

The Achievement Evaluation Based on Collaborative Filtering

Yubiao Dai^{1, a} and Xueli Ren^{1, b*}

¹ School of Computer Science and Engineer, Qujing Normal University, Yunnan 655011, China

^aabiaodai@163.com, ^boliveleave@126.com

Keywords: Principal component analysis; Teaching; Knowledge points; Classroom

Abstract. Curriculum is a main way to acquire knowledge, and achievement is only a test of learning the curriculum, so in the teaching activities, the teaching process is vital important, and which determines the result. Collaborative filtering is a general recommendation technology, which has been applied to personalized recommendation successfully. In the paper collaborative filtering is used in the teaching process to predict learning achievement of student according to homework scores, which not only lay timely warning for students who may not pass the test, but also lay a solid foundation for further study

Introduction

The management orienting result emphasizes the results of management and the work, performance management and daily work ability, attitude to comply with the requirements of the result, otherwise there is no value and meaning[1-2]. Process management uses a set of practical methods, techniques, and tools to plan, control, and improve the effectiveness, efficiency, and adaptability of a process. Results and processes coexist in everything. In the course of teaching activities, the teaching process is the process of starting, developing, changing and ending of teaching activities. It has gone through a long process of historical development in understanding of the teaching process for people. With the development of time and research, people gradually realize the complexity and diversity of the teaching process[2-3]. The process of teaching is not only the process of cognition, but also the process of psychological activities and the process of socialization. Therefore, the teaching process is a compound process of cognition process, psychological process and socialization process. The teaching process is a special cognitive process, and also a process of promoting the development of students' mind and body. In the process of teaching, teachers guide students to actively plan for awareness activities, consciously adjust their own interests and emotions, gradually grasp the cultural and scientific knowledge and basic skills in order to promote the development of students' intelligence, physical strength and socialist morality, aesthetic taste, and lays a solid foundation for students' scientific world outlook. The result of teaching is the promotion of knowledge and ability gained through certain teaching activities. For each course of instruction, the process is the experience of learning the course, and the result is academic achievement. The particularity of education requires that the result should be pay attention, but the process should be pay more attention. In order to make the teaching reach a certain effect, we should pay attention to the process of teaching and monitor the teaching process. Collaborative filtering is a typical way to make use of collective intelligence. This paper applies it to the prediction of students' achievement.

Collaborative Filtering

Collaborative Filtering is an important estimation technique in the information retrieval research domain. It has been successfully applied in information filtering and E-commerce. Collaborative filtering is a method of making automatic predictions about the interests of a user by collecting preferences or taste information from many users. The underlying assumption of the collaborative filtering approach is that if a person A has the same opinion as a person B on an issue, A is more likely to have B's opinion on a different issue than that of a randomly chosen person[4-6]. In the method, similarity is the basis; the common methods to count similarity are Euclidean Distance, Cosine, Modify cosine and Pearson correlation[7-11].

Euclidean Distance. If uses rating look as the points in Euclidean space, then the distance in the points is similarity for them. If the common item set is I_{ij} which include the items rated by user i and user j , $R_{i,c}$ and $R_{j,c}$ are the rate which are rated separately by user i and j , then the similarity between user i and user j is computed used Formula 1.

$$sim(i, j) = \frac{1}{1 + \sqrt{(\sum_{c \in I_{ij}} (R_{i,c} - R_{j,c})^2)}} \quad (1)$$

Where $R_{i,c}$ is the rate of item c by user i ; $R_{j,c}$ is the rate of item c by user j .

Cosine. If uses rate look as the vectors in n space, then the similarity between one user and the other user is defined as cosine between one vector and the other vector. If \vec{i} and \vec{j} are rating vectors by user i and user j , then the similarity between user i and user j is computed used Formula 2.

$$sim(i, j) = \cos(\vec{i}, \vec{j}) = \frac{\vec{i} \cdot \vec{j}}{\|\vec{i}\| \|\vec{j}\|} \quad (2)$$

Modify cosine. As the different user's rating scale does not considered in the cosine similarity, the modified cosine similarity is used to improve the defect by minus the average score of user rating for the project. If I_{ij} is the common item set that are rated by user i and user j , R_i and R_j are separately the rate which is rate by user i and j , then the similarity between user i and user j is computed used Formula 3.

$$sim(i, j) = \frac{\sum_{c \in I_{ij}} (R_{i,c} - \bar{R}_i)(R_{j,c} - \bar{R}_j)}{\sqrt{\sum_{c \in I_i} (R_{i,c} - \bar{R}_i)^2} \sqrt{\sum_{c \in I_j} (R_{j,c} - \bar{R}_j)^2}} \quad (3)$$

Where $R_{i,c}$ is the rate of item c by user i ; \bar{R}_i and \bar{R}_j are respectively the average rate for the whole items by user i and user j .

Pearson correlation. If the common item set is I_{ij} which include the items rated by user i and user j , then $sim(i, j)$ of the Pearson correlation similarity of two users i and j is defined as Formula 4.

$$sim(i, j) = \frac{\sum_{c \in I_{ij}} (R_{i,c} - \bar{R}_i)(R_{j,c} - \bar{R}_j)}{\sqrt{\sum_{c \in I_i} (R_{i,c} - \bar{R}_i)^2} \sqrt{\sum_{c \in I_j} (R_{j,c} - \bar{R}_j)^2}} \quad (4)$$

Where $R_{i,c}$ is the rate of item c by user i ; $R_{j,c}$ is the rate of item c by user j . \bar{R}_i and \bar{R}_j are average value of rate for the whole items by both user x and user y .

Application in Achievement Prediction of Collaborative Filtering

The achievement forecast has the vital significance in the teaching process, we may predict the examination result according to the usual study situation, thus promptly solves the question which exists in the study, to improve study effect of students. Collaborative filtering is a commonly used recommendation technology. This paper applies it to the prediction of performance, and the specific calculation process is shown in Fig.1.

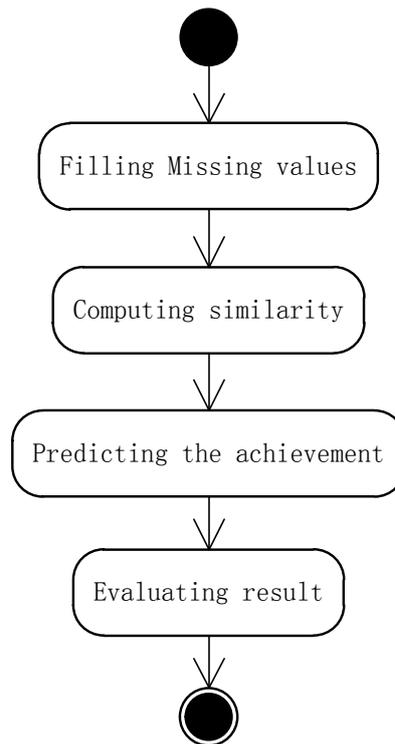


Figure.1. the estimation process

Filling Missing values. One problem in collecting data is that the historical grade data usually contains substantial numbers of missing values. Especially, process metrics contain larger numbers of them since they are collected by hand. The missing values can give bad influences to the accuracy of estimation, so some complementary techniques have been developed for dealing with missing values. The techniques are: listwise deletion, mean imputation and some types of hot-deck imputation [12-13]. Listwise deletion is the simplest technique to ignore data sets that have missing values. Mean imputation is a technique to fill the missing values on a variable with the mean of data sets that are not missing. Hot-deck imputation is alternative forms of imputation that are based on estimates of the missing values using other variables from the subset of the data that have no missing values. The listwise deletion is used to deal with missing value in the paper.

Computing similarity between student and other student. The assignments of students are get according to analysis the homework, and the similarities between student and other students are calculated, then the student achievement is predicted based on similar. The similarity between students is calculated by cosine, the specific formula for calculating reference formula (3).

Predicting the achievement by similar sets. The K nearest neighbors is identified according to the similarity between students. Then, student's achievement is predicted according to the nearest neighbor set, the concrete formula is carried out:

$$\hat{g}_{ab} = \frac{\sum_{i \in k\text{-nearest}} g_{ib} \times sim(s_a, s_i)}{\sum_{i \in k\text{-nearest}} sim(s_a, s_i)} \quad (5)$$

Where k-nearest students denotes set of k students chosen (called neighborhoods) that have highest similarity with s_a .

Evaluating result. The mean absolute error (MAE) is a quantity used to measure how close forecasts or predictions are to the eventual outcomes [16], so it is used to measure forecast error in the paper. MAE is given by formula 6.

$$MAE = \frac{\sum_{i=1}^n |g_i - \hat{g}_i|}{n} \quad (6)$$

In that, g_i is the actual value of course i, \hat{g}_i is the evaluation value, and n is the number of student evaluated.

Experiment

To verify the feasibility of the method, to reduce human performance is not the same problem, the choice of college basic computer course assignments and final grades as the experimental data, because of the course assignments using basic examination system, all the exercise of judgment by the machine automatically, greatly reduces the subjectivity of artificial marking. In order to improve the accuracy of calculation, the results are normalized, and the normalized part results are shown as shown in the table:

Table 1. the partial data

	h1	h2	h3	h4	h5	h6	h7	h8
1	99	89	98	95	94	77	88	68
2	89	85	87	90	64	73	65	73
3	95	72	94	93	81	73	72	43
4	81	78	80	82	88	73	89	77
5	90	84	89	90	95	68	60	77
6	93	100	92	91	96	76	93	76
7	95	100	94	94	90	75	81	73
8	95	82	94	93	94	64	79	68
9	90	100	89	90	89	77	47	88
10	84	95	84	86	79	77	91	68

According to the normalized data, the similarity between students is calculated, and then the student's neighbor is determined according to the size of similarity, and the results are predicted according to the formula. Figure 2 shows the prediction of the first 10 students in comparison with the predicted and true values. The MAE result is shown in Figure 3.

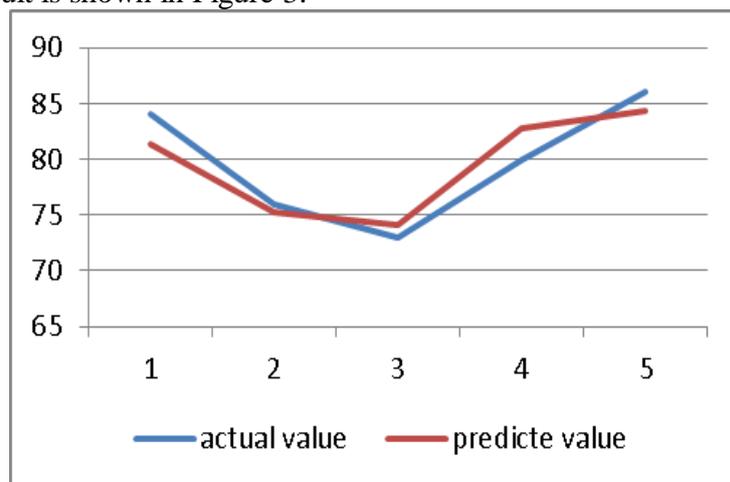


Figure 2. score comparison chart

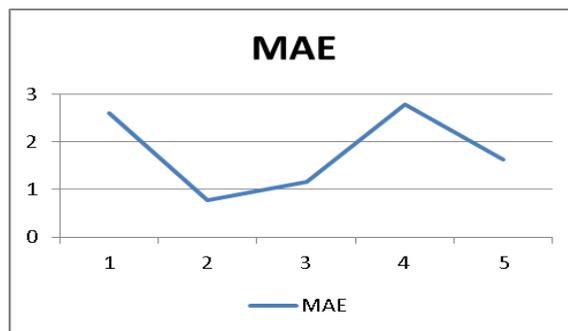


Figure 3. MAE

Summary

Collaborative filtering is a general recommendation method, this paper applied to forecasts of student achievement, the students usually study the nearest neighbor set, then according to the prediction of student achievement, the experimental results show that the prediction accuracy can be achieved.

References

- [1] Wang Weiping. Analysis of the relationship between teaching process and results and Its Implications [J]. Chinese and foreign entrepreneurs,2016/08
- [2] Yang Shengrong. Discussion on the relationship between teaching management process and teaching results [J]. Chinese education off campus. 2013(23)
- [3] Wu Xiaohong,Dai Pingbo. Analysis of process teaching and result Teaching [J]. Journal of Xuzhou Normal University,2000/01
- [4] Collaborative filtering[EB/OL]. https://en.wikipedia.org/wiki/Collaborative_filtering,2017.5.28
- [5] Fuzhi Zhang, Yuanli Lu, Jianmin Chen. Robust collaborative filtering based on non-negative matrix factorization and R 1 -norm. Knowledge-Based Systems,2016-11-26
- [6] Hamidreza Koohi, Kourosh Kiani. User Based Collaborative Filtering using Fuzzy C-Means.Measurement, 2016-05-13
- [7] Taehyun Ha, Sangwon Lee. Item-network-based collaborative filtering: A personalized recommendation method based on a user's item network. Information Processing and Management, 2017-06-1.
- [8] F. Petroni, L. Querzoni, R. Beraldi.LCBM: a fast and lightweight collaborative filtering algorithm for binary ratings. The Journal of Systems & Software,2016-04-26.
- [9] Advances in collaborative filtering. YEHUDA K,BELL R. Recommender Systems Handbook . 2011.
- [10]O'Connor M,Herlocker J.Clustering Items for Collaborative Filtering. http://www.cin.ufpe.br/~idal/rs/oconner_m.pdf . 2017-3-6
- [11]Wu M L,Chang C H,Liu R Z. Integrating Content-Based Filtering with Collaborative Filtering Using Co-clustering with Augmented Matrices. Expert Systems With Applications . 2014
- [12]Fatukasi,O,Kittler,J,Poh,N. Estimation of Missing Values in Multi modal Biometric Fusion. Biometrics Theory Applications and Systems . 2008
- [13]Allison,Paul D.Missing Data Techniques for Structural Equation Models. Journal of Abnormal Psychology . 2003