An overview of Machine Learning Technologies and their use in E-learning

Ramzi FARHAT
LaTICE research laboratory
Université de Tunis
Tunis, Tunisia
ramzi.farhat@esstt.rnu.tn

Yosra MOURALI
Faculty of Economics and Management
Université de Sfax
Sfax, Tunisia
yosra.mourali@etu-uphf.fr

Mohamed JEMNI
LaTICE research laboratory
Université de Tunis
Tunis, Tunisia
mohamed.jemni@fst.rnu.tn

Houcine EZZEDINE
Univ. Polytechnique Hauts-de-France
CNRS, UMR 8201 - LAMIH
Valenciennes, France
Houcine.ezzedine@uphf.fr

***Abstract*—Thanks to new technologies, internet, connected objects we produce a phenomenal amount of data. Putting these data in context, organizing them to be able to perceive, understand and reflect them is very important. Traditionally, human have analyzed data. However, as the volume of data surpasses, human increasingly turn to automated systems that can imitate him. Those systems able to learn from both data and changes in data in order to solve problems are called machine learning. Artificial intelligence has a major impact on e-learning research and the machine learning based methods can be implemented to improve Technology Enhanced Learning Environments (TELE). This paper is an overview of the recent findings in this research field. At first, we introduce the key concepts related to machine learning. Then, we present some recent works using machine learning in e-learning context.**

Keywords—E-learning, Technology Enhanced Learning Environments, Data, Learners' traces, Machine Learning, Deep Learning

I. INTRODUCTION

Almost everything we do today leaves a digital trace that describes our activities, specifies our location, and provides many other information about what we say, what we buy, etc. Thanks to both data storage capacity and digitalization of society, most of devices, machines and everything we use, produce data. We can, as example, extract information from pay stations, parking, smart phone, social networks, videos, photos, etc. It is necessary to benefit and find meaning to all this collected data.

Analyzing data makes it possible to understand phenomena, to model behaviors and to make predictions. Before, humans analyze data, wrote algorithms and the machine applied them to solve problems. Today, humans introduce data and allows the machine to learn on its own from these data without being explicitly programmed. We talk about the power of data. This is the principle of machine learning.

In reality, there is an awareness of the richness that data can hold and the importance of valuing it. Actually, analyze of complex data through machine learning methods has emerged as an important era in several scientific research domains such as medicine [1] [2], e-commerce [3], industry [4][5], education [6][7], social networks [8][9], economics and finance [10], etc.

Figure 1 shows machine learning relationships to some other concepts of data science and artificial intelligence. In fact, data mining use statistics to extract hidden information (patterns) from raw data [11]. However, ma-chine learning as a subfield of computer science and artificial intelligence, learns from patterns to predict. The Deep Learning is one of the main technologies of ma-chine learning and artificial intelligence. We can say that it is the new generation of machine learning which characterized by learning by layer and on each layer the ma-chine has to learn a little more.

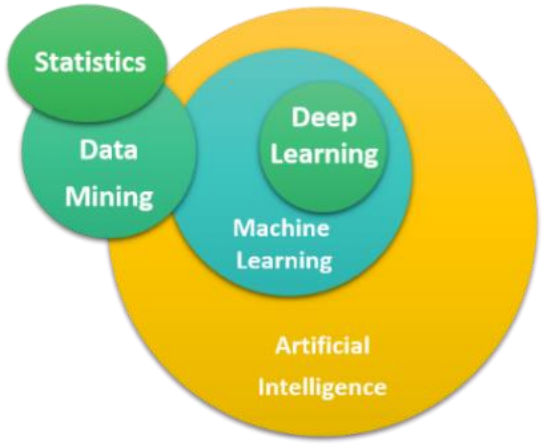

Fig. 1 Machine Learning relationships to other related fields

II. MACHINE LEARNING

In machine learning, a computer learns from example data how to perform tasks. We know that if we give more experiences (E) with a defined task (T) to a machine, its performance (P) improves [12]. For example, let suppose that we want an email client to classify emails as spam or not. The experience E in this case should be a set of emails already classified as spam or not. The Task T perform is to classify automatically new emails. The performance P that should increase is the accuracy rate of the classification made by the machine on a set of new emails.

A. Machine learning process

The generic machine learning process consists of seven steps as described next [13]. The first step is to collect data. It is a very important task because it will determine how good

978-1-7281-6403-8/20/$31.00 ©2020 IEEE

Authorized licensed use limited to: University of Zagreb. Downloaded on October 16,2024 at 11:14:33 UTC from IEEE Xplore. Restrictions apply.

predictive model can be. But, data we gathered are, in most times, unstructured, contain a lot of noise or have to take other forms to be useful for our ma-chine learning. So, data need to be cleaned and pre-processed.

After that we can begin building our machine learning model. For this, we start by the feature engineering in which we choose the most relevant features from data, then we try to select the best machine learning algorithm for the problem at hand. It is imperative in getting the best possible results.

The next task is training. In this step we use a part of our data to incrementally improve machine learning ability to predict. Once training is complete, it is time to test the model and observe how it might perform against the other part of data unseen. The performance evaluation is measured by various parameters like accuracy, precision and recall. Sometimes, it is possible to go back and improve training then test again. The last step is the result given by the machine learning. It can be a prediction or inference.

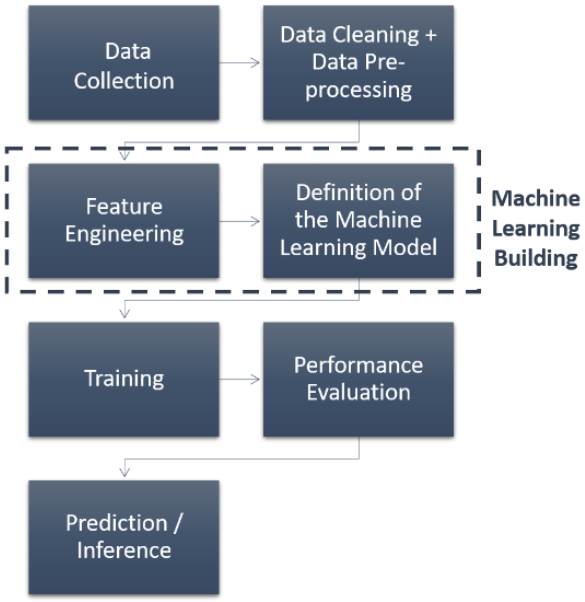

Fig. 2 Components of a Generic machine learning model

B. *Machine learning paradigms*

Machine learning can be classified based on the approach used for the learning process. Four main categories were identified: supervised, unsupervised, semi-supervised, and reinforcement learning [12].

In supervised learning, we have a set of training data or labelled data in which we know the structure and the outcome. We take this data and train a machine learning model, so it can understand patterns in the data. Once the model has been trained, we can use it to predict results of data in which outcomes are unknown [14].

Conversely, unsupervised learning methods learn structure from the data itself without the need for prior label-ling [15]. That is mean we can apply unsupervised ma-chine learning to find patterns that exist within labelled data.

However, full label information is not available at all times. Semi-supervised learning provides a powerful framework for leveraging unlabeled data when labels are limited or expensive to obtain [16].

The last machine learning approach, serves when we know what we are looking for but we do not know how to get it. The principle is to test several solutions and then we see which ones make it possible to obtain the desired result. Reinforcement learning problem can be formalized as an agent that has to make decisions in an environment. The agent learns a good behavior. This means that it modifies or acquires new behaviors and skills incrementally. Thus, the reinforcement agent does not require complete knowledge or control of the environment, but it only needs to be able to interact with the environment and collect information [17].

III. MACHINE LEARNING E-LEANRING APPLICATIONS

Nowadays, everyone wants to learn and develop his knowledge in many fields, students, employers, etc. With the spread of lifelong learning, education systems are seriously facing the modernization and e-learning is be-coming more and more popular. All this leads to a massive growth in the number of Technology Enhanced Learning Environments (TELE) offering open or private online courses and other different types of services. Ana-lysing the large amount of data produced by TELE through machine learning methods has emerged. It is useful to study how to exploit this powerful, new technology to enhance e-learning.

A. *Sentiment analysis*

Recently, Massive Open Online Course (MOOC) success is considered as the extent of student satisfaction with the course [18]. Sentiment analysis can be used to identify complex emotions [19] aiming the prediction of learner satisfaction. In [19], researchers want to deter-mine throw forum messages in MOOC the polarity of learners' sentiments, positive sentiments and negative sentiments. They compare five supervised machine learning algorithms which have been used more frequently in contributions related to prediction in MOOCs: Logistic Regression, Support Vector Machine, Decision Tree, Random Forest and Naïve Bayes. Results show that the most reliable technique was Random Forest.

Understanding the role of emotions in MOOC students' learning experiences is very important. In one hand, according to [20] the control of achievement emotions may serve to improve learning engagement. Based on SVM, [20] build a supervised machine learning model in order to automatically categorize achievement emotions. SVM was adopted as it gives better performance results than Naïve Bayes, Logistic Regression and Decision Tree. On the other hand, [21] track the emotional tendencies of learners in order to analyze the acceptance of the courses using big data from homework completion, comments and forums. Based on semantic analysis and machine learning, [21] investigate the relationship be-tween emotional tendencies and learning effects.

B. *Student behaviour prediction*

An interesting literature review [22] has addressed the question of machine learning use in predicting student behavior. Tow research goals were identified: student classification and dropout prediction.

- Student classification:

Certainly, personalities, backgrounds knowledge, skills and preferences play a crucial part in the learning process. Recommender systems serve to give most suitable content to

Authorized licensed use limited to: University of Zagreb. Downloaded on October 16,2024 at 11:14:33 UTC from IEEE Xplore. Restrictions apply.

each learner. Profiling and classifying learners is a primordial task not only to personalize learning but also to identify abandonment factors and many other purposes. We summarize in table 1, some recent works focusing on the student classification using machine learning.

TABLE I. STUDENT CLASSIFICATION

Paper	Machine Learning Algorithm	Classification goal	Results
[23]	k-means Support Vector Machine (SVM) Naïve Bayes	Classification of engaged and disengaged faces of students with dyslexia	accuracy with 97–97.8%
[24]	Backpropagation (BP), Support Vector Machine (SVM), Gradient Boosting Classifier (GBC)	classification of student performance	Accuracy: BP = 87.78%, 83.20%= 83.20%, GBC= 82.44%
[25]	Decision Tree, Logistic regression, k-nn, SVM, random forest algorithms	Classification of successful and unsuccessful students	K-nn gives the higher accuracy = 85%
[26]	K-modes clustering algorithm Naive Bayes classifier	Classification of learner's learning style	Accuracy = 89%

- Dropout prediction:

Various machine learning techniques have been applied to analyze interactive behavior traces left across TELE. According to [27] who focuses on learners' clickstream data, Logistic regression (LR) has been the most frequently used technique to predict student dropout in MOOC environment achieving 89% as accuracy. SVM and Decision Tree occupy the second position, however, Natural Language Processing Technique come in the third place.

C. Self-Regulated Learning

With the little external teacher's monitoring in majority of TELE, learners are required to make decisions related to their own activities [28]. In that case, individuals with strong self-regulated learning (SRL) skills, characterized by the ability to plan, manage and control their learning process, can learn faster and better than those with weaker SRL skills [29]. As it is one of e-learning platforms supporting SRL strategies [30], MOOC aims learners to self-evaluate the quality or the progress of their work, to set goals and plan and give them the possibility to reread notes, logs, tests, or learning materials to prepare for testing, etc. De-spite all those features, it remains important for many re-searchers to enhance student SRL based on machine learning approach.

Based on learners' log traces and responses to a survey, [31] contribute to enhance understanding of how students learn, and how instruction should be designed to support SRL in an asynchronous online course at a women's university in South Kore. In this study, researchers proceed to the discovery of student profiles and the examination of student SRL process overtime. At first, they suggested three key SRL attributes; time investment in content learning, study regularity and help-seeking that apply to asynchronous online courses to serve as the basis for the analytics of SRL, and guided the selection of log variables. Second, they identified student subpopulations using K-medoids clustering algorithm by silhouette method. After discovering existing clusters and their learning patterns, [31] use random forest classification as a decision tree-based machine learning algorithm to predict cluster membership by referring to each week's log variable.

IV. CONCLUSION

E-learning researchers have spent considerable effort on analyzing learners' data through machine learning methods to enhance learning experience. This seems to be wise since the learner is considered as the main component in the e-learning sphere. However, no research work, the best of our knowledge does carry out in order to use learning data in order to measure content quality in order to improve it.

Thus, in our future work will focus on e-learning con-tent evaluation by using machine learning. The main objective is to help course designers in the educational reengineering process based on machine learning finding and based on many factors, especially past learners' interactions.

REFERENCES

[1] Rakhmetulayeva, S. B., Duisebekova, K. S., Mamyrbekov, A. M., Kozhamzharova, D. K., Astaubayeva, G. N., & Stamkulova, K. (2018). Application of classification algorithm based on SVM for determining the effectiveness of treatment of tuberculosis. Procedia computer science, 130, 231-238.

[2] Kabyshev, M. V., & Kovalchuk, S. V. (2019). Development of personalized mobile assistant for chronic disease patients: diabetes mellitus case study. Procedia Computer Science, 156, 123-133.

[3] Zhu, G., Wu, Z., Wang, Y., Cao, S., & Cao, J. (2019). Online Purchase Decisions for Tourism E-commerce. Electronic Commerce Research and Applications, 100887.

[4] Brik, B., Bettayeb, B., Sahnoun, M. H., & Duval, F. (2019). Towards Predicting System Disruption in Industry 4.0: Machine Learning-Based Approach. Procedia Computer Science, 151, 667-674.

[5] Han, Y., Zeng, Q., Geng, Z., & Zhu, Q. (2018). Energy management and optimization modeling based on a novel fuzzy extreme learning machine: Case study of complex petrochemical industries. Energy conversion and management, 165, 163-171.

[6] Hew, K. F., Hu, X., Qiao, C., & Tang, Y. (2019). What predicts student satisfaction with MOOCs: A gradient boosting trees supervised machine learning and sentiment analysis approach. Computers & Education, 103724.

[7] Hmedna, B., El Mezouary, A., & Baz, O. (2019). How Does Learners' Prefer to Process Information in MOOCs? A Data-driven Study. Procedia computer science, 148, 371-379.

[8] Birjali, M., Beni-Hssane, A., & Erritali, M. (2017). Machine learning and semantic sentiment analysis based algorithms for suicide sentiment prediction in social networks. Procedia Computer Science, 113, 65-72.

[9] Kumari, K. V., & Kavitha, C. R. (2019). Spam Detection Using Machine Learning in R. In International Conference on Computer Networks and Communication Technologies (pp. 55-64). Springer, Singapore.

[10] Ghoddusi, H., Creamer, G. G., & Rafizadeh, N. (2019). Machine learning in energy economics and finance: A review. Energy Economics, 81, 709-727.

[11] Liu, J., Kong, X., Zhou, X., Wang, L., Zhang, D., Lee, I., ... & Xia, F. (2019). Data Mining and Information Retrieval in the 21st century: A bibliographic review. Computer Science Review, 34, 100193.

[12] Portugal, I., Alencar, P., & Cowan, D. (2018). The use of machine learning algorithms in recommender systems: A systematic review. Expert Systems with Applications, 97, 205-227

Authorized licensed use limited to: University of Zagreb. Downloaded on October 16,2024 at 11:14:33 UTC from IEEE Xplore. Restrictions apply.

[13] Alzubi, J., Nayyar, A., & Kumar, A. (2018, November). Machine learning from theory to algorithms: an overview. In Journal of Physics: Conference Series (Vol. 1142, No. 1, p. 012012). IOP Publishing.

[14] Schrider, D. R., & Kern, A. D. (2018). Supervised machine learning for population genetics: a new paradigm. Trends in Genetics, 34(4), 301-312.

[15] Rodriguez-Nieva, J. F., & Scheurer, M. S. (2019). Identifying topological order through unsupervised machine learning. Nature Physics, 1.

[16] Oliver, A., Odena, A., Raffel, C. A., Cubuk, E. D., & Goodfellow, I. (2018). Realistic evaluation of deep semi-supervised learning algorithms. In Advances in Neural Information Processing Systems (pp. 3235-3246).

[17] François-Lavet, V., Henderson, P., Islam, R., Bellemare, M. G., & Pineau, J. (2018). An introduction to deep reinforcement learning. Foundations and Trends® in Machine Learning, 11(3-4), 219-354.

[18] Hew, K. F., Hu, X., Qiao, C., & Tang, Y. (2019). What predicts student satisfaction with MOOCs: A gradient boosting trees supervised machine learning and sentiment analysis approach. Computers & Education, 103724.

[19] Moreno-Marcos, P. M., Alario-Hoyos, C., Muñoz-Merino, P. J., Estévez-Ayres, I., & Kloos, C. D. (2018, April). Sentiment Analysis in MOOCs: A case study. In 2018 IEEE Global Engineering Education Conference (EDUCON) (pp. 1489-1496). IEEE.

[20] Xing, W., Tang, H., & Pei, B. (2019). Beyond positive and negative emotions: Looking into the role of achievement emotions in discussion forums of MOOCs. The Internet and Higher Education, 100690.

[21] Wang, L., Hu, G., & Zhou, T. (2018). Semantic analysis of learners' emotional tendencies on online MOOC education. Sustainability, 10(6), 1921.

[22] de Souza, V. F., & Perry, G. (2019). Identifying student behavior in MOOCs using Machine Learning. International Journal of Innovation Education and Research, 7(3), 30-39.

[23] Hamid, S. S. A., Admodisastro, N., Manshor, N., Kamaruddin, A., & Ghani, A. A. A. (2018, February). Dyslexia adaptive learning model: student engagement prediction using machine learning approach. In International Conference on Soft Computing and Data Mining (pp. 372-384). Springer, Cham.

[24] Sekeroglu, B., Dimililer, K., & Tuncal, K. (2019, March). Student performance prediction and classification using machine learning algorithms. In Proceedings of the 2019 8th International Conference on Educational and Information Technology (pp. 7-11). ACM.

[25] Fedushko, S., & Ustyianovych, T. (2019, January). Predicting pupil's successfulness factors using machine learning algorithms and mathematical modelling methods. In International Conference on Computer Science, Engineering and Education Applications (pp. 625-636). Springer, Cham.

[26] EL AISSAOUI, O., EL MADANI, Y. E. A., OUGHDIR, L., & EL ALLIOUI, Y. (2019). Combining supervised and unsupervised machine learning algorithms to predict the learners' learning styles. Procedia computer science, 148, 87-96.

[27] Dalipi, F., Imran, A. S., & Kastrati, Z. (2018, April). MOOC dropout prediction using machine learning techniques: Review and research challenges. In 2018 IEEE Global Engineering Education Conference (EDUCON) (pp. 1007-1014). IEEE.

[28] Wong, J., Baars, M., Davis, D., Van Der Zee, T., Houben, G. J., & Paas, F. (2019). Supporting self-regulated learning in online learning environments and MOOCs: A systematic review. International Journal of Human–Computer Interaction, 35(4-5), 356-373.

[29] Kizilcec, R. F., Pérez-Sanagustín, M., & Maldonado, J. J. (2017). Self-regulated learning strategies predict learner behavior and goal attainment in Massive Open Online Courses. Computers & education, 104, 18-33.

[30] Garcia, R., Falkner, K., & Vivian, R. (2018). Systematic literature review: Self-Regulated Learning strategies using e-learning tools for Computer Science. Computers & Education, 123, 150-163.

[31] Kim, D., Yoon, M., Jo, I. H., & Branch, R. M. (2018). Learning analytics to support self-regulated learning in asynchronous online courses: A case study at a women's university in South Korea. Computers & Education, 127, 233-251.

Authorized licensed use limited to: University of Zagreb. Downloaded on October 16,2024 at 11:14:33 UTC from IEEE Xplore. Restrictions apply.