Student's Performance Prediction Using Hybrid Optimization Algorithm-based Map Reduce Framework

Mahendra Sawane[1*], Dipak Bhosale[2]

[1] Department of Information Technology, JSPM's BSIOTR, Wagholi, Pune, MH, India.

[2,] Department of Computer Science & Engineering, KEC, Shelve, Pandharpur, MH, India.

*Corresponding author

ABSTRACT

Learning analytics (LA) is a growing research area, which aims at selecting, analyzing and reporting student data (in their interaction with the online learning environment), finding patterns in student behaviour, displaying relevant information in suggestive formats; the end goal is the prediction of student performance, the optimization of the educational platform and the implementation of personalized interventions. According to the Society of Learning Analytics Research1, LA can be defined as "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs". The topic is highly interdisciplinary, including machine learning techniques, educational data mining, statistical analysis, social network analysis, natural language processing, but also knowledge from learning sciences, pedagogy and sociology; up-to-date overviews of the area are provided in. Various educational tasks can be supported by learning analytics, as identified in analysis and visualization of data; providing feedback for supporting instructors; providing recommendations for students; predicting student's performance; student modelling; detecting undesirable student behaviours; grouping students; social network analysis; developing concept maps; constructing courseware; planning and scheduling. Similarly, seven main objectives of learning analytics are summarized in: monitoring and analysis; prediction and intervention; tutoring and mentoring; assessment and feedback; adaptation; personalization and recommendation; reflection.

Keywords: Student's performance prediction, Map Reduce, Optimization Algorithm, Data Science.

1   Introduction

The prediction of students' performance is one of the most popular goals of Learning Analytics (LA), which aims to estimate future learning outcomes and identify indicators for learning success, more specifically, the objective of the project is to develop a model which can infer the students' academic performance (i.e., the predicted variable, generally in the form of grades or scores) from a combination of various indicators (i.e., predictor variables) from the educational dataset. The predictive information is extremely valuable, as it can offer instructors the ability to monitor the learning progress and provide students with personalized feedback and interventions; in particular, the instructor can be advised about students at-risk, who are in need of more assistance [1]. It predicts the future performance of a student after being enrolled in an educational institute. Hence, determining who would do well and who would receive poor scores. Such results would help to make admission decisions more efficient and improve the quality of academic services. Administrators can use predictive results to evaluate the performance of students in the upcoming academic semesters. Lecturers can select suitable learning strategies for students depending on their scores and estimate how they would help the students improve to a certain of extent. The benefits lead to the development of computerized methods that could predict performance with high accuracy [5]. In addition, individualized strategies for improving participation may also be suggested. Furthermore, the automatic

prediction mechanism may be used for a formative assessment tool, which has the potential to decrease the instructors' assessment loads. Finally, providing prediction results and personalized feedback can foster students' awareness [1].

Performance prediction has been extensively studied in web-based educational systems and, in particular, in Learning Management Systems. This is due to the availability of large amounts of student behavioral data, automatically logged by these systems, such as visits and session times, accessed resources, assessment results, online activity and involvement in chats and forums, etc. Thus, student performance prediction models based on Moodle log data have been proposed in multiple previous studies. Additionally, log data from (ITS) intelligent tutoring systems have also been used for performance prediction. In contrast, the student's engagement with social media tools in emerging social learning environments has been less investigated as a potential performance predictor [1]. Literature has suggested the clustering-based schemes for the prediction since the application of the clustering schemes in the fields, such as bioinformatics, image analysis, etc., have produced significant results. The clustering scheme divides the database into various groups by calculating the similarity measures between the data points in the database. One of the important challenges involved in clustering is dividing the database with vagueness. The literature [19] has used the fuzzy rule-based classifiers (FRBCs) for constructing the prediction system for the students and [20] has used the particle swarm optimization (PSO) for selecting the suitable centroids for clustering. Moreover, the existing works have suggested the prediction schemes based on the neural networks (NN), decision tree, support vector machine (SVM) [16], and Naive Bayes [17] suitable for the EDM [18]. Several works [15] suggested the use of the academic performance of the students in the high school to predict the performance of the students in the future. This can be studied by the prediction scheme to identify the study pattern of the students [7].

2 Literature Survey

Table 1: *Summary of studied related papers published*

Authors	Methods	Advantages	Disadvantage
Elvira Popescu and Florin Leon [1]	LMNNR algorithm	Good correlation coefficients	It did not compared courses with different internal and external conditions
Jie Xu *et al.* [2]	Machine Learning Approach	It provided valuable information for academic advisors to recommend subsequent courses to students and carry out pedagogical intervention measures if necessary	performance prediction to elective courses was not evaluated
Raheela Asif *et al.* [3]	Educational Data Mining	Effective performance prediction	generalizability of the results should be deepened

Mrinal Pandey and S. Taruna [4]	heterogeneous multiple classifier-based framework	High performance accuracy	Had overfitting problems
Le Hoang Son and Hamido Fujita [5]	Neural-fuzzy with representative sets	better accuracy and computational time	dynamic neurofuzzy models and the convergence rate analysis were not examined
M. R. M. VeeraManickam *et al.* [6]	Map-reduce architecture based cumulative dragonfly based neural network (CDF-NN)	Low MSE and RMSE value	Map reduce required high processing time
L. Ramanathan *et al.* [7]	Lion–Wolf based deep belief network (LW-DBN)	lower error performance	clustering and optimization were not done
Asmaa Elbadrawy *et al.*[8]	Personalized multiregression and matrix factorization	forecasting student performance in a timely and accurate manner	It was not incorporated into real time planning

3   Challenges

The challenges faced during student performance evaluation are,

- The clustering scheme divides the data into multiple groups by calculating the similarity measures between the data points in the database. One of the major challenges involved in clustering is dividing the database with vagueness [21].
- For that time, predicting student performance in a particular course relies on the student's past performance in other courses, a key challenge for training an effective predictor is how to handle heterogeneous student data due to different areas and interests [2].
- In [5], although better accuracy and computational time are achieved, dynamic neuro fuzzy models and the convergence rate analysis were not examined, which is an important challenge.
- The main challenge in the Educational Data Mining method is deepening the generalizability of the results for improving the prediction performance [3].
- In [2], the Machine Learning Approach provided valuable information for academic advisors to recommend subsequent courses to students and carry out pedagogical intervention measures but the performance prediction to elective courses was not evaluated.

4   Proposed Methodology

The main aim of the research is to design and develop a student performance prediction model. The main contribution of the research will be the development of new prediction strategy based on the optimization algorithm. The proposed method consists of two phases: Training phase and the testing phase. The input to the proposed method will be the student assessment marks. In the training phase, the input will be given to the mapper, in which the features will be selected from the input data using holoentropy. Then, the intermediate data will be generated using the features. After that, the intermediate data will be provided to the reducer function, which is built with hybrid NFS- BCS (Neuro Fuzzy system- Bat based chicken swarm)

algorithm. The hybrid NFS-BCS algorithm is the hybridization of Neuro fuzzy system, chicken swarm [22] and bat algorithm [23]. Then, a training model will be generated. At the testing phase, the information of the students will be provided to the map function which will be built with the trained hybrid NFS- BCS algorithm and the intermediate data will be generated. Then, the intermediate data is provided to the reducer function which will provide the estimated marks of the students. The proposed method will be implemented in Python and the analysis will be enabled. The performance analysis of the proposed method will be done based on the metrics, such as average error, Root Square Mean Error. The comparative analysis will be done with respect to the performance metrics, and the results of the methods will be compared among each other. Specifically, the proposed method of student performance prediction model will be compared with the existing methods [2], [5], and [6]. Figure 1 shows the block diagram of the proposed student performance prediction model.
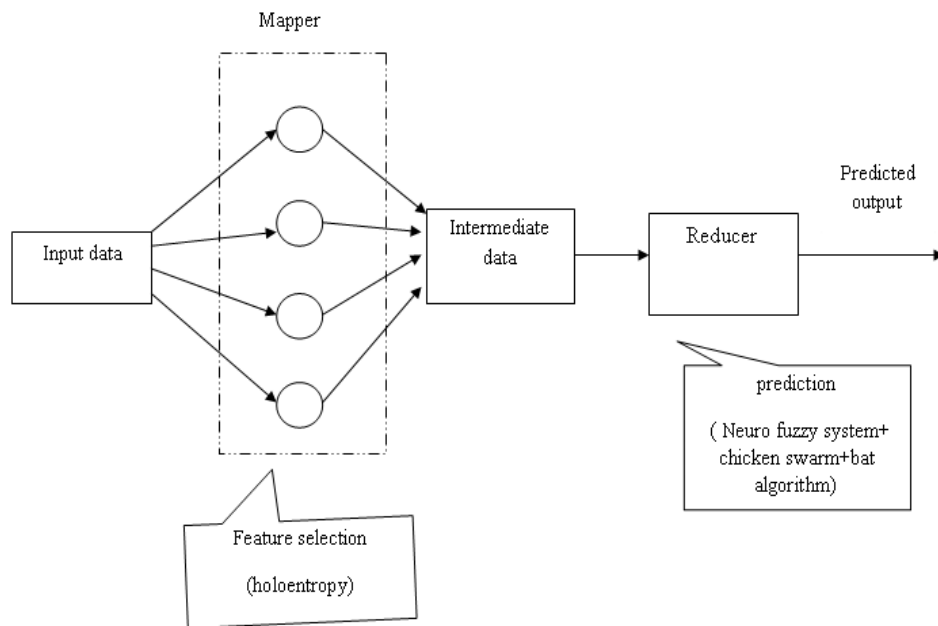

Figure 1: *Block diagram of proposed student performance prediction model*

5 Conclusions

So, at the end of this paper we can conclude that the proposed technique of map reduce algorithm is helpful to predict the students' performance analysis and plan the training activity to them to get motivated in their interested domain as per their choice.

References

[1] Elvira Popescu and Florin Leon," Predicting Academic Performance Based on Learner Traces in a Social Learning Environment, "IEEE Access, 1–1, 2018.

[2] Jie Xu, Kyeong Ho Moon, and Mihaela van der Schaar, " A machine learning approach for tracking and predicting student performance in degree programs," IEEE Journal of Selected Topics in Signal Processing, 11(5), pp.742-753, 2017.

[3] Asif, R., Merceron, A., Ali, S. A., & Haider, N. G. (2017). Analyzing undergraduate students' performance using educational data mining. Computers & Education, 113, 177–194.

[4] Mrinal Pandey and S. Taruna,"Towards the integration of multiple classifier pertaining to the Student's performance prediction,"Perspectives in Science, 8, 364–366, 2016.

[5] Le Hoang Son and Hamido Fujita," Neural-fuzzy with representative sets for prediction of student performance," Applied Intelligence, 2018.

[6] M. R. M. VeeraManickam, M. Mohanapriya, Bishwajeet K. Pandey, Sushma Akhade, S. A. Kale, Reshma Patil, M. Vigneshwar, " Map-Reduce framework based cluster architecture for academic student's performance prediction using cumulative dragonfly based neural network. Cluster Computing, 2018.

[7]   L. Ramanathan, G. Parthasarathy, K. Vijayakumar, L. Lakshmanan, S. Ramani," Cluster based distributed architecture for prediction of student's performance in higher education,"Cluster Computing, 2018.

[8]   Elbadrawy, A., Polyzou, A., Ren, Z., Sweeney, M., Karypis, G., & Rangwala, H. (2016). Predicting Student Performance Using Personalized Analytics. Computer, 49(4), 61–69.

[9]   Ferguson, R. (2012). The State of Learning Analytics in 2012: A Review and Future Challenges. Technical report, Milton Keynes, UK: Knowledge Media Institute, The Open University.

[10]   Steiner, C., Kickmeier-Rust, M., & Türker, M.A. (2014). Review Article about LA and EDM Approaches. Project deliverable, available at: http://css-kmi.tugraz.at/mkrwww/leas-box/downloads/D3.1.pdf.

[11]   Ellis, C. (2013). Broadening the scope and increasing the usefulness of learning analytics: the case for assessment analytics. British Journal of Educational Technology, 44 (4), 662-664.

[12]   Baker, R.S., & Inventado, P.S. (2014). Educational data mining and learning analytics. In Larusson J., White B. (Eds.), Learning Analytics, Springer, pp. 61-75.

[13]   Romero, C., & Ventura, S. (2010). Educational data mining: a review of the state of the art. IEEE Transactions on Systems, Man, and Cybernetics – Part C: Applications and Reviews, 40, 601-618.

[14]   Chatti, M.A., Dyckhoff, A.L., Schroeder, U., & Thüs, H. (2012). A reference model for learning analytics. International Journal of Technology Enhanced Learning, 5, 318-331.

[15]   Asif, R., Merceron, A., Ali, S.A., Haider, N.G.: Analyzing undergraduate students' performance using educational data mining. Comput. Educ. 113, 177–194 (2017).

[16]   Domingos, P., Pazzani, M.: On the optimality of the simple Bayesian classifier under zero-one loss. Mach. Learn. 29(2), 103– 130, 1997.

[17]   Nürnberger, A., Pedrycz, W., Kruse, R.: Data mining tasks and methods: Classification: neural network approaches. In: Handbook of Data Mining andKnowledge Discovery,NewYork, pp. 304–317, 2002.

[18]   Costa, E.B., Fonseca, B., Santana, M.A., de Araújo, F.F., Rego, J.: Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses. Comput. Hum. Behav. 73, 247–256, 2017.

[19]   Wolff, A., Zdrahal, Z., Herrmannova, D., Knoth, P.: Predicting student performance from combined data sources. In: Alejandro, P.-A. (ed.)Educational datamining, pp. 175–202. Springer, Cham, 2014.

[20]   Yahya, A.A.: Swarm intelligence-based approach for educational data classification. J. King Saud Univ. Comput. Inf. Sci, 2017.

[21]   Liu, L., Sun, L., Chen, S., Liu, M., Zhong, J.: K-PRSCAN: a clustering method based on PageRank. Neurocomputing 175, 65–80, 2016.

[22]   Meng, X., Liu, Y., Gao, X., & Zhang, H. (2014). A New Bio-inspired Algorithm: Chicken Swarm Optimization. Advances in Swarm Intelligence, 86–94.

[23]   Yang, X.-S. (2010). A New Metaheuristic Bat-Inspired Algorithm. Studies in Computational Intelligence, 65–74.