the Chinese Academy of Sciences, vol. 2, **(2015)**, pp. 229-237.

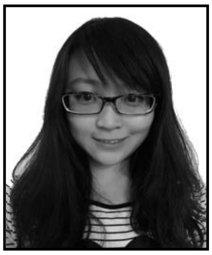
[6] Y. Ming, Q. Wei, Y. Xiangbin and L. Yijun, “Journal of Management Sciences in China analysis of online product reviews the utility”, vol. 5, **(2012)**, pp. 65-75.

[7] H. Xinming and X. huosong, “Online reviews in the preference recognition method of commodity attribute”, User information magazine, vol. 9, **(2012)**, pp. 197-201.

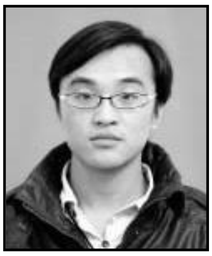
[8] C. dazhi and J. Lee, “Online review sentiment corpus construction”, The Chinese Academy of Social Sciences Journal of the graduate school, vol. 4, **(2010)**, pp. 119-123.
[9] T. X. Bo and Z. Zhao, “Based on mixed graphs online social network personalized recommendation system”, Information theory and practice, vol. 2, **(2013)**, pp. 91-95.
[10] W. Wei, H. Xiaodan, S. Yanguang and Z. Jijun, “Study on the affective recommendation system”, information and control, vol. 2, **(2013)**, pp. 218-228.
[11] L. Zhiyu, “Study on the utility order model of online commodity review”, Modern library and information technology, vol. 4, **(2013)**, pp. 62-68.
[12] X. Wei and G. Yangjie, “The text mining of hotel customer online reviews”, Journal of Beijing International Studies University, vol. 11, **(2013)**, pp. 38-47.
[13] L. Yuming, W. Xiaoling, Z. Tao and Z. Aoying, “Review of the research on the quality control and detection of user reviews”, Journal of software, vol. 3, **(2014)**, pp. 506-527.
[14] Z. Lei, C. Zhenxiang and Y. Bo, “Analysis and prediction of the personality of social network users”, Journal of computer science, vol. 8, **(2014)**, pp. 1877-1894.
[15] Z. Yujie, D. Yulu and M. Xiangwu, “Group recommendation system and its application”, Journal of computer science, vol. 4, **(2016)**, pp. 745-764.
[16] X. Xia, “Collaborative filtering algorithm in personalized recommendation technology”, Ocean University of China, **(2011)**.
[17] T. Xiaohong, “Personalized Course recommendation and genetic algorithm”, Shanghai Jiao Tong University, **(2013)**.
[18] C. Jiemin, T. Yung, L. Jianguo and C. Yibin, “Research on personalized recommendation algorithm”, Journal of South China Normal University (Natural Sscience Edition), vol. 5, **(2014)**, pp. 8-15.
[19] Z. Zhijun, “The model and algorithm of personalized recommendation in social network”, Shandong Normal University, **(2015)**.
[20] L. Haixia, “Electronic commerce personalized recommendation algorithm based on ant colony clustering”, Shandong Normal University, **(2014)**.
[21] Y. Li, “Research on Personalized Recommendation Algorithm of collaborative filtering based on clustering”, Huazhong Normal University, **(2014)**.
[22] Z. Huaqing, W. Hong, T. Zhaoming and M. Xiaohui, “The personalized recommendation algorithm in multi dimension weighted social network”, computer application, vol. 9, **(2011)**, pp. 2408-2411.
[23] Y. Ting, “Personalized recommendation based on collaborative filtering algorithm”, Beijing Institute of Technology, **(2015)**.
[24] C. Yibo, “Research on Key Technologies of personalized knowledge service based on association data and user ontology”, Wuhan University, **(2012)**.
[25] Y. Junwei, H. Xun, L. Min, Z. Yanbo and N. Haibin, “Ontology model of user interest based personalized recommendation algorithm. Computer integrated manufacturing systems”, vol. 12, **(2010)**, pp. 2757-2762.
[26] H. Qiuting, “Research on the personalized recommendation algorithm based on statistics”, Xiamen University, **(2014)**.
[27] W. Wang, “Recommendation algorithm based on the location of social networks in the personalized location”, Xiamen University, **(2014)**.
[28] L. Xueqin, “Research on Personalized Recommendation Algorithm Based on location service”, South China University of Technology, **(2012)**.
[29] Y. Jing, “User interest model and real-time personalized recommendation algorithm research”, Nanjing University of Posts and Telecommunications, **(2013)**.
[30] W. Wenhao, “Research and application of personalized recommendation algorithm based on mobile library”, China University of Geosciences (Beijing), **(2015)**.
[31] W. Cong, “Recommended. Research on Algorithm of personalized user behavior”, Harbin Institute of Technology, **(2015)**.
[32] F. Pengcheng, “Research on Personalized Recommendation Algorithm Based on context aware”, Donghua University, **(2014)**.
[33] Z. Lan, “Research on Personalized Recommendation Algorithm Based on collaborative filtering”, Huazhong Normal University, **(2009)**.
[34] L. Qingwen, “Research on the recommendation algorithm based on collaborative filtering”, University of Science & Technology China, **(2013)**.

[35] S. Guangfu, W. Yue, L. Qi, Z. Chen and C. Enhong, "A collaborative filtering recommendation algorithm based on sequential behavior", Journal of software, vol. 11, **(2013)**, pp. 2721-2733.
[36] S. Geng, "Research on E-commerce Recommendation Algorithm Based on collaborative filtering", Shandong Normal University, **(2014)**.
[37] H. Yang, "Research on Collaborative Filtering Recommendation Algorithm Based on item clustering and preference categories", Zhejiang Sci-Tech University, **(2014)**.
[38] J. Herlocker, J. Konstan and L. Terveen, "Evalutaing Collaborative Filtering Recommender System", ACM Trans on Information System (TOIS), vol. 22, no. 1, **(2004)**, pp. 5-53
[39] R. Burke, "Hybird System for Personalized Recommendations", Intelligent Techniques for Web Personalization. Springer Berlin Heidelberg, **(2005)**, pp. 133-152.
[40] H. Yongwen, "Research on Key Technologies of Chinese product reviews mining", Chongqing: Chongqing University, **(2009)**.

Authors



Min Yang, She received her B.S degree from Inner Mongolia Normal University and received her M.S degree from Yan Shan University. She is a Lecturer in Department of Computer Science and Technology, Bao Tou Medical College. Her research interests include Graphics and Image Processing.



Yong Ma, He received his B.S degree from Tai Shan College and received his M.S degree from Inner Mongolia Technology University. He is a lecturer in Department of Computer Science and Technology, Bao Tou Medical College. His research interests include Smart Campus and Internet of things.

Junlan Nie, She received her B.S degree from Hebei University and received her M.S degree from Hebei University of technology. She got her Ph.d from Hebei University of technology. She is a professor in College of Information Science and Engineering, Yanshan University. Her research interests include Scientific visualization, visual analysis.

