

Detecting Students Gifted in Mathematics with Stream Mining and Concept Drift Based M-Learning Models Integrating Educational Computer Games

<https://doi.org/10.3991/ijet.v16i12.21925>

Petar Juric ^(✉), Marija Brkic Bakaric, Maja Matetic
University of Rijeka, Rijeka, Croatia
petar_jur@yahoo.com

Abstract—One of the problems of individualized classes which adapt contents and methods of teaching to students of different cognitive capabilities is early and widely available detection of students gifted in certain educational fields. The paper proposes models which are based on stream mining and which can detect students gifted in Mathematics solely on the basis of their interaction with the m-learning system using educational computer games and with no access to any other feature except for student age. Classification accuracy and time-efficiency of different feature selection methods are examined in order to make the models more interpretable, hence less complex. Stream mining classification accuracy in the utilized models is evaluated on new (yet unseen) records, while the concept drift detection analyses at which point of time new models should be built.

Keywords—E-learning, m-learning, game-based learning, Mathematics, stream mining, concept drift, gifted students

1 Introduction

Educational data mining (EDM) is an interdisciplinary field which is concerned with the development of methods for conducting research on educational data, which can also be applied in analysing data from different e-learning systems. Besides applying traditional techniques for data mining, the approach requires certain adaptations specific to education-related problems, such as improving the educational process or predicting student success [1].

Transformation and adaptation of e-learning systems to m-learning facilitates their wide availability. It makes learning available anytime and anywhere [2]. In the last four years, research in the field of m-learning has grown almost 67.6% [3].

In order to improve the educational process by adapting to specific needs of students gifted in certain fields, it is necessary to detect these students as early as possible. One of the ways to accomplish this is to use decision trees for classifying students and content [4]. This paper focuses on detecting primary school students gifted in Mathematics.

In order to facilitate the development of mathematical creativity with students in the educational process, one has to make sure that the choice of learning problem

tasks coincides with their interests, understanding, skills, and that creativity is credited for as soon as it is spotted [5].

All the components of big data knowledge discovery are utilized in this research. However, since classical data mining models are not appropriate for data streams generated by the interaction with educational computer games in m-learning systems, stream mining algorithms are employed in the research [6].

2 Related Work

Feature selection enables reducing complexity of data mining models by decreasing the number of dimensions, facilitating generalization, speeding up the process of learning, and facilitating model interpretability [7].

Majority of machine learning algorithms already entail feature selection. However, it has been shown in practice that adding just one additional feature with random binary values decreases classification accuracy of decision trees (C4.5) for 5-10%. This is due to the inner workings of the splitting in decision trees. With each new tree depth level, less data is available for choosing a feature for splitting the data, which is why even a newly added feature might seem important to the algorithm. This issue is even more pronounced when the dataset contains a great number of irrelevant features [8].

Besides the methods already integrated in classification algorithms, two additional approaches to feature selection can be used in order to get the reduced dataset [9]:

1. Independent assessment based on general data characteristics (filter method)
2. Applying machine learning algorithm which will eventually be used for training (envelope method).

The first approach takes place prior to learning the prediction model and filters the feature set in order to obtain the reduced dataset which is as good as possible before training, while the second approach filters features while learning a predictive model by selecting a subset of features which gives the best accuracy.

Static datasets are finalized datasets which can be read several times either thoroughly or in different ratios, dependent on the model evaluation strategy, i.e. the partitioning of data into training and test sets. Unlike finalized datasets, data streams are being continuously updated and modified and may never end, which is why the following requirements and limitations are imposed [10]:

1. Every stream is observed only once
2. Processing time of a record must be short
3. Memory usage must be small
4. Results have to be available at any point of time
5. Streams evolve over time and hence represent non-stationary data sources.

Additionally, data in streams are analysed sequentially, in order of arrival. Random access is not available, and a record cannot be reused once it was processed. In order to reduce the processing time, it is necessary to define the upper limit which refers to the number of processes needed for processing a record. For real-time data analysis,

algorithms have to process data at the same speed at which the data arrive, or faster. Memory used by the algorithms refers to the stored statistical data and the stored model, the two of which are often combined so that statistical data is an elementary part of the model while making predictions. A model has to be built irrespective of the number of data records processed, and a prediction needs to be made even when only a small number of records was processed [11].

In order to develop decision making trends related to teaching and its advancement based on data mining, a data driven culture needs to be developed within teachers and institutions. One of the currently active areas of research is game learning analytics (applying data-mining and visualization techniques to player interactions) [12].

Outcomes are influenced by affective states, i.e. positively by concentration (engaged), and negatively by confusion and boredom, as evident from the analysis of learning outcomes in Online Mathematical Problem Solving. With the aim to reduce students' boredom, Chiu in [13] suggests including interesting problem-solving designs, such as games, into online mathematical problem-solving platforms.

Decision trees are one of the most often used algorithms in the data analysis of computer games (intended for fun) and in building predictive models of players, with the goal to enhance player experience [14]. Besides some other algorithms, they are also used with computer games which are primarily intended for learning. Findings of such research studies are important to teachers, students, parents, as well as game developers, and researchers [15].

Data stream processing supplemented with feature selection and analysis can be used in secondary schools for predicting STEM careers from student usage of an Intelligent Tutoring System [16].

In this paper, a learning model which uses computer games is implemented, and data stream mining is utilized for a preliminary selection of primary school students gifted in Mathematics. The feature selection and analysis in the computer game is important for building predictive models in order to reduce the time needed for building the model and to increase the interpretability of models. The proposed learning model enables monitoring motivation for using the system and learning outcomes achieved. It provides equal user experience in e-learning and in m-learning. The learning model integrates a social network adapted for primary school use.

3 Research Questions

For the purpose of this research, students gifted in Mathematics are defined as those primary school students who take additional school classes in Mathematics.

The research analyses whether it is possible to use data stream mining for detecting students gifted in Mathematics with a preliminary evaluation based on assessing motivation and acquired knowledge.

The following hypothesis is formulated: it is possible to apply data stream mining techniques for a preliminary student evaluation based on assessing motivation and acquired knowledge in a mobile-friendly model which uses educational computer games for learning Mathematics.

The hypothesis will be accepted if stream mining enables building models which detect students gifted in Mathematics based on their interaction with the mobile

learning system which uses educational computer games and implements features which refer to motivation and acquired knowledge.

Since the research additionally questions how to reduce the complexity of the obtained models, effects of feature selection methods on processing time, dimensions, and model accuracy are also explored. The question of how distributional changes can be observed within the recorded data is answered by the concept drift detection.

4 Methodology

Since the research presented in this paper deals with the classification of students based on additional school classes in Mathematics, classification decision trees are used to ease the interpretation and visualization of the model for the teachers of Mathematics. These machine learning algorithms can be applied with data streams, and besides being easily interpretable, they are flexible, which means that it is possible to obtain different versions of a model with the aim to achieve higher accuracy. The disadvantage of using classification decision trees refers to situations in which slight changes in feature values can affect accuracy and the model [17]. This research, therefore, utilizes feature selection methods in order to reduce the number of features.

4.1 Feature selection methods

For the selection of relevant features, besides the envelope method which uses classification evaluators, this research also uses Hoeffding trees for data streams [18], and filter methods implemented in WEKA [19] which work with nominal class values: Correlation, Information Gain, Chi-squared, Gain Ratio. Filter methods assign weight to each feature dependent on its relevance for the class [20].

4.2 Stream mining algorithms

Classic decision trees process data by keeping all the data in memory, which is why it is not possible to apply them efficiently to data streams.

Classification decision trees which are used in this research and which process data by satisfying all the data stream requirements are Very Fast Decision Trees (VFDT), or the so-called Hoeffding trees. The Hoeffding tree exploits the assumption that a small sample is good enough to choose an optimal splitting feature. The Hoeffding bound is used for selecting the best splitting feature out of n samples [21].

With data streams, first records are used for selecting the root of the tree. After selecting the root feature, subsequent records are used for branching to leaves. The Hoeffding bound is used for determining the smallest number of examples needed at a node for selecting a splitting feature.

For a random variable r in range R with n independent observations, the mean value \bar{r} is calculated, and then the Hoeffding bound determines with a probability of $1 - \delta$ that the real average of variable values is at least $\bar{r} - \epsilon$, where $\epsilon = \sqrt{\frac{R^2 \ln(1/\delta)}{2n}}$. Since the Hoeffding tree processes data streams, generated trees can grow infinitely,

which collides with the memory usage requirement, which is why the VFDT algorithm limits the number of nodes. The algorithm estimates the error reduction achievable by each active leaf. When the memory limit is reached by the model, the algorithm deactivates less promising nodes. VFDT keeps observing inactive leaves and can re-activate them when they become more promising than those currently active. Besides, the algorithm eliminates features which do not look promising after splitting. It can also delete from the memory the statistics (class distributions for each value) related to these features.

This research also uses two extensions of the Hoeffding tree: the Hoeffding option tree and the Hoeffding adaptive tree.

The Hoeffding option tree is built out of a unique structure of multiple Hoeffding trees because it contains additional nodes which enable optional splitting and can thus change the tree building path [22].

The Hoeffding adaptive tree uses the ADWIN (Adaptive Windowing) algorithm based on the sliding window model which decides on the active parts and the ones to be discarded by analysing the window size. When their accuracy decreases, old branches are replaced with the new ones [23].

The evaluation of the Hoeffding tree-based models built within this research is conducted by calculating the algorithm accuracy in a way that every single model is always tested on yet unseen records since every single record is first used in testing and only then in training.

4.3 Concept drift

Unlike stationary data, data streams can contain different concept drifts, dependent on whether the reason of a change or the speed of a change is considered. A concept drift refers to the situations in supervised learning in which the relationship between input data and target outcome changes over time. A real concept drift refers to the cases in which the function separating different classes changes (function determines the probability of class affiliation), while a virtual concept drift occurs when there is a change in the distribution of data. With respect to time, a concept drift can occur suddenly, incrementally, gradually, or repeatedly, while anomalies do not have to lead to a concept drift if they occur only once or by mere chance [24].

In this research, the Drift Detection Method (DDM) is chosen for the analysis of concept drifts by comparing accuracy over different data streams, since it is considered best suited for datasets which exhibit sudden concept drifts [25].

The Drift Detection Method is independent of the learning algorithm used. It analyses whether the classifier makes a correct or an incorrect class prediction in the form of bimodal distribution, where the probability v_i of an incorrect classification in the i^{th} position of the sequence is calculated with the standard deviation s_i . A significant increase in classification errors signifies a change in the distribution of classes, which is why the algorithm used for building the model is no longer appropriate. Warnings and detection values are analysed, and in case of a concept drift, the learned model is reset to its initial state, as well as the values v_{min} and s_{min} . A new model is build based on the records stored after the last warning [26].

Sudden concept drifts (change in the classification accuracy) in this research occur in cases in which, instead of incremental classification over the whole dataset,

window-based classification is made. Dependent on the window size, different ranges of students are analysed. Therefore, ratios between classes and distributions of records of different students significantly vary, especially in cases of smaller window sizes.

5 Dataset

The dataset is collected from the student interaction with the game “Zagonetke mudrog lisca” (Riddles of the wise fox), which deals with mathematical problem tasks presented through the Roman numerals on 15 pre-defined levels ordered by increasing difficulty. Throughout the interaction with the game, the data related to the following features are stored: current level, element holding time (time taken to think), element played, hit, number of moves at the level, number of remaining moves, number of times played, scores gained by the move, session (the number of games played in succession), total number of moves, total score prior to the move. Besides the listed features, the system also stores timestamps of records and platforms used for accessing the system (desktop/mobile).

The student data are anonymized by nicknames and based on the model Anonymized Social Network-based Mobile Game System for Learning Mathematics [27]. Besides nicknames, student data includes also age and information whether students attend additional school classes in Mathematics. Additional school classes in Mathematics are used as preparations for competitions, and cover more advanced content in comparison to the regular classes and with respect to the age. Additional school classes in Mathematics are attended only by students gifted in Mathematics.

The sample on which the research is conducted consists of 104 students aged 11-14 years and enrolled from fifth to eighth grade of a primary school in the Republic of Croatia. Besides taking school classes in Mathematics, the total of 23 students out of 104 students attend also additional school classes in Mathematics.

The time period includes one week prior to winter holidays, three weeks of holidays, and two weeks after holidays in the academic year 2017/2018.

The system was accessed by 73 students (70.19% of students).

The data is analysed with MOA, which is a tool for data stream mining with the concept drift detection [28].

6 Data Analysis and Results

6.1 Anomaly in the hoeffding tree accuracy at the first level of the game

During the course of the game, the students made 33081 move. Due to the unexpected deviation observed in the model accuracy, a total of 29212 moves is used for the analysis. More precisely, a significant deviation is detected in accuracy related to the records which refer to the first level. In the model of students gifted in Mathematics based on the Hoeffding tree, the problem is most pronounced for the first 5000 records with the window size set to 500, grace period of 200 records (the number of instances a leaf should observe between splits) and sample frequency of 1000. The whole dataset contains an abrupt decrease in accuracy on the initial records

(which contain first level data) compared to the dataset without the first level records (Fig. 1). This is due to the observation that some students skip the first level by purposefully making mistakes (the game ends after three wrong moves) in order to access the menu at the end of the game which enables them to see the top list (Top 20) and compare their accomplishments with other students, and to check the number of medals won and whether anyone beat them. Additionally, the first level is used by the teachers for providing instructions and explaining different ways of solving tasks. The dataset is thus pre-processed to discard records related to the first-level moves.

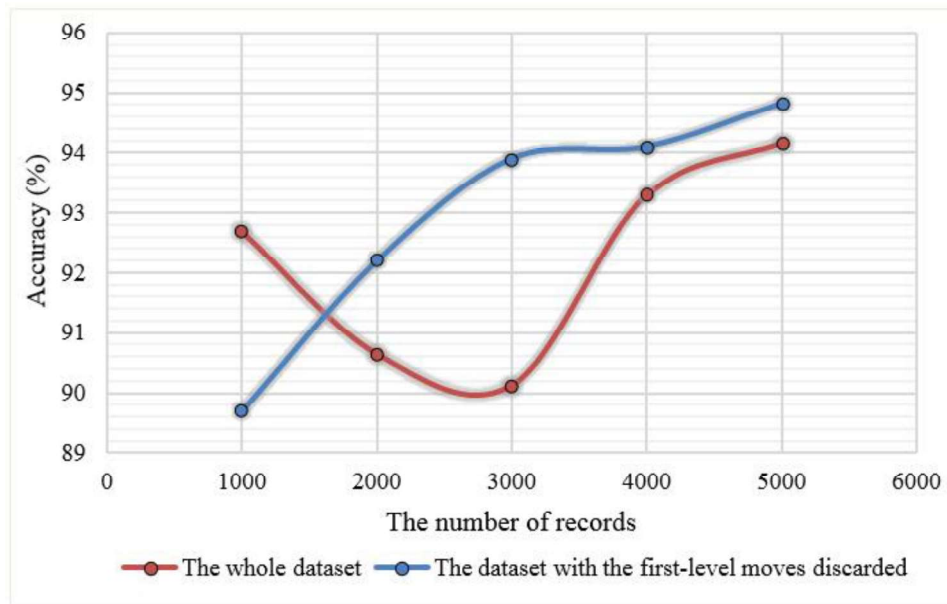


Fig. 1. Accuracy of the Hoeffding tree on the whole dataset and the dataset with the first-level moves discarded in the game “Zagonetke mudrog lisca”

6.2 Applying feature selection methods

Table 1 shows feature selection methods and the respective rank values of features related to student activities in the game “Zagonetke mudrog lisca” with respect to the class of taking additional school classes in Mathematics.

Table 1. Feature ranks in the game „Zagonetke mudrog lisca”

Feature selection methods and feature rankings				
<i>Correlation</i>	<i>Information gain</i>	<i>Chi-square</i>	<i>Gain ratio</i>	<i>Classification attribute evaluator</i>
0.121 Element holding time	0.083 Total score prior to the move	2904.8 Total score prior to the move	0.021 Element holding time	0.061 Element holding time
0.119 Number of times played	0.064 Element holding time	2530.0 Element holding time	0.018 Number of times played	0.037 Session
0.085 Number of remaining moves	0.042 Number of times played	1508.5 Scores gained by the move	0.016 Total score prior to the move	0.028 Total score prior to the move
0.067 Session	0.040 Scores gained by the move	1385.6 Number of times played	0.015 Session	0.026 Number of times played
0.066 Hit	0.033 Session	1316.7 Session	0.015 Scores gained by the move	0.017 Scores gained by the move
0.049 Scores gained by the move	0.006 Number of remaining moves	242.6 Number of remaining moves	0.007 Number of remaining moves	0.000 Element played
0.049 Number of moves at the level	0.005 Element played	201.5 Element played	0.005 Hit	0.000 Total number of moves
0.039 Current level	0.003 Hit	127.7 Hit	0.004 Number of moves at the level	0.000 Number of remaining moves
0.029 Total score prior to the move	0.002 Number of moves at the level	82.3 Number of moves at the level	0.002 Element played	0.000 Number of moves at the level
0.026 Total number of moves	0.001 Current level	38.4 Current level	0.001 Current level	0.000 Hit
0.022 Element played	0.001 Total number of moves	20.9 Total number of moves	0.001 Total number of moves	0.000 Current level

Relative ratios of all features which have values greater than zero according to the classification attribute evaluator, i.e. five features which have the greatest influence on the resulting classification model, are shown in Fig. 2.

In case of a correct answer, the scores gained by the move correlate with the element holding time (the amount of scores gained by answering correctly equals the time left after answering). That feature is therefore discarded in the reduced dataset of the game “Zagonetke mudrog lisca”. Considering the fact that areas of Mathematics taught in different grades differ, student age is also used as a feature.

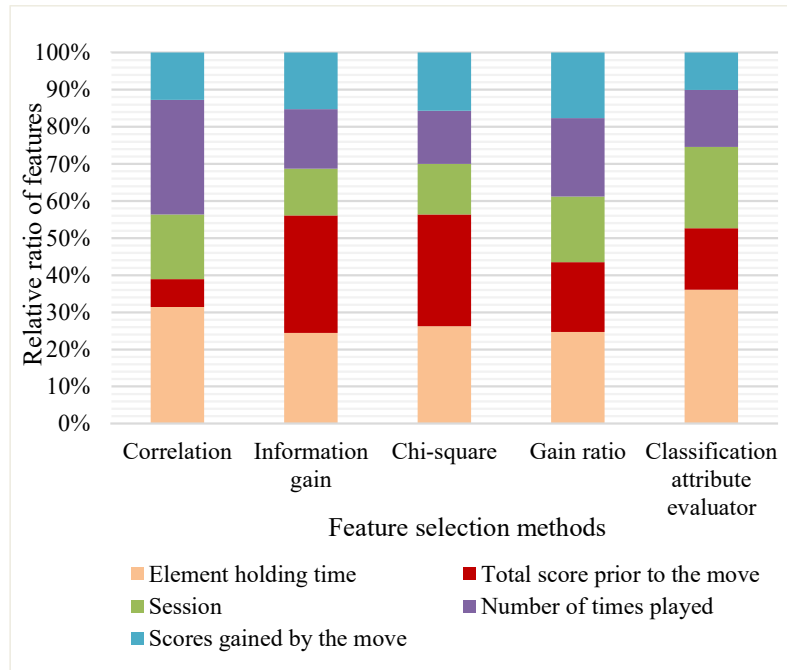


Fig. 2. The relationship between relevant features in the game “Zagonetke mudrog lisca” with respect to the selection method

The reduced set of features, which is compiled out of the ranked feature set, and which contains only 43% of features compared to the original feature set, contains the following features: number of times played, total score prior to the move, element holding time, session, age, and additional Mathematics (class).

6.3 Effects of feature selection on model accuracy

Classification models for the detection of students gifted in Mathematics (students who take additional school classes in Mathematics) are built with stream mining algorithms. Average accuracies of algorithms over the original and reduced feature sets are given in Table 2. Different splitting confidence values (error allowed) are taken into account. Higher coefficient values allow for wider splitting, i.e. trees of greater dimensions. The values of 0.1 and 0.5 are considered as the splitting confidence in calculating the Hoeffding bound. The same values are used for the secondary splitting with the Hoeffding option tree.

Table 2. Effects of feature selection on average accuracies of student classification models in the game “Zagonetke mudrog lisca”

Feature set	Accuracy (%) of algorithms based on Hoeffding trees / splitting confidence					
	<i>Regular</i>		<i>Option</i>		<i>Adaptive</i>	
	0.1	0.5	0.1	0.5	0.1	0.5
Original	93.86	94.30	94.88	95.01	94.91	95.59
Reduced	93.19	93.98	94.82	94.83	96.07	96.12

Since the applied algorithms achieve similar results on the original and the reduced feature set (differences range from -0.67% to 1.16%), the complete model is built with the dataset of smaller dimensionality, i.e. the reduced dataset with the higher coefficient as the splitting confidence. Additionally, experiments show that training and model evaluation last twice shorter on the reduced dataset, i.e. it is more time efficient. The training times range from 0.22:0.45 for the reduced feature set up to 0.34:0.69 seconds for the original feature set, dependent on the algorithm. The highest accuracy model, which is also the most simple regarding the depth of the tree, is the adaptive one. That model represents students who take additional school classes in Mathematics as those who play the game more frequently with less games played in succession (session < 7) compared to those who do not take additional school classes in Mathematics.

By considering only features from the reduced dataset which refer to the learning speed, i.e. time taken to think dependent on the number of attempts, the resulting model describes students who take additional school classes in Mathematics as those who in less attempts spend less time to think. The time taken to think by the students who take additional school classes in Mathematics is over 2.8 seconds only for those students who play less times (number of times played <= 13), while students who do not take additional school classes in Mathematics are slower in solving tasks even afterwards. Taking into consideration that the number of played games refers to the motivation, and time taken to think to the level of acquired knowledge, the obtained model enables preliminary evaluation (additional Mathematics class YES and NO) based on assessing motivation and acquired knowledge.

Therefore, the following hypothesis is accepted: stream mining methods can be used for preliminary student evaluation based on assessing motivation and acquired knowledge in the model which is mobile-friendly and which uses educational computer games for learning Mathematics.

6.4 Concept drift in relation to the size of data window

Algorithm accuracy can also be analysed from the aspect of window sizes, which are determined in the following ways in the research:

- 20 (maximum number of possible moves in one session, which therefore includes data related to at least one student)
- 100 (obtained from the average number of sessions per student multiplied by the average number of moves made)
- 1200 (for the maximum number of student sessions multiplied by the maximum number of moves made)
- 5200 (for the maximum number of games played multiplied by the maximum number of moves made).

The learning curve in stream mining does not increase with the increase in the number of records, but oscillates and depends on the window size. Windows greater in size balance the curve (oscillations in accuracy are reduced).

The Hoeffding adaptive tree exhibits smallest standard deviations compared to the regular and option trees, due to discarding branches which cause a decrease in accuracy and replacing them with new branches with every window shift.

Unlike the incremental models in which the increase in the number of records is followed by the increase in complexity, the models based on the regular and Hoeffding option trees can be also built by including the concept drift. The concept drift detection has similarities with the Hoeffding adaptive tree. While the concept drift discards feature values which cause a decrease in accuracy on a certain number of records, the adaptive algorithm continuously adapts model to new values by making evaluations on each new record in order to achieve higher accuracy, which eventually leads to a complete model. Unlike with the adaptive tree, the utilization of the concept drift on the regular Hoeffding tree, which is not adaptive, leads to discarding a model whenever there is an accuracy shift and building a new model from the current drift up until a new concept drift.

Fig. 3 gives a cumulative display of the depth of the regular Hoeffding tree before and after each concept drift for all the window sizes considered. The red dashed lines are used to label records at which concept drifts are detected for each window size.

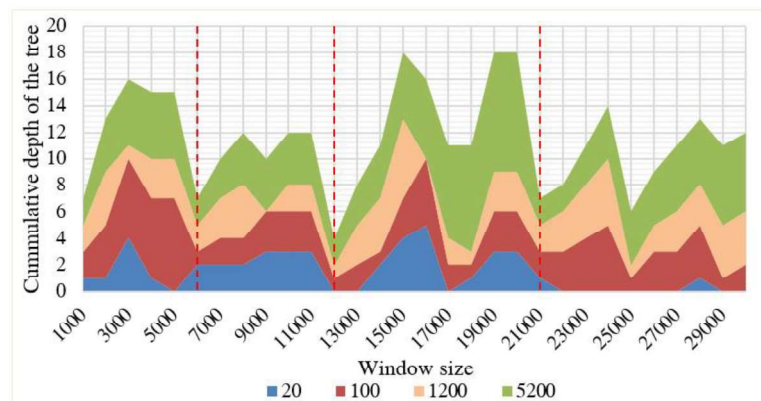


Fig. 3. Cumulative size of the Hoeffding tree before and after each concept drift in the game “Zagonetke mudrog lisca”

The complete (incremental) model has a tree depth of 10, while the model with the same algorithm settings which includes concept drifts gives a detailed tree ($\text{depth}_{\max} = 9$) only with the largest window size. The smallest window size (20) does not result in greater changes due to frequent concept drifts ($\text{depth}_{\max} = 5$).

On the given dataset, the obtained results show that larger window sizes enable training the model which includes concept drifts and which, in terms of complexity and accuracy (93.26%), approaches the complete model (93.98%).

7 Conclusion and Future Work

Models based on stream mining can be simplified by utilizing feature selection methods in order to retain only data relevant for the analysis. The sizes and training times of models obtained within this research are twice as small at no expense in accuracy. The dimensionality reduction, due to smaller number of participating features, leads to a reduction in the model complexity in terms of its interpretability.

This research shows that stream mining can be used for the detection of students gifted in Mathematics in mobile systems which are based on educational computer games.

Since data streams used for training the model often include concept drifts, up-to-date models based solely on valid data can be obtained with the concept drift detection and inspection of different window sizes. Windows larger in size enable training models which include concept drifts, and, at the same time, approach complete models in terms of complexity and average accuracy.

Predictive classification models trained within this research can be used in pre-selecting a large number of students who are gifted in Mathematics, without knowing any other feature except for their age, i.e. students can be classified based solely on their interaction with the m-learning system which uses educational computer games.

In the presented implementation of the learning system, stream mining algorithms based on Hoeffding trees detect students gifted in Mathematics with 93.26% accuracy in case of the model with the concept drift detection, and 93.98% accuracy in case of the complete model. By considering and analysing features which reflect motivation and the level of knowledge acquired, i.e. number of times played and time taken to think, respectively, a model which enables preliminary student evaluation can be built.

The implemented m-learning system supports also other types of games, e.g. those with levels based on random variable values. This is left for our future work.

8 Acknowledgement

This research has been supported under the Grant No. uniri-drustv-18-122 of the University of Rijeka and by the Centre for Artificial Intelligence and Cybersecurity of the University of Rijeka.

9 References

- [1] Suhirman, S.; Zainn, J.M.; Chiroma, H. & Herawan, T. (2014). Data Mining for Education Decision Support: A Review. *International Journal of Emerging Technologies in Learning (iJET)*, 9(6): 4-19. <https://doi.org/10.3991/ijet.v9i6.3950>
- [2] Dirin, A. & Nieminen, M. (2015). mLUX: Usability and User Experience Development Framework for M-Learning. *International Journal of Interactive Mobile Technologies (iJIM)*, 9(3): 37-51. <https://doi.org/10.3991/ijim.v9i3.4446>
- [3] Sobral, S.R. (2020). Mobile Learning in Higher Education: A Bibliometric Review. *International Journal of Interactive Mobile Technologies (iJIM)*, 14(11): 153-170. <https://doi.org/10.3991/ijim.v14i11.13973>
- [4] El Mabrouk, M.; Gaou, S. & Rtili, M.K. (2017). Towards an Intelligent Hybrid Recommendation System for E-Learning Platforms Using Data Mining. *International Journal of Emerging Technologies in Learning (iJET)*, 12(6): 52-76. <https://doi.org/10.3991/ijet.v12i06.6610>
- [5] Pham, H.L. & Cho, S. (2018). Nurturing Mathematical Creativity in Schools. *Turkish Journal of Giftedness and Education*, 8(1): 65-82.
- [6] Han, J.; Kamber, M. & Pei, J. (2012). *Data Mining Concepts and Techniques*, 3rd Edition, Morgan Kaufmann.
- [7] Xu, G.; Zong, Y. & Yang, Z. (2013). *Applied Data Mining*, CRC Press.
- [8] Witten, I.H.; Frank, F.; Hall, M.A. & Pal, C.J. (2016). *Data Mining: Practical Machine Learning Tools and Techniques*, 4th Edition, Morgan Kaufmann.
- [9] Guyon, I. & Elisseeff, A. (2003). An Introduction of Variable and Feature Selection. *Journal of Machine Learning Research*, 3: 1157-1182.
- [10] Bifet, A.; Gavaldà, R.; Holmes, G. & Pfahringer, B. (2018). *Machine Learning for Data Streams: with Practical Examples in MOA*, MIT Press. <https://doi.org/10.7551/mitpress/10654.001.0001>
- [11] Bifet, A.; Holmes, G.; Kirkby, R. & Pfahringer, B. (2011). *Data Stream Mining: A Practical Approach*, Center for Open Software Innovation.
- [12] Romero, C. & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(3): e1355. <https://doi.org/10.1002/widm.1355>
- [13] Chiu, M.-S. (2020). Gender differences in Predicting STEM Choice by Affective States and Behaviors in Online Mathematical Problem Solving: Positive-Affect-to-Success Hypothesis. *Journal of Educational Data Mining*, 12(2): 48-77. <https://doi.org/10.5281/zenodo.4008056>
- [14] Su, Y.; Backlund, P. & Engström, H. (2020). Comprehensive review and classification of game analytics. *Service Oriented Computing and Applications*. <https://doi.org/10.1007/s11761-020-00303-z>
- [15] Alonso-Fernández, C. et al. (2019). Applications of data science to game learning analytics data: A systematic literature review. *Computers & Education*, 141. <https://doi.org/10.1016/j.compedu.2019.103612>
- [16] Makhlof, J., & Mine, T. (2020). Analysis of Click-Stream Data to Predict STEM Careers from Student Usage of an Intelligent Tutoring System. *Journal of Educational Data Mining*, 12(2): 1-18. <https://doi.org/10.5281/zenodo.4008050>
- [17] James, G.; Witten, D.; Hastie, T. & Tibshirani, R. (2013). *An Introduction to Statistical Learning with Applications in R*, Springer Texts in Statistics, Springer. https://doi.org/10.1007/978-1-4614-7138-7_1
- [18] Domingos, P. & Hulten, G. (2000). Mining HighSpeed Data Streams. *Proceedings of the 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 71-80. <https://doi.org/10.1145/347090.347107>

- [19] Frank, E.; Hall, M.A. & Witten, I.H. (2016). The WEKA Workbench, Online Appendix for Data Mining: Practical Machine Learning Tools and Techniques, 4th Edition, Morgan Kaufmann. <https://doi.org/10.1016/b978-0-12-374856-0.00017-1>
- [20] Al Janabi, K.B.S. & Kadhim, R. (2018). Data Reduction Techniques: A Comparative Study for Attribute Selection Methods. *International Journal of Advanced Computer Science and Technology*, 8(1): 1-13.
- [21] Hulten, G.; Spencer, L. & Domingos, P. (2001). Mining Time-Changing Data Streams. Proceedings of the 7th ACM SIGKDD international conference on Knowledge discovery and data mining, 97-106. <https://doi.org/10.1145/502512.502529>
- [22] Pfahringer, B.; Holmes, G. & Kirkby, R. (2007). New Options for Hoeffding Trees. Proceedings of 20th Australian Joint Conference on Artificial Intelligence AI 2007: Advances in Artificial Intelligence (eds: Orgun, M.A. & Thornton, J.), Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 4830: 90-99. https://doi.org/10.1007/978-3-540-76928-6_11
- [23] Bifet, A. & Gavaldà, R. (2007). Learning from Time-Changing Data with Adaptive Windowing. Proceedings of the 2007 SIAM International Conference on Data Mining (eds: Apte, C.; Skillicorn, D.; Liu, B. & Parthasarathy, S.), Minneapolis, Minnesota, USA, 443-448. <https://doi.org/10.1137/1.9781611972771.42>
- [24] Gama, J. et al. (2014). A Survey on Concept Drift Adaptation. *ACM Computing Surveys*, 46(4): 44:1-44:37. <https://doi.org/10.1145/2523813>
- [25] Gonçalves, P.M.Jr.; Carvalho Santos, S.G.T.; Barros, R.S.M. & Vieira, D.C.L. (2014). A Comparative Study on Concept Drift Detectors. *Expert Systems with Applications*, 41(18): 8144-8156. <https://doi.org/10.1016/j.eswa.2014.07.019>
- [26] Gama, J.; Medas, P.; Castillo, G. & Rodrigues, P. (2004). Learning with Drift Detection. Proceedings of Advances in Artificial Intelligence – SBIA 2004, 17th Brazilian Symposium on Artificial Intelligence, Sao Luis, Maranhao, Brazil, 286-295. https://doi.org/10.1007/978-3-540-28645-5_29
- [27] Jurić, P.; Bakarić, M.B. & Matetić, M. (2018). Design and Implementation of Anonymized Social Network-based Mobile Game System for Learning Mathematics. *International Journal of Emerging Technologies in Learning (IJET)*, 13(12): 83-98. <https://doi.org/10.3991/ijet.v13i12.8762>
- [28] Bifet, A.; Holmes, G.; Kirkby, R. & Pfahringer, B. (2010). MOA: Massive Online Analysis. *Journal of Machine Learning Research*, 11: 1601-1604.

10 Authors

Petar Juric earned his PhD degree at the Department of Informatics of the University of Rijeka. He works as Head of ICT Department of Primorje-Gorski Kotar County in Croatia. Email: petar_jur@yahoo.com

Marija Brkic Bakaric is an assistant professor and a senior research associate at the Department of Informatics of the University of Rijeka.

Maja Matetic is a full professor at the Department of Informatics of the University of Rijeka.

Article submitted 2021-02-10. Final acceptance 2021-03-15. Final version published as submitted by the authors.