A Survey on Recommender System in Learning Analytics

S. Ranjeeth[1]
Department of Computer Science and Engineering
VFSTR (Deemed to be University),
Guntur, Andhra Pradesh -522213, India

T. P. Latchoumi [2,*]
Department of Computer Science and Engineering
VFSTR (Deemed to be University),
Guntur, Andhra Pradesh -522213, India

Abstract: Education Data mining is an important field that derives insights from the education data and one of the famous applications in data mining is recommender system. In the field of learning analytics recommender system have become common in recent years. These are widely used as a tool which can take input of various selection criteria & user preferences and yields to improve performance of student in education sector. The user's style and preferences should be constructed accurately to supply most relevant suggestions. Researchers proposed various types of Recommender System in learning analytics (RSL). In this paper, authors studied various current state of recommendation system models in learning analytics and discussed their preference criteria. As a part of that, authors studied various important preference factors in RS (Recommender system) and categorized them based on their likeness. This article reports RSL (Recommender system in learning analytics) model's future directions and compiling a comprehensive reference list to assist researchers.
Keywords: Recommender system, learning analytics, collaborative filtering, education data mining

I. INTRODUCTION

Data Mining (DM) is used to discover information from the company's various databases and re-construct it for uses other than the databases which were initially planned for data mining implementation is different for variant organizations depending upon the nature of data and organization. It specifies to juice or mining insight from enormous data, mining knowledge from data is called data mining in this paper we are mainly focusing on recommender system in learning analytics so to understand from scratch we explained about recommender system and its working, its techniques (machine learning methods) and in next section explained about learning analytics and its techniques finally recommender system in learning analytics.

II. RECOMMENDER SYSTEM

The system which can recommend something you are maybe interested in that you haven't a try is called Recommender Systems. For example, if you bought a book about Artificial Intelligence it gives a recommendation list including some books about pattern recognition, data mining, etc. Recommendations such as things to be useful to a client are offered by Rs (Recommender system) software tools and techniques.

2.1 Recommendation System – Working Model

The number of choices is vast on the internet; the filter is required, important and efficiently delivers appropriate data in order to reduce the issue of data overwhelm, which has made a future complication to numerous online clients. Recommender systems figure out this problem by finding through big amount of aggressively developed data to provide users with personalized info and services.

Working off the recommendation system in five steps, the following are

Step 1: Data collection

Collecting data is an initial phase in making a proposal engine. The type of information can be either implicit or explicit data. Information like ratings and comments which are given by the user on products is treated as explicit knowledge. Request history and return history, Cart events, Page sees, Click through and find log treated as implicit data. When every user visits the site this data set will be created.

Step 2: Storing

We can get good recommendations by providing more data to our algorithms. In a recommendation project if data is more quickly turn into big data projects. The database tool is decided by the type of data that is used to create recommendations. Examples of database tools area standard SQL database, NoSQL database or even some sort of item repository.

Step 3: Analyzing

How will we notice things that have the same user engagement information? So as to try to do, therefore, we tend to clean the info by utilizing totally variant analysis ways. If you wish to produce emergence suggestions to the client as they're seeing the item then you may want an additional quick-thinking kind of analysis.

Some of the ways by which we can break down the information are:

- Real-time systems: Created data can process by it. Processing and analyzing streams of events tools usually involve in a real-time system. This system would be needed to provide in the bit recommendations.
- Batch analysis: Processing the data periodically is a demand for batch analysis. This concern suggests enough information needs to be created in order to form the analysis closely, such as daily sales capacity. This system might work well to send an e-mail on the next date.
- Near-real-time analysis: you can refresh the analytics for every few minutes or seconds if you collect information quickly. Working in the near-real-time system is best in giving recommendations during the clone (same) browsing period.

Algorithm

For each item in the product catalog, I1
For every client C who purchased I1
For every item I2 purchased by client C
The record that a client purchased I1 and I2
For every item I2
Calculate the similarity between I1 and I2

Step 4: Filtering

The final step is data filtering it is used to bring consistent information need to give suggestions to the client. We must select a method which is a better suit for proposal motor (engine).

The following are the algorithms.

- Content-based**:** A famous, suggested product has the same features as what a consumer likes or views.
- Cluster: Recommended items go together fine; it doesn't care about what other users have done.
- Collaborative: Other customers, who like the similar item as another customer views or likes, will also call a recommended product.

In this article we are mainly focusing on collaborative filtering in learning analytics.

2.2 Recommendation System algorithms in the recommendation system three main filter techniques, these techniques contain algorithms. Techniques like content, collaborative, hybrid-based techniques. Collaborative technique divided into two categories: model-based and memory-based

2.2.1 Model-Based Techniques: The description model develops from the database and an active user will get predictions by this model. Model building processes done by variant pursuing algorithms such as the following are.

- Artificial Neural Network (ANN): Structure of numerous associated neurons that are organized in layers in efficient ways is called ANN. The associations between neurons have weights related to them relying upon the measure of impact one neuron has on another. In a few special problem situations, we can use neural networks there are having some advantages. For instance, because of the way that it contains numerous neurons and furthermore assigned a weight to every association, unnatural neural systems are very strong in dealing with the disturbance and mistaken Data sets. Suppose nonlinear capacities and catching difficult connections in information groups can also done by artificial neural networks, they can be proficient and even work if part of the system fails. Finding the perfect system topology for given issue is major disadvantage of ANN and it acts as lower headed for the group mistake, once topology is decided.

- Decision Tree Technique (DTT): Decision Tree Induction. It is a structure that incorporates a root hub, branches, and leaf hubs. Each interior hub indicates a test on quality, each branch means the result of a test, and each leaf hub holds a class mark. The decision tree is built by breaking down an arrangement of preparing cases for which the class names are known, and it depends on the philosophy of tree diagram. They are then connected to characterize already unseen instances. If trained on top quality information, they can make exceptionally exact predictions [13]. Compare to SVM (support vector machine) and ANN (artificial neural network) classifiers decision trees are more interpretable since they join basic inquiries regarding information in an understandable way. In handling products with mixture of real-valued and categorical options decision tree is very flexible and in handling objects that have few specific missing options

- Link analysis: Connection Analysis is the way toward working up systems of interconnected items in order to investigate patterns and trends [14]. It is used to improve the accomplishment of web searches by presenting its great potentials. HITS algorithms and page rank also present in link analysis, a website page as a solo hub in the web graph [16] handle by mostly link analysis algorithms

- Matrix completion techniques: Within the user-item matrices to predict unknown values we can use the matrix completion technique. In collaborative filtering recommendation systems [15] Correlation-based K-nearest neighbor is one of the significant methods utilized. They lean on widely on the chronicled rating information of clients on things. Many Items that are represented within the matrix [13] are not given rate by the user because often rating grid is enormous and sparse. This issue dependably prompts the failure of the framework to give solid and precise suggestions to clients. In matrix completion for practice variations different of low-rank models have been utilized especially in collaborative filtering toward application is used.

- Bayesian Classifiers: Classification problems can solve by the probabilistic framework. It depends on the definition of Bayes theorem and conditional probability. Class label and each attribute are considered as random variables by Bayesian classifiers [12]. The primary advantages of Naive Bayes classifiers are that they are powerful to isolated commotion points and insignificant properties, and they handle missing qualities by disregarding the instance during likelihood evaluate computations.

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}$$

P =Probability

H =Posterior Probability of a Hypothesis

X = training data

Figure 1 Bayesian classifier probability formula

2.2.2 Memory Based Technique: Prediction generation can do by the entire client item database; we can find the neighbor using statistical techniques**,** it is also known as the nearest neighbor. Through item, user depends skills can achieve Memory-based CF (collaborative filtering). By looking at client evaluations on a similar thing client based collaborative filtering techniques calculates likeness between clients, and it then calculates the predicted rating for a thing by the dynamic client as a weighted normal of the evaluations of the thing by clients like the dynamic client where weights are the likenesses of these clients with the objective thing. To calculate predictions in item-based filtering techniques use similarity between things but not clients. It assembles a model of thing likenesses by recovering all things appraised by a functioning client from the client thing network (matrix); to calculate similarity between item/user many kinds of likeness metrics are utilized. Pearson correlation coefficient and cosine are the 2 famous likeness metrics.

- Pearson correlation coefficient**:** It will take a range (r) of values from +1 to -1. No association between the two variables is indicated with 0. If the no is greater than 0 is called positive association, if it is less than 0 is called negative association. In positive association value of 1 variable increment other variable value increments but in negative association variable value increments another variable value reduces

$$r = \frac{N(\sum XY) - (\sum X)(\sum Y)}{\sqrt{[N(\sum X^2) - (\sum X)^2][N(\sum Y^2) - (\sum Y)^2]}}$$

N = number of pairs of scores

$\sum XY$ = sum of the products of paired scores

$\sum X$ = sum of X scores

$\sum Y$ = sum of Y scores

$\sum X^2$ = sum of squared X scores

$\sum Y^2$ = sum of squared Y scores

Figure2 Pearson correlation coefficient formula

- Cosine based: Compare to Pearson-based measure cosine similarity is different in that it is a vector-space show which depends on linear algebra-based math rather than a statistical approach. Between 2 n-dimensional vectors depend on the angle between them relativeness measured by it. Text mining (TM) to compare 2 text files and fields of information retrieval uses the cosine-based measure, in this case, vectors of terms represents the documents. The comparability between two things closeness measure is also alluded to as similitude metric, and their techniques used to figure the scores that express how comparable clients or things are to each other. Thing or Client based suggestion developing use scores as a foundation. Based on the condition of use, distance measure or correlation metrics can also be alluded to as similarity measures

There are many domains using recommendation systems, for example, tourism and travel, e-commerce, movie and music, health and Sensex, etc. This paper reviews about recommendation system in learning analytics.

III. LEARNING ANALYTICS

Learning analytics (LA) [11] is an emerging field in the education sector. Mainly focused on learning process [8], it is used to assess the relevant data on understudies and instructors at a little scope level which targets solitary understudies and the courses are taken in order to understand student execution and advance understudy accomplishment [9] with LA strategies and tools of analytics which are advance, student performance and examining results can be improved by redesigned centering of help and intercession, along these lines advancing learning and education [10]. The research and further LA incorporate the development, use, and blend of new methodology and tools in order to improve the demonstration of learning and teaching for singular understudies and instructors.

Table 1. Learning analytics and its related fields

S.NO	Field	Stake Holders	Objectives	Methods	Data
1	AA (Academic Analytics)	Institutions of Education	Managing of admissions, Prediction, Marketing, Decision making	Techniques of statistics	Data of education sector
2	EDM (Education Data Mining)	Faculty, Understudies	Improving learning process by changing over information into significant data	Techniques of DM (data mining)	Data of education sector
3	LA (Learning Analytics)	Faculty, Learners, Institutions of education	Suggestion, prediction, enrollment, syndication, domestication, personalization	Methods of quantitative and DM techniques	Data of education sector

3.1 Learning Analytics Techniques

Regression, classification, clustering, artificial neural network are the techniques used in learning analytics field. LA is a part of education data mining in previously, but now treating as different sectors like AA (academic analytics), LA (Learning analytics) and education data mining. To get clear knowledge about learning analytics [11]

IV. RECOMMENDER SYSTEM IN LEARNING ANALYTICS

Table-2 Literature survey of recommender system in learning analytics

S.No	Author	Proposed Work and techniques used
1	Faisal M. Almutairi [1] 2017	In this article three approaches are developed under collaborative filtering framework; two approaches are built by using coupled matrix factorization with latent matrix factor third approach was build using tensor factorization to model grades. These three methods are used to incorporate additional information in the context of collaborative filtering (CF), slow learners' problem also handle when predicting for next semester, and author evaluated these three proposed models on grade data which is collected from university of Minnesota.
2	Nguyen Thai-Nghe [2] 2010	In this article, recommender system methods for learning analytics in education data mining are used to propose a novel approach for predicting student performance, common regression techniques such as logistic/linear regression are used to validate the proposed approach on different dataset

3	**Sanjog Ray [3] 2011**	**In this article, author proposed a course recommender system to student which is used for improving their grade with the help of collaborative filtering technique; It is evaluated on real time data set.**
4	**S.JothiLakshmi [4] 2018**	**In this article, author proposed a recommender system called Integrated Recommender Educational Data mining (IRED) for higher educational institution target marketing, it is a unique design which provides solutions for target marketing in higher education institutions. Techniques used for building proposed work are collaborative filtering as filtering agent and C4.5 as pattern discovery model.**
5	**Phung Do [5] 2017**	**In this article, e-learning material recommendation system is proposed using collaborative filtering and knowledge based reasoning techniques, proposed method evaluated on three datasets which are taken from cognitive tour and compared model performance with other three different techniques called MF (Matrix factorization), RBR (Rules based reasoning) and CBR (case-based reasoning).**
6	**Alexandre L [6] 2018**	**In this article, student gets the recommendation based on the assessment taken by him. The assessments used in this article are calculus, hand-on, remote lab, simulations. After completion of assessment student will get analysis and suggestion.**
7	**Alexandre L. Gonçalves [7] 2018**	**In this article, on remote laboratories activities suggestions provided to students in order to scaffold their performance using learning analytics and recommender system techniques. It is also providing the performance analysis and possible errors done by student during lab experiment**

V. CONCLUSION

This article reported on the current state of recommendation system in the learning analytics. We also discussed a novel recommendation system working flow with sample algorithm, types of recommendation algorithms and this paper focused on recommender system in learning analytics and its importance in education field, giving recommendations to student or learner or teacher or education organizations. Based on preference options with demographic profiles, online courses attended moocs data etc.

REFERENCES

[1] Faisal M. Almutairi, Nicholas D. Sidiropoulos, George Karypis, “Context-Aware Recommendation-Based Learning Analytics Using Tensor and Coupled Matrix Factorization”, IEEE journal of selected topics in signal processing,11(5), 2017.

[2] Nguyen Thai-Nghe, Lucas Drumond, Artus Krohn-Grimberghe, Lars Schmidt-Thieme, “Recommender System for Predicting Student Performance”, Procedia Computer Science, pp: 2811–2819, 2010.

[3] Sanjog Ray, Anuj Sharma, “A Collaborative Filtering Based Approach for Recommending Elective Courses”, International Conference on Information Intelligence, Systems, Technology and Management ICISTM 2011: Information Intelligence, Systems, Technology and Management, pp 330-339, 2011.

[4] S. JothiLakshmi, M.Thangaraj, “Design and Development of Recommender System for Target Marketing of Higher Education Institution Using EDM”, International Journal of Applied Engineering Research, 13(19), pp. 14431-14437, 2018.

[5] Phung Do, Kha Nguyen, Thanh Nguyen Vu, Tran Nam Dung, and Tuan Dinh Le, Integrating Knowledge-Based Reasoning Algorithms and Collaborative Filtering into E-Learning Material Recommendation System, Springer International Publishing AG 2017 T.K. Dang et al. (Eds.): FDSE 2017, LNCS 10646, pp. 419–432, 2017.

[6] A. L. Gonçalves, L. M. Carlos, J. B. da Silva, G. R. Alves, "Personalized Student Assessment based on Learning Analytics and Recommender Systems," 2018 3rd International Conference of the Portuguese Society for Engineering Education (CISPEE), Aveiro, pp. 1-7, doi: 10.1109/CISPEE.2018.8593493.2018

[7] Latchoumi, T. P., Ezhilarasi, T. P., Balamurugan, K. “Bio-inspired weighed quantum particle swarm optimization and smooth support vector machine ensembles for identification of abnormalities in medical data”. SN Applied Sciences, 1(10), 1137, 2019.

[8] G. Siemens, P Long, “Penetrating the fog: analytics in learning and education.” Educause Rev. 46(5), pp:30–40, 2011.

[9] Loganathan, J., Janakiraman, S., & Latchoumi, T. P. “A Novel Architecture for Next Generation Cellular Network Using Opportunistic Spectrum Access Scheme”, Journal of Advanced Research in Dynamical and Control Systems ,(12), pp:1388-1400, 2017.

[10] A van Barneveld K.E Arnold J.P Campbell, “Analytics in higher education: establishing a common language”. EDUCAUSE Learn. Initiative. 1, pp:1–11, 2012.

[11] S Ranjeeth, T.P. Latchoumi, P Victer Paul. "A Survey on Predictive Models of Learning Analytics", Procedia Computer Science, Volume 167, Pages 37-46, 2020 https://doi.org/10.1016/j.procs.2020.03.180.

[12] Ranjeeth, S., Latchoumi, T. P., & Victer Paul, P. “Optimal stochastic gradient descent with multilayer perceptron based student's academic performance prediction model”. Recent Advances in Computer Science and Communications. https://doi. org/10.2174/2666255813666191116150319.2019.

[13] Caruana R, Niculescu-Mizil A., “an empirical comparison of supervised learning algorithms. In: Cohen W. Moore AW, editors. Machine Learning”, Proceedings of the twenty-third international conference, ACM, New York; pp. 161–8, 2003.

[14] Loganathan, J., Latchoumi, T. P., Janakiraman, S., parthiban, L., “A novel multi-criteria channel decision in co-operative cognitive radio network using E-TOPSIS”. In Proceedings of the International Conference on Informatics and Analytics, pp. 1-6, 2016.

[15] Koren, Y., Bell, R., Kolinsky, C.” Matrix factorization techniques for recommender systems”. IEEE Comput, 8, pp.30–7, 2009.

[16] Ranjeeth, S., Latchoumi, T. P., & Paul, P. V. “Role of gender on academic performance based on different parameters: Data from secondary school education”. Data in brief, 29, 105257, 2020.