

Unsupervised Learning Path Extraction in Massive Open Online Course Systems

Mohammad Ali Alipour and Dr. Kamran Layeghi,

Department of Computer Engineering, Tehran North Branch, Islamic Azad University, Tehran, Iran

Abstract

Studying a little about human behavior regarding various activities, we can conclude that each person has a different approach for doing activities. Studies shows that these approaches are somehow similar among people with similar characteristics and the pattern of behavior of similar people can be extracted. There is such approach regarding gaining knowledge and it can be studies, because the path which each person selects to gain knowledge about certain subjects can be studied and improved however the differences there are. the main objective is in fact the possibility of measuring and finding a method to present a way for learning without supervision, regarding the similarities among people's pattern of behavior in gaining knowledge.

Keywords:

learning Path, recommender systems, the slope of learning curve, k nearest neighbor, the latent factor.

1. Introduction

Regarding numerous backgrounds and extensive dimensions of knowledge nowadays, one of the important necessities to pick up knowledge is to look for it regarding the information background of individuals and their familiarities with the subject in which they want to gain knowledge. A learning process for individuals should be designed and recommended to them in a way that they can be informed about prerequisite of knowledge they want to gain in the shortest time and assist them in reaching the desired point. [1]. The problem here is presenting a method which focuses on the learning path and study its impact on the learning curve and recommend the path which individuals should pass in order to gain knowledge. Knowing people using the machine learning methods and studying similarities among people are very salient and many researchers have studied them. [2]. Therefore, the reason of this study which is less studied about, is the degree of similarities among people in knowledge acquisition.

This similarity can result in individuals not doing numerous and futile searches and using useless ways and byways and instead using their energy about the main problem which is gaining knowledge, because using similar individuals' experience and evaluating the results of the studies can result in individuals not repeating

similar mistakes and have better efficiency. [3]. Based on this, it can be expected that the performance of individuals improves and they find out the best way to reach desired point. [4]

2- Relative work and Implementation

In order to study and find out the similarities, we use a sample data set. Various algorithms are used to find out similarities. The algorithms used in this article are k nearest neighbor and the latent factor. Other types AI algorithms and machine learning can also be used. [5]. In pattern recognition, the k-nearest neighbors algorithm (k-NN) is a non-parametric method used for classification and regression. In both cases, the input consists of the k closest training examples and it is an optimization problem for finding the nearest points in metric spaces. The goal of this algorithm is predicting about unknown items for users. For example, look at the following data whose rows are users and columns are items.

	P ₁	P ₂	P ₃	P ₄	P ₅
U ₁	3	4	-	-	4
U ₂	3	5	3	4	5
U ₃	4	2	-	5	4
U ₄	3	-	4	5	2
U ₅	1	-	4	2	1
U ₆	3	4	-	2	5

Figure 1 sample recommended items in a MOOC

The information about U₁ shows that he has given points to P₁, P₂ and P₅ and given no points to P₃, P₄. The goal is finding out anticipation for the possible points for items for which points is given. In order to calculate these anticipations, regarding the similarities among other users similar to this one, we try to create a set of similar users with the degree of similarities. At the first observation, we see that U₂'s points are approximately similar to the ones of U₁, while U₅ has given completely different marks to the items. Therefore, we can conclude that the weight of

U2 and U5 are different compared to the one of U1 and the trend of U2 is relatively closer.

The sample data set in this study shows attended courses by users in EDX massive open online course system. This set has a category column which is determining the subject of the students' study. For example, a student has studied about a programming language like Java and passed courses regarding it or another one has studied about designing algorithms. so this category is an important index for categorizing educational courses. The reason is that usually; an individual has no general information about a subject [6]. For example, he wants to gain knowledge about math, while there are various courses with different preceding and followings regarding to this subject.

Another salient feature existing in this data set is the time label T in each row which shows the time of passing courses by students. In fact the time label indicates the order of the passed course compared to the other passed ones by the user, because it is possible that a user passes more than one course to learn a subject. (This can be found from category index). Regarding the points, they achieve in time periods, we can find out how those courses were useful and how they should be ordered about a certain subject for each person [7].

Another example is that a user wants to gain knowledge about machine learning so that he completely dominates its subject. If we search through the data set in the massive open online course system, we can find numerous courses for him. But it is not clear which one of them should precede the other ones or which are prerequisites and which not for gaining the optimal results. Therefore, there is no clear learning path for users. The user should pass numerous courses using the trial and error method. It is possible that he passes courses which need previous knowledge. But prerequisite courses are not clearly stated. For example, subjects like statistics and probabilities considered prerequisites.

For Extracting a learning path using the k nearest neighbor on a data set, we can find the closest user to him regarding language, age, education, previous background, geographical situation and other features. We can also observe the courses which each user has passed. [8]. It is assumed that the first user has passed the course of the theory of linear regression, factor analysis and classification to learn about machine learning. The second user has passed the courses like machine learning with python, learning basic machine, engineering and the decision tree. We observe the other close neighbors to the user and regarding the points of courses passed by each user, they gain a rank and we achieve different learning paths for users. In fact, these ways are found out via searching the raw data. We call this phase of algorithm, the learning path Extraction. [9]

Therefore, we can recommend optimal learning ways to current users, using the calculation of Hamming distance of various learning ways of different users.

For example, we suppose that a user wants to gain knowledge about computer. His user name is Bins. First we calculate the k of users who passed courses of computer sciences and closer to this user. For this, first, we make a pivot table with rows and columns from users and their points to various courses, as follows:

Course Title	A Global History of Architecture	A New National Culture	Advanced Introductory Classical Mechanics	Advanced St
Username				
Bins	3.500000	NaN		2.500000
Franecki	1.000000	1.0		3.000000
Gusikowski	4.000000	1.0		4.000000
Heathcote	1.000000	2.0		4.000000
Kemper	3.500000	1.0		2.500000
Kilback	1.000000	1.0		1.000000
Kunze	3.000000	1.0		4.000000
Larkin	4.000000	2.0		1.000000
Legros	1.000000	2.0		2.500000
Love	4.000000	2.0		NaN
Medhurst	3.500000	4.0		2.500000
Nicolas	NaN	4.0		3.000000
Nienow	1.000000	NaN		3.000000
Nolan	2.000000	2.0		NaN
Pacocha	3.000000	2.0		4.000000
Parker	2.000000	1.0		2.000000
Paucek	NaN	2.0		2.000000
Schaden	3.000000	1.0		3.000000
Schmidt	NaN	1.0		3.500000
Schuppe	2.000000	4.0		3.500000
Steuber	3.000000	NaN		NaN

Figure 2. pivot tables of users and their points to courses

The first column shows the users and the others indicate the presented courses in the system and the number in any junction of columns and rows is the points of users. NAN shows that a user has not passed the course. In the next stage, regarding the descriptions about the calculation method of distance among users, we must calculate the users' distance from the active user. Among the distance calculation methods, the Hamming distance is a more useful method in solving these problems. Therefore, this method is chosen and the Hamming distances of other users from the active one are counted. and the arrays are arranged in an ascending way.

	Username	Distan
1	Franecki	0.0265
2	Gusikowski	0.0270
3	Heathcote	0.0275
4	Alex	0.0275
5	John	0.0275
6	Fibi	0.0275

Figure 3 The distance among users can be seen here

Afterwards the users, are chosen from the subject group that have passed the desired subject and their learning path is Extracted. In the next stage the common feature of these extracted paths are measured with each other based on their time label and preceding and followings. [11]

With Implementing k nearest neighbor algorithm on train set and finally comparing it with 25 % of test data, we achieve a learning curve like the following:

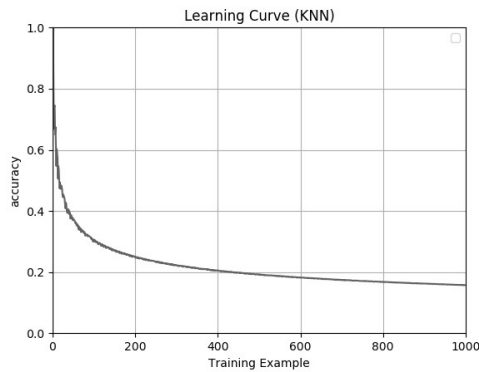


Figure 4. the learning curve - k nearest neighbor

The above chart is the machine learning curve using the k nearest neighbor in the data set. This curve, in fact shows that how algorithm calculations and their results were accurate. The vertical axis, here, is the accuracy of calculations and the horizontal one is the number of examples. [12].

In order to achieve the learning path, we can use public wisdom. The system must observe which courses precede or follow out of the chosen useful courses and afterwards, using the Hamming distance, we can find out common features of learning path and present the results as the learning path [13]. Path Extraction can always be improved - the public wisdom is always changing- and various users give different points to each course in different time periods. another algorithm which we use is the latent factor which is used technique in the correlation filter.

In this algorithm, the main objective is the analysis of data set to main factors which are indicators of behavior patterns of the data set. Therefore, the data set is divided into two separate matrices. The first one indicates the separate features for each user and the second one states the features of each items. The multiplying of these two matrices results in a third one which indicates the data set itself. Therefore this primary information gives us chance to get results about the unknown users. [14].

We must note that in this method the studied features in the user matrix and the one of items must be similar. The user matrix shows the number of columns and the one of items shows the number of rows. This reverse situation is only for making multiplying the matrices possible and the reverse situation of this is also possible.

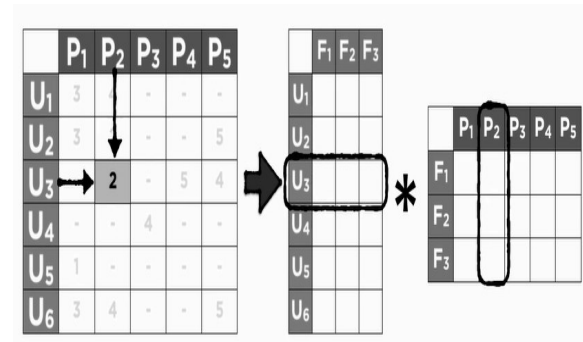


Figure 5. Latent factor

Consequently, we can find out factors that are about similarities of users and use them to calculate their degree of similarities. With Finding out the degree of similarities of users via the latent factor algorithm, we can find the similar users and Extract all users' learning path. Also we can find out the proper learning path for users in all subjects, regarding the common features in the learning paths of similar users and time labels. In fact, the process of extracting the learning path in this algorithm is completely similar to the k nearest neighbor algorithm. The only difference is that the latent factor is used to find similar factors. Implementing this algorithm on test set and finally comparing it with 25 percent of test data, we achieve the following learning curve.

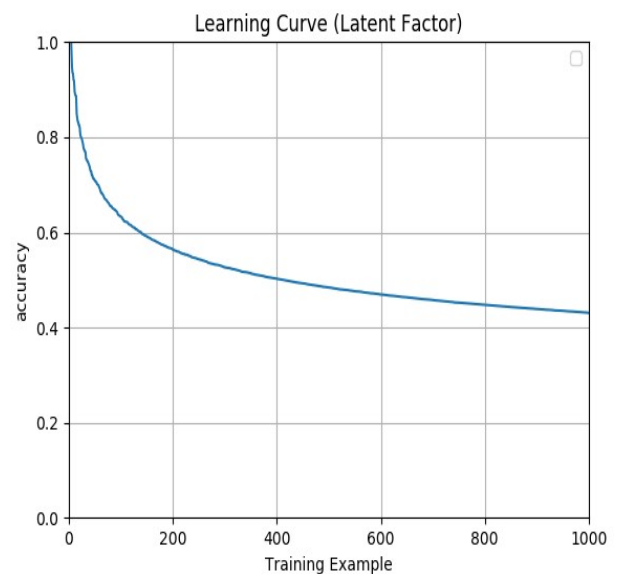


Figure 6 The learning curve - Latent factor

3. Conclusion

The findings of this study which was regarding the improvement of the slope of learning curve in massive open online course system, are as follows in the two sections of descriptive and explanatory results.

3.1 Descriptive Results

The descriptive results show that the learning path has a salient role in the process of gaining knowledge and its understanding for a user who does not know about it. traditionally representing a learning path was made manually by an expert in which personalization is very unlikely or even impossible. One of the most salient results of this study is the possibility of measuring and presenting a personalized learning path by systems without an expert. The observations in this study shows that the learning path of each person about a certain subject can be related to the ones of other ones who are close to him regarding the geographical, lingual, educational backgrounds and other features. Via finding the users who are close to the current one, and extracting a path which they used to gain knowledge about a certain subject and linking these paths and extracting a learning path, which is the result of users' information and other similar ones, we can present a personalized learning path about a certain subject. It must be noted that the path of learning without supervision is presented by the system and given to users. This result, at first can bring more satisfaction to users, because a personalized learning path can have major effect on the method of knowledge achievement, the time for gaining it and the cost it takes.

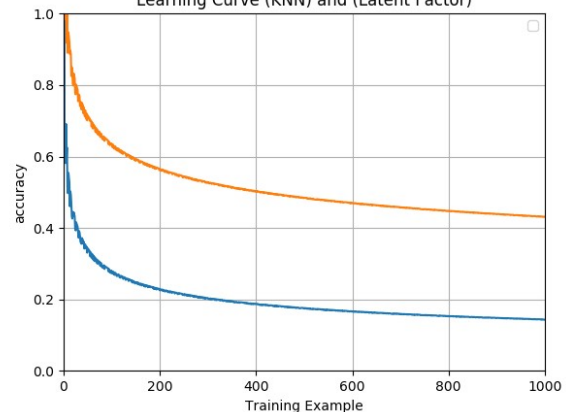
Secondly, it results in increasing the market share for massive open online course system which represent personalized learning path.

3.1 The Explanatory Results

Previously, the method of Extracting information based on the type of algorithm for stating recommendations and differences of each algorithm was stated in order to establish a recommending system, without experts, for making a learning path. The two main algorithms for this activity, the k nearest neighbor and the latent factor were explained and the method of their use in studying data of users' educating and learning was studied. Also the optimal learning path to recommend to users was studied. Regarding both, the results were studied regarding the behavior based on increasing the number of samples in the data set.

Studying the used algorithm and their comparison show that the latent factor algorithm has more consistency compared with the k nearest neighbor when it comes to increasing the number of samples. Although the two algorithms' accuracy decreases drastically when the number of samples increases, and the slope of their curves, first is steep and then it becomes moderate, this decrease is consistent when it reaches 40 percent in latent factor algorithm. This is a valuable accuracy for results. in case of k nearest neighbor, when the number of samples increases, the results reaches lower than 20 percent. Nonetheless, the two algorithms can be used in categorizing users and recommending them an a recommending system without experts in determining the learning method can be made based on them. The second algorithm is more depended on than the first because of its more consistency.

Figure 7 – The curve of learning – k nearest neighbor and latent factor
Learning Curve (KNN) and (Latent Factor)



References

- [1] Francesco Ricci and et. al., "Introduction to Recommender Systems Handbook", Recommender Systems Handbook, Springer, pp. 1-35, 2011
- [2] V. R. Raghuveer et. al., "Reinforcement learning approach towards effective content recommendation in MOOC environments", International Conference on MOOC, Innovation and Technology in Education (MITE), IEEE, pp. 285-289, 2014
- [3] Li-An Ho, Tsung-Hsien Kuo, "How can one amplify the effect of elearning? An examination of high-tech employees' computer attitude and flow experience", pp.23-31, 2010
- [4] L. Pappano, "Massive open online courses are multiplying at a rapid pace", in Education Life,

- The New York Times, 2014
http://www.nytimes.com/2012/11/04/education/edlife/massiveopen-online-courses-are-multiplying-at-a-rapid-pace.html?_r=0.
 Accessed: Jul. 24, 2016
- [5] Zhang, Yongfeng, et al. Explicit factor models for explainable recommendation based on phrase-level sentiment analysis. Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval. ACM, 2014.
- [6] Cui Peng, "Learning curve: The Design and Implementation of Learning Path Management Module of Online Education System", Chinese Control and Decision Conference (CCDC), IEEE, pp. 5711-5714, 2015
- [7] Lakiotaki, K; Matsatsinis; Tsoukias, A, "Multicriteria User Modeling in Recommender Systems", IEEE Intelligent Systems 26 (2): 64–76.doi:10.1109/mis, 2011
- [8] O. R. Zaiane, "Building a recommender agent for e-learning systems", in Proceedings of 2002 International Conference on Computers in Education, pp. 55-59, 2002
- [9] J. Lu, "Personalized e-learning material recommender system", in Proceedings of the 2nd International Conference on Information Technology for Application (ICITA 2004), pp. 374-379, 2004
- [10] Grouping in Collaborative E-Learning Environment Based On Interaction among Students D.Jagadish M.Tech (Information Technology) MIT Campus, Anna University Chennai, India." International Conference on Recent Trends in Information Technology", 978-1-4799-4989-2/14/\$31.00 © 2014 IEEE, 2014
- [11] Campos, Guilherme O.; Zimek, Arthur; Sander, Jörg; Campello, Ricardo J. G. B.; Micenková, Barbora; Schubert, Erich; Assent, Ira; Houle, Michael E. "On the evaluation of unsupervised outlier detection: measures, datasets, and an empirical study". Data Mining and Knowledge Discovery. ISSN 1384-5810. doi:10.1007/s10618-015-0444-8, 2016
- [12] Gong Huang,Wang Man," Learning curve: principle, application and Limitation", International Conference on E-Business and E-Government, IEEE, pp. 1840–1843, 2010
- [13] Juan Carlos Roca, Marylene Gagne," Understanding e-learning continuance intention in the workplace: A self-determination theory perspective", pp.1585–1604, 2008
- [14] Koren, Yehuda, R. Bell, and C. Volinsky. Matrix Factorization Techniques for Recommender Systems, pp. 30-37, 2009
- [15] A.Safia,T.Mala, "Ascertaining the More Knowledgeable Other among peers in Collaborative E-Learning Environment", Fourth International Conference on Advanced Computing (ICoAC),13-15 Dec, 2012



Mohammad Ali Alipour

received the M.S. degrees in Software Engineering from Tehran north Branch of Islamic Azad University in 2017.He is now Project Manager in Rahyab Rayaneh Gostar Inc. and works on edge of software development technology.