https://doi.org/10.48009/2_iis_2019_56-61

Issues in Information Systems
Volume 20, Issue 2, pp. 56-61, 2019

THE USE OF MACHINE LEARNING IN HIGHER EDUCATION

Frank Swiontek, University of North Dakota, frank.swiontek@und.edu
Assion Lawson-Body, University of North Dakota, alawsonbody@business.und.edu
Laurence Lawson-Body, University of North Dakota, laurence.lawsonbody@business.und.edu

ABSTRACT

The ability to have machines draw accurate conclusions has had powerful implications in today's education setting. Despite the powerful applications of machine learning, the higher education industry has been slow to adapt its usage. The acceptance and proper use of machine learning techniques are still in their infancy stages within the higher education industry. The objective of this paper is to examine the proper usage of machine learning in higher education from both students' perspective and institutional perspective. The Bayesian Modeling Averaging (BMA) method was used to study the usage of machine learning at a student level whereas, the Monte Carlo Simulations (MCS) method was used to study the usage of machine learning at an institutional level. From the students' perspective, the results allowed a comparison between actual and projected retention. From the institutional perspective, the results allowed establishing a competitive strategy to improve the university's rank among other research universities possible.

Keywords: Artificial Intelligence, Machine Learning, Higher Education, Bayesian Modeling Average, and Monte Carlo Simulations

INTRODUCTION

Machine learning techniques and algorithms have been used in several fields such as healthcare, smart grids, vehicular communications, smart cities, education, and so on (Ud Din, Guizani, Rodrigues, Hassan & Korotaev, 2019). These artificial learning techniques helped provide pervasive connections for wireless nodes (Ud Din, Guizani, Rodrigues, Hassan & Korotaev, 2019). Recently, many authors have studied machine learning and identified its value in higher education (Gray & Perkins, 2019; Aldowah, Al-Samarraie & Fauzy, 2019; Chung & Lee, 2019). As a matter of fact, machine learning, under the umbrella of artificial intelligence, has been utilized to find solutions to the problem of student retention (Gray & Perkins, 2019), to students' dropouts (Chung & Lee, 2019), and to the potential influence of data mining analytics on the students' learning processes (Aldowah, Al-Samarraie & Fauzy, 2019). Machine learning becomes one of the emerging technologies that has grabbed the attention of academicians and industrialists (Ud Din, Guizani, Rodrigues, Hassan & Korotaev, 2019).
.
The higher education industry can trace its roots all the way back to 1696 (Geiger, 2014). Hundreds of universities and colleges have stood the test of time, and some have not. Today, we see a blending of contrasting values that make the modern education system a truly unique and powerful force. However, lucrative universities have used strategies to reshape the playing field for many of these enduring universities. Longstanding tradition and forward thinking are the pillars to the success the higher education industry has experienced, but now the speed at which information travels is unparalleled. Universities have been pushed to use the techniques they teach, to sustain a competitive advantage in today's market. Analytics and predictive modeling techniques, which have been commonplace in the business sector for years, have now taken hold within the higher education communities.

The highest level of these techniques falls under the category of machine learning. Machine learning is defined as, "A computer program [learning] from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E. (Mitchell, 1997)". Simply put, a computer program improves its ability to perform a task over repetition by developing associations to the information.

In one of the world's most competitive industries, universities can use machine learning to assist in almost all areas of operations and the successful use of the techniques could lead to powerful and cascading advantages never before

witnessed in the industry. Despite the importance of the powerful applications of machine learning, the higher education industry has been slow to adapt its usage. That means, the acceptance and proper use of machine learning techniques are still in their infancy stages within the higher education industry. This poses challenges for not only the developers and students, but also for institutional leadership. The objective of this paper is to examine the proper usage of machine learning in higher education from both students' perspective and institutional perspective. The rest of the paper is organized as follows: the next section presents the literature review and the evolution of machine learning followed by the research methodology, discussion and conclusion.

LITERATURE REVIEW AND EVOLUTION OF MACHINE LEARNING

The term, "machine learning," popped up for the first time in 1959, when researchers at IBM began exploring neural networks (Russell & Norvig, 1995). The early stages of neural networks used an algorithm called the perceptron. The idea behind the perceptron was to give binary decision points throughout a task to formulate weights for a specific function. However, the work of Marvin Minsky and Seymour Papert's *Perceptrons* argued that these algorithms would not improve their function by adding multiple layers, and ultimately led to the assumption that the algorithm had reached its limitations in computational learning. Minsky and Papert's work led to a decrease in research in machine learning until around the early 1990s. In late 1980's and 90's new work was being done to allow for multiple decision layers, and Minsky and Papert's work was proven incorrect with the introduction of deep learning algorithms (Bishop, 2016).

Now a number of models, algorithms, and processes exist in the machine learning field. Uses range from chat programs to procurement systems and forecasting to phone app suggestions. The last few years, many authors have expressed the need to apply machine learning to the evolution of our society. Ud Din, Guizani, Rodrigues, Hassan & Korotaev (2019) studied different Internet of Things-based machine learning mechanisms used in the higher education. Internet of Things is a network that supports communications among various devices without human interactions (Ud Din, Guizani, Rodrigues, Hassan & Korotaev, 2019). Chung & Lee (2019) combined machine learning and big data in education to develop a predictive model showing performance metrics that can be used to analyze students' dropouts. Aldowah, Al-Samarraie & Fauzy (2019) found that specific educational data mining and learning analytics techniques could offer the best means of solving certain learning problems. Zhang, Meng, Ordonez de Pablos & Sun (2019) used learning analytics in IT environment to understand the whole process of students' learning in groups, analyze problems in collaborative learning, and find ways to improve group-learning effects. Filva, Forment, Garcia-Penalvo, Escudero & Casan (2019) developed Learning Analytics (LA) and Educational Data Mining (EDM). Both are analytical approaches to enhance and optimize learning environments. EDM is a branch of data mining and machine learning to analyze educational data, and its origin is in educational software and student modeling (Filva, Forment, Garcia-Penalvo, Escudero & Casan, 2019). The EDM approach is focused on computing and automation system (Filva, Forment, Garcia-Penalvo, Escudero & Casan, 2019). On the other hand, the discipline of LA is more human-centered (Filva, Forment, Garcia-Penalvo, Escudero & Casan, 2019). It allows for analyzing student behavior patterns when they interact with tools and online learning environments, set them in context with their learning outcomes and draw conclusions to enhance the evaluation or improve the learning process through human judgment (Filva, Forment, Garcia-Penalvo, Escudero & Casan, 2019).

Machine learning has become so ingrained in our day-to-day life that it has almost become a second member of our society and our education systems.

RESEARCH METHODOLOGY

The following two sections will cover the practical use of machine learning in today's higher education industry from a first-hand perspective. The Midwestern research university small team's efforts were able to provide accurate, timely, and unique information. As we will discuss the sections, despite being accurate, the data proved difficult to implement due to its unfamiliar nature to staff and faculty.

Bayesian Modeling Averaging (BMA)
Our team was given the opportunity to assess the value of potential student ACT name buys. These are names, emails, and test score lists available for purchase from the ACT testing agency. These names were then used for recruitment mailing lists. Throughout the years, the university had purchased upwards of 400,000 names with only a 4% yield. As costs were being cut for budgetary purposes, these name buys were subjected to a value analysis. With only a 4% yield, a simple regression analysis highlighted that the majority of the actual enrollees came from the region or for the universities most popular major. However, because the correlations with region and major were so strong, the team looked to further the assessment by creating a tool for recruitment. Initially, the name buy lists were used to establishing a predictive model. Due to the findings of the previous analysis, the team turned its attention to using application data to outline who had the highest probability of enrolling.

The team utilized an ensemble method of Bayesian Modeling Averaging (BMA) with a machine learning technique called random forest. BMA allows multiple models to be ran, and weights accordingly to the likelihood of accuracy. Random forest is a type of decision tree machine learning that establishes observations that contribute to the intended end result. In other words, the program sets out to find various combinations of variables to establish its own formula to predict the outcome given. In the team's case, students from past years were given a binary dummy variable if they enrolled or did not enroll. That data was then used as training data for the model to use on students who were actively in the admissions cycle for the upcoming year. BMA allowed the random forest to create multiple models and weight those models accordingly based off the accuracy each model had in comparison with the others.

The difficulty lies in the lack of previous year data for admissions, as well as the inclusion of the new practice of having students deposit $200 into their accounts. Ninety three percent (93%) of students who deposited ended up enrolling, and while the variable was/is telling, it did offset the results until the model could be fine-tuned. Interestingly, part of the fine-tuning was a result of data runs, and the process "learning" that was occurring.

One hundred thousand (100,000) simulations were ran each week in parallel to the enrollment probability scores to give an estimated enrollment count for the fall semester. These final numbers are used for budgetary purposes and financial forecasting as they have been 99% accurate since the model's inception.

The team then applied the same methodology with alternate data to predict retention. By using BMA and random forest, the team used data like course grades, housing status, dining center frequency, and various other data points to establish a similarly successful model that gave students a probability of returning the following semester. A sample of past data was used to predict the likelihood of returning with a high degree of accuracy.

Unfortunately, the implementation of the information was not as successful as the results of the modeling. While the projected enrollment fall number is still in use today by the university finance department, the probability scores proved to be a difficult sell to the departments that the information was intended for. The team still struggles with why those departments had such difficulty even accepting the data, after the team proved its validity. Machine learning can be difficult to understand, and the notion that the product was not created by a human can be difficult to accept. That inability to understand how the information was created may act as a barrier for the information's use more often than not.

Monte Carlo Simulations (MCS)
The second application of machine learning came in the way of trying to establish a competitive strategy to improve its ranking among other research universities. As the ranking system, Carnegie Classification of Institutions of Higher Education, uses numerous variables to finalize on rankings; the direction of where to place resources to improve one's position is vague due to the number of universities ranked.

The team set out to establish (a) MCS that used previous year data from the Carnegie ranking database. Essentially, the team looked to use the MCS in accordance with game theory to project what all other universities were likely to do under the assumption that those same universities were trying to improve their rank. Research had to be done in accordance with how much maneuverability each university had in respect to the weights used in the ranking calculations (e.g., internal research expenditures, number of Ph.D. programs, etc…). The universities were given tiers in order to account for what would be a realistic investment for each area.

Once the constraints for each area of the ranking formula had been assigned, the team elected to pair the simulations with reinforcement learning. The idea behind reinforcement learning is have the program develop methodology through uncertainty to maximize a reward. Specific actions, or investments, would change the states that may have occurred in the environment as seen in Figure 1. This process is also sometimes referred to as a Markov Decision Process.

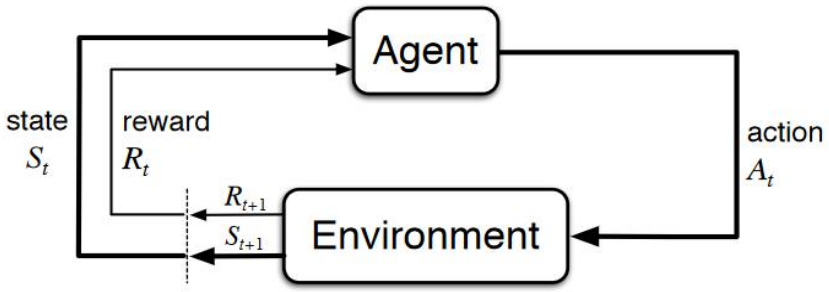

Figure 1. The agent–environment interaction in reinforcement learning (Sutton & Barto, 2017)

In the team's case that reward was to maximize the rank of the university while accounting for random variables from the Monte Carlo simulations used to create the simulated ranks. Because each individual ranked area would ultimately change the overall ranking of the same university, those areas would also affect the rankings of all other universities due to the calculation utilizing a universities ranking of each area as a weight in the final formula. Thus, small changes in a single area could have a profound impact on all other universities. The university intended for the maximized reward would need to adjust its actions based off the various states that could occur. This required the program to be able to learn from various scenarios, judge what was the most likely outcome, and act in a manner that maximized its reward while considering all other factors.

While the intent of the team was sound, and the model itself functioned, the team was unable to acquire the vast amount of information needed to give an accurate assumption for the section actions each university would and/or could take. Furthermore, the difficulty in predicting unique events, such as one time funding or events that lead to an increase in university attendance, were to obscure to model. Due to the nature of the model's intention, millions of dollars would have to be refocused and/or acquired to act on the educated guess the program was performing. Even with 20 years of data for the program to learn from, the team did not feel comfortable allowing its results to be used even at the informational stage.

Unlike the reception of the admissions and retention models, the reinforcement learning results were sought out by university leadership. Ironically the team lobbied to only show the initial results and not current data as it may have skewed decisions within the organization. With enough time, data, and resources the model would certainly be unique as the concept of using a reinforcement learning model could be seen as excessive for a rankings project.

CONCLUSION

Looking ahead, the utilization of this type of modeling would fit in well to predicting and guiding students in degree paths and course selections. If the model was formatted in a manner that gave suggestions for optimal semester registrations, students could find more palatable course load outs that improved their likelihood of success and graduation.

Despite the notorious stereotype of being an industry hesitant to adapt and change, higher education has found itself in a competitive field that is now seemingly changing as fast as the technology industry of twenty years ago. With college enrollment decreasing, and birth rates slowing, the industry will soon be starved for customers, and use whatever means to maintain their survival.

A new use of machine learning that utilizes unstructured data recognition patterns to identify questions, and even interact with students, is seemingly taking the industry by storm. Chatbots have existed in service and sales

industries for years, but higher education's new found need to be lean and agile has given rise to the adoption of these machine learning tools. Schools have begun using these bots not only to handle billing and registration requests, they have also started integrating the bots into actual coursework to allow for self-guided learning (Jia, 2009). The ability to provide stakeholders personalized communication at a fraction of the cost, allows those same institutions to funnel funds into areas that will allow them to grow rather than just maintain the status quo.

Course instruction and assignment construction is also becoming augmented with machine learning, and in some cases in real-time. Various machine learning methods allow the programs to give suggestions on things like what chapter should be revisited, what assignments should or should not be given, and even construct optimal learning methods for students based on the performance of previous material (Siddique, Durrani, & Naqvi, 2018). It should be noted that these methods have caused some controversy within the industry; as in some cases, the program elects to add or remove material from the curriculum at the expense and/or benefit of the student.

As society moves forward with online coursework, technical "boot-camps," and the election of not pursing a secondary education; the higher educational system, in its current form, is in doubt. However, no matter where the higher education industry ends up, machine learning is here to stay. A shocking sentence when one considers machine learning's age in comparison to higher education's age.

The industry has begun to adapt, and will continue to do so or may it be left to the vulnerabilities of the previous paragraphs alternatives. The implementation of chatbots and adaptive courseware are wonderful examples of how machine learning can benefit the students, but the a real test will be how the higher education industry uses machine learning to sustain a competitive advantage in today's economy, let alone within their own industry.

A tool that can give us a suggested music based of our listening history to vehicles that drive themselves will certainly change our day-to-day lives. But will an industry built on tradition and formalities change its very identity to use the very tools it most certainly had a hand in creating? We believe they will, and we believe they will become better for it. However, there will be those who do not adapt quick enough, and they may find their inability to learn will ultimately lead to their inability to stay relevant.

REFERENCES

Aldowah, H., Al-Samarraie, H., & Fauzy, M. W. (2019). Educational data mining and learning analytics for 21st century higher education: A review and synthesis. *Telematics and Informatics*. 13-49.

Bishop, C. (2016). Artificial Intelligence, the History and Future. London, United Kingdom https://www.youtube.com/watch?v=8FHBh_OmdsM. Accessed Fall 2018.

Chung, J. Y & Lee, S. (2019). Dropout early warning systems for high school students using machine learning. *Children and Youth Services Review*. 346-353.

Filva, D. A., Forment, M. A., Garcia-Penalvo, F. J., Escudero, D. F. & Casan, M. J. (2019). Clickstream for learning analytics to assess students' behavior with Scratch *Future Generation Computer Systems*, 673-686.

Geiger, R. L. (2014). The History of American Higher Education Learning and Culture from the Founding to World War II. Princeton University Press.

Gray, C. C. & Perkins, D. (2019) Utilizing early engagement and machine learning to predict student outcomes. *Computers & Education Journal*, 22-32.

Jia, J. (2009). CSIEC: A computer assisted English learning chatbot based on textual knowledge and reasoning. *Knowledge-Based Systems*, 249-255.

Mitchell, T. M. (1997). Machine Learning. McGraw-Hill International Edition Computer Science Series. 1st Edition.

Russell, S. J., & Norvig, P. (1995). *Artifical Intelligence: A Modern Approach.* Englewood Cliffs: Prentice Hall. 3rd Edition.

Siddique, A., Durrani, Q. S., & Naqvi, H. A. (2018). Developing Adaptive E-Learning Environment Using Cognitive and Noncognitive Parameters. *Journal of Educational Computing Research*, 1-35.

Sutton, R. S., & Barto, A. G. (2017). *Reinforcement Learning: An Introduction.* Cambridge: The MIT Press.

Ud Din, I., Guizani, M., Rodrigues, J. J.P.C., Hassan, S. & Korotaev, V. (2019). Machine learning in the Internet of Things: Designed techniques for smart cities. *Future Generation Computer Systems.* https://doi.org/10.1016/j.future.2019.04.017

Zhang, X., Meng, Y., Ordonez de Pablos, P. & Sun, Y. (2019). Learning analytics in collaborative learning supported by Slack: From the perspective of engagement. *Computers in Human Behavior*, 625-633.