A Big-Data Oriented Recommendation Method in E-Learning Environment

https://doi.org/10.3991/ijet.v17i10.27861

Outmane Bourkoukou[1](✉), Essaid El Bachari[2]
[1] Moroccan School of Engineering Sciences, Marrakesh, Morocco
[2] Department of Computer Science, Cadi Ayyad University, Marrakesh, Morocco
o.bourkoukou@emsi.ma

Abstract—E-learning is increasingly gaining popularity in organizational and institutional learning for its several benefits to learn anywhere, anytime, and anyplace. Therefore, explosive growth of E-learning has caused difficulty of locating appropriate learning activities for learner in this context, and it becomes relatively widespread learning method for learner. Several research in e-learning mainly focused on improving learner achievements based on recommendation technique. An ideal recommender system in e-learning environment should be built with both accurate and learning goals. To address this challenge, we propose a recommendation method based on machine learning technique. Based on this tool, a learning approach is designed to achieve personalized learning experiences by selecting the most appropriate learning activities. Moreover, some experiments were conducted to evaluate the performance of our approach. The results demonstrate that our method outperforms other state-of-the-art methods and reveals suitability of using recommender system in order to support online learning activities to enhance learning.

Keywords—e-learning, big data, recommender system, learning object, collaborative filtering

1 Introduction

As the volume of learning objects on e-learning platforms continues to grow at an exponential rate, it provides difficulty of locating appropriate learning objects for learner in this environment, and it becomes relatively widespread learning method for learner [1,2]. In E-learning environment, the accuracy of acquiring useful and optimal personalized learning from Massive Open Online Courses (MOOC) has become a prior issue. Big data providers, Coursera, edX, and Udacity serve several thousands of registered learners. MOOC provides a plenty of opportunities: 1) advanced and continued learning, 2) low cost, 3) life-long learning, 4) new skills and improve knowledge of learners and teachers. MOOCs can include thousands of learners and it is expected to grow in number and influence within the next years [1, 3].

However, to a certain extent, the creation of personalized learning path has solved the problem between content diversity and learners' requirements specialization [4,5].

It is meaningful that recommender systems can help learners to get useful and valuable learning objects from massive data. Most of e-learning platforms, have applied various kinds of recommender systems to improve the service quality. In real applications, traditional recommender systems usually adopt collaborative filtering algorithms based on similarity between learners or learning objects to generate suggestions [6,7,8]. This approach guarantees the accuracy of recommendation results while causes diversity loss.

In this paper, we design a novel recommendation method which consists of an improved collaborative filtering algorithm to achieve optimal personalized learning experiences by selecting the most appropriate learning objects. Firstly, we construct a similarity computational model according to resource diffusion principle which considers more factors including explicit and implicit feedbacks, preference and social relationship between learners. Then, we are adapted a new formula for prediction preference of learners, to sequence and structure a personalized learning scenario. Finally, a revisited module is added, to adjust learner' recommendation according his scores obtained after each learning experience.

The remainder of this paper is organized as follows: Section 2 describes the related works; In Sections 3, an improved Recommender Model based on CF is proposed; Section 4 provides experimental results on several data sets; Finally, conclusions are presented in Section 5.

2 E-learning recommender systems

In the last decade, Technology-enhanced learning (TEL) has become one of the most important ways that helps learners by providing them the material that can help them study everywhere and anytime [1,4, 27]. TEL aims to design, to develop, and to test socio-technical innovations that will support and enhance learning practices of both individuals and organizations. Recently, recommender systems (RS) have been researched extensively and applied to technology-enhanced learning in order to identify suitable learning objects and to deliver a variety of learning activities to the learners [10,11]. However, several previous observations have argued that classical RS system in e-learning still lacks intelligence that may not fit for each learner characteristic [1, 4]. The main features of the RS system are mainly focused on interactivity and its capability to be customized according to a typical learner model [5, 6].

Therefore, according to [12], learning object is being regularly produced, organized, and published in different types of TEL environments such as Learning object repositories (LOR), A massive open online course (MOOC), Learning management systems (LMS). This excessive amount of LOs merges various opportunities but also causes difficulties for learner to locate appropriate learning objects.

In the last decade, a number of many researchers motivated on building a RS system in order to support learners to achieve specific learning needs since it is very useful to design and to implement especially in informal learning [13].

One of the first attempts to develop a collaborative filtering system for digital learning objects has been the Altered Vista system [14]. This system supports discovery and automatic filtering for relevant learning resources that addresses needs of learners and educators. Another system that has been proposed for the recommendation of learning objects is the RACOFI system (Rule Applying Collaborative Filtering) [15]. The RACOFI system assists and recommends online users audio learning objects. Imran et al. [16] proposed PLORS system supports learners by providing them recommendations about which learning objects within the course are more useful for them. The recommendation mechanism uses association rule mining to find the association between LOs. The CYCLADES system has proposed by Avancini and Straccia [17] for allowing users and communities search, share and organize their information space according to their own view and evaluate learning resources available in Open Archives Initiative (OAI). The system is able to give recommendations of several types based on user and community profiles. Dascalua et al. [18] use recommender agents for recommending online learning activities or shortcuts in a course web site based on a learner's web logs using association rule algorithm.

However, in last few years, many researchers suggest that recommender system should combine more than technique in order to provide a better selecting, and sequencing recommendation list of learning objects to fit the specific learner's needs and interests [1, 5]. As examples, an evolving learning management system has been developed by Tang and McCalla [19] to store, and to share digital learning resources using a hybrid recommendation process based on a clustering and collaborative filtering approach to classify students with similar interests and tastes. In his work [4] Klasnja-Milicevic et al. have developed a system called PROTUS (PRogramming TUtoring System) which can recommend relevant links and activities for learners, by considering the Felder-Silverman Learning styles Model and the learner's level of knowledge. This system has been designed based on hybrid recommendation using the collaborative filtering and the sequential pattern mining. Li et al. [20] present a general architecture of learning recommender system for the smart learning environment. By constructing learner models and resource models, the proposed recommender system aims to recommend learning resources by using the clustering and association rule mining and to recommend peers via social interaction computing. Bourkoukou et al. [1], propose a recommender model for e-learning environment to achieve personalized learning experiences by selecting and sequencing the most appropriate learning objects. By using a hybrid recommender system based on collaborative filtering technique and association rule mining algorithm.

3 Recommender model

Unlike to E-commerce RSS, E-learning RSs have their own characteristics, and simply transferring a recommender system from e-commerce to e-learning systems may not accurately meet the achievements of the targeted learners. Compared with resource recommendation in e-commerce systems, learners in e-learning systems have topic preferences in e-learning systems. However, learners' behaviors in learning systems are in a more consistent and coherent way [21]. Learning objects have some intrinsic orders in learner' learning experience. For example, a learner will probably learn from easy resources to difficult ones; for a single knowledge point, a learner will probably learn from theoretical to practical. In another example, the learning resource access sequence in a special learning process is: {Lecture notes, Reference, Exercise book}. Therefore, the time-dependency relationship between learning resources in a learning process can reflect learner's latent resource access pattern and preference. However, conventional CF approaches cannot reflect these characteristics. Therefore, it is necessary to find a method to represent the sequences of the e-learning resources. We can mine learner's historical access records for discovering the resource access sequential patterns. Then using these sequential patterns, we can predict the most probable resource that a learner will access in the near feature to further improve quality of recommendations and solve new user problem. The recommender process for E-learning context is depicted in Figure 1.

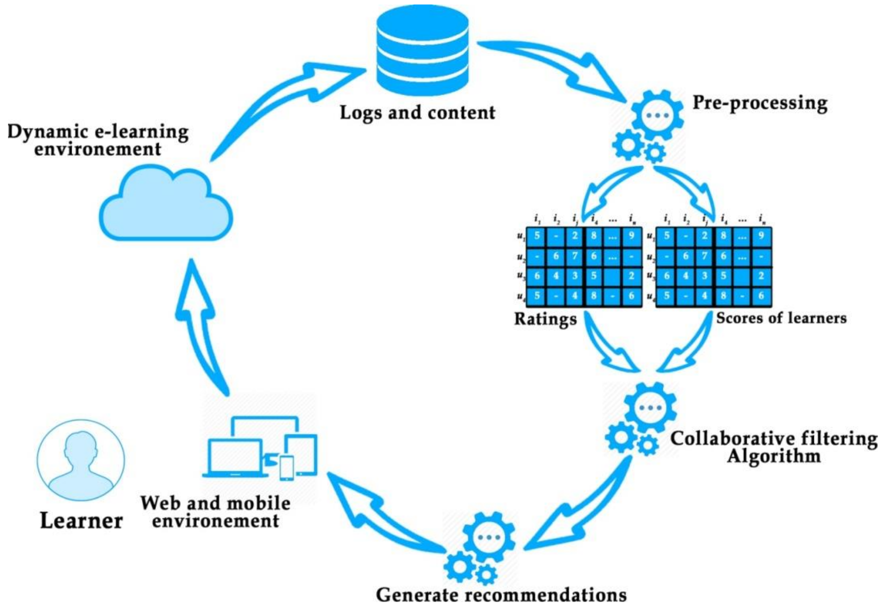

Fig. 1. Recommender process

3.1 Framework and principal of MapeReduce

Processing and analyzing big data requires special tools and systems that make available a computing environment to satisfy the requirements of analytics for large datasets. Therefore, we can conclude that for any big data platform there must be an application development framework which can make simpler the process of development, execution, testing, and debugging appropriate educational software components and building blocks.

There are several technologies to handle big data such as Map-Reduce [22], NoSQL [23], PIG [24], and Hive [25].

One of the most known and significant tools commonly used by most data management systems is Hadoop. Hadoop is an open-source software framework, written in Java, intended for distributed storage, and processing of very large data sets, using simple programing models. Principally, it achieves two tasks: large data storage and faster processing. Hadoop is responsible for the strong Hadoop Distributed File System (HDFS), encouraged by Google's file system, as well as parallel programing model using the Map-Reduce paradigm [4, 22]. Therefore, program execution is divided into a Map and a Reduce phases, separated by data transfer between nodes in the cluster [26]:

1. Map phase. A node performs a Map function on a segment of the input data and converts it into another set of data, where individual elements are broken down into tuples (key/value pairs). The records for any given key, possibly spread across many nodes, are grouped at the node running the Reducer for that key. This involves data transfer between machines.
2. Reduce phase. This second Reduce stage takes the output from a Map as an input and combines those data tuples into a smaller set of tuples. As the sequence of the name Map-Reduce implies, the reduce task is always performed after the map job.

The key contributions of the MapReduce framework are not the actual map and reduce functions, but the scalability and fault-tolerance achieved for a variety of applications by optimizing the execution engine once.

3.2 Improved collaborative filtering

In E-learning context, learning content is divided into several courses, each of which consists of several chapters, A chapter can be represented as a tree of learning units or concepts. Figure 2 shows the structure of our suggested domain model.

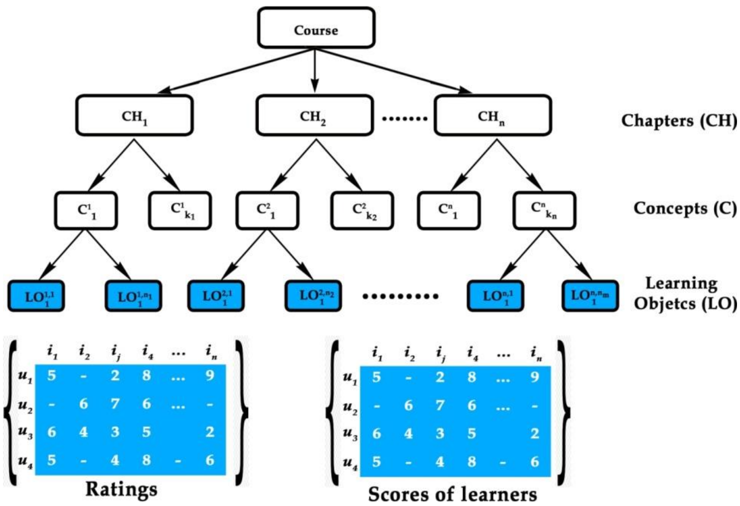

Fig. 2. Domain model

A learning unit holds one unit of knowledge and presents different aspects of it with different types of learning object which constitutes multiple external representations such presentations, questions activities, examples, exercises, glossary.

After Pre-processing step, we obtained two matrix of learner-item-ratings with n rows, where *n* denotes the number of learners $L = \{l_1, l_2, \ldots, l_n\}$, and *m* columns, where *m* denotes the number of learning objects $I = \{ i_1, i_2, \ldots, i_m\}$.

We apply a novel method based on CF in order to build virtual community of learners sharing the same interests and preferences. This technique allows finding predictions by computing the similarities between learners.

The critical step in memory-based CF methods is to define similarity and dissimilarity between learners or learning objects. Indeed, various approaches are proposed to compute similarities and dissimilarities, the most used are as follows: Pearson's correlation, Cosine similarity. We propose a novel learner similarity computational method based on learning experience of learners to tackle with the deficiencies of traditional methods.

The fundamental idea that we would like to formalize is the weighting of the recommendations which stem from the learner details, not only like the traditional similarity between their ratios and those of the rest of the learners, but also to take into account that the recommendations of the learners with better scores have a greater weight than the recommendations of the learners with lower scores.

In order to evaluate the knowledge level of a learner *u* (*KLu*) over the recommendations that will be received from him a learner v with knowledge level *KLv*, a large number of metrics can be established.

In this paper it has been decided to use a simple and asymmetric metric that can be established by means of the function f (1), although, using other metrics such as those shown in (2) would also be feasible. The choice of one metric or another comes from the manner in which it is desired to weight the relationship of knowledge level demonstrated by each pair of learners and by the nature of the RS details themselves.

$$f = \begin{cases} KLv - KLu, & KLv > KLu \\ 0, & KLv \leq KLu \end{cases} \tag{1}$$

Thus, in the metric (1), if the knowledge of learners v is 0.7 (on a scale of 0–1) and that of learner u is 0.2 (on the same scale), the weighting of the knowledge level of learner u to learner v would be 0.5, while the weighting of knowledge level of learner v to learner u would be zero.

The new measurement for similarity between the learners' u and v can be established as defined in (3). The first term of the equation refers to the knowledge level scores, while the second term refers to the similarity of the learners according to their preferences and applying some of the traditional metrics (Pearson, Cosine, MSD. . .).

The sum serves to discover the arithmetic mean of the *T* scores that evaluate the knowledge level of the learner; a score that is not evaluated must be initialized with the minimum score (0 on the scale of 0–1). *KLt* represents the knowledge of the learner u on the t subject, test, etc.

$$S(u,v) = \frac{1}{T}\sum_{t=1}^{T} f(KLu, KLv) * sim_{I(u,v)}) \tag{2}$$

The values of similarity obtained between the pairs of learners serve to obtain the desired k-neighborhoods of each learner, just as is done with the traditional metrics of CF, and in this way, recommendations can be made based on the evaluations given to the k learners most similar to each other. Here we assume that the preference to learning object j of target learner u is p_{uj}. The set of equations that express these ideas mathematically are:

$$\overline{KL}_u = \sum_{t=1}^{T} KL_{u,t}\,, KL_{u,t} \in [0,1] \tag{3}$$

$$w_{u,j} = \frac{1}{\mu} \sum_{u=1}^{\breve{U}} \overline{KL}_u r_{u,j} \; \forall j \;\big|\; \exists\, r_{u,j} \neq \emptyset, \breve{U} \in U \big| \exists\, r_{u,j} \neq \emptyset, \mu = \sum_{u=1}^{\breve{U}} \overline{KL}_u \tag{4}$$

$$pu,j = \frac{1}{|\breve{K}|} \sum_{k=1,k\# u}^{|\breve{K}|} S_{u,l} * \beta w_{k,j} \tag{5}$$

In equation (4), $w_{u,j}$ denotes the rating estimation for learner u and ratings of all learners that have rated learning object j.

4 Experimentation and results

In order to verify the effectiveness of the proposed approach, we conduct several experiments in real data sets in E-learning context.

4.1 Data sets

A real-world data set are used in our experiments, namely Geometry 2006–2007 (Geo) which is extracted from the Cognitive Tutor System and published by PSLC DataShop [28]. This available data set contains the implicit information about interactions between learners with the tutoring system and learning resources. To evaluate the performance of our algorithm, the data set needs to be partitioned into two sections: training set (80%) and testing set (20%). The specifications of the data sets are summarized in Table 1.

Table 1. Experimental data sets

Data Set	Learners	LOs	Transactions	Sparsity (%)
Geo	567	5181	2,441,583	93.01

4.2 Evaluation metric

We mainly focus on testing the prediction accuracy of our proposed method, and we used the mean absolute error (MAE), which is the most widely used technique to compare the deviation between predictions and the real user-specified values. MAE can be defined as:

$$MAE = \frac{\sum_{u,j} |Pu,j - Ru,j|}{\mathrm{m}} \tag{6}$$

Where *m* is the total number of ratings over all learners, p_u, *j* is the predicted rating for learner *u* on learning object *j*, and r_u, *j* is the actual rating. Obviously, the smaller MAE is the better performance of the algorithm will be.

4.3 Experiment Process and result

The experiments are conducted specifically to find out the following questions: (1) How parameters like similarity metrics, number of neighborhood and data sets size could influence results?

In the experiments, we have used KNN algorithm with different similarity metrics using formula (2) in order to find the best value of K-neighbors in static data set and in incremental data set, depicted in Figure 3.

In Figure 3(a), to determine the optimal K-value of neighbors for KNN method on Geo data set, the number of neighbors is chosen between 10 and 200. By increasing the number of neighbors, the performance of the KNN algorithms using different similarity metrics obtains better prediction accuracy. In Figure 3(b) by varying K-value, it can be seen that take into account knowledge level of learners outperforms better than other KNN metrics without. The optimal performance is achieved when K is approximately equal to 190 for Geo data set.

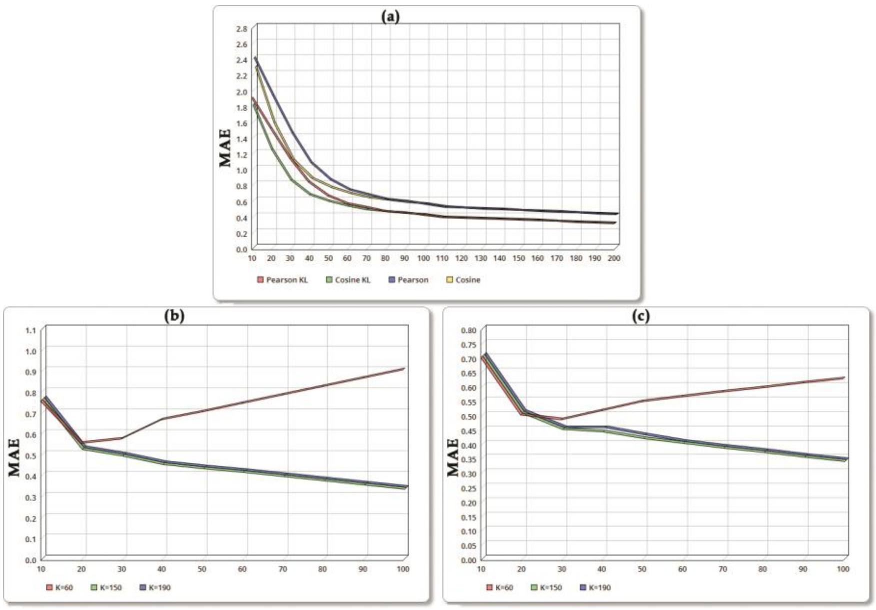

Fig. 3. Comparison between traditional and personalized learning strategy

In Figure 3(c), the experiment was carried for each of the following values 60, 150, and 190. It can be seen that by increasing the number of users with varying the K-value, we can obtain an optimal prediction except when K = 60 for the most similarity metrics.

The value K = 150 can be considered the best value for KNN algorithm with different similarity metrics since the corresponding MAE value is the smallest one.

5 Conclusion

In this last decade, recommender systems are one of the recent and most important technologies used to improve individual and personalized learning in E-learning context. However, there are several limitations when applying the existing recommendations algorithms. To address these limitations in this paper, we propose a personalized E-learning environment based on learning identification and collaborative filtering approach. The main idea is to deliver a personalized teaching strategy for each learner by selecting and sequencing the most appropriate learning objects into a coherent, focused organization in online distance education.

Moreover, the availability of open data sets in E-learning seems, until now, to be a real challenge since there are not enough experiences in real scenario using many learners and transactions. In order to evaluate the prediction accuracy of our proposed recommendation model, we used external data sets in E-learning environment. Results show that using the proposed approach could improve the performance of predictions.

In the future, we plan to refine the recommender model to deal with several inherent issues such as data sparsity and cold start. Since CF methods are known to be vulnerable to these problems in recommendation. In addition, we will consider more complex recommendation approaches, by including other factors such as learner motivation, learning styles, and apply other intelligent artificial techniques.

6 Acknowledgment

We used the "Geometry 2006–2007" data set accessed via DataShop (www.pslcdatashop.org).

7 References

[1] Bourkoukou, O., El Bachari E., E. El Adnani M. (2017). A Recommender Model in E-learning Environment. Arab J Sci Eng, vol. 42, pp. 607–617. https://doi.org/10.1007/s13369-016-2292-2

[2] Klašnja-Milićević, A., Ivanović, M., Budimac Z. (2017). Data science in education: Big data and learning analytics. Comput Appl Eng Educ. Vol. 25, pp. 1066– 1078. https://doi.org/10.1002/cae.21844

[3] Papamitsiou, Z. and Economides A. A. (2014). Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence, J. Edu. Technol. Soc.17.

[4] Klasnja-Milicevic, A., Vesin, B., Ivanovic, M. and Budimac, Z. (2011). E-Learning personalization based on hybrid recommendation strategy and learning style identication. Comput. Educ. Vol. 56, pp. 885–899. https://doi.org/10.1016/j.compedu.2010.11.001

[5] Dorca, F. A, Araujo, R.D, Carvalho, V., Resende, D.T, and Cattelan R.G, (2016). An automatic and dynamic approach for personalized recommendation of learning objects considering students learning styles: an experimental analysis. Inf. Educ. Vol. 15(1), pp. 45–62. https://doi.org/10.15388/infedu.2016.03
[6] Salehi, M. (2013). Application of implicit and explicit attribute based collaborative filtering and bide for learning resource recommendation. Data Knowl. Eng. Vol. 87, pp. 130–145. https://doi.org/10.1016/j.datak.2013.07.001
[7] Khribi, M.K, Jemni, M., and Nasraoui, O. (2009). Automatic recommendation for e-learning personalization based on web usage mining techniques and information retrieval. Educ. Technol. Soc. Vol. 12(4), pp. 30–42.
[8] Zhou, W., Zhang, M., M. et al. (2018). Incorporating Social Network and User's Preference in Matrix Factorization for Recommendation. Arab J Sci Eng, vol. 43, pp. 8179–8193. https://doi.org/10.1007/s13369-018-3380-2
[9] Duo, S., and Ying, Z.C. (2012). Personalized E- learning System Based on Intelligent Agent, Physics Procedia, vol. 24, pp.1899-1902. https://doi.org/10.1016/j.phpro.2012.02.279
[10] Anaya, A.R., Luque, M., Garcia-Saiz, T. (2013). Recommender system in collaborative learning environment using an influence diagram. Expert Syst. Appl. Vol. 40, pp. 7193–7202. https://doi.org/10.1016/j.eswa.2013.07.030
[11] Manouselis, N., Drachsler, H., Verbert, K. and Santos, O.C. (2010). Proceedings of the 1st Workshop on Recommender Systems for Technology Enhanced Learning (RecSysTEL 2010), Procedia Computer Science, 1(2). https://doi.org/10.1145/1864708.1864797
[12] Ochoa, X. (2011). Modeling the macro-behavior of learning object repos- itories. Interdiscip. J. E-Learn. Learn. Objects, vol. 7, pp. 25–35. https://doi.org/10.28945/1343
[13] Manouselis, N., Drachsler, H., Vuorikari, R., et al. (2011). Recommender systems in technology enhanced learning. In: Kantor, P.B., Ricci, F., Rokach, L., Shapira, B. (eds.) Recommender Systems Handbook. Springer, US, pp. 387–415. https://doi.org/10.1007/978-0-387-85820-3_12
[14] Recker M., Walker, A., Wiley, D. (2000). An interface for collaborative filtering of educational resources, in In Proceedings of the 2000 International Conference on Artificial Intelligence (IC-AI'2000): 317-323.
[15] Anderson, M., Ball, M., Boley, H., Greene, S., Howse, N. et al. (2003). RACOFI: A Rule-Applying Collaborative Filtering System, in Proceedings of IEEE/WIC international conference on web intelligence/intelligent agent technology, Halifax, Canada. 13-23.
[16] Imran, H., Belghis-Zadeh, M., Chang, T., Graf, K. S. (2016). PLORS: a personalized learning object recommender system. Vietnam J. Comput. Sci. vol. 3(1), pp. 3–13. https://doi.org/10.1007/s40595-015-0049-6
[17] Avancini, H., Straccia, U. (2005). User Recommendation for Collaborative and Personalised Digital Archives, Int. J. Web Based Communities, vol. 1(2), pp. 163-175. https://doi.org/10.1504/IJWBC.2005.006061
[18] Dascalua, M., Bodea, J. M., Moldoveanuc, I., et al. (2015). A recommender agent based on learning styles for better virtual collaborative learning experiences. Comput Hum Behav, vol. 45, pp. 243–253. https://doi.org/10.1016/j.chb.2014.12.027
[19] Tang, T., McCalla, G. (2005). Smart Recommendation for an Evolving E- Learning System: Architecture and Experiment., International Journal on E-Learning, 4 (1), 105-129.
[20] Li, Y., Zheng, Y., Kang, J. and H. Bao. (2016). Designing a Learning Recommender System by Incorporating Resource Association Analysis and Social Interaction Computing. In: Li Y. et al. (eds) State-of-the-Art and Future Directions of Smart Learning. Lecture Notes in Educational Technology. Springer, Singapore. https://doi.org/10.1007/978-981-287-868-7_16

[21] Bourkoukou, O., EL Bachari, E. and El Boustani, H. (2020). Building Effective Collaborative Groups in E-Learning Environment, Advances in Intelligent Systems and Computingthis, 1102 AISC, pp. 107–117. https://doi.org/10.1007/978-3-030-36653-7_11
[22] EMC. (2012). Data science and big data analytics. EMC education services, pp. 1–508.
[23] Hecht, R. and Jablonski, S. (2011). Nosql evaluation. In International conference on cloud and service computing, pp. 336–341.
[24] Olston, C. and Pig latin, Al. (2008). a notso-foreign language for data processing, In: SIGMOD '08: Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data, ACM, Vancouver, Canada, 1099–1110. https://doi.org/10.1145/1376616.1376726
[25] Thusoo, A., et Hive, Al. (2009). A warehousing solution over a map-reduce framework, Proceedings of the VLDB Endowment, 2 (2009), pp. 1626–1629. https://doi.org/10.14778/1687553.1687609
[26] Bakshi, K. (2012). Considerations for big data: Architecture and approaches, In Proceedings of the IEEE Aerospace Conference., pp. 1–7, 2012. https://doi.org/10.1109/AERO.2012.6187357
[27] Maslova, I., Burdina, G., & Krapotkina, I. (2020). The Use of Electronic Educational Resources and Innovative Educational Technologies in University Education. International Journal of Emerging Technologies in Learning (iJET), 15(16), pp. 68–79. https://doi.org/10.3991/ijet.v15i16.14909
[28] Koedinger, K.R.; Baker, RSJd; Cunningham, K.; Skogsholm, A.; Leber, B.; Stamper, J. (2010). A data repository for the EDM community: the PSLC DataShop. In: Romero, C., Ventura, S., Pechenizkiy, M., Baker, R.S.J.D. (eds.) Handbook of Educational Data Mining. CRC Press, Boca Raton.

8 Authors

Outmane Bourkoukou received his PhD in Computer Science from Cadi Ayyad University Morocco in 2017. He is currently a Professor at Computer Science. Moroccan School of Engineering. Most of his scientific activities are devoted to computer science especially e-learning, recommender systems and engineering. He has been general chair and co-PC Chair of number of international conferences. He is the author of numerous publications related to his research interests.

Essaid El Bachari is a Professor of Computer Science at Cadi Ayyad University, Morocco since 2004. He received his PhD in Mathematics from Paris VI University France in 1998. He is responsible for the development and presentation of open learning courses, which include the investigation of various modes of course presentation and tutor development. He is the author of numerous publications related to his research interests.

Article submitted 2021-10-26. Resubmitted 2022-01-10. Final acceptance 2022-01-10. Final version published as submitted by the authors.