Analysis of document recommendation based on Hierarchical analysis method and BP neural network

Rongyu Zhang

350 West, 88th Street, New York, NY 10024, USA

Nanjing, China

rzhang11@nyit.edu

Keywords: Document recommendation, Correlation analysis; Analytic hierarchy process; BP neural network; Document scores.

Abstract. At present, with the development of the information technologies and the Internet, evaluation and ,the recommendation of all kinds of information is increasingly concerned. We can get User' evaluation on the books for 5 points. Considering the effect of different factors on the books' score then the label, social friends, books browsing amount of three groups of data were analyzed by bivariate correlation analysis respectively，so that we can get the number of users on the books, scores and label users good friends, the history of the book number of page views a positive correlation. Impact of users to book score were history views > the number of users' good friends > tag number of books. This paper tries to analyze the effects on document remarks by users by using the data. Attached data are the information of users' behavior at a popular on-line bookstore, including remarks on books, tags of books, and the users' social networks.

Introduction

With the rapid growth of Internet, people around the world can easily distribute, browse, and share as much information as possible through the Internet.

The enormous amount of information, however, causes the information overload problem that is beyond users' limited information processing ability. When we want to search documents information, we may not describe the requirements precisely. For example, consumers find it is harder to have access to information they are interested in, with explosively growing information around them; and producers also find it is harder to make their information focused by readers.

To solve this kind of conflict problem, document recommendation system may arise to help all the users (including consumers and producers) to look for some useful document information. Document recommendations system is widely used in network products and Apps. Typical styles include the relevant search, topics recommendations, products push in the e-commercial, and 'recommended friends' search in various SNSs.

Analysis of the Problem

Scores and document recommendation are mainly based on the amount of statistical data processing. Therefore, we need to seize the key and useful date to solve the problem and transform, screen, analyze, summarize the data to analyze the factors affecting the users' score to documents, based on which we complete the user score prediction of document recommendations. through the establishment of a model of user's score of document books.

The problem requires us to analyze the factors that affect scores of books. It is a comprehensive analysis of the data in the annex. Firstly, we screen raw data of user_book_score.txt and we get the data of 1-5 points by users; take into account the different factors impact on scores of books, and then filter the data for other analyzes, the relationship between the various stages of books initially been evaluated scores and the number of labels, the relationship with social friends, relations with views of books. Finally, the data obtained were analyzed and summarized the scientific obtain affect users of books score factors.

Establishment of the Problem

Topic requires us to analyze the factors that affect scores of books, using SPSS statistical software[1] Annex user_book_score.txt (User rating data) on scores of books sorted and then filtered to obtain rated 5 stars praise the corresponding books.

Taking into account the number of books tagged books score results for the impact of the existence of certain relationships, so to conduct a comprehensive analysis on the number of tags. First, using SPSS statistical software accessories book_tag.txt (books tag data) for data filtering, get a different book rated 5 stars appearing several frequency; then counted using EXCEL statistical software, to give the score to 5 points for all books tag number. SPSS software again its bivariate correlation analysis [2], to obtain descriptive statistics table under the label number factors including mean, standard deviation, the observed number of samples，As shown in Table 1.

Table 1 the number of tags and praise of descriptive statistics

Descriptive Statistics			
	Mean	Standard deviation	N
Tag Number	15.16	7.352	4738
Good Reputation	15.16	7.352	4738

Correlation analysis of the number of tags and ratings，As shown in Table 2.

Table 2 The analysis of the results of label and praise correlation

Correlation			
		Tag	Score
Tag Number	Pearson Correlation	1	0.012
	Significant (Bothside)		0.414
	Square plus cross-products	256030.232	3042.232
	Covariance	54.049	0.642
	N	4738	4738
Favorable comment	Pearson correlation	0.012	1
	Significant (Bothside)	0.414	
	Square plus cross-products	3042.232	256030.232
	Covariance	0.642	54.049
	N	4738	4738

When analyzing Table 2, Pearson correlation coefficient between the number of tags with the books score is 0.012, which means an incomplete correlation and a positive correlation between the two. Bilateral significant value between unrelated 0.414, denied the two are not related assumptions. Therefore, it can be concluded according to Table 2: There is a positive correlation between the user's score of books and books tag number, the more the number of tags, books, the higher the rating. Based on consideration of social relations, to a certain extent the number appears on the user's friends praise fifth the number of books rated frequency, the first use of SPSS software accessories user_social.txt (user's social relational data) for data screening, each user Books rated 5 stars appearing several frequency; then counted using EXCEL statistical software, to give the score to 5 points the number of user's friends. SPSS software using them again bivariate correlation analysis, the number of users affected concern for descriptive statistics Friends scores under the influence of factors including mean, standard deviation, observing the number of samples，As shown in Table 3.

Table 3 the number of users and friends from descriptive

Descriptive Statistics			
	Mean	Standard deviation	N
Tag Number	8.06	6.682	3661
Good Reputation	16.71	29.426	3661

Correlation analysis and scoring of the number of the user's friends are shown in Table 4.

Table 4 Correlation analysis of friends and score results table

Correlation			
		Tag	Score
Number of Friends	Pearson Correlation	1	0.164
	Significant (Bothside)		0
	Square plus cross-products	163411.043	118097.177
	Covariance	44.648	32.267
	N	3661	3661
Favorable comment	Pearson correlation	0.164	1
	Significant (Bothside)	0	
	Square plus cross-products	118097.177	3169049.279
	Covariance	32.267	865.860
	N	3661	3661

When analyzing Table 4, the book ratings and the user number of the user's friends Pearson correlation coefficient was 0.164, which means an incomplete correlation and a positive correlation between the two. Bilateral significant value between unrelated 0, denied the two are not related assumptions. Therefore, it can be concluded according to Table 4: User score between books and book the user's friends there is a positive correlation between the number, the more the number of the user's friends, the higher the rating the user of books.

We use SPSS statistical software accessories user_read_history.txt (user data read books) to screen data of annex, in this annex different books rated 5 stars appearing several frequency; then use EXCEL statistical software for counting process, get rated as 5 points of views of all the history books. SPSS software using them again bivariate correlation analysis, considering only the history books descriptive statistics table views under this factor include the mean, standard deviation, the observed number of samples, as shown in Table 5.

Table 5 Historical views and praise descriptive statistics table

Descriptive Statistics			
	Mean	Standard deviation	N
Tag Number	579.44	2242.964	4738
Good Reputation	12.91	35.435	4738

Correlation analysis of the history books views and ratings are shown in Table 6.

Table 6 Browsing history correlation analysis and results table

Correlation			
		Tag	Score
By browsing quantity	Pearson correlation	1	0.527
	Significant (Bothside)		0
	Square plus cross-products	23831319985.228	198426991.949
	Covariance	5030888.745	41888.746
	N	4738	4738
Favorable comment	Pearson correlation	0.527	1
	Significant (Bothside)	0	
	Square plus cross-products	198426991.949	5947795.210
	Covariance	41888.746	1255.604
	N	4738	4738

To analyze Table 6, Pearson correlation coefficients views the history of books and books score between 0.527, which means an incomplete correlation and positive correlation between the two. Bilateral significant value between unrelated 0, denied the two are not related assumptions. Therefore, according to Table 6, it can be concluded: the user visits scores of books and books there is a positive correlation, the more traffic history books, books, the higher the rating.

According to the solution process, we analyze the factors that affect user's evaluation from the aspects of books labels, social relations, history, the results were positively correlated with a user score of books. User ratings impact of books were historical views> number of user's friends> number of books tag.

References

[1] C.H. Yu, SPSS and statistical analysis, Beijing, Publishing House Of Electronics Industry,2007.

[2] S.K. Cheng, SPSS statistical analysis from entry to the master, Beijing, Tsinghua university press,2010.

[3] B.J. Zhang, AHP and its Applications, Beijing: Publishing House Of Electronics Industry，2014.

[4] Wind power prediction problem http://wenku.baidu.com/view/8fd6ed00eff9aef8 941e06e2.html

[5] S. Yan, Q. Li. Han, Examples of neural network analysis and design methods,Beijing: Beijing University of Posts and Telecommunications Press,2010

[6] X.C. Wang, F. Shi, Y. Lei, MATLAB neural network analysis of 43 cases, Beijing, Beihang University Press,2013.

[7] L.F. Wang, S.B. Wang, AHP Introduction, Beijing, China Renmin University Press Co. LTD,1989.

[8] Q.Y.Jiang, J.X. Xie, J. Ye, Mathematical Model, Beijing, Higher Education Press,2011.

[9] D.H. Liu, Ning Li,SPSS15.0 statistical analysis from entry to the master, Beijing, Tsinghua university press,2008.