

# Chapter 3

## Algorithms for the Development of Deep Learning Models for Classification and Prediction of Learner Behaviour in MOOCs



José Edmond Meku Fotso, Bernabe Batchakui, Roger Nkambou, and George Okereke

### 3.1 Introduction

#### 3.1.1 Background

The demand for training has grown rapidly in recent years. This is evidenced by the high number of learners at every level and type of education: not only in universities but also at the primary and secondary levels as well as in technical and vocational education, both formal and informal. In order to address the issue of massification in education and to encourage sharing of knowledge, distance learning, and especially e-learning, appears to be a suitable approach. This trend is supported by the international agenda, including UNESCO’s “Education for All” initiative, the United Nations’ fourth sustainable development goal (SDG4) (UNESCO et al., 2016; UNESCO, 2005), the Incheon declaration on SDG4-Education 2030, and the Continental Education Strategy for Africa (CESA 16-25) (AFRICAINE, 2016). In particular, SDG4’s objective is to provide “inclusive and equitable quality education and promote lifelong learning opportunities for all” (UNESCO et al., 2016). Massive open online courses (MOOCs) appear to be a suitable approach to support such an initiative. However, MOOCs face the unsolved

---

J. E. M. Fotso (✉) · B. Batchakui  
National Advanced School of Engineering of Yaounde, University of Yaounde 1, Yaounde, Cameroon

R. Nkambou  
Computer Science Department, University of Quebec Montreal, Montreal, QC, Canada

G. Okereke  
Computer Science Department, University of Nigeria Nsukka, Nsukka, Nigeria  
e-mail: [george.okereke@unn.edu.ng](mailto:george.okereke@unn.edu.ng)

major problems of high dropout, low completion, and a low success rate. Around 90% of students who enrol in a MOOC fail to complete it (Andres et al., 2018). In addition, while slightly more than half of students intend to receive a certificate of completion from a typical MOOC, only around 30% of these respondents achieve this certification (Brooks et al., 2015).

Much research has addressed the problems of dropout and failure in MOOCs. Because of the high number of learners and their heterogeneity, a huge volume of data is generated by learners' activities. Many models have been designed to predict dropout, completion, certification, and/or success, with dropout prediction being the most common (UNESCO, 2005). However, the concept of dropping out and success needs to be reconsidered in the context of MOOCs because not all the learners enrolled in MOOCs intend to get the certificate or even complete the course. Therefore, for people who enrol in MOOCs for other purposes, than to get the certificate or to complete the course, not completing the course, and/or not getting the certificate should not be considered a failure and thus should not be classified as such. Nevertheless, whatever the learners' objectives, they need to take part in the learning process in order to truly fulfil those objectives.

### ***3.1.2 Interest***

Our work aims to predict learner participation in the course learning process. To do this, our objective is to design a model to classify and predict learner behaviour, more specifically learners' interaction in the learning process, including with course activities and resources. Such classification and prediction of learner behaviour can serve many purposes. It can be used to improve personalized support and interventions by course instructors and managers; it can also guide the development of adaptive content and learner pathways for learners (Gardner & Brooks, 2018). Altogether, it can then be used to help predict and prevent dropout, thus improving the completion and success rate.

### ***3.1.3 State of the Art***

Prior academic and commercial studies of MOOCs have established that there is a strong correlation between student dropout, student general learning outcomes, and student's behaviour vis-a-vis course activities such as attempting quizzes, posting in forums, submitting homework, and utilizing course resources such as videos, audio lectures, and downloadable files (Brown et al., 2015). These behaviours can be divided into two main types: pure learning behaviour, which involve student–system interactions (e.g., completing quizzes, watching videos) (Sinha et al., 2014), and social behaviours, which involve student–student interactions (e.g., posting or commenting on messages in a forum) (Pursel et al., 2016).

Many research works on learner performance prediction using activity logs have demonstrated that learner's behaviour during the learning process can efficiently serve as learner performance predictor (Brown et al., 2015). Activity logs and social metrics include various aspects of learner behaviour. Therefore, by combining the features extracted from those two main sources with other features, like features from geographical, academic, and socio-professional background, we may be able to obtain a very comprehensive understanding of the learner behaviour and then improve our ability to predict learner behaviour or learner participation in the learning process.

### ***3.1.4 Our Contribution***

This chapter describes the development and implementation of more accurate behaviour prediction models for learners enrolled in MOOCs, which are based on deep learning algorithms. Since time series data is involved, recurrent neural networks (RNNs) are used. We compare three RNN architectures: simple RNNs, gated recurrent unit RNNs (GRU RNNs), and long short-term memory (LSTM). We find that simple RNNs provide best prediction performance. Finally, we propose a tool to support efficiently the course designers in their process of supporting and guiding the learners in the learning process.

### ***3.1.5 Structure of the Document***

This document is organized as follows. Related prior research is described in Sect. 3.2. The research methodology, including the approach, context, and method, is presented in Sect. 3.3, as well as presentation and analysis of results. Section 3.4 gives the conclusion and the future work or perspectives.

## **3.2 Related Work**

### ***3.2.1 Outline***

This section aims to explore the concept, concepts, as well as the state of the art related to prediction of student success, and for other learning outcomes. The learning analytics (LA) techniques are designed to analyse learning-driven data. LA includes (1) descriptive analysis (what happened?), (2) predictive analysis (what will happen next?), (3) diagnostic analysis (why did it happen?), and (4) prescriptive analysis (what should be done to improve?). In our current study related to learner

behaviour in MOOCs, the focus is done on the predictive analysis. The section starts with an introduction to learning analytics, and then address the importance of student success predictive models in MOOCs, after that we explore the diverse types of inputs that are used by student success predictive models as well as the various data sources providing those inputs/features. The section also explores the features engineering, helping to extract features from data sources, and the section presents the relation between types of model and the outcome predicted, as well as the algorithms used for predictive models and metrics for their evaluation. The section ends with lessons learned from this exploratory exercise.

### ***3.2.2 Overview of Learning Analytics***

The subject matter of this chapter falls under the area of learning analytics (LA) in e-learning on MOOC platforms. LA includes predictive, diagnostic, and prescriptive analysis in addition to descriptive analysis. According to the Society for Learning Analytics Research, “LA is 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” (Siemens & Baker, 2012). LA utilizes raw data extracted from any learning system, including learning management systems (LMSs), open educational resources, online libraries, e-portfolio systems, and student services systems. The analysis of this data highlights relation between variables in log files that are related to the learning process and generates new knowledge about students’ behaviour.

The field of LA is itself a subset of educational data mining (EDM) and consists of four main categories of analysis: (1) descriptive (what happened?), (2) predictive (what will happen next?), (3) diagnostic (why did it happen?), and (4) prescriptive (what should be done to improve?) (Rokach, 2005). Tasks and methods are drawn from the areas of statistics, classification, clustering, visualization, and data mining (Rokach, 2005). Machine learning techniques (including deep learning) are also utilized.

Educational data mining techniques can be divided into two main categories: verification-oriented techniques, which rely on traditional statistical techniques such as hypothesis tests and analysis of variance, and discovery-oriented techniques that are used for prediction and categorization, such as classification, clustering, web mining, and others (Rokach, 2005). Those two categories may employ similar techniques, but for different purposes. For example, in the first case, a logistic regression model might be constructed with the aim of understanding its parameters (e.g., Kizilcec & Halawa, 2015), while in the second case the same modeling technique could be used for a purely predictive goal (e.g., Whitehill et al., 2015). Figure 3.1 below further breaks down the machine learning methods used for classification, which may also be used for other purposes.

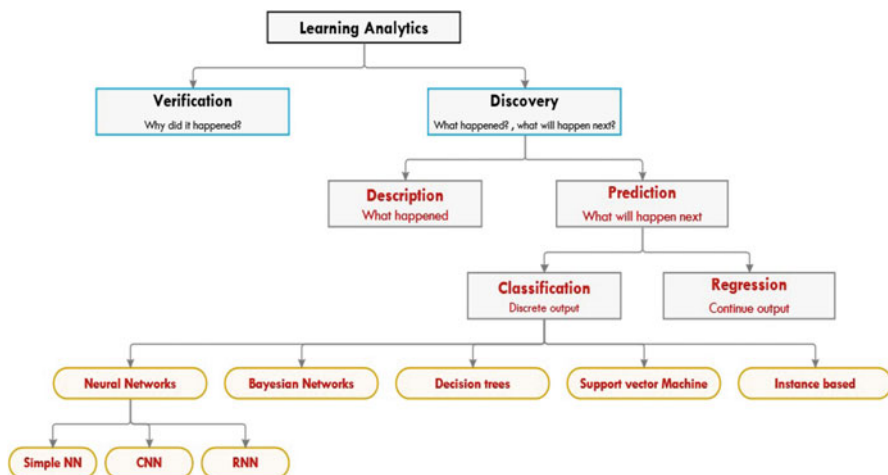


Fig. 3.1 Taxonomy of learning analytics for classification (adapted from Rokach, 2005)

Many research works have used features extracted from pure learning behaviour, as well as features related to social behaviour or interactions. Pure learning activities include visiting pages, watching videos, downloading files, and taking quizzes, while social interaction includes posting in forums, participating in discussions, sending private messages, participating in social networks, and so on. Data related to these activities may be obtained from sources like activity logs, data bases, and external sources to describe or predict outcomes like dropout, completion, success, and certification, using statistics, machine learning and deep learning algorithms. In the area of descriptive learning analytics, Cocea and Weibelzahl (2009) has proposed models to describe learner behaviour based on learning data. Many other projects have also led to the development of plug-ins used in Moodle for descriptive analysis (Mwalumbwe & Mtebe, 2017). Concerning predictive learning analytics, many authors have also produced valuable research works (Kotsiantis et al., 2013). There are also some plug-ins developed for predictive analysis (Mwalumbwe & Mtebe, 2017).

### 3.2.3 Importance of Student Success Predictive Models in MOOCs

In MOOCs, predicting student achievement is beneficial for a wide range of tasks. Many authors agree on three major reasons for constructing predictive models of student achievement, as described in the following subsections. Our model for the classification and prediction of learner behaviour in MOOCs addresses all three of these main purposes.

### ***3.2.4 Personalized Support and Interventions***

Identifying students who are more likely to or fail offers the potential to improve the student experience by enabling focused and personalized interventions to those students who are most likely to need help. This is the stated motivation for many previous works, which frequently refer to these pupils as “at-risk” students (a term adopted from the broader educational literature). Because of the large number of students enrolled in MOOCs compared to the amount of the instructional support personnel, clearly identifying difficult students is critical for delivering focused and timely assistance.

While a teacher in a regular in-person higher education course, or even a moderately sized e-learning course, may be able to directly monitor students, such observation is not possible to support MOOC instructors at scale, and predictive models can help with (a) identifying which students require these resources and (b) intervening by forecasting which resources will best support each at-risk student. Predictive models that can identify these individuals with high confidence and accuracy are necessary, especially when instructor time and resources are limited. Furthermore, many interventions would be superfluous or even harmful to the learning of students who are engaged or otherwise successful. A predictive model must generate accurate and actionable forecasts in order to provide individualized support and actions.

### ***3.2.5 Adaptive Content and Learner Pathways***

Predictive models in MOOCs have the potential to optimize the delivery of course content and experiences for projected student performance. In MOOCs, there has been very little research into adaptivity or true real-time intervention based on student success forecasts in any manner. For example, dropout prediction is used by Whitehill et al. to improve learner response to a post-course survey (this work optimizes for data collection, not learner performance) (Whitehill et al., 2015), and He et al. propose a hypothetical intervention based on projected dropout rates (but only implements the predictive model to support it, not the intervention itself) (He et al., 2015). Kotsiantis et al. offer a predictive model-based support tool for a 354-student distance learning degree programme, which is far smaller than most MOOCs. Pardos et al. implement a real-time adaptive content model in an edX MOOC. However, their approach is geared around increasing time spent on the page rather than improving student learning (Pardos et al., 2017).

To some extent, the paucity of research on adaptive content and learner pathways backed by accurate, actionable models at scale is owing to a lack of consensus on the most effective strategies for developing predictive models in MOOCs, which we address in the current study.

### 3.2.6 *Data Understanding*

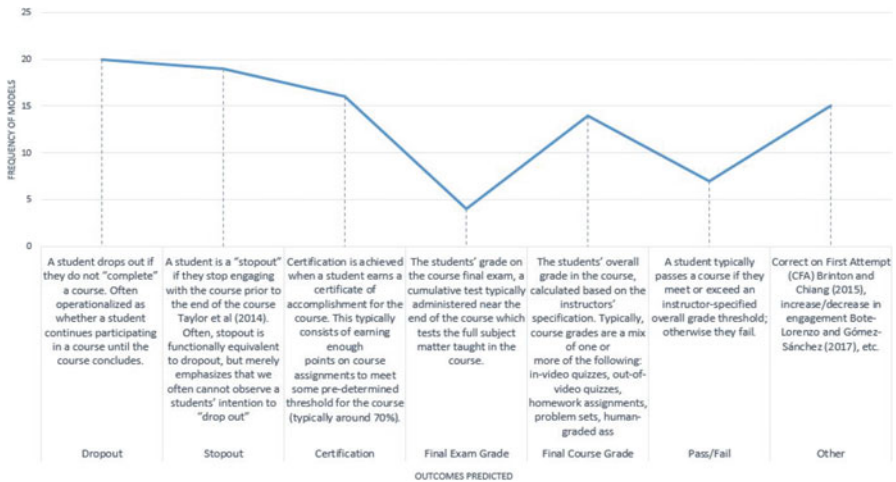
Predictive models can be used as exploratory or explanatory tools, assisting in the understanding of the mechanisms underlying the desired outcome. Predictive models can also be used to detect learner behaviours, learner qualities, and course attributes related with MOOC performance, rather than just offering predictions to enable targeted interventions or adaptive material. These findings can help us enhance the content, pedagogy, and platform, as well as gain a better understanding of the underlying elements that affect student success in these settings. They also make a more direct contribution to theory by offering a better understanding of the complicated interactions between predictors and outcomes derived from the predictive modeling.

Certain types of models are more useful than others from this perspective. Models with simple interpretable parameters (like linear or generalized linear models, which provide interpretable coefficients and  $p$ -values, and decision trees, which generate human-readable decision rules) are far more useful for human understanding of the underlying relationship than models with many complex parameters (such as a multilayer neural network). Unfortunately, the latter are usually (but not always) more effective in making predictions in practice, so interpretability and predictive performance are frequently trade-offs. Some major developments in making increasingly sophisticated models interpretable suggest that this trade-off may be decreased in the future (e.g., Baehrens et al., 2009), but for predictive models in MOOCs, this “fidelity-interpretability trade-off” is still a major concern (Nagrecha et al., 2017).

### 3.2.7 *Common Metrics for Student Success in MOOCs*

Much research in the area of MOOCs has focused on analysing different learning outcomes including dropout, stopout, and certification, as well as measures such as final exam grade or final course grade. These outcomes are taken as measures of student success. Figure 3.1 shows the distribution of measures used in recent studies of success prediction (Gardner & Brooks, 2018).

The issue of evaluating student success in MOOCs is particularly problematic since indicators from traditional educational contexts—such as dropout, achievement, participation, and enrolment—can mean various things or appear illogical in the context of a MOOC. The authors frequently employ different definitions for these concepts, in the context of MOOCs. Furthermore, the abovementioned measures all focus on final outcomes and ignore the level of engagement of the student during the course. A balanced picture of student success should provide a broad collection of measures that assess course completion, engagement, and learning outcomes. Having various different measures to quantify MOOC effectiveness and outcomes allows us to test the robustness of models by potentially checking their



**Fig. 3.2** Trend of predictive models according to outcomes predicted (adapted from Gardner & Brooks, 2018)

ability to predict multiple different outcomes. Furthermore, it allows us to capture metrics such as course completion, certification, and career advancement as shown in Fig. 3.2.

### 3.2.8 Inputs Used by Student Success Predictive Models

Besides the variety of outcomes predicted, there is also a variety of inputs used in MOOC predictive models. Those outcomes include logging status and frequency, attendance status, dropout status, completion, final grade, success, learning outcomes, learner behaviour, and so on. Table 3.1 below categorizes predictive models according to input types, with their associated data sources and usual outputs.

The figure below shows the tendency, in 2017, of predictive models in MOOC by types of inputs used.

### 3.2.9 Activity-Based Models

Activity-based models use inputs related to learner behaviour to evaluate learner behavioural outcomes such as dropout, failure, retention, success, and certification. From Fig. 3.3, it is evident that activity-based models are the most commonly found in the literature. This may be attributed to the fact that many outcomes predictable are more driven by activities. In addition, in MOOC platforms, activity

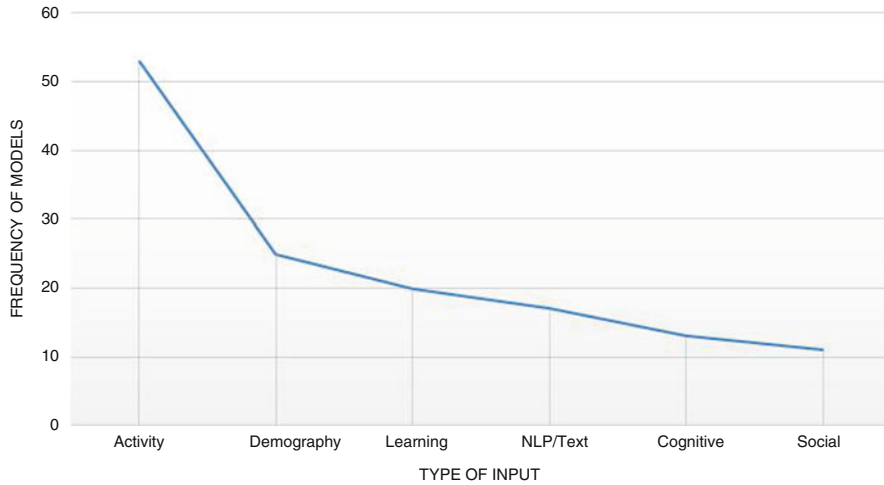


**Table 3.1** Synthesis of predictive model categories from types of features perspectives

Predictive model categories	Type of input/features	Sources	Types of predictions/outcomes
Activity-based models	Use inputs data from learner behaviour to evaluate learner behavioural outcomes (dropout, failure, retention, success, certification)	Clickstream files data bases	Dropout, grade, certification
Demographics-based models	Use learner attributes which remain static (age, sex, academic level, town, marital status, country, town, . . .) over the interval of a course to predict student success	Data bases	Dropout, grade, certification
Learning-based models	Use observed student learning or performance on course assignments or theories of student learning as the basis for predictive modeling	Clickstream files data bases	Pass/fail, final grade, assignment, or exam prediction
Discussion forum and text-based models	Use natural language data generated by learners, as well as linguistic theory as the basis of student models	Clickstream-based activity features and natural language processing features	Completion
Cognitive models	Cognitive models incorporate observed or inferred cognitive states or rely on theories of cognition	Biometric tracking, contemporaneous questionnaires	Dropout, grade, certification
Social models	Social models use observed or inferred social relationships, or theories of social interaction, as the foundation for student models	For a data, external sources	Dropout, grade, certification

data are more abundant and more granular than any other data. Clickstream files, for example, provide detailed and granular interaction-level data on users' engagement or interaction with the platform.

In addition, Brinton and Chiang (2015) established that activity-based features not only predict activity-based outcomes but also appear to provide reasonable predictive performance even in non-activity-based prediction tasks, such as in grade prediction. Activity-based features may include simple counting-based features (like the number of posts in forum, the number of quizzes completed) (e.g., Xing et al., 2016), as well as more complex features such as temporal indicators of increase/decrease of course engagement (date of last connexion to the course



**Fig. 3.3** Trend of predictive models in MOOC according to the types of inputs used (based on data from Gardner & Brooks, 2018)

platform) (Bote-Lorenzo & Gómez-Sánchez, 2017), sequences (Fei & Yeung, 2015), and latent variable models (Qiu et al., 2016). Despite the variety of types of activity-based inputs, all are obtained from the same underlying data source, namely the course clickstream log (which may be reformatted as a relational database consisting of extracted time-stamped clickstream events). Base features are drawn from a relatively small and consistent set of events, including page views, activity views, forum posts and views, and quiz completion: these features reflect the structure of courses available across the dominant MOOC platforms, such as edX and Coursera. Recently, however, due to the gamification of learning, different types of features have also emerged, as shown in Table 3.2 below.

### 3.2.10 Demographics-Based Models

Demographics-based models use learner attributes which remain static over the interval of a course to predict student success. Examples of demographic variables include age, sex, academic level, marital status, and so on. Several works have investigated the relationship between learner demographics and success in MOOCs. For instance, Qiu et al. (2016) examine the impact of both gender and level of education on forum posting, total active time, and certification rate for a sample of XuetangX MOOCs. They found that in non-science courses, females had higher rates of forum activities (posting and replying), more time spent on video and assignment activities and higher certification rates—but in science courses, the reverse was true.

**Table 3.2** Synthesis of the common types of activity-based features

Types of activity-based features	Description	Example of work
Counting-based activity features	Predictive model, built on simple counting-based features extracted entirely from clickstream events	Predictive model, utilizing a support vector machine (Ramos & Yudko, 2008)
Early course activity features	Students' behaviour early in the course might be particularly predictive of their final performance in a MOOC. The fact that "early warning" systems anticipate based on limited amounts of data means that their projections can alter drastically throughout the early stages of a journey.	A simple logistic regression classifier based only on week 1 behaviour which effectively predicts certification in a MOOC offered to university students. Within the first two weeks of an e-learning history course, early access to course resources, such as a textbook and its integrated formative assessments, provides precise predictions of success or failure. Exercise in the first week of a nine-week course is a significant predictor of persistence in the course (Stein & Allione, 2014)
Temporal and sequential activity features	Rely on cumulative percentage of available lecture videos watched, the number of threads viewed on the forum, the number of posts made on the forum, the number of times the course progress page was checked), compiled over each week of the course. A particularly novel aspect of this work is the use of students' checking of their course progress page as an input feature.	Hidden Markov Model (HMM) to predict dropout. Students who never check their course progress have a dropout rate of 20-40% at each week of the course, while students who check their progress four or more times have a dropout rate of less than 5% each week. The use of an LSTM with sequential data is a promising technique, but it has to be replicated across a wider sample of courses (Fei & Yeung, 2015)
Latent activity features/variable	Because of its ability to infer complicated associations between variables in a data-driven manner, latent activity features have been widely used in predictive models of student achievement	To model student performance, probabilistic soft logic (PSL) is used to a set of activity- and natural language-based features. The latent variable assignments from this PSL method are used as predictors in a survival model (Kizilecc & Halawa, 2015)

(continued)

**Table 3.2** (continued)

Types of activity-based features	Description	Example of work
Course metadata features	There are few researches into elements of courses (course metadata) that may be important to student activity inside the courses	Early engagement (registering earlier or completing a pre-course survey) is the strongest predictor of completion. Individual lecture titles are linked to different levels of engagement. Course subject interacts with learner demographics (i.e., gender) in predictive models (Qiu et al., 2016)
Higher order activity-based features	Other studies in MOOC research have starting to emerge, leveraging more complicated feature types. Explorations of higher order n-gram representations of learner activity data, which have shown promise predictive performance, are part of this	In activity-based n-gram models, counts of unique sequences of events or behaviours are used to create features, which are then used to build supervised learning models. Sequences of behaviour, irrespective of the time gaps between them, contain richer information than individual events or counts of these events without considering the context of other neighbouring events in time (Brooks et al., 2015)

Other works established links between demographic features like age, prior education, and prior experience with MOOCs and both dropout and achievement. Greene et al. (2015) identified those three features as significant predictors of both dropout and achievement. Khalil and Ebner (2014) identified several features that explained a large proportion of dropout in MOOC, including lack of time, lack of motivation, and “hidden costs” (textbooks needed for reference or paid certificates not clearly mentioned at the beginning).

On the other hand, Brooks et al. (2015) found that adding demographics provided only minimal improvement over the performance of activity-based predictive models for academic achievement of learners enrolled in MOOCs. Brooks et al. (2015) demonstrate that demographics-based models underperform activity-based models in MOOCs, even during early stages of the course when activity data is minimal. Additionally, demographic features provide no discernible improvement when added to activity-only models—on the contrary demographic features tend to degrade the performance of activity-only models in the second half of the course, as activity data accumulates (Brooks et al., 2015).

### ***3.2.11 Learning-Based Models***

Learning-based models use observed student performance on learning tasks (including course assignments) or theories of student learning as the basis for predictive modeling. Learning is obviously the basic objective of any MOOC—however, learning-based features are limited and are only used to predict a limited set of outcomes, such as pass/fail, final grade, assignment, or exam prediction. Bayesian Knowledge Tracing (BKT) (Mao, 2018; Pardos et al., 2013) has been widely used in intelligent tutoring systems to predict homework scores. However, Ren et al. (2016) found that “personalized linear regression” for predicting student quiz and homework grade outperformed an item-level variant of BKT (IDEM-KT) across two MOOCs.

Garman et al. (2010) applied pre-existing learning assessment to online courses by administering a commonly used reading comprehension test (the Cloze test) to students in an e-learning course. He found that reading comprehension is positively associated with exam performance and overall course grade but found no association between reading comprehension and open-book quizzes or projects. Kennedy et al. (2015) evaluated how prior knowledge and prior problem solving abilities predict student performance in a discrete optimization MOOC with relatively high prior knowledge requirements, drawing on robust learning theory results from in-person courses. The prior knowledge variables alone account for 83% of the variance in students’ performance in this MOOC. The relationship between prior knowledge and student performance is well documented in traditional education research but is largely unexplored in MOOCs, despite the potential presence of many more students who lack prerequisite prior knowledge in MOOCs relative to traditional higher education courses.

Other works focusing on time-on-task and task engagement are also student performance concepts which have been applied extensively to educational contexts outside of MOOCs. Champaign et al. (2014) evaluate how learner time dedicated to various tasks within the MOOC platform (assignment problems, assessments, e-text, check point questions) correlates with their learning gain and skill improvement in two engineering MOOCs. They find negative correlations between time spent on a variety of instructional resources and both skill level and skill increase (i.e., improvement in students' individual rate of learning), using assessments calibrated according to Item Response Theory.

On the other hand, DeBoer and Breslow (2014) find that time spent on homework and labs in a Circuits and Electronics MOOC on edX predict higher achievement on assignments, while time spent on the discussion board or book is less predictive or not statistically significant. Additionally, time on the ungraded in-video quiz problems between lecture videos is found to be more predictive of achievement than time on lecture videos themselves.

Moreover, peer learning and peer assessment are also important theoretical concepts in education but have seen only limited applications in MOOCs to date. Ashenafi et al. (2016) examine models for student grade prediction which only use peer evaluation. They apply these models to traditional courses with web-based components but argue that their findings are also applicable to MOOC contexts. Peer assessment is used extensively in MOOCs, and its predictive capacity remains largely unexplored (Jordan, 2015).

### ***3.2.12 Discussion Forum and Text-Based Models***

Discussion forums are an embedded feature in every major MOOC platform and are widely used in most courses. A detailed analysis of the data from discussion forums provides the opportunity to study many dimensions of learner experience and engagement, which could not be identified elsewhere. This includes a rich set of linguistic features (derived from the analysis of the textual content of forum posts), social features (measured by the networks of posts and responses or actions such as likes and dislikes), and some behavioural features not available purely from the evaluation of clickstream data. Typically, discussion forum and text-based models use natural language processing (NLP) applied to data generated by learners, as well as linguistic theory as the basis of student models.

Crossley et al. (2016) compared the predictiveness of clickstream-based activity features and natural language processing features. They found that clickstream-based activity features are the strongest predictors of completion but discovered that NLP features were also predictive. Their work established that the addition of clickstream-based activity features improves the performance over a linguistic-only model by about 10%.

Other works (Tucker et al., 2014) found a moderate negative association between students' mood in posts regarding individual tasks and their performance on those

assignments in an art MOOC. They also discover a little upward trend in forum post mood throughout the course of the study. Other predictive work in the area of sentiment analysis shows a link between sentiment and attrition that appears to be different depending on the course content (Wen et al., 2014).

### 3.2.13 *Cognitive Models*

As the foundation for student models, cognitive models incorporate observed or inferred cognitive states or rely on theories of cognition. Despite the fact that MOOCs are ultimately concerned with influencing learners' cognitive states (learning is a cognitive activity), there has been surprisingly little research on the use of cognitive data in MOOCs. This could be due to the particular difficulties of obtaining this data, especially when compared to other rich data sources (activity, forum postings, etc.). Novel data collection approaches, ranging from biometric tracking (e.g., Xiao et al., 2015) to contemporaneous questionnaires, are used in much of the research on cognitive states in MOOCs (Dillon et al., 2016).

Other authors like Wang et al. (2015) look at behaviours related to higher order thinking as displayed in student discourse and examine their relationship to learning using data from discussion forums. Several learning outcomes are evaluated using hand-coded data and a learning activity classification framework based on cognitive science research. The authors discovered that students who used "active" and "constructive" behaviours in the discussion forum, behaviours that demonstrate higher level cognitive tasks like synthesis rather than simply paraphrasing or defining, produced significantly more learning gains than students who did not use these behaviours. Furthermore, they established that useful cognitive data relevant to student performance may be retrieved from discussion forum posts and applied to simple models using techniques such as bag of words and linear regression. Furthermore, cognitive strategies if they can be discovered properly appear to be associated with student performance in MOOCs and that cognitive theory can be used to inform MOOC prediction models.

Novel data collection approaches are used in much of the work in this subject. Future research should move increasingly beyond questionnaires and self-reports as the sole source of cognitive data from learners. The type of data required for this type of research should become more available for researchers as sensing technology gets more affordable and consumers' devices (such as smartphones and tablets) become increasingly outfitted with sensors. Many canonical cognitive findings in educational research have yet to be investigated or replicated in a MOOC environment, and further study is needed to establish the limitations of these findings when applied to MOOCs.

### 3.2.14 *Social Models*

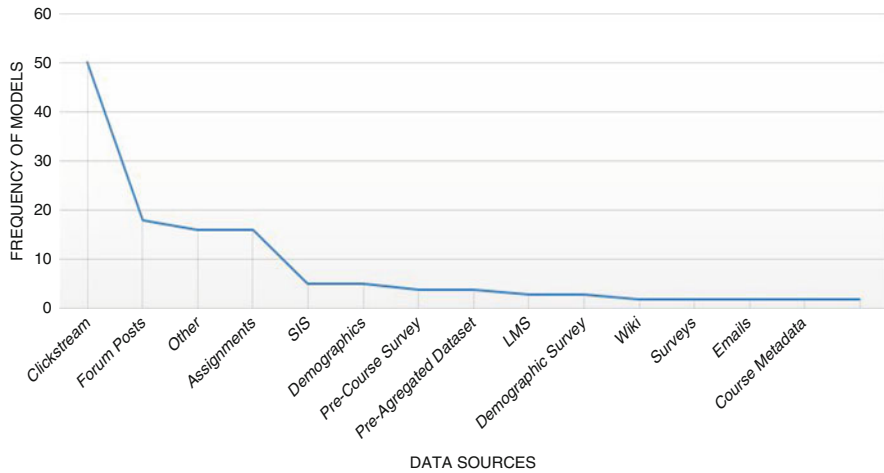
Social models of learning are built on the foundation of observed or assumed social relationships or ideas of social interaction.

Many studies employ discussion forums to build social networks in which students serve as nodes and varied response relationships serve as edges. Joksimović et al. (2016), for example, use two sessions of a programming MOOC, one in English and one in Spanish, to assess the association between social network ties and performance (specifically, non-completion vs. completion vs. completion with distinction). Students who received a certificate or distinction were more likely to interact with one another than non-completers. Jiang et al. (2014) discovered that learners in different performance groups tend to communicate with one another in different types of MOOCs. Joksimović et al. (2016) found that weighted degree centrality was a statistically significant predictor of completion with distinction in the both the English and Spanish courses mentioned above, as well as a significant predictor of basic completion in the Spanish language course. On the other hand, closeness and betweenness centrality had more variable and inconsistent effects across courses. They get to the conclusion that structural centrality in the network is related to course completion (Joksimović et al., 2016). The findings are similar to those of Russo and Koesten (2005), who found centrality to be a statistically significant predictor of student achievement in a short online course. The findings are similar to those of Russo and Koesten (2005), who found centrality to be a statistically significant predictor of student achievement in a short online course.

In a related study, Dowell et al. (2015) examine how text discourse features can predict social centrality and that discourse features explain about 10% of the variance in performance (compared to 92% with a model using discourse + participant features). The explained variance increased to 23% for the most active participants in the forums.

There is a need for more research into the impact of social networks in MOOCs, as well as more exploration of external social network data. Although social networks appear to play a major role in students' learning, they are difficult to quantify using existing MOOC data, especially with small single-course samples. Despite the richness of these data sources, MOOCs rarely incorporate external digital social networks (such as data from Facebook or LinkedIn). Existing research, on the other hand, appears to be unduly reliant on discussion forums as sources of social network data. The evaluation of new data sources on social issues has the potential to have a significant impact on the scholarly consensus in this field.





**Fig. 3.4** Trend of predictive models in MOOC according to the data sources providing inputs used (based on data from Gardner & Brooks, 2018)

### 3.2.15 *Data Sources Providing Inputs/Features for Predictive Models*

The data sources employed in MOOC predictive modeling research have received little attention. Knowing which data sources are valuable for prediction and which are unexplored is a good starting point for future research. Recognizing which data sources are most valuable can also increase the efficiency of predictive modeling work in practice because feature extraction is costly in terms of both development and computing time.

The figure below displays the common data sources used in predictive models designing in MOOCs. Moreover, this figure confirms that clickstreams are the most common raw data source for predictive modeling research in MOOCs, out of the raw data sources outlined in Fig. 3.4 above.

In some ways, the much used of clickstream data is unsurprising: clickstreams give rich granular data that the field is only just beginning to understand how to capture in its entirety. Clickstreams, on the other hand, are unstructured text files that need a lot of human and computational work to parse. Due to faults in platform server logging, their forms are complex and sometimes inconsistent, and a single item can have multiple levels of aggregate applied to it. The various data types in the figure above are usually given as organized relational databases that may be accessed using simple SQL commands. The fact that clickstreams are so frequently used, despite the difficulties in acquiring and using this data, demonstrates their utility in predictive modeling. The authors (Gardner & Brooks, 2018) compare the predictiveness of clickstream features vs. forum- and assignment-based features when predicting dropout across the entire population of learners in a large state

university in the USA. This work verifies that clickstream features are more effective predictors than forum- or assignment-based features when predicting dropout across the entire population of learners.

While clickstreams contain complicated, potentially relevant temporal information regarding learner behaviour across time, most modeling of these temporal patterns has been limited to simple counting-based representations (with few exceptions; i.e., Fei & Yeung, 2015; Brooks et al., 2015). Much of the complexity seen in these contact logs is unlikely to have been caught using current study methodologies.

### ***3.2.16 Features Engineering in Predictive Models***

Boyer and Veeramachaneni (2015) show how good feature engineering may be paired with effective statistical models to produce performant student success predictors in a series of papers. Several unique approaches to developing activity-based models of student success in MOOCs are demonstrated in these studies, which combine crowd-sourced feature extraction, automatic model tweaking, and transfer learning.

Boyer and Veeramachaneni (2015) employ crowd-sourced feature extraction to create behavioural features for stopout prediction, using members of a MOOC to apply their human skills and domain knowledge. For all four cohorts studied, the authors find that these crowd-proposed characteristics are more complex and have greater predictive performance than simpler author-proposed features (passive collaborator, wiki contributor, forum contributor, and fully collaborative). The predictive model in this study is based on a basic regularized logistic regression, revealing that many good predictive models of student achievement in MOOCs have depended on creative feature engineering rather than complicated algorithms.

Boyer and Veeramachaneni conclude that a posteriori models, which are built retrospectively using labeled data from the target course, provide an “optimistic estimate” and “struggle to achieve the same performance when transferred.” The same researchers discovered that an in situ prediction architecture transfers well, with performance comparable to a model that takes into account a user’s whole history (which is not actually possible to obtain during an in-progress course).

The two sessions above helped to explore the data sources commonly used and to address the issue of extracting features from those sources. It appears that clickstreams are the most common raw data source used for predictive modeling research in MOOCs. In the session above, we saw that many types of activity-based features are embedded in clickstreams. Activity-based features are the basics of activity-based models that are the most commonly used predictive models. From the explanations provided above, it appears that the model we are developing for the classification and prediction of learner behaviours is an activity-based model.

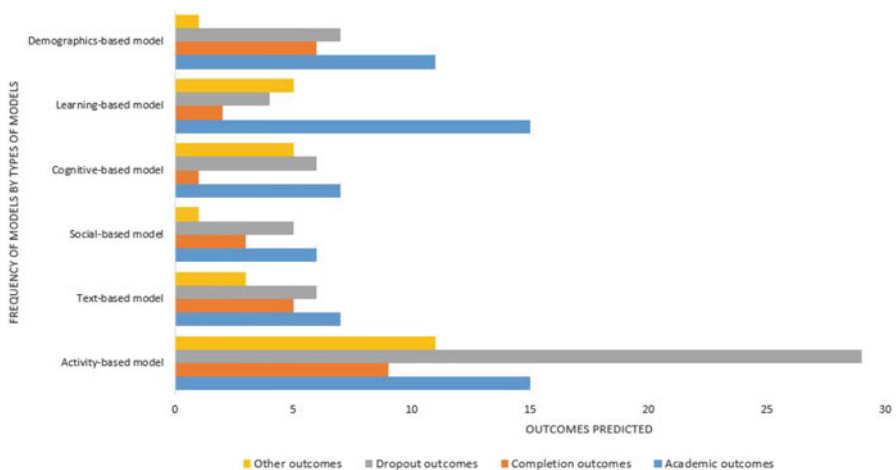
### 3.2.17 *Relation Between Types of Model and the Outcome Predicted*

Above we presented various types of models, ranging from activity-based to social model. We also presented the most common types of outcomes that have been predicted in MOOCs, and they include dropout, completion, and others.

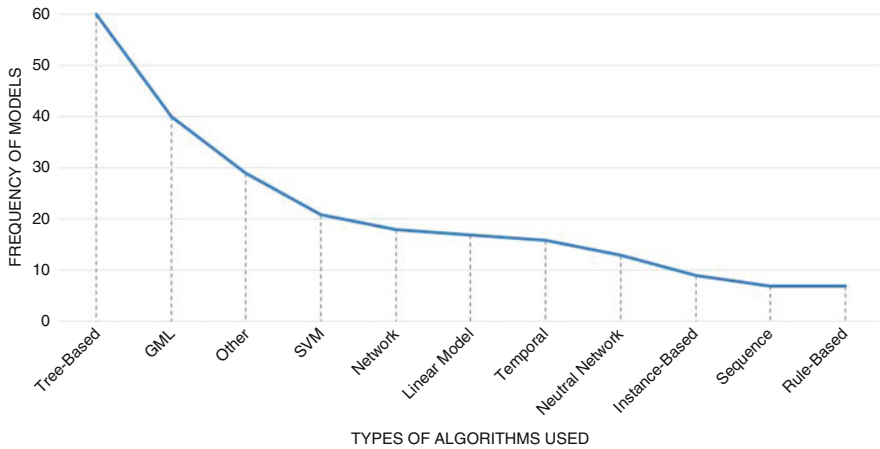
The figure below is adapted (Gardner & Brooks, 2018), and the results presented were established by experiments that included a predictive model that could be classified as many categories or predicted numerous outcomes were included in each category in this table, resulting in cell totals that surpass the total number of works assessed. Pass/fail, final grade, assignment grade, and exam grade are examples of “academic” outcomes.

All measures of course completion, such as certification and participation in the final course module, are included in the term “complete.” They established that academic outcomes are the most predicted, while completion and other types of outcomes are the least predicted. On the other hand, while activity-based model is the most common, cognitive-based models have not been well explored.

The figure below groups several outcomes into a single “academic” outcome category. “Pass/Fail” indicates whether a learner met a predetermined final grade threshold to pass the course, and “Certification/Completion” indicates whether a student successfully completed all course requirements and received an official certificate of completion. In addition, those official certificate sometimes requires payment and identity verification; Fig. 3.5 shows the various types of models according to the outcomes predicted.



**Fig. 3.5** Trend of types of predictive models (inputs used) in MOOC according to the outcomes predicted (adapted from Gardner & Brooks, 2018)



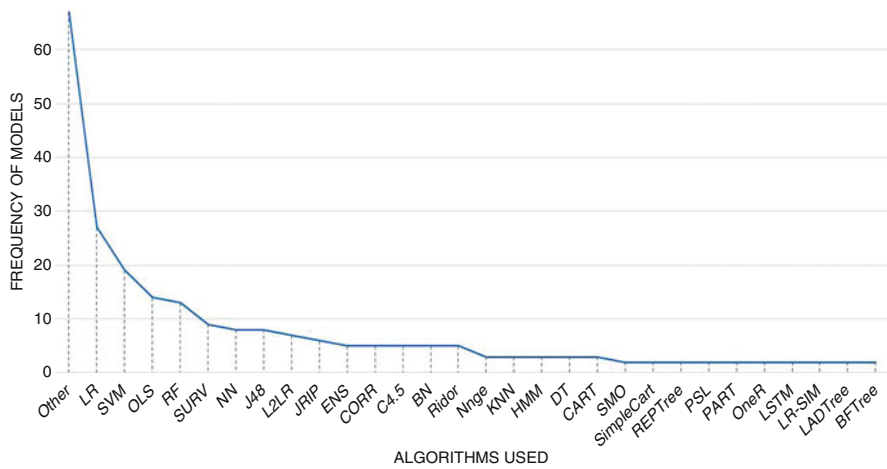
**Fig. 3.6** Trend of predictive models in MOOC according to the types of algorithms used (adapted from Gardner & Brooks, 2018)

### 3.2.18 Algorithms for Predictive Models and Metrics for Their Evaluation

Once the outcomes to be predicted are identified, as well as the type of model, the features and the sources to extract the features from, the remaining tasks include selecting relevant algorithms to build/train the model and the suitable metrics for its evaluation. According to outcomes predicted and the model types, a number of algorithms have been commonly used to build models, and various metrics have been used to evaluate those models.

### 3.2.19 Algorithms for Predictive Models

Predictive student modeling in MOOCs relies heavily on statistical models to translate features to predictions. Figure 3.6 represents the frequencies of different classes of statistical algorithms used to develop MOOC predictive models (Gardner & Brooks, 2018). The figure shows that tree-based models and generalized linear models are the two most common types. The popularity of tree-based algorithms can be attributed to several advantages: tree-based models can handle a variety of data types (categorical, binary, and continuous), they are less prone to multicollinearity than linear models, they are nonparametric and make few assumptions about the underlying data, and their outputs may be interpreted through visualization, inspection of decision rules, variable importance metrics, and other methods. On the other hand, GLMs are quick and easy to fit to data, requiring little or no hyperparameter tuning. Unlike tree-based models, they produce regression coefficients that can be



**Fig. 3.7** Trend of predictive models in MOOCs according to the specific algorithms used (adapted from Gardner & Brooks, 2018)

directly interpreted as expressing the relative contribution to overall accuracy of the different predictors.

Figure 3.7 further disaggregates Fig. 3.6 by showing the specific algorithms utilized. The figure shows that how the prevalence of tree-based algorithms obscures the lack of uniformity in the algorithms utilized. It appears that, of all the tree-based algorithms studied, only random forests were used in more than ten of the works. As a result, evaluating the effectiveness of any particular tree method across their survey is challenging. In contrast, there are few GLM algorithms used in the literature; practically, all GLM algorithms are logistic regression (LR) and L2-penalized logistic regression (“ridge” regression, L2LR). Despite their high parametric assumptions about the underlying data, GLMs, and L2LR in particular, often achieve outstanding performance when applied with large and robust feature sets.

Finally, Fig. 3.7 shows a “long tail” of modeling methodologies, with about half of the work using customized, individualized algorithms, indicated by “Other” in Figs. 3.6 and 3.7. This reflects a focus on innovation in academic research, as well as a new area with limited consensus on the optimal strategy to solve prediction problems. We notice that none of the methods in the assessed work consistently outperforms all other algorithms, implying that there is no one “best” algorithm for a given job or dataset (Wolpert & Macready, 1997). At this juncture, future work comparing and evaluating the fitness of various predictive modeling algorithms for various objectives in MOOC research would be suitable. It also appears that supervised learning approaches dominate the literature, with few examples of unsupervised approaches; this is likely due to the fact that many of the outcomes (i.e., dropout, certification, pass/fail, grades) are observable for all learners, making unsupervised techniques unnecessary for many of the prediction tasks addressed by research to date.

### ***3.2.20 Metrics for Model Evaluation***

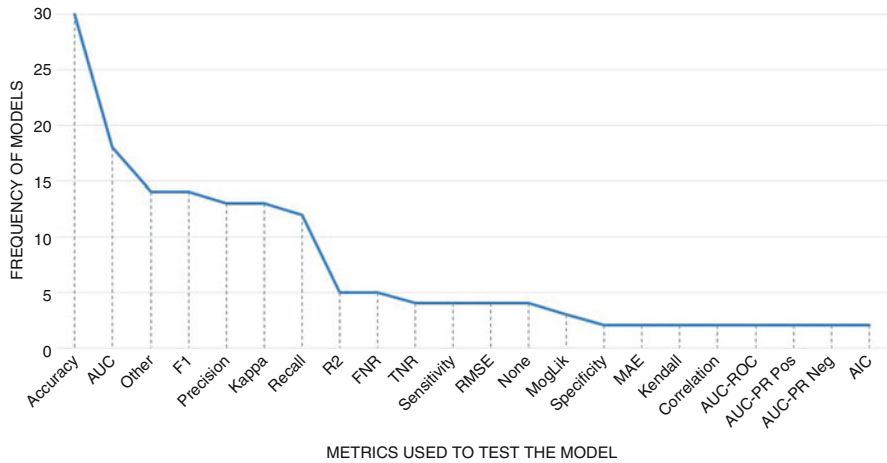
The figure below shows the distribution of evaluation measures commonly used as highlighted in Gardner and Brooks (2018). There is a general consensus on a set of evaluation metrics, including accuracy (ACC), area under the Receiver Operating Characteristic curve (AUC), precision (also known as positive predictive value) (PREC), recall (REC) (also called true positive rate, sensitivity, or probability of detection), F1, and kappa. Different measures evaluate different aspects of predictive quality, which change based on the task and study goals. However, sometimes readers are unable to compare performance across otherwise-similar studies that report different performance metrics due to the lack of a common baseline. Reporting multiple measures would frequently provide a fuller view of model performance and make cross-study comparisons easier, while still allowing researchers to look at performance using their preferred metric(s). Open data or open replication frameworks would enable more detailed comparisons and shift the burden of proof from the researcher to reviewers and critical readers, who would be able to evaluate results using any performance indicator of interest.

Classification accuracy is reported as the only model performance metric in ten of the studies surveyed (more than 10%). But although classification accuracy is easily interpretable, but it can be a misleading measure of prediction quality when outcome classes are of unequal size, as is the usual case in MOOCs (i.e., most students dropout, do not certify, etc.). The same data that is used to compute accuracy can also be used to obtain more informative performance metrics, such as sensitivity, specificity, F1, Fleiss' kappa, and so on. Other measures, such as the AUC, assess performance for all potential thresholds, thus taking into account the fact that performance is threshold dependent, as shown in Fig. 3.8 below.

The best model evaluation metric is typically determined by the outcome being measured as well as the specific objectives of a predictive modeling project. For example, recall may be an appropriate model evaluation metric in a dropout modeling experiment where the goal is to provide an inexpensive and simple intervention to learners (such as a reminder or encouragement); however, precision may be a better choice when the goal is to provide an expensive or resource-intensive intervention to predicted dropouts.

#### ***3.2.21 Lessons Learned from Related Work***

The accuracy dimension of predictive student models described in the figure above is reflected by the data source, feature extraction method, statistical modeling algorithm, and assessment metric taken together. Research into and methodological development in each of these domains (feature extraction, modeling, and evaluation) stands to significantly increase the accuracy of future predictive MOOC models.



**Fig. 3.8** Trend of predictive models in MOOC according to the evaluation metrics used (adapted from Gardner & Brooks, 2018)

We have observed the various aspects of the related works and the key observations made are the following: (1) input/features/predictors have been focusing most on exploring activity-based data through click stream logs. (2) Outcomes predicted or the prediction have been focusing on (a) dropout, (b) completion, (c) success, and (d) certification, not providing the actor with a good understanding of what is going on in the MOOC platform which remain a “Black Box,” Course instructor is not provider with a tool clear understanding of the learner behaviour in order to support him in the learning process. (3) Algorithm models have been focusing most on the descriptive models; in addition, predictive models used so far have not explore enough the power of deep learning for better prediction.

Given the previous observations, we conclude that the following actions would provide actors with tools for better management of the learning process in MOOCs: (1) exploring more features, (2) exploiting the power of deep learning, and (3) classifying and predicting learner behaviour.

### 3.2.22 Approach

Traditional educational researchers and practitioners used methods such as (1) surveys, (2) interviews, and (3) observations, those methods are (1) time consuming, (2) costly, and (3) do not provide the course instructor with timely and useful information to understand and manage the teaching and learning process, and in addition, a delay always occurs between data collection and prediction. So educational researchers progressively switched to learning analytics (LA) for real-time analysis of data generated by the learning process.

LA includes both descriptive and predictive analyses. As described above, our goal is to construct a predictive model that takes into account a variety of features of a learner life cycle (activity-based features, demographics-based features, learning-based features, Discussion Forum and Text-Based Features, Cognitive Features, Social Features) in a MOOC platform, for the purpose of enabling timely, targeted, and personalized intervention to promote retention and successful course and programme completion, by classifying and predicting the learning behaviour in the MOOC. For this purpose, it is necessary to take into account the characteristics of data to be analysed such as volume, variability, velocity, veracity, as well as the highly imbalanced nature of dropping out over retention or failure over success. The large volume of data available, as well as the complicated interaction of factors involved, indicates that deep learning algorithms can potentially provide far better prediction performance and accuracy than traditional algorithms (Xing & Du, 2019). The experimentation will start by activity-based features and then will progressively consider other types of features.

Our overall methodology consists of (1) exploring the potential of deep learning techniques to provide learner behaviour predictive model which can potentially outperform the traditional-used Machine Learning approaches, (2) analysing personalized interventions using individual's learner behaviour prediction, and (3) checking whether deep learning models can better personalized and prioritised interventions to support learner in the learning process than other algorithms. In the current chapter, we deal exclusively with point (1). In order to achieve predictions that are actionable in a real-world context, and given time series types of data generated by learner behaviour, the training technique is L2 regularization on models using RNN architectures (Che et al., 2018).

While training our model, we initially play on (1) the number of epochs, (2) the type of RNN, (3) the number of hidden layers, (4) the learning rate, and (5) the regularization parameter. Accuracy is the most suitable model evaluation metric to work, in view of the binary nature of our outputs and the dataset. So, we choose accuracy as the metric to evaluate the model. During the training, we keep constant the parameters related to the Adams optimizer, as shown in Fig. 3.9 below.

### **3.2.23 Context/Dataset**

The context for this work is a gender-sensitive STEM Education course deployed on Moodle. The course took place in September 2018 and lasted for 6 weeks. The course consisted of 6 modules and 3617 students registered. The course consisted of involved many activities and resources, including forums, quizzes, assignments, videos, audios, wiki, downloadable files, lecture content pages, announcement, calendar, and gradebook, whereas Moodle has many data sources, for this work our dataset will be extracted from clickstream logs, which is a file provided directly by Moodle containing historical information about the learners' and instructors' interactions in the platform, such as pages visited including when and how many



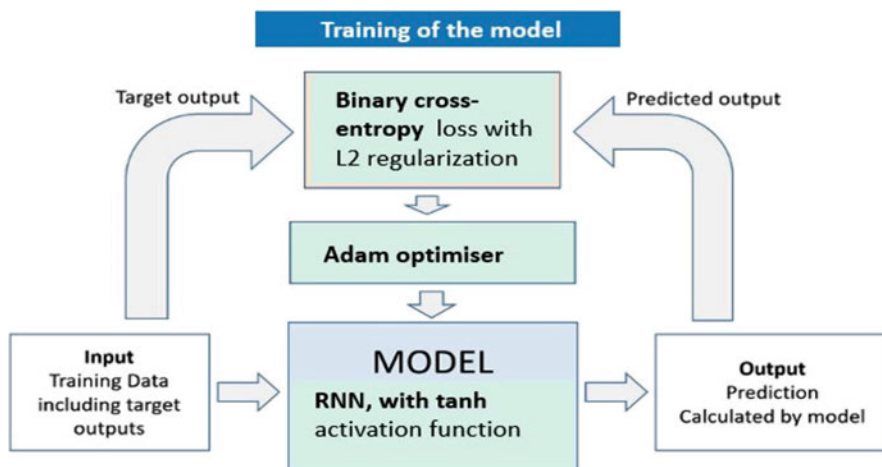


Fig. 3.9 Architecture of the training of the model

students and teachers interacted with the course content. The dataset also includes demographic data. Those sources are enough for the current preliminary study, because we do not intend at this level to go deep in the contents of learners' interactions.

### 3.2.24 Method

#### 3.2.24.1 Features Generation

In order to build our predictive model, several activity-based features related to learners' interaction with the learning process or course content (activities and resources) were generated. The features and their descriptions are given in Table 3.1. These are all activity-based features and were chosen according to the MOOC prediction literature (Rosé et al., 2014) and our previous work (Sinha et al., 2014). Since feature engineering is not a principal objective of this work, we used a flat feature structure containing clickstream, directly provided by the log file. Our work adopts a classical 80/20 train/test split because we did not have a very large amount of data, as shown in Figs. 3.10 and 3.11 below.

### 3.2.25 Building the Model

Our model works on time series data (learner behaviour in the learning process), so RNN (Recurrent Neural Network) is a suitable deep learning algorithm to use

Feature	Description
TTL_DISC_VIEW	Total views on discussion forums
TTL_DISCPOSTCOUNT	Total number of discussion posts
TTL_VIDEO_VIEW	Total views on review-video pages
TTL_QUIZATTEMPT	Total number of quiz attempts
TTL_QUIZTIMESPENT	Total time spent on quizzes
TTL_QUIZSCORE	Total quiz score
TTL_INTRO_VIEW	Total views on introduction pages
TTL_LECCONT_VIEW	Total views on lecture pages
TTL_ASSIGN_VIEW	Total views on assignment pages
TTL_ASSIGN_SBM	Total number of assignment submissions
TTL_ANNOUN_VIEW	Total view on announcement pages
TTL_WIKI_VIEW	Total view on wiki pages
TTL_GRADEBOOK_VIEW	Total view on gradebook
TTL_CALENDAR_VIEW	Total view on calendar
TTL_AUDIO_VIEW	Total view on audios
TTL_FILE_VIEW	Total view on files
TTL_ACTIVITY	Total number of activity taken by the learner
LAST_ACTIVITY	Last activity taken by the learner
TTL_DAY_LAST_ACTIVITY	Total number of days since the learner took the last activity
....	....

Fig. 3.10 Overview of features generated to serve as inputs for the predictive mode

Page: 1 2 3 4 5 6 (Next)

Time	IP address	User full name	Action	Information
Wed 8 January 2020, 6:50 PM	196.207.229.154	José Meku	forum view forum	Forum Introductif / présentation
Wed 8 January 2020, 6:50 PM	196.207.229.154	José Meku	forum view forum	Forum Introductif / présentation
Wed 8 January 2020, 6:50 PM	196.207.229.154	José Meku	forum add post	Re: Forum Introductif / présentation
Wed 8 January 2020, 6:49 PM	196.207.229.154	José Meku	forum view forum	Forum Introductif / présentation
Wed 3 July 2019, 2:45 PM	196.183.38.195	Célestine BACON	forum view forum	Forum Introductif / présentation
Mon 17 December 2018, 3:51 PM	41.203.147.4	Idrissa MOUSSA	forum view forum	Forum Introductif / présentation
Sun 9 December 2018, 9:48 PM	154.68.5.144	Akoua Desire KOUAME	forum view forum	Forum Introductif / présentation
Sun 2 December 2018, 10:47 PM	41.219.50.234	Alimatou Absa DIARRA	forum view forum	Forum Introductif / présentation

Fig. 3.11 Example of clickstream data extracted from the Moodle platform/clickstream data having 8844 records

(Che et al., 2018). The most common RNN architectures used in predictive models are simple RNNs, GRUs, and LSTMs. All the RNN models work with sequenced data and feed information about previous states or time steps into each next state. Practically, LSTMs require the most memory, followed by GRUs. Simple RNNs have the smallest memory capacity (Anani & Samarabandu, 2018).

This chapter implements all the three aforementioned architectures. The model consists of a recurrent layer with 200 hidden layers coupled with a tanh activation function, and then the output is given to a layer, having a sigmoid activation function to produce the probability that the learner will take the next activity in the learning process. This probability will be used to classify and predict their behaviour. The model is built in such a flexible way to allow easy switching from one RNN architecture to another. This enables us to compare the models built using the three types of RNNs. The model also makes it possible to adapt the dimension of the RNNs to suit the input dataset.

In our experimentation, some learner's activities or interactions with the learning environment have been identified, and separate feature values are defined for each of those activities.

Learner's activities are related to the various tasks of a learner in an online learning environment. Those tasks include and are not limited to: (1) Log into the platform, (2) Using forums, (3) Using quiz. If we consider learners activities associated with the task (1) Log into the platform, then the features we can use, concerning learners, are "never log in", "log in on time". So for this activity, the model classifies and predicts learners who could log into the platform or not, as well as those who could log into the platform on time or not. If we consider learners activities associated with the task (2) Using forums, then the features we can use, concerning learners, are "view forum messages", "post messages in forum". So for this activity, the model classifies and predicts learners who could view forum messages or not, as well as those who could post messages in forum or not. If we consider learners activities associated with the task (3) Using quiz, then the features we can use, concerning learners, are "viewing quizzes", "completing quizzes", "validating quizzes". So for this activity, the model classifies and predicts learners who could view quizzes or not, those who could complete quizzes or not, as well as those who could validate quizzes or not . . .

- Training algorithm: ADAM optimizer
- ADAM uses the running averages of the previous gradients to adjust parameter updates during the training phase
- For the cost function  $J$ , represented below, we use binary cross-entropy loss with L2 regularization (helps model to fit the training data well and generalize better)

This cost function is constructed based on binary cross-entropy loss coupled to L2 regularization. Our model is performing a binary classification (learner takes the next activity or not) with estimated probabilities, compared to currently used training techniques like cross-validation, in situ or gradient clipping, so L2 regularization is appropriated because it helps smoothing oscillations in the training loss. The Adam optimization algorithm was used to train the model; Adam is an optimizing algorithm that uses the running averages of the previous gradients to adjust the parameters of the model during the training phase, and we play on (1) the number of epochs, (2) the type of RNNs, (3) the number of hidden layers, (4) the learning rate, and (5) the L2 regularization parameter. During the training, we keep constant the parameters related to the Adams optimizer. In addition, the learning

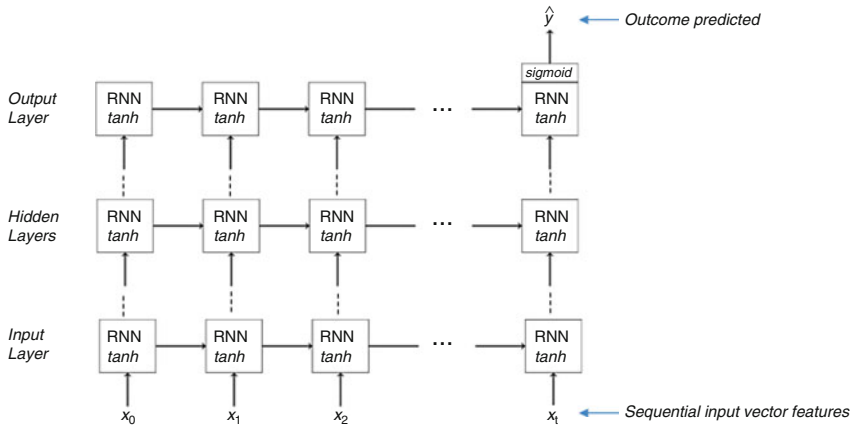


Fig. 3.12 Structure of the model with hidden layers

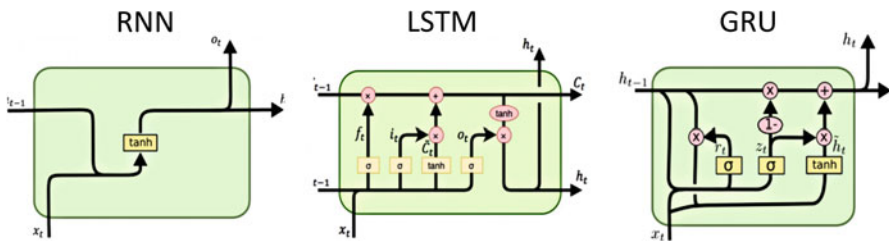


Fig. 3.13 Structure of any single type of RNN tested (simple RNN, LSTM, and GRUs)

$$J = -\frac{1}{m} \sum_{i=1}^m (y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})) + \frac{\lambda}{2m} \sum_{l=1}^L \|w^{[l]}\|_F^2,$$

- ✓  $\lambda$  = L2 regularization parameter
- ✓  $m$  = number of inputs
- ✓  $W^{[l]}$  = the weight matrix for the  $l$ th layer
- ✓  $L$  = number of layers

Fig. 3.14 Cost function: binary cross-entropy loss coupled to L2 regularization

rate used by the model is automatically reduced during the training, if the validation loss did not decrease after a few numbers of epochs, as shown in Figs. 3.12, 3.13 and 3.14 above.

**Table 3.3** Comparison of model accuracy for different types of the RNNs used

Inputs type	Simple RNN	GRU	LSTM
All features	0.8920	0.8801	0.8870

### 3.3 Experimental Results and Discussion

We experimented the model with many sets of hyper parameters, by try and error, with the three RNN architectures (simple RNNs, GRUs, and LSTMs) before concluding that simple RNNs perform best, as shown in Table 3.3, using a regularization parameter of  $\lambda = 0.01$ , which produced the best accuracy of 89.2% for simple RNNs.

With any of the three architectures, 200 hidden layers appear to be offering the best balance in terms of speed and accuracy compared to models with 64 or 256 hidden layers. In our work, we found that simple RNNs produced the best accuracy for the model on the dataset used. This was not expected as experience and previous works tend to predict that LSTMs should perform better than other types of RNNs. One potential explanation of this situation is that having long memory seems not to be so important in this dataset as expected. In fact, some previous works suggest that there might be a transition point in the learning process where learner behaviour far likely changes. Furthermore, past studies on learner behaviour suggest that learner's activities are high at the beginning of the learning process, then decrease during the process, and then slightly increase at the end of the process. The previous explanations imply that there is not a real need of advanced memory capacities for the required model, and then LSTMs may not necessarily be the best option.

### 3.4 Conclusion and Future Work

Our main objective in this chapter was to study relevant algorithms for the development of deep learning model to classify and predict learner's behaviour in MOOC. Given the time series type of dataset, we tested three architectures of RNNs to find out that simple RNN with input features offers the best precision (performance and accuracy) in classifying and predicting learning behaviour in the learning process. One of the key benefits of this model is the fact that, by giving a good understand of learner's behaviour, the model might guide teachers to provide personalized support and interventions to learners in the learning process. This would give a tool to the course instructor/teacher who is the main tutor of any learning process, who mastered the course content, and who can better assist learners so that they actively benefit from the entire course. This analysis also reveals that learner's behaviours concerning video viewing, posting in discussion forums, viewing discussion forums and quizzes could help predict learner behaviour about other types of content. This model could also be used to support adaptive content

and learner pathways, by suggesting the revision or restructuring of the content and/or training path, closure of a course, or the launch of a new course. This model could also support data understanding, by providing insight information, exploratory or explanatory tools, assisting in the understanding of the mechanisms underlying the desired outcome, and then helping to understand the mechanisms behind the outcomes.

The main challenge of this study was the imbalanced nature of our dataset in the single MOOC used for this first experimentation. In addition, the learner behaviour or interaction with a given content is not easily and deeply measurable. We used common behaviours like clicking, viewing, downloading, uploading, attempting, and posting which are not always suitable and easy to measure.

Future research would address the following main issues: (a) develop a method to measure the quality of learner behaviours, (b) test the model in other MOOCs and explore methods to further improve the deep learning behaviour prediction model performance in MOOCs by increasing the hidden layers in the network, (c) the current study only shows statistical validity of the model, (d) further researchers could examine its validity by implementing the model in ongoing MOOC courses to assess it in real-time prediction, and (e) subsequent research should also include designing personalized interventions based on the model predictions.

## References

- AFRICAINE, U. (2016). Stratégie continentale de l'éducation pour l'Afrique.
- Anani, W., & Samarabandu, J. (2018). Comparison of recurrent neural network algorithms for intrusion detection based on predicting packet sequences. In *2018 IEEE Canadian Conference on Electrical & Computer Engineering (CCECE)* (pp. 1–4). IEEE.
- Andres, J. M. L., Baker, R. S., Gašević, D., Siemens, G., Crossley, S. A., & Joksimović, S. (2018). Studying MOOC completion at scale using the MOOC replication framework. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge* (pp. 71–78).
- Ashenafi, M. M., Ronchetti, M., & Riccardi, G. (2016). Predicting student progress from peer-assessment data. In *International Educational Data Mining Society*.
- Baehrens, D., Schroeter, T., Harmeling, S., Kawanabe, M., Hansen, K., & Müller, K.-R. (2009). How to explain individual classification decisions. Preprint, arXiv:0912.1128.
- Bote-Lorenzo, M. L., & Gómez-Sánchez, E. (2017). Predicting the decrease of engagement indicators in a MOOC. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference* (pp. 143–147).
- Boyer, S., & Veeramachaneni, K. (2015). Transfer learning for predictive models in massive open online courses. In *International Conference on Artificial Intelligence in Education* (pp. 54–63). Springer.
- Brinton, C. G., & Chiang, M. (2015). MOOC performance prediction via clickstream data and social learning networks. In *2015 IEEE Conference on Computer Communications (INFOCOM)* (pp. 2299–2307). IEEE.
- Brooks, C., Thompson, C., & Teasley, S. (2015). A time series interaction analysis method for building predictive models of learners using log data. In *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge* (pp. 126–135).
- Brooks, C., Thompson, C., & Teasley, S. (2015). Who you are or what you do: Comparing the predictive power of demographics vs. activity patterns in massive open online courses

- (MOOCs). In *Proceedings of the Second (2015) ACM Conference on Learning@ Scale* (pp. 245–248).
- Brown, R., Lynch, C. F., Wang, Y., Eagle, M., Albert, J., Barnes, T., Baker, R. S., Bergner, Y., & McNamara, D. S. (2015). Communities of performance & communities of preference. In *EDM (Workshops)*. Citeseer.
- Champaign, J., Colvin, K. F., Liu, A., Fredericks, C., Seaton, D., & Pritchard, D. E. (2014). Correlating skill and improvement in 2 MOOCs with a student's time on tasks. In *Proceedings of the First ACM Conference on Learning@ Scale Conference* (pp. 11–20).
- Che, Z., Purushotham, S., Cho, K., Sontag, D., & Liu, Y. (2018). Recurrent neural networks for multivariate time series with missing values. *Scientific Reports*, 8(1), 1–12.
- Cocca, M., & Weibelzahl, S. (2009). Log file analysis for disengagement detection in e-learning environments. *User Modeling and User-Adapted Interaction*, 19(4), 341–385.
- Crossley, S., Paquette, L., Dascalu, M., McNamara, D. S., & Baker, R. S. (2016). Combining click-stream data with NLP tools to better understand MOOC completion. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge* (pp. 6–14).
- DeBoer, J., & Breslow, L. (2014). Tracking progress: Predictors of students' weekly achievement during a circuits and electronics MOOC. In *Proceedings of the First ACM Conference on Learning@ Scale Conference* (pp. 169–170).
- Dillon, J., Bosch, N., Chetlur, M., Wanigasekara, N., Ambrose, G. A., Sengupta, B., & D'Mello, S. K. (2016). Student emotion, co-occurrence, and dropout in a MOOC context. In *International Educational Data Mining Society*.
- Dowell, N. M., Skrypyk, O., Joksimovic, S., Graesser, A. C., Dawson, S., Gašević, D., Hennis, T. A., de Vries, P., & Kovanovic, V. (2015). Modeling learners' social centrality and performance through language and discourse. In *International Educational Data Mining Society*.
- Fei, M., & Yeung, D.-Y. (2015). Temporal models for predicting student dropout in massive open online courses. In *2015 IEEE International Conference on Data Mining Workshop (ICDMW)* (pp. 256–263). IEEE.
- Gardner, J., & Brooks, C. (2018). Student success prediction in MOOCs. *User Modeling and User-Adapted Interaction*, 28(2), 127–203.
- Garman, G., et al. (2010). A logistic approach to predicting student success in online database courses. *American Journal of Business Education (AJBE)*, 3(12), 1–6.
- Greene, J. A., Oswald, C. A., & Pomerantz, J. (2015). Predictors of retention and achievement in a massive open online course. *American Educational Research Journal*, 52(5), 925–955.
- He, J., Bailey, J., Rubinstein, B., & Zhang, R. (2015). Identifying at-risk students in massive open online courses. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 29).
- Jiang, S., Williams, A., Schenke, K., Warschauer, M., & O'dowd, D. (2014). Predicting MOOC performance with week 1 behavior. In *Educational Data Mining 2014*.
- Joksimović, S., Manataki, A., Gašević, D., Dawson, S., Kovanović, V., & De Kereki, I. F. (2016). Translating network position into performance: Importance of centrality in different network configurations. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge* (pp. 314–323).
- Jordan, K. (2015). MOOC completion rates. *Recuperado de*. <http://www.katyjordan.com/MOOCproject.html>
- Kennedy, G., Coffrin, C., De Barba, P., & Corrin, L. (2015). Predicting success: How learners' prior knowledge, skills and activities predict MOOC performance. In *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge* (pp. 136–140).
- Khalil, H., & Ebner, M. (2014). MOOCs completion rates and possible methods to improve retention—a literature review. In *EdMedia+ Innovate Learning* (pp. 1305–1313). Association for the Advancement of Computing in Education (AACE).
- Kizilcec, R. F., & Halawa, S. (2015). Attrition and achievement gaps in online learning. In *Proceedings of the Second (2015) ACM Conference on Learning@ Scale* (pp. 57–66).

- Kotsiantis, S., Tselios, N., Filippidi, A., & Komis, V. (2013). Using learning analytics to identify successful learners in a blended learning course. *International Journal of Technology Enhanced Learning*, 5(2), 133–150.
- Mao, Y. (2018). Deep learning vs. Bayesian knowledge tracing: Student models for interventions. *Journal of Educational Data Mining*, 10(2).
- Mwalumbwe, I., & Mtebe, J. S. (2017). Using learning analytics to predict students' performance in Moodle learning management system: A case of Mbeya University of Science and Technology. *The Electronic Journal of Information Systems in Developing Countries*, 79(1), 1–13.
- Nagrecha, S., Dillon, J. Z., & Chawla, N. V. (2017). MOOC dropout prediction: Lessons learned from making pipelines interpretable. In *Proceedings of the 26th International Conference on World Wide Web Companion* (pp. 351–359).
- Pardos, Z., Bergner, Y., Seaton, D., & Pritchard, D. (2013). Adapting Bayesian knowledge tracing to a massive open online course in edX. In *Educational Data Mining 2013*. Citeseer.
- Pardos, Z. A., Tang, S., Davis, D., & Le, C. V. (2017). Enabling real-time adaptivity in MOOCs with a personalized next-step recommendation framework. In *Proceedings of the Fourth (2017) ACM Conference on Learning@ Scale* (pp. 23–32).
- Pursel, B. K., Zhang, L., Jablokow, K. W., Choi, G. W., & Velegol, D. (2016). Understanding MOOC students: Motivations and behaviours indicative of MOOC completion. *Journal of Computer Assisted Learning*, 32(3), 202–217.
- Qiu, J., Tang, J., Liu, T. X., Gong, J., Zhang, Q., & Xue, Y. (2016). Modeling and predicting learning behavior in MOOCs. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining* (pp. 93–102).
- Ramos, C., & Yudko, E. (2008). “Hits” (not “discussion posts”) predict student success in online courses: A double cross-validation study. *Computers & Education*, 50(4), 1174–1182.
- Ren, Z., Rangwala, H., & Johri, A. (2016). Predicting performance on MOOC assessments using multi-regression models. Preprint, arXiv:1605.02269.
- Rokach, L. (2005). *Data mining and knowledge discovery handbook*. Springer Science+ Business Media, Incorporated.
- Rosé, C. P., Carlson, R., Yang, D., Wen, M., Resnick, L., Goldman, P., & Sherer, J. (2014). Social factors that contribute to attrition in MOOCs. In *Proceedings of the First ACM Conference on Learning@ Scale Conference* (pp. 197–198).
- Russo, T. C., & Koesten, J. (2005). Prestige, centrality, and learning: A social network analysis of an online class. *Communication Education*, 54(3), 254–261.
- Siemens, G., & Baker, R. S. d. (2012). Learning analytics and educational data mining: Towards communication and collaboration. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 252–254).
- Sinha, T., Li, N., Jermann, P., & Dillenbourg, P. (2014). Capturing” attrition intensifying” structural traits from didactic interaction sequences of MOOC learners. Preprint, arXiv:1409.5887.
- Stein, R. M., & Allione, G. (2014). Mass attrition: An analysis of drop out from a principles of microeconomics MOOC.
- Tucker, C., Pursel, B. K., & Divinsky, A. (2014). Mining student-generated textual data in MOOCs and quantifying their effects on student performance and learning outcomes. In *2014 ASEE Annual Conference & Exposition* (pp. 24–907).
- UNESCO. (2005). Rapport mondial de suivi sur l'épt 2005. *UNESCO* (4), 57–84.
- UNESCO, Mundial, G. B., UNICEF, et al. (2016). Education 2030: Incheon declaration and framework for action: Towards inclusive and equitable quality education and lifelong learning for all.
- Wang, X., Yang, D., Wen, M., Koedinger, K., & Rosé, C. P. (2015). Investigating how student's cognitive behavior in MOOC discussion forums affect learning gains. In *International Educational Data Mining Society*.
- Wen, M., Yang, D., & Rose, C. (2014). Sentiment analysis in MOOC discussion forums: What does it tell us? In *Educational Data Mining 2014*.



- Whitehill, J., Williams, J., Lopez, G., Coleman, C., & Reich, J. (2015). Beyond prediction: First steps toward automatic intervention in MOOC student stopout. Available at SSRN 2611750.
- Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, 1(1), 67–82.
- Xiao, X., Pham, P., & Wang, J. (2015). Attentivelearner: Adaptive mobile MOOC learning via implicit cognitive states inference. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction* (pp. 373–374).
- Xing, W., Chen, X., Stein, J., & Marcinkowski, M. (2016). Temporal predication of dropouts in MOOCs: Reaching the low hanging fruit through stacking generalization. *Computers in Human Behavior*, 58, 119–129.
- Xing, W., & Du, D. (2019). Dropout prediction in MOOCs: Using deep learning for personalized intervention. *Journal of Educational Computing Research*, 57(3), 547–570.