ISSN: 0975-766X
CODEN: IJPTFI
Available Online through **Research Article**
www.ijptonline.com

COLLABORATIVE FILTERING APPROACH IN ADAPTIVE LEARNING

Kamil Khadiev*, MiftahutdinovZulfat, MarselSidikov
Kazan Federal University, 420008, Provectus Inc. Kazan office, 420133.

Received on 14-07-2016 *Accepted on 11-08-2016*

Abstract

Nowadays an adaptive approach in education is gaining in popularity. But what does this adaptive approach mean? Adaptive learning (also known as Adaptive education) means that the education system has a personal approach for each student or for groups of students that fits to the students' abilities. Teacher must pick up the most relevant topic for explanation, exercises and tests for such an education process. Also the teacher should adapt the order of learning topics for the current student. It is a very big and hard job, but machine learning algorithms can solve some of these tasks instead of teacher.

What do we have? We have a set of lessons and students. In each step of the education process we select one lesson that fits best for a current student. This problem can be solved by recommender systems of algorithms. Recommender systems predict rating or "preference"' that a user would give to the item, and by similar way an adaptive education system also predict lessons "ratings" for user. In the paper we will define "rating" of lessons and what does it mean "fits the best". Also we give some explanations of a chosen machine learning algorithm.

Keywords: Adaptive education system, adaptive learning, machine learning, recommender system, collaborative filtering,

Introduction

Education for all ages has been the main part of our life. In fact, we can say that our life is a never-ending educational process. Therefore, we always tried to optimize the learning process. We created schools, universities and other institutions that help us to organize the educational process. Nowadays the most popular trend is online courses. Online courses are a very powerful educational instrument that help to connect a lot of students worldwide to the process of education. But mostly they have a linear structure that is not adapted to specific students. Here we have some kind of dual problem. "Classroom" (live) education cannot include a large number of participants, but in small groups the teacher can provide an adaptive approach to everyone. On the contrary online courses can include large number of participants but cannot provide an adaptive approach. In fact, we can formulate the problem wider: if we increase number of participants in the education process, then providing an adaptive approach become harder.

The adaptive learning systems were born together with artificial intelligence (AI).The educational system SCHOLAR that offered adaptive learning for the topics of geography of South America, can be called the first adaptive educational system [1]. And as AI transformed to the multiple disciplines such as machine learning, data mining, etc. approaches in adaptive learning also changed. The machine learning algorithms are already actively used in this area and different authors apply it in different ways. For example, Gaudioso and Boticario in [2] use it for analyzing web activity and build an educational process with respect to it.

Another good example of Machine Learning algorithm usage is the Knewton system. Authors of system use a probabilistic graphical model (PGM) in their work. As they say: “one of the ways in which Knewton applies PGM is by using a student's known proficiencies to determine which other topics he may be ready to master” [3].

In our paper we will show that an educational system can be simulated by a recommender system. Then we will apply a collaborative filtering algorithm from recommender systems to our problem.

The theory of recommender systems began from the mid-1990s papers [4], [5], [6], [7]. After that many scientists explored this, for example, in papers [8], [9], [10]. In the last decade this kind of problem has become very popular, because it useful in practice, for example for online stores.

The paper is organized in the following way. In Section **Adaptive Learning System Overview**

We describe the meaning of an adaptive approach in education, introduce some concepts of our system and show similarity between recommender systems and adaptive educational systems. Then in Section **Adaptive Learning in Terms of Collaborative Filtering** we will show how the problem is described in the framework of collaborative filtering. In Section **Ratings for Lessons** we describe what would be the “rating” of a lesson. In Section **Cold Start Problems** we offer some methods to solve a cold start problem. In Section **Education trajectory** we will explain some problems of educational trajectory and give some solutions of them. In Section **Conclusion** we discuss open problems and feature works.

Adaptive Learning System Overview

As it was mentioned teachers can provide an adaptive approach only in small groups of students. But online courses are more powerful in the sense that they can provide a large scale educational process. Our aim is to combine these two advantages into one adaptive educational system. At first, we need to determine the meaning of adaptive learning. In terms of “classroom” education it means that a teacher has an individual approach for each student.

The differences in those approaches can be based on following items:

1. Each topic can be explained in different ways to different people,
2. Each student can have their own educational trajectory,
3. Each student has different abilities and different knowledge.

This is just an intuitive definition, and there is no unambiguous definition for adaptive learning. Knewton also uses the intuitive definition [11]. In the work “Leveraging the Semantic Web for Adaptive Education” [12] adaptation of hypermedia decomposed to:

1. Adaptive presentation (content level adaptation),
2. Adaptive navigation support (link level adaptation),
3. Adaptive content selection (content level adaptation).

To simulate such an approach in an adaptive educational system we just need to pick up the right explanation of a topic and correctly estimate student knowledge. It means that we have the set of lessons L and the set of students S. For each step of the educational process we select one lesson that fits best for a certain student. It seems like a typical recommender system.

The recommender system predicts the “rating” or “preference” that a user would give to an item and then choose an item with maximum rating [10]. The adaptive learning system predicts which lesson will be more comprehensible for the student. If we introduce some method for comprehensibility measuring in both systems, then it will be necessary to predict and then take the maximum of “rating” for pairs (s, l) where s and l are user and item respectively [13], [14], [11].

Adaptive Learning in Terms of Collaborative Filtering

This is how a recommendation task looks like. We have user ratings or preferences or some measurable activity for some products and we need to predict ratings or preferences for the rest of the products and choose the best of them. Formally: let S be the set of users and L - the set of products. Let u be the utility function that measures usefulness of product l to the user s.

$$u: S \times L \to R,$$

where R is a totally ordered set. Now for each user s we choose product l that maximizes utility function:

$$\forall s \in S, l_s = \underset{l \in L}{argmax} u(l, s).$$

This formal explanation was introduced in paper [9]. Usually a utility function is presented as ratings. In the following part of the paper we will consider only ratings because it's more appropriate for our aims. The information about ratings can be presented by simply a matrix of size $m \times n$ where at the insertion of i-th row and j-th column the rating of i-th user for j-th product. But if i-th user never rated j-th product $x_{ij} = 0$. The problem is that we have to predict unknown ratings.

$$\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{pmatrix}$$

To reduce the adaptive learning system to a recommender system we just need to suppose that S is a set of students and L is a set of lessons and correctly define the meaning of rating.

Ratings for Lessons

Our aim is to compose a curriculum more understandable for each student. So we need to improve the level of comprehensibility. That means our ratings should be the level of comprehensibility. And to measure the level of comprehensibility we use the popular method, this is a test with one or more right answer selections. Thereby we need to mention a pair of lessons and tests (l, t). If a student has a high mark on the test that means that he has learned the lesson well. And our curriculum should be tailored to each student.

At first, it means that a smart student should receive a high level of knowledge. In this case we need to separate the pairs of lessons and tests by a level of difficulty. We can do that by assigning the weights for each question. In this way we provide that each j-th pair lesson, test has its own level of difficulty. The test is a set of n_j questions where each question has a difficulty level of q_{ij}. So the difficulty level of the pair would be

$$T_j = \sum_{i=1}^{n_j} q_{ij},$$

where $T_j \in [0; \infty]$ and $q_{ij} \in [0; \infty]$.

Secondly, that means that we need take into account that a student can have some gaps in knowledge on some of the topics, explanatory style, difficulties in understanding some structures of a subject, etc. That means that we have to explain in more detail topics that a student understands poorly and choose the correct style of explanation. In this case we should know information about each factor that we want to control for adaptability. It may seem like a content-based approach. But we can use the SVD (singular value decomposition) algorithm. Simply the SVD algorithm approximates the initial matrix of ratings by a product of three matrix

$$X = U \cdot D \cdot V^T = U \cdot \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \sigma_k \end{pmatrix} \cdot V^T$$

Where U, D and V are real number matrices. V^Tmeans transposed V. In this case i-th user is represented by k factors $u_i = (u_{i1}, u_{i2}, \dots, u_{ik})$ and j-th product is represented also by $k$factors $v_j = \left(v_{j1}, v_{j2}, \dots, v_{jk}\right)$. In simple recommender systems we can say that i-th element of a user's factors vector shows us his preference of i-th factor and the product factors vector shows how those factors impressed in product [15]. For more information about the SVD algorithm see [16], [15], [17]. So in terms of adaptive education our lesson's factors vector can be comprehensibility of some topics in that lesson or more complex structures. Similar to the student factors vector. That means that we can control not only the level of difficulty of the lesson in adaptability but also some other factors that that should be explored. And as factors can differ from subject to subject, the student-lessons matrix can different for each subject. But also as our approach is general (for each subject we got the matrix of ratings) we can use one common SVD algorithm and make a cross-disciplinary adaptive learning system.

Cold Start Problems

As in a recommender system, in our system we have a "cold start"" problem: what should we do if a new student comes to our system? Or what should we do if a teacher added a new lesson to the system? Solutions to these questions are described in many papers [18], [19], [20].

For Students

To recommend a lesson to a student we firstly need to know something about that student. In the SVD algorithm we know the ratings of a student, but the new student did not rate on any lessons (never been tested). It means that in matrix U his row vector equals 0-vector. And that means that we can't receive any recommendations. Thus we need some extra information about the student. This information could be demographic, e.g. age, academic degree, gender, etc. Or we could ask for the student to solve some initial test. The second variant is more appropriate because cognitive abilities do not depend heavily on demographics, but an initial test helps to estimate more properly those abilities. Anyways the student is presented as vector s that contains information about him. On that information we find similar users and recommend a lesson that most fits each student in a group.

Example of recommendation group:

$$r_{jk} = \sum_{i=1}^{s_j} w_{ij} \cdot r_{ik}$$

where w_{ij} is weight of i-th student relatively to j-th one or the degree of similarity j-th and i-th students, $w_{ij}$is a real number that varies in range $[-1;1]$. And $r_{ik}$is i-th student predicted a rating for k-th lesson, $r_{ik}$is a non-negative real number. Weights can be calculated in different ways. In [20] author uses a Pearson Correlation Coefficient, it is calculated as:

$$w_{ij} = \frac{cov(s_i, s_j)}{\sigma_{s_i} \cdot \sigma_{s_j}}$$

In [19] author use cosine similarity

$$w_{ij} = \frac{\langle s_i, s_j \rangle}{|s_i| \cdot |s_j|}$$

where $\langle s_i, s_j \rangle$ is a dot product. Also sometimes as weights standard the Euclidean distance is used.

For Lessons

The same situation appears in adding a new lesson to the adaptive learning system. And we use the same problem solution. But the question is which characteristics as extra information should we consider in clustering. The idea is that the lesson could be represented by n-dimensional vector, if the subject has n topics, where each component of a vector is complexity or detalization of a corresponding topic in a lesson. This approach is time-consuming on the part of the teacher that composes the course. But it can provide a more accurate problem solution. Because this approach takes into account not only the lesson's total complexity, but also each topic's complexity in a lesson.

Education trajectory

At the moment we have a set of “rated” lessons. Ratings are presented by two values r_1 andr_2, where r_1 is the total score for correctly answered questions, and r_2 is the mark or percentage of correctly answered questions from all of the questions.

$$r_1 = \sum_{i \epsilon Q_R} q_{ij}$$

$$r_2 = \frac{r_1}{\sum_{i=1}^{n_j} q_{ij}}$$

Where Q_R is a set of rightly answered questions, $r_1$and r_2 are predicted ratings.

We want to give lessons or sets of lessons to a student, which should well understand a given topic, with respect to the student's capabilities. That means we should firstly take the maximum by r_2 and then byr_1. But the easier the lesson the higher the mark. It means that the system will attend to give easier lessons to everyone. To solve that problem we can identify several intervals forr_1, such that we do not maximize by r_2 but maximize by r_1. Intervals could be:$0-55$, $56-70$, $71-85$,$86-100$. In this case a student will receive a higher level of knowledge inside the intervals. But sometimes it may be better to get in interval 71-85 and receive more knowledge. Formally: we need to find dependency between r_1 and r_2 (plotted blue) which represents the best selection. And if we find this dependency then our task in each step would be only to find the closest lesson to the curve. Another question in choosing an educational trajectory is "May be some lessons should be learned before others". That means we got a partially ordered set of lessons. To solve that problem, we can associate with the lesson$l$, set of lessons $L^r(l)$, which are required to be learned before lesson$l$. And in each step we choose a lesson to learn from a set of available lessons: $L^a(L^l) = \{l: L^r(l) \subset L^r\}$, where $L^l \subset L$ is a set of learned lessons.

Discussion

In this paper we discuss applying of recommender system algorithms and methods for adaptive learning. We show that one of main algorithms, which is named collaborative filtering, give us good results. Also, we discuss metric of learning quality. It was results of test. The metric is discussable and this is way to improve our method. Finally we had problem with "Could start" and we debate only one method of solving this problem. Our solution can give good results for huge data about previous activities, but for small data set it can be better to use another approach.

Conclusion

Machine learning and data analysis have been widely used in the last couple of years in different areas as well as in adaptive learning. As an example it may be called the Knewton system which uses PGM to make predictions. In this paper we mentioned that a distance adaptive learning process can be reduced to the collaborative filtering and that the SVD algorithm makes sense to be applied. But the open questions are:

1. How many factors should we use?
2. What do they mean?
3. It would be interesting to compare results of different approaches in adaptive learning. But we don't have any criteria to do that. So the question is how to compare two models of adaptive learning?
4. How to choose the best or at least a good educational trajectory for a student?

Acknowledgements: We thank Stepan Pushkarev, General Manager of Provectus Inc. Kazan Office, who told us about the Adaptive Learning problems. Also we thank the students of Kazan's Federal University Institute of

Computer Mathematics and Information Technologies, who passed tests and prepared data for our algorithm. The work is performed according to the Russian Government Program of Competitive Growth of Kazan Federal University. KamilKhadiev is partially supported by Russian Foundation for Basic Research, Grant 14-07-00557.

References

1. Jaime R Carbonell. Ai in cai: An artificial-intelligence approach to computer-assisted instruction. Man-Machine Systems, IEEE Transactions on, 11(4):190–202, 1970.
2. E. Gaudioso and J.G. Boticario.Towards web-based adaptive learning communities.In 11th International Conference on Artificial Intelligence in Education (AIED2003). Sidney, Australia, July 20- 24,, 2003.
3. Daphne Koller and Nir Friedman. Probabilistic graphical models: principles and techniques. MIT press, 2009.
4. David Goldberg, David Nichols, Brian M Oki, and Douglas Terry.Using collaborative filtering to weave an information tapestry. Communications of the ACM, 35(12):61–70, 1992.
5. Will Hill, Larry Stead, Mark Rosenstein, and George Furnas. Recommending and evaluating choices in a virtual community of use. In Proceedings of the SIGCHI conference on Human factors in computing systems, pages 194–201. ACM Press/Addison-Wesley Publishing Co., 1995.
6. Paul Resnick, NeophytosIacovou, MiteshSuchak, Peter Bergstrom, and John Riedl. Grouplens: an open architecture for collaborative filtering of netnews. In Proceedings of the 1994 ACM conference on Computer supported cooperative work, pages 175–186. ACM, 1994.
7. Upendra Shardanand and Pattie Maes. Social information filtering: algorithms for automating word of mouth. In Proceedings of the SIGCHI conference on Human factors in computing systems, pages 210–217. ACM Press/Addison-Wesley Publishing Co., 1995.
8. Jonathan L Herlocker, Joseph A Konstan, Loren G Terveen, and John T Riedl. Evaluating collaborative filtering recommender systems. ACM Transactions on Information Systems (TOIS), 22(1):5– 53, 2004.
9. G. Adomavicius and A. Tuzhilin. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. Knowledge and Data Engineering, IEEE Transactions on, 17(6):734–749, June 2005.
10. Francesco Ricci, LiorRokach, BrachaShapira, and Paul B Kantor.Recommender systems handbook, volume 1.Springer, 2011.
11. Liu David. The knewton blog.Building with Knewton, 5.

12. MiloˇsKravˇc´ık and DraganGaˇsevi´c. Leveraging the semantic web for adaptive education. Journal of Interactive Media in Education, 2007(1):Art–6, 2007.
13. ˇ Kateˇrina Kostol´anyov´a, Jana Sarmanov´a, and Ondˇrej Tak´acs. Clas- ˇ sification of learning styles for adaptive education. The New Educational Review, 2011(23):199–212, 2011.
14. Peter Brusilovsky et al. Adaptive and intelligent technologies for web-based eduction. KI, 13(4):19–25, 1999.
15. ArkadiuszPaterek. Improving regularized singular value decomposition for collaborative filtering. In Proceedings of KDD cup and workshop, volume 2007, pages 5–8, 2007.
16. BadrulSarwar, George Karypis, Joseph Konstan, and John Riedl.Application of dimensionality reduction in recommender system-a case study.Technical report, DTIC Document, 2000.
17. L De Lathauwer, B De Moor, J Vandewalle, and Blind Source Separation by Higher-Order. Singular value decomposition. In Proc. EUSIPCO-94, Edinburgh, Scotland, UK, volume 1, pages 175–178, 1994.
18. Andrew I Schein, AlexandrinPopescul, Lyle H Ungar, and David M Pennock.Methods and metrics for cold-start recommendations.In Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval, pages 253– 260.ACM, 2002.
19. Zi-Ke Zhang, Chuang Liu, Yi-Cheng Zhang, and Tao Zhou. Solving the cold-start problem in recommender systems with social tags. EPL (Europhysics Letters), 92(2):28002, 2010.
20. ShaghayeghSahebi and William W Cohen. Community-based recommendations: a solution to the cold start problem. In Workshop on Recommender Systems and the Social Web, RSWEB, 2011.

Corresponding Author:

KamilKhadiev*,