

Using Data Mining and Machine Learning to Improve Education Experience

Introduction

The objective of this review is to analyze possible uses of data mining and machine learning to provide a better learning experience for students. Using these tools, it is possible to more accurately analyze student performance, make education more personalized, and assist in the education process of learning.

Analyzing Student Performance

In modern society, education is now sought after more than ever. Georgetown University estimates that 35 percent of employment opportunities will require a bachelor's degree, 30 percent of jobs will require some college, and 36 percent will not require more than a high school diploma or equivalent [1]. With the majority of the population needing to achieve some college it will be important to identify student's potential and weaknesses early so that the proper education path can be curated. Data mining has been found to be a quite effective tool in analyzing student performance.

Ma et al. [2] found when using data mining they were able to accurately identify students that needed this additional help. They were able to reduce the remedial class size by 20 percent when compared to a traditional method of selecting students while maintaining a close recall score [2]. These results were produced using a newly created SBA (Scoring Based on Association) scoring technique. Association rule data mining can be explained by how items purchased from customers are associated [2]. For example, if a customer buys chocolate milk syrup it could be associated with milk to show how often a customer will buy the other item. However, Ma et al. [2] faced issues since their data was a relational table. To avoid this issue, they converted numeric values into intervals and treated each record as a set.

Dekker et al. [3] also found that using data mining they were able to predict students drop out rate with an accuracy between 75 to 80 percent. However, they approached their study with a different data mining technique in which they used decision tree algorithms. Dekker et al. [3] used a wide array of Weka machine learning algorithms such as J48, CART, BayesNet, SimpleLogistic, JRip, RandomForest, and OneR to classify their data. First, pre-university data and university grades data were classified separately with each algorithm or technique. Next, classification was then performed on the dataset that contained both of these which produced results much closer to the university grades dataset results. After a discussion with a counselor, they found misclassifications which were then justified. Dekker et al. [3] note that even with minimal data it is possible to produce accurate results when using machine learning cost-sensitive learning.

Personalizing Education

To further better the learning experience it could also be beneficial to personalize the education material that is delivered. Kalloo, Kinshuk, and Mohan explain in [4] that students may have vastly different ways of learning and it can be useful to provide different tools. There have

already been several attempts at creating systems that can help produce better material or more personalized content.

For example, Head et al. [5] describe the process of creating Tutoron, a tool that autogenerates explanations of online code using data mining. In [5] Tutoron was built in three rather simple steps which contain a detection stage, parsing stage, and lastly an explanation stage that produced confident results. These micro-explanations led to only 6 percent of users needing to access additional resources compared to the 64 percent that used additional resources without the micro-explanations.

Another example of personalizing education is a recommender system that is used to let students explore interest. Romero et al. [6] describe the assembly of two recommender systems which suggest links based on content-based filtering and collaborative-based filtering. Using sequential mining algorithms such as AprioriAll, GSP, and PrefixSpan Romero et al. [6] were able to produce suggestions to just a of couple links at 55 percent confidence. Since online education and resources are growing rapidly a robust recommender system could help students continue their learning or interest in a topic.

Conclusion

These studies have shown that there are many applications of data mining and machine learning which can be applied to improve the educational experience. As shown in [2] it is possible to identify students that are in greater need of a remedial class. This makes it easier to provide more one-on-one time with students that need additional instruction. More accurately predicting students that are at-risk could also help ease the stress on educators which often relied on to make these decisions. Also, by using machine learning techniques discussed in [3] it would be possible to start weighing these decisions from an early moment in the student's educational journey. Machine learning could better portray information about students so that funds and enhanced focus could be targeted to at risk groups. For example, if there were results that showed a particular subset of students that are more likely to drop out more resources and directed attention could be concentrated towards them.

It is shown in [4] that some students learn in a vastly different way and keeping a student's attention it could increase their performance. Accompanied by the results from [5] you could develop a diverse tool that could keep a student's attention and minimize distraction or motivation problems by giving a more personalized experience. I think it will be important to realize that students learn topics at different speeds and the use of data mining can create content that will better suit their needs. Using data mining we have the ability to find and produce an abundance of relatable content which can save the time and headache of searching for similar additional resources. Complemented by another technology such as the tool described in [6] could better adapt to an individual learner. Imagine a dedicated system that could determine what your interest or educational needs are and suggest links based on that information.

I believe that we are just on the cusp of starting to integrate tools like data mining and machine learning to provide a well-rounded learning experience. These studies have shown that there are a wide array of areas and techniques that can be explored moving into the future. I believe that the studies and figures above are crucial to the furthered research and development that will enable student's learning retention and educational success.

References

- [1] “Recovery: Job growth and education requirements through 2020 - CEW Georgetown,” *Georgetown.edu*, 27-Jun-2013. [Online]. Available: <https://cew.georgetown.edu/cew-reports/recovery-job-growth-and-education-requirements-through-2020/>. [Accessed: 15-Nov-2020].
- [2] Y. Ma, B. Liu, C. K. Wong, P. S. Yu, and S. M. Lee, “Targeting the right students using data mining,” in *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '00*, 2000.
- [3] G. W. Dekker, M. Pechenizkiy, and J. M. Vleeshouwers, *Predicting Students Drop Out: A Case Study*. International Working Group on Educational Data Mining.
- [4] V. Kalloo, Kinshuk, and P. Mohan, “Personalized game based mobile learning to assist high school students with mathematics,” in *2010 10th IEEE International Conference on Advanced Learning Technologies*, 2010.
- [5] A. Head, C. Appachu, M. A. Hearst, and B. Hartmann, “Tutorons: Generating context-relevant, on-demand explanations and demonstrations of online code,” in *2015 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, 2015.
- [6] C. Romero, S. Ventura, J. A. Delgado, and P. De Bra, “Personalized links recommendation based on data mining in adaptive educational hypermedia systems,” in *Creating New Learning Experiences on a Global Scale*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 292–306.