International Journal of Computer Applications (0975 – 8887)
Volume 179 – No.3, December 2017

Identification of Wheel Spinning Cases while Learning and Retaining a Skill in Intelligent Tutoring Systems

N. Sharada
Research Scholar
Dept. of CS & SE
Andhra University

M. Shashi
Professor
Dept. of CS & SE
Andhra University

Xiaolu Xiong
Data Researcher
Worcester Polytechnic Institute
Worcester MA01609

ABSTRACT
Wheel Spinning, by definition [1] means a situation where in the students may not reach mastery in reasonable amount of time. The Wheel Spinning is unproductive and may lead to frustrating experience to some students, in particular dyslexics, since they have difficulty in memorizing and thinking skills. If the wheel spinning can be identified early, then they may be offered some other mode of instruction such as remedial intervention by the teacher, peer tutoring method or incorporating personalized styles of instruction. The aim of this work is to extend wheel spinning concept to enhanced mastery cycle and to design more accurate personalized retention schedules. This study was conducted using real world data from Personalized Adaptive Scheduling System(PASS) a newly introduced module ASSISTments, web based tutoring system. Application of the state-of-the-art machine learning approaches such as deep learning and random forest are investigated on the extracted features for modeling wheel spinning cases. Experiments demonstrate that Random Forest model can predict mastery or wheel spinning at an early stage with an AUC of 0.87.

General Terms
Educational Data Mining, Intelligent Tutoring Systems

Keywords
Wheel Spinning, Personalized Adaptive Scheduling Systems, Deep Learning, Random Forests

1. INTRODUCTION
Intelligent Tutoring Systems (ITS) are effective computer based learning environments designed to provide one to one personalized instruction to learners. ITS primarily perform tutoring function by asking questions related to specific topic or skill to be learned and interactively offers feedback or hints upon request. From the performance on each problem, the student's knowledge model [2] will have information whether the skill is mastered by the student or not. This process is referred to as mastery learning. The most commonly used mastery threshold for a skill is three consecutive correct answers. In mastery learning, the students are presented with as many problems as needed for practice, with the hope that they might utilize these opportunities to master the skill. The students with learning issues or low knowledge however could keep failing to learn the skill, while the system monotonously presents the student with even more practice problems. Yue et al., [3] introduced the concept of wheel spinning, when the student gets trapped in the mastery learning cycle and if the student fails to achieve mastery in timely manner. The previous study [4] revealed that students spend 28% of their time wheel spinning in ASSISTments, a non profit, web based tutoring system. Neural Network [5] based wheel-spinning detector has been proposed to perform binary classification of mastery learning data from Cognitive Tutor. A study [6] investigated whether, the performance of pre-requisite skills influenced the ability to learn post-requisite skills and subsequently lead to wheel spinning. The shallow learners (learning that does not support transfer) were identified [7] using step regression technique from features such as the speed of student responses, student responses after receiving bug message and slips during performance. The student's attitude towards learning and amount learned and perception [8]of the system was modeled using combination of machine learning methods and classical statistical analysis.

A recently introduced feature in ITS is enhanced mastery cycle [9], which involves periodically retesting students of the mastered skills, which helps them relearn/recollect forgotten skills and thereby enhance long term retention performance. A system of Personalized Adaptive Scheduling System (PASS) [10,11,12] was developed to make decisions on when to review the initially mastered skills in ASSISTments. PASS will automatically reassess the student with retention tests at expanding intervals spread across a schedule of at least three months from initial mastery. The first level of reassessment test is personalized, based on the student's mastered speed. The longest period of time or delay for first level retention test is seven days when the students' mastery speed is as good as three, while the shortest time period is one day for students who require seven or more opportunities to achieve initial mastery. The second level of retention test takes place 14 days after successfully completing the first level retention tests, followed by third and fourth retention tests with much longer delay of 28 days and 56 days. When a student gives correct responses he will be promoted to the next level. However, if a student answers incorrectly in any one of the retention tests, ASSISTments will give him an opportunity to relearn this skill before redoing the same level of test.

In enhanced mastery cycle framework, as implemented in ITS, there is no upper limit for the number of retention test attempts, but continues to present the student with retention tests until he successfully completes the tests at all the four levels as per the retention schedule. Consequently, the system keeps giving him more spaced retrieval practices in the hope that he might utilize these new opportunities to retain the skill. The student however could keep failing to retain the skill, which triggers the system to present even more problems to the student. Thus the student can possibly become trapped in the spaced retrieval cycle if he fails to achieve long term retention of the skill. In tutoring system this phenomenon is termed as wheel-spinning, wherein the tutor is providing spaced retrieval practice opportunities to students and there appears to be a productive work, but the students are not making progress towards long term mastery.

In the existing enhanced mastery cycle, the medium and slow learners had to spend approximately four weeks to reach Level 2 retention test while students with better learning potential needed approximately 18 days. Although, the slow learners are presented with more spaced retention tests and relearning assignments, but this didn't stop the decay of their retention levels. These students had spent approximately 3 additional relearning assignments on the same skill, but there is only slight improvement on their retention performance. This indicates that wheel spinning phenomenon is observed even after initial mastery of the skill and is reflected in terms of additional relearning attempts. This work aims to devise a classifier to accurately identify wheel spinning cases during skill retention cycle.

The reminder of the paper is organized as follows. In section II the problem statement is defined. In section III the theoretical aspects of Feed Forward Neural Networks and Random Forests techniques and methodology are discussed. In section IV experimental results on the data set with 6,23,904 student learning experiences collected from PASS retention test performance are analyzed. Section V offers conclusions and future directions.

2. PROBLEM STATEMENT

The Wheel Spinning is unproductive and may lead to frustrating experience to some students, in particular dyslexics, since they have difficulty in memorizing and thinking skills. If the wheel spinning learners can be identified early, then they may be offered some other mode of instruction such as remedial intervention by the teacher, peer tutoring method or incorporating personalized styles of instruction in ITS itself. The aim of this work is to identify the most reliable features for predicting the wheel-spinning cases or in trouble students at an early stage in the enhanced mastery cycle.

Recently, there is a major advancement in training densely connected, neural networks with many hidden layers. The deep networks thus resulting learn a hierarchy of nonlinear features, which can capture complex patterns in data. These advances triggered authors interest in developing machine learning models based on deep learning techniques and random forests for predicting whether the student will master the skill or wheel spin.

3. METHODOLOGY

3.1 DATASET FROM PASS Module

The Retention Test Performances spread across a span of 3 months from PASS module in ASSISTments platform is considered for this study. The Web based Tutoring System is mostly used for urban school districts of the Northern United States for 4th through 10th grade mathematics. The data is collected from school year 2014 to 2015 which comprises of learning experiences of about 14,512 unique students while solving problems related to 154 mathematics skills within ASSISTments. In total there are 6,23,904 data records, the description of the dataset is in Table1. Each row of the data set represents the retention performance, at different levels after a student mastered a skill. The logged information includes the identity of the student, the class to which he/she belong to, the identity of the teacher, the skill identity, the mastery speed, number of days after which retention test is conducted, the difficulty of the question that was asked in the test, the level at which the test is conducted, repeated times, response time and the result of the retention test in terms of 0 and 1.

3.2 Preprocessing

Since learning and long term retention are complex cognitive processes, feature extraction is crucial in student modeling. The features to model wheel-spinning cases were chosen from three perspectives: student performance, attentiveness of the learner through the speed of responding to a problem and general information about learning as detailed below.

Attributes	**Description**
Student_id	Student identity
Student_class_id	Class identity
Mastery_speed	Number of practice problems required to obtain three consecutive correct answers
Problem_id	Problem identity
Delay_days	Number of days after which the retention test is scheduled
Teacher_id	Teacher identity
Skill_id	Skill identification
Correct	Retention test performance
Easiness	Difficulty of the question asked in the retention test
Teacher _id	Teacher identity
Current_Grade	Grade of the student
Response Time	Time taken to attempt the question in seconds
Reassesement Level	The current level of the Reassesment Test
Repeated Times	Number of times the current level is repeated

Table 1: Retention Performance Description Table

Student Performance

"Prior_correct_count" – The prior number of problems for this skill, solved correctly by the student, without taking hint support.

"Mean_retention_performance" - The forgetfulness nature of the student is captured by computing average retention performance across all the skills practiced in the tutor.

"Mean_mastery_speed" – The learning speed of new skill or capacity to apply acquired knowledge and skills from one problem solving situation to another is estimated by means of average mastery speed of the student.

Attentiveness of the Learner

"Prior_Fast_Count" – The number of prior problems solved correctly and one standard deviation less than average response time for that skill.

"Prior_Slow_Count"- The number of prior problems solved correctly and one standard deviation more than average response time for that skill.

"Prior_Normal_Count" – The prior number of problems solved correctly and within one standard deviation of average response time for that skill.

General Information

"Prior_Problem_Count" - The maximum number of problems the student attempted for this skill. This feature is included to investigate the relationship of this parameter with wheel spinning.

"Skill_Mean_Retention" – Mathematical learning involves acquisition of a number of component skills such as logical and analytical skills, procedural skills and computational skills. The skills that require cognitive resources and effort require good amount of practice because they are more likely to easily forgotten. The nature of the skill is assessed from the average retention ability of the skill across all the students.

"Skill_Mastery_Speed" – The average number of practice problems required to master the skill, which indicates the difficulty level of the skill to be mastered.

"Class_Mean_Retention" – The class mean retention parameter represents the effectiveness of teaching learning process experienced by the students within the class.

"Class_Mastery_Speed" - The teaching approaches that engage and enhance student learning in the class as well as the effectiveness of the course material, are analyzed by computing the class average mastery speed.

3.3 Machine Learning Models

Several Machine Learning algorithms each one with its own purposes and capabilities, have been proposed for classification and regression tasks. To investigate the influence of constructed features three binary classifier models using Deep Learning, Random Forest and Logistic Regression technique are build. The dependent variable is binary variable with two possible values of successful retention or wheel spinning.

3.4 Deep Learning

The deep neural networks used for predictive modeling of wheel spinning cases are based on multi-layered feed forward networks with multiple layers of interconnected neurons. The computational models [12] with multiple processing layers transform representation of data at one level to data representations at much higher abstract level. With the composition of many such transformations, complex patterns can be learned. Deep learning improved many aspects of modern society from speech recognition, visual object detection to recommendations on e-commerce websites. The multiple layers of neurons present in the feed forward neural network constitute the depth of the network and the number of neurons in each layer represent the width of the network. The weights linking the neurons and biases from other neurons in addition to the width and depth of the network determine the output of the entire network.

The deep learning model is trained with three hidden layers and each layer comprised of 100 neurons. The hyperbolic tangent function is used to transmit the input information, through hidden layers until it reaches the output nodes. The hyperbolic tangent (tanh) function is a rescaled version of the sigmoid function, whose output ranges from -1 to +1. This allows the algorithm to converge faster. Binomial distribution function is used along with cross entropy or log-loss for the response variables in the classification.

3.5 Random Forests

A Random Forest model [13] with 100 unpruned classification trees is built each of which, with a random subset of candidate features. Each tree of the Random Forest is constructed with a random replacement bootstrap sample from the data. To reduce the correlations among trees, the square root of the number of variables in the data set are randomly sampled for classification. The Random Forest model predicts the category and supported by the majority of trees for classifying a test example.

4. EXPERIMENTAL RESULTS

All the experiments reported in this study were conducted using R scripting functionality for H2O, an open source environment for big data that facilitates the use of parallel distributed Machine Learning algorithms. R is a free and high-level programming language with a powerful suite of tools for statistical and data analysis.

The deep feed forward model is trained with an input layer, three hidden layers (each layer with 100 neurons), and an output layer. The distribution function of response variable is set to binomial and the cross entropy loss function is chosen for model estimation. Random forest (RF), a powerful ensemble classification algorithm is built with 100 trees as base classifiers.

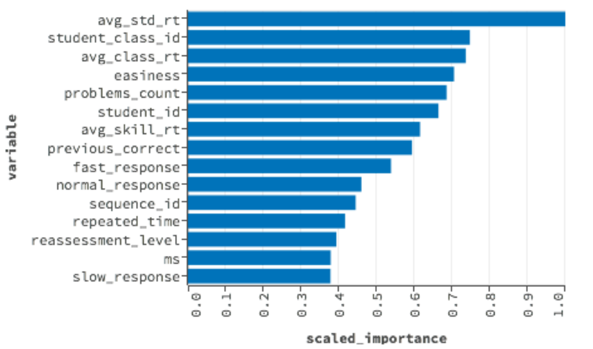

Figure 1. Importance of the features

In Figure 1 the most important features for predicting whether the student will wheel spin or master are consistent across all three classification models. As expected the student performances have a high impact in all the models. Among all the student performance features "avg_std_rt" is the most relevant feature and is negatively associated with wheel spinning. In the features which reflect the attentiveness of the learner "fast_response" is negatively associated with wheel spinning. In the general features "problems_count" is positively associated with wheel spinning and "avg_class_rt" is negatively associated with wheel spinning. Nevertheless, there are other relevant factors, such as easiness (difficulty of the question asked in the test), repeated time (number of times the level is repeated) and reassessment level (the time interval after which review test is conducted) variables .

The test dataset is analyzed on three models. In classification the model goodness is evaluated in terms of log loss, Root Mean Square Error, R-Square, Area under Curve, and mean per class error. Each evaluation metric account for different aspects of the model and the data, hence a combination of metrics is ideal for comparing models and assessing the quality of predictions. In general, the complex hierarchal Deep Learning and Logistic Regression are outperformed by the Ensemble Random Forests.

Logarithmic loss quantifies the accuracy of the classifier by penalizing wrong classifications. Minimization of log loss leads to maximization of classifier accuracy. In Figure 2 & 3 the log loss error is analyzed with respect to the number of epochs in Feed Forward Neural Network and number of trees in Random Forest respectively. The Random Forest with a minimum logloss of 0.235 is marginally better than Deep Neural Network with a log loss of 0.247.

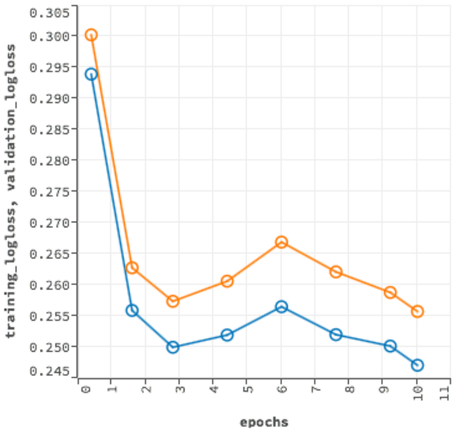

Figure 2: Log loss for Deep Learning

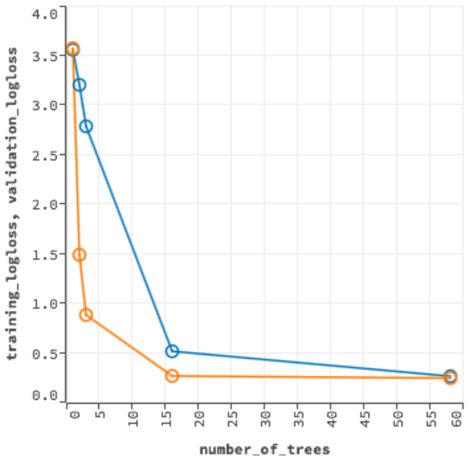

Figure 3: Logloss for Random Forests

The Area under the receiver operating characteristics curve , known as AUC is a standard method for evaluating the predictive accuracy of classifier systems, with higher scores indicating higher accuracies. The model performances are also assessed using the error metrics such as MSE, RMSE and coefficient of determination (R Square).

Classification Metrics	Logistic regression	Deep Learning	Random Forest
MSE	0.081	0.0916	**0.072**
RMSE	0.432	0.382	**0.268**
R2	0.40	0.45	**0.46**
Logloss	0.341	0.247	**0.235**
AUC	0.84	0.86	**0.87**
Mean_per class_error	0.346	0.195	**0.186**

Table 2: Evaluation metric for classification.

The Root Mean Square Error is a measure of average deviation of the predicted values from the observed values, whereas R Square measure the variability in the depended variable explained by the regression model. The Random Forest is the best choice in terms of all metrics, followed by the Deep Learning, which obtains next best results. Statistically significant differences between Random Forests and all other algorithms are indicated in bold in Table 2.

5. CONCLUSIONS

The prediction of wheel spinning cases or in trouble students in enhanced mastery cycle, is addressed in this work by using real-world dataset from Personalized Adaptive Scheduling System. Three sets of features were constructed and their influence on predicting wheel spinning cases are analysed. The state-of-the-art machine learning approaches such as Deep Learning, Random Forests and Logistic Regression to predict whether the student will master a problem or wheel-spin were explored. The results obtained from the best predictive models identified significance of the features such as forgetfulness and teaching learning experiences of the student. These findings have practical implication, in the development of personalized tutoring systems for high school students. When the student is likely to wheel-spin, then there is no point in testing for long term retention knowledge of the student. The implementation of corrective measures such as remedial intervention by the teacher, peer tutoring, or training of missing related prerequisite skills need to be considered to avoid frustrating learners.

6. FUTURE WORK

This research work proposes the usage of features such as student forgetfulness and teaching learning process as a basis for modeling and predicting the wheel-spinning or mastery classes. It may be beneficial to consider other aspects of the student such as hint usage, pre-requisite skills of the skills under consideration. Future Research involves analysis of how much time otherwise spend in wheel spinning can be reduced by ensuring that students mastered the prerequisite skills. Future work should also consider other Machine Learning approaches for addressing the class imbalance problem in the dataset, the wheel-spinning cases being minority class.

7. REFERENCES

[1] Beck, Joseph E., and Yue Gong. "Wheel-spinning: Students who fail to master a skill." International Conference on Artificial Intelligence in Education. Springer, Berlin, Heidelberg, 2013.

[2] Corbett, Albert T., and John R. Anderson. "Knowledge tracing: Modeling the acquisition of procedural knowledge." User modeling and user-adapted interaction 4.4 (1994): 253-278.

[3] Beck, Joseph, and Ma Mercedes T. Rodrigo. "Understanding wheel spinning in the context of affective factors." International Conference on Intelligent Tutoring Systems. Springer, Cham, 2014.

[4] Gong, Yue, and Joseph E. Beck. "Towards detecting wheel-spinning: Future failure in mastery learning." Proceedings of the Second (2015) ACM Conference on Learning@ Scale. ACM, 2015.

[5] Matsuda, Noboru, Sanjay Chandrasekaran, and John C. Stamper. "How quickly can wheel spinning be detected?." EDM. 2016.

[6] Wan, Hao, and Joseph Barbosa Beck. "Considering the Influence of Prerequisite Performance on Wheel Spinning." International Educational Data Mining Society (2015).

[7] Baker, Ryan, et al. "Towards automatically detecting whether student learning is shallow." Intelligent Tutoring Systems. Springer Berlin/Heidelberg, 2012.

[8] Arroyo, Ivon, and Beverly Park Woolf. "Inferring learning and attitudes from a Bayesian Network of log file data." AIED. 2005.

[9] Xiong, Xiaolu, Shoujing Li, and Joseph E. Beck. "Will You Get It Right Next Week: Predict Delayed Performance in Enhanced ITS Mastery Cycle." FLAIRS Conference. 2013.

[10] Xiong, Xiaolu, Yan Wang, and Joseph Barbosa Beck. "Improving students' long-term retention performance: a study on personalized retention schedules." Proceedings of the Fifth International Conference on Learning Analytics And Knowledge. ACM, 2015.

[11] Xiong, Xiaolu, and Joseph Barbosa Beck. "Improving Long-Term Retention Level in an Environment of Personalized Expanding Intervals." EDM. 2015.

[12] Sharada, N., M. Shashi, and Xiaolu Xiong. "Enhanced Retention Performance Modeling for Intelligent Tutoring System." International Journal of Computer Applications 151.10 (2016).

[13] LeCun, Yann, YoshuaBengio, and Geoffrey Hinton. "Deep learning." Nature 521.7553 (2015): 436-444.

[14] Breiman, Leo. “Random Forests,” Machine Learning 45.1(2001)