

Teaching Quality Evaluation and Scheme Prediction Model Based on Improved Decision Tree Algorithm

<https://doi.org/10.3991/ijet.v13i10.9460>

Sujuan Jia^(✉), Yajing Pang
Hebei University of Science and Technology, Shijiazhuang, China
107662109@qq.com

Abstract—Vast data in the higher education system are used to analyse and evaluate the teaching quality, so that the key factors that affect the quality of teaching can be predicted. Besides, the learner's personalized behaviour can also become the data source for teaching result prediction. This paper proposes a decision tree model by taking the teaching quality data and the statistical analysis results of the learner's personalized behaviour as inputs. This model was based on the improved C4.5 decision tree algorithm, which used the FAYYAD boundary point decision theorem for effectively reducing the computation time to the most threshold. In this algorithm, the iterative analysis mechanism was introduced in combination with the data change of the learner's personalized behaviour, so as to dynamically adjust the final teaching evaluation result. Finally, according to the actual statistical data of one academic year, the teaching quality evaluation was effectively completed and the direction of future teaching prediction was proposed.

Keywords—decision tree algorithm; statistical analysis; teaching quality evaluation; teaching direction prediction

1 Introduction

College teaching is different from basic education. It's concentrated in the aspects of professionalism, stage, creativity, openness and autonomy etc. Therefore, it is much more complicated than basic education in terms of ways and means for students to achieve learning and development. Along with the rapid development of computer information technology, an important breakthrough in the current education and teaching reform is to integrate cutting-edge technology and education technology [1]. As a result, more and more researchers are beginning to pay attention to the hybrid teaching model. This teaching model combines the advantages of the traditional teaching model with the advantages of efficient and intelligent cutting-edge technology, and also advocates a personalized customized teaching program for each student's characteristics; besides, it can predict the future teaching direction by detecting the feedback of teaching quality.

In the efficient teaching system, there are a large number of teaching data, such as online course resource downloading, courseware on-demand, student evaluation

information, and course test scores. How to effectively mine these data and use valuable data has become a research hotspot for researchers at home and abroad [2].

Decision tree algorithm is widely used in popular research directions such as machine learning and data mining. It is characterized by the ability to use known probability of occurrence of various situations, construct a decision tree for evaluating the risk of the solution, determine the feasibility of the program, and make the final decision. The decision tree itself is a predictive model that exhibits a mapping between object attributes and object values, so the root node of the decision tree indicates the most informative attribute of all sample objects. The idea of decision tree can also be integrated into the choice of modern educational programs [3]. With the individualized behaviour of students and the evaluation information of educational quality as the object attributes, an optimal teaching plan covering diversified information is predicted.

In this paper, a statistical analysis scheme was proposed for students' learning behaviours, and by introducing the learner's personality characteristics, learning content, learning strategies, relevance analysis, classification of results, etc. into this program, the customized behavioural analysis model for the investigation result was formed according to the iterative analysis. Then, in order to better introduce the concept of hybrid teaching, the statistical analysis of teaching data was conducted by taking curriculum design, teaching content, teaching ability, teaching effect and teaching attitude as introducing parameters, collecting subject and object evaluation information and quantify it, to achieve the final evaluation of teaching quality. Finally, the parameters of the above two statistical schemes were introduced according to the algorithm logic of the decision tree algorithm to achieve the prediction of the final teaching program. In the decision tree algorithm, the FAYYAD boundary point decision theorem was also used to reduce the mining time and improve the algorithm efficiency, and the prediction scheme was modified according to the dynamic update of the statistical data.

2 Research status at home and abroad

Nowadays, the hybrid teaching mode that integrates frontier computer technology into traditional teaching concept has become a hot spot for many researchers and educators. In China, Yu Hongtao, Ren Jun et al. introduced a hybrid teaching training system, which was implemented in the practical teaching scenes of colleges and universities, and the effect verification was carried out at different stages of teaching development. Chen et al. [4] constructed one three-stage hybrid teaching model based on micro-courses for the analysis of mixed teaching practice and effect. Zhang Chenglong, Li et al. [5] in order to study students' learning adaptability, proposed to apply the factors such as learning attitude, self-learning ability, teaching evaluation, curriculum management, teaching feedback, learning environment, etc. to statistically analyse the students' learning status through the model. Based on the above hybrid teaching research, in addition to the statistical model of multi-element comprehensive consideration, this paper also introduces the final model statistical data into the decision tree algorithm, and then apply the improved decision tree algorithm to evaluate the final teaching quality and predict the future teaching direction.

2.1 Research status of students' personalized statistics at home and abroad

The cycle model of learning analysis proposed by George Siemens et al. [6] includes seven parts: collection, storage, data cleansing, data integration, analysis and visualization presentation, and action. The data sources for the model are from learning management system data, sensory data, manual input data, and data markets etc. Those such as intervention, optimization, guidance, and warnings are incorporated into the idea of a linear loop as parameters, enabling all learning links to be integrated closely. Katrien Verbert et al. [7] focused on the data visualization and fluidity, so as to think about the questions raised by users in the process of learning analysis and the evaluation of the validity and relevance of these problems, which ultimately affects users' self-innovation and practice to further optimize the self-learning ability. Tanya Elias et al. [8] proposed an improved model that can analyse the learning process and related stakeholders; this model includes data collection, data processing and data application, the resources of which come from the learning and guidance organization, computers, stakeholders and theory, and can effectively make data selection and mining, information extraction and integration and widely use of the data.

However, the student behaviour analysis model proposed above has the following insufficiency: 1) Be lack of computational model with students as the core analysis object; 2) Only focus on the diversification of big data, but ignoring the hierarchical analysis of the conceptualization of learning behaviour; 3) Make no good use of the analysed data. For the student's personalized statistical analysis proposed in this paper, according to the nature of the decision tree algorithm to be introduced later, the data-mation for student learning behaviour was made as much as possible, so as to effectively pass the collected data into the algorithm for decision making and prediction.

2.2 Statistical analysis of teaching data

W. Greller et al. [9] applied the C4.5 algorithm to the teaching evaluation field through the data collection of the teaching evaluation system; with the redesign of the teaching evaluation system, the data processing speed was improved to optimize the evaluation effect. Suthers [10], during the construction of teaching evaluation model, considered its nonlinear relationship with the final evaluation system in the context of complex factors, and then proposed that the analytic hierarchy process achieved the effect of teaching evaluation to a certain extent. However, for the above two methods of teaching data statistics, the expert judgment method can only be adopted in terms of various secondary indicators such as teacher dressing, teaching guidance and teaching interaction etc., which leads to the subjective randomness of statistical evaluation; thus, there will be certain difference between the final statistical results and the actual value. In this paper, all the above information was quantified to establish the nonlinear mapping. Through data training, it was then introduced into the decision tree model, to make the final evaluation of teaching quality.

2.3 Decision tree algorithm

Decision nodes, branches, and leaves are the basic components of the entire decision tree. The beginning of the decision tree is the root node at the top, and each new decision is a branch or a leaf which represents the object attribute or certain classification result. Through the traversal of the decision tree from top to bottom, the final bottom leaf node is called the process of classification using the decision tree. The C4.5 decision tree algorithm is a widely used decision tree algorithm. Quinlan [3] proposed this algorithm as the improved one of the previous ID3 algorithm in 1993. It continues to adopt information entropy as the criterion for attribute selection, and also increases the features such as discretization of continuous attributes, the processing of unknown attributes and the generation rules. In the C4.5 algorithm, the recursive method is chosen when constructing branches, and the maximum information gain rate is used as the reference standard for the classification attribute. But the traditional C4.5 algorithm hasn't made good clipping of invalid information. The improved decision tree algorithm proposed in this paper introduces the FAYYAD boundary point decision theorem, so as to reduce the mining time and improve the algorithm efficiency.

3 Student's personalized statistical analysis

For the statistical analysis of student personalization, the two questions of how to reflect individual differences in students and establish links with later learning outcomes must be solved firstly. This paper adopts the method of iterative analysis. Iterative statistical analysis was carried out from three aspects: student's personal characteristics, learning content and student learning approach, by using the time as the measuring standard to select the iteration cycle within the specified time. Table 1 lists the attributes of the student's personalized learning behaviour.

According to the learning factors in the above table, this paper combines the actual statistical data of teaching, and selects 15,496 students from the school. Through research analysis and data sorting, the learning model of A student was randomly selected, and then the learning results of this student during the first semester was summarized by iterative scheme [11] and weight analysis. In the individualized iterative analysis (Fig.1), the learning elements in Table 1 were iteratively convolved using the convolution model layer according to the four factors of learning personal characteristics, social network, emotional state and online activities. Then, the final learning results were calculated and derived.

Considering that the student learning is a dynamic behaviour changing with time, this paper proposes an iterative statistical model based on time granularity. Fig.2 shows the schematic diagram of the model. Taking one semester as the final evaluation period of the learning results, the data at three times was selected for collection, and the personal iterative learning result calculation model in Figure 1 was used so that the final learning outcome index was obtained after three-time statistics.

Table 1. Student’s personalized learning behaviour attributes

	Personal characteristics	social networks	Emotional state	Online activity
feature analysis	learning interest	Self network	Will quality	Learning trajectory
	Learning style	social networks	Emotional management	learning evaluation
	learning strategy	Social tendencies	Learning state	Learning platform
content analysis	Learning interest module	Network structure analysis	Autonomy and endurance	Individualized learning activities
	Style of learning	Peer analysis	Participation and slack	Monitoring of learning process
	cognitive strategy	Application media analysis	Learning attention	Department of learning management
behavior analysis	Learning video analysis	Analysis of communication content	Module access analysis	Sequence analysis of learning activities
	Analysis of learning resources	Social network analysis	Access time analysis	Evaluation and analysis of learning

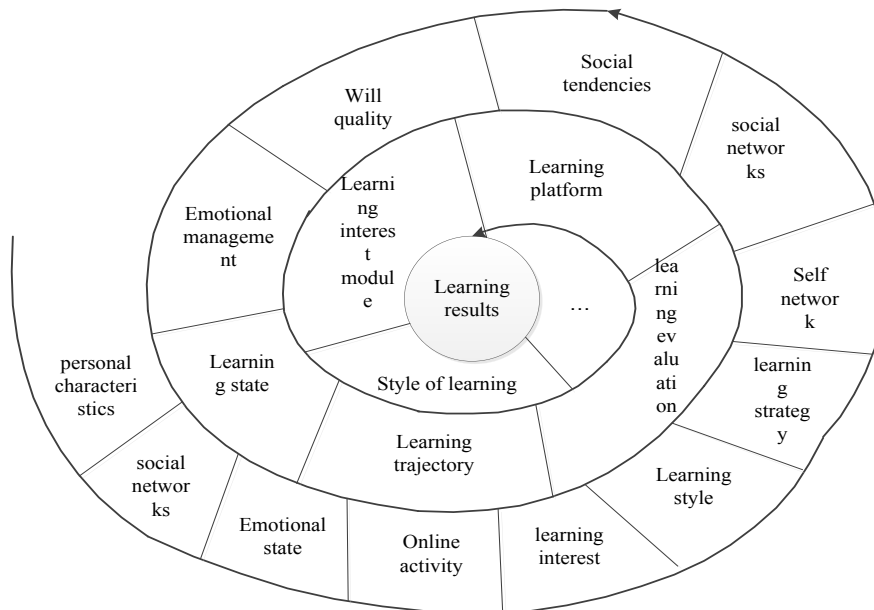


Fig. 1. Iterative analysis of student’s individual learning

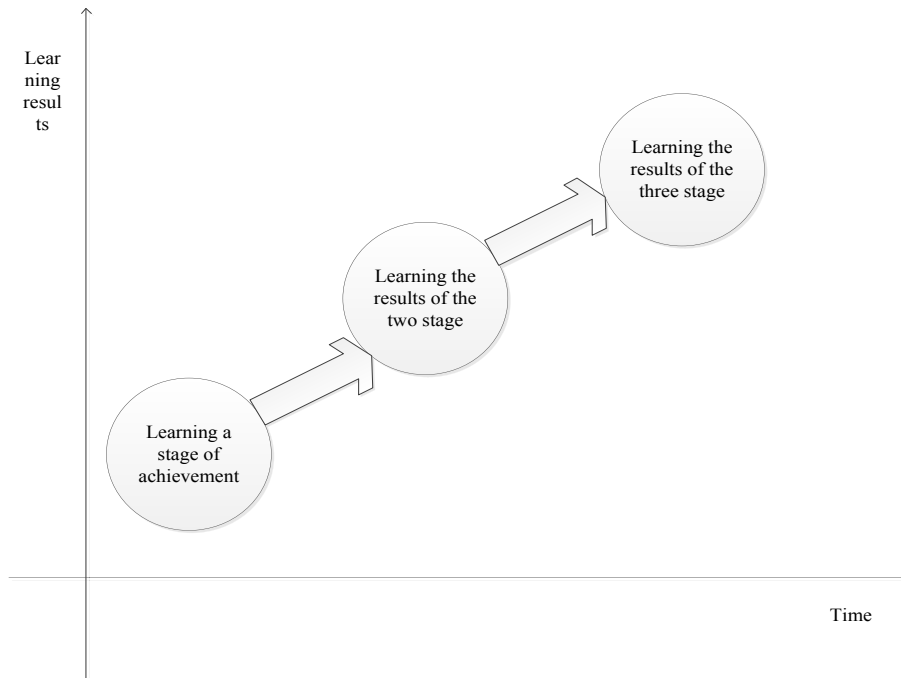


Fig. 2. Iterative analysis for one-semester learning results of students

4 Statistical analysis of teaching data

The previous statistical analysis of teaching data was mostly measured by non-quantitative indicators, resulting in lack of certain objectivity in the teaching evaluation. Based on the research results of Verbert et al. [12], this paper determines a set of teaching evaluation index system and introduces the index evaluation parameters into formula 1.

$$f(x) = 1 / (1 + \lim_{x \rightarrow \infty} m\sigma_x^2) \quad (1)$$

Then, the quantitative benchmark ultimately applicable to the objectives of teaching quality evaluation was derived. Based on the summarization of 10-year data, the teaching measurement parameters such as teaching attitude, teaching content and teaching methods were quantified. The quantified parameters are shown in Table 2.

Table 2. Quantification of teaching indicators

Root index	Sub index	Evaluation parameters	Evaluation goal
Teaching attitude	Teaching guidance x1	5	4.35
	Quantity of teaching information x3	6	6.45
	Teaching interaction x5	4	3.75
	The state of teaching and answering questions x2	7.5	8
	Teaching practice ability x9	8.5	6.55
	Job correction state x4	4.5	8
Content of courses	Focus on the content of the control x3	6	3
	Complex problem expression x5	9	5.65
	Curriculum system arrangement x7	4	7.75
	bilingual education x8	2	8
	Teaching plan x5	8	4.35
Teaching conditions	Investment in Teaching x2	4	6
	practice base x5	7.5	4.85
	State of literature x6	6	5
Teaching management	Teaching evaluation and examination x8	4	65
	Teaching system x3	8	8
	Feedback quality standard x5	9.5	7.85
	The structure of teaching team x9	5	4
	Research findings x5	5.5	6
teaching staff	Proportion of Senior Professional Title Teachers x6	6	5.75
	Proportion of doctorate teachers x8	4	4.35
	Proportion of young teachers x2	7.5	6
	Teacher planning training x 5	8	8.45

5 Improved decision tree algorithm

In Schmidt’s view [13], the teaching quality directly affects the student’s learning state, and the learning state of the student during certain learning time will also reflect the more detailed evaluation of the teaching quality. Based on this point of view, this paper conducts hierarchical research on the individualized learning state of students and the fine-grained evaluation of teaching quality, using the decision tree algorithm. Teaching evaluation is the root node of the decision tree, because the teaching evaluation includes both direct evaluation factors such as teaching status and indirect factors such as the individualized learning status of students. Teaching evaluation as the root node of the decision tree can comprehensively include the maximum attributes of each sample set. In the improved decision tree algorithm of this paper, the student learning state and the teaching target were used as the sample set denoted by S, the statistics of teaching state was used as the sample set denoted by A, the current candidate attribute set was represented by T, and the final generation result, i.e. the leaf node, was indicated

by G. In order to introduce the concept of boundary clipping, this paper proposes the relevant definitions and theorems:

Definition 1: Select I as a value in the attribute T, if and only if there are two values $e_1, e_2 \subseteq T$ in the instance sequence sorted by T, and have different attributes, so that $T(e_1) < T(e_3) < T(e_2)$, and there are no other values satisfying $e_3 \subseteq T$, to ensure $T(e_1) < T(e_3) < T(e_2)$. $T(e_3)$ represents the attribute value of T in the instance e_3 .

Theorem 1: The information entropy of the training sample set T is expressed as $H(T)$, according to the formula (1)

$$H(T) = \sum_{i=1}^n p(x_i) \mathcal{I}(x_i) = -\sum_{i=1}^n p(x_i) \sum_{i=1}^n p(x_i^2) \quad (2)$$

Where $x_i \in T$, $p(x_i) \in S$. If x_i is minimized, x_i is one certain threshold point. For continuous T, it is possible to minimize its information entropy and always be able to stay between two different adjacent values.

According to the above definitions and theorems, this paper gives the code logic of the tree structure and core algorithm for the improved decision tree.

```

typedef struct tree
{
    int pro // is the leaf node (0) or internal node (1)
    int value // If it is a leaf node, it means the specific classification result. If it is an
    internal node, it means a certain feature.
    int parentpro // If the node has a parent node, the value represents the specific prop-
    erty value of the feature represented by the parent node
    tree child[] // represents the subtree array of the node
    Attribute Tested ; // Node attribute
    Set * Subset ; // subset of discrete properties
    Tree * Branch ; // branch of the node
}
    
```

It can be seen from the above procedure that the improved decision tree algorithm is a doubly linked list structure. This structure can solve the omission issue of attribute value update in the decision tree algorithm during the incremental self-learning. Child[] and Branch can dynamically improve the knowledge update of leaf nodes. In the calculation process, the process adds a property array to the leaf node for the minimum entropy selection, that is, it only allocates memory space at this time. The improved decision tree algorithm code is as follows:

```

Input: Enter a code instance, with the teaching information as the total set;
function tree = maketree(featurelabels,trainfeatures,targets,epsino)// Use the in-
stance properties to create a basic decision tree structure, and create basic decision tree
for teaching information
tree=struct('pro',0,'value',-1,'child',[],'parentpro',-1); // Add a node class in the deci-
sion tree structure to complete filling in the student's personalized information in the
node
Child[n,m] = size(trainfeatures);
cn = unique(targets);
    
```



```

l=length(cn); // calculate the total depth of the decision tree
// The following code implements the information entropy of the update node and
all nodes on its parent node, leaf node decision information, etc.
if l==1
tree.pro=0;
tree.value = cn; //where,

$$cn = H(T) = \sum_{i=1}^n p(x_i) \log(x_i) = -\sum_{i=1}^n p(x_i) \sum_{i=1}^n p(x_i^2)$$

tree.child=[];
return;
end
// On the decision node tree, dynamically update the attribute information, and com-
bined with the definition 1 and the theorem 1 clipping, re-track back to find the optimal
test attribute set segmentation threshold, to output the final prediction rating infor-
mation, and end the learning.
if n==0% if feature number equals 0
H = hist(targets, length(cn));
tree[ma, largest] = max(H);
tree.pro=0;
tree.value=cn(largest);
tree.child=[];
return
end

```

6 Verification examples and results analysis

In this paper, 15,496 students' learning statistical information and 10-year teaching evaluation data were selected, and then one-semester teaching quality evaluation and program prediction was carried out on the basis of the improved C4.5 decision tree algorithm. Besides, the prediction scheme generated by the decision tree algorithm was carried out for one-semester teaching research. By comparing the final two groups of teaching quality and student learning achievement data, it can be seen that the teaching quality and student learning results have been significantly improved and optimized.

According to the decision tree algorithm, the information entropy was first completed for all attributes in Experiment 1. The results are shown in Table 3.

The data in Table 3 were analysed by bringing them into the algorithm logic. This paper summarizes the status information in Fig.3.

It can be seen from Fig.3 that the final evaluation of the quality of teaching is carried out for the selected student samples, in which teaching guidance, key content control and bilingual teaching are at lower standards; job correction and document status are moderate evaluation; teachers with doctor degree, teaching interactions and research results are at the higher evaluations. According to the evaluation criteria, resource scheduling was carried out for low-standard evaluation, and program improvement of low-standard indicators was strengthened. This paper selects the one-year program

adjustment period and takes the final student evaluation results as the reference indicator. Fig.4 shows the significant improvement of the evaluation results.

Table 3. Sample attribute information entropy

attribute	Information entropy
Personal characteristics	0.63464
Social networks	0.33425
Emotional state	0.53489
Online activity	0.18273
Teaching attitude	0.37181
Content of courses	0.48592
Teaching conditions	0.64728
Teaching management	0.58291
Teaching staff	0.46294

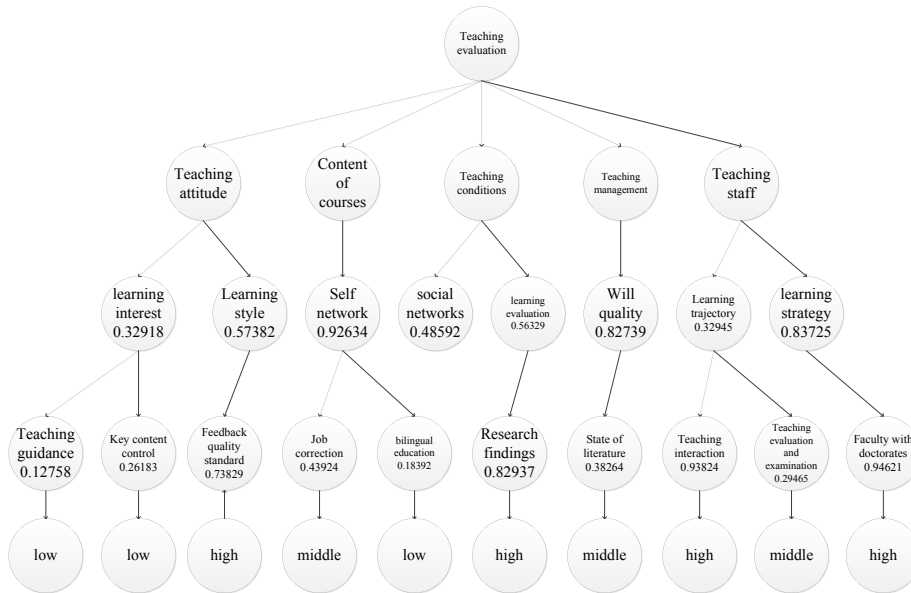


Fig. 3. State diagram of teaching quality evaluation decision tree

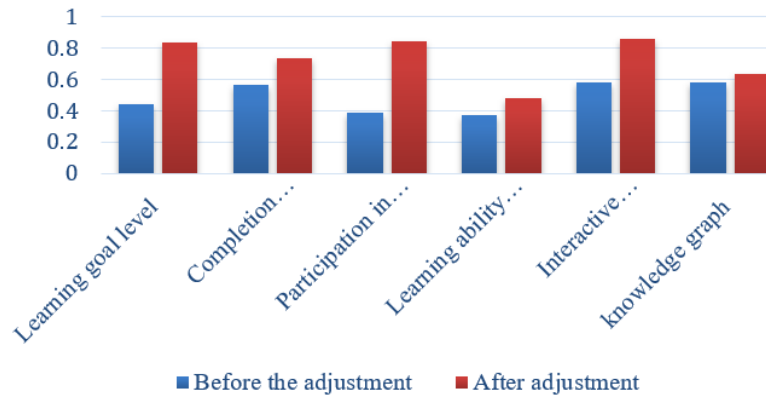


Fig. 4. Comparison of learning results before and after the adjustment of teaching program

7 Conclusions

Based on the personalized learning behaviour of students, this paper summarizes the factors affecting the learning results, and quantifies 10-year teaching data into teaching indicators through iterative statistics. Then, taking the individualized learning behaviour and teaching evaluation index as the sample set of the decision tree, the decision tree algorithm was used to calculate the information entropy equivalence of each attribute in the sample set. At the same time, effective boundary clipping rule was adopted when calling the decision tree algorithm, which effectively improves the computational efficiency of the algorithm. Finally, the teaching evaluation criteria were calculated and the prediction plan was proposed. The predicted educational scheme was put into practice for one school year to compare the learning results of the students. It is found that the adjusted teaching program can effectively improve the students' learning results, thus verifying the effectiveness of the prediction scheme.

8 References

- [1] Jayaprakash, S.M., Moody, E.W., Lauría, E.J.M., Regan, J.R., Baron, J.D. (2014). Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative. *Journal of Learning Analytics*, 1(1): 6-47. <https://doi.org/10.18608/jla.2014.11.3>
- [2] Romero, C., Ventura, S., Pechenizkiy, M., Baker, R.S. (2010). *Handbook of educational data mining*. Data Mining and Knowledge Discovery Series. Boca Raton, FL: Chapman and Hall/CRC Press.
- [3] Quinlan, J.R. C4.5 (1993) Programs for machine learning. San Ma. Teo: Morgan Kaufmann Publishers Inc: 17-42.
- [4] Chen, J.F., Hsieh, H.N., Do, Q.H. (2015): Evaluating teaching performance based on fuzzy AHP and comprehensive evaluation approach. *Applied Soft Computing Journal*, 28: 100-108. <https://doi.org/10.1016/j.asoc.2014.11.050>
- [5] Zhang, C.L., Li, L.J., Li, J.F. (2017). Student Learning Adaptability in MOOC-based Blended Teaching in Y University. *China Educational Technology*, 36(4): 60-65.

- [6] Siemens, G. (2013). Learning Analytics: The Emergence of a Discipline. *American Behavioral Scientist*, 57(10): 1380-1400. <https://doi.org/10.1177/0002764213498851>
- [7] Verbert, K., Duval, E., Klerkx, J., Govaerts, S., Santos, J.L. (2013). Learning Analytics Dashboard Applications. *American Behavioral Scientist*, 57(10): 1500-1509. <https://doi.org/10.1177/0002764213479363>
- [8] Elias, T. (2011). Learning Analytics: The Definitions, the Processes, and the Potential [EB/OL]. <http://learninganalytics.net/LearningAnalyticsDefinitionsProcessesPotential.pdf>, 2011-12-26.
- [9] Greller, W., Drachsler, H. (2012) Translating learning into numbers: A generic framework for learning analytics. *Educational Technology & Society*, 13(3):42-57, 2012.
- [10] Suthers, D.D., Vatrappu, R., Medina, R., Joseph, S., Dwyer, N. (2008). Beyond threaded discussion: Representational guidance in asynchronous collaborative learning environments. *Computers & Education*, 50(4): 1103-1127. <https://doi.org/10.1016/j.compedu.2006.10.007>
- [11] Kalakoski, V., Ratilainen, H., Drupsteen, L. (2015). Enhancing learning at work. how to combine theoretical and data-driven approaches, and individual, behavioural, and organizational levels of data? In Proc. of the European Symposium on Artificial Neural Networks.
- [12] Verbert, K., Manouselis, N., Drachsler, H., Duval, E. (2012). Dataset-Driven Research to Support Learning and Knowledge Analytics. *Journal of Educational Technology & Society*, 15(3): 133-148.
- [13] Schmidt, H.G., Van Der Molen, H.T., Te Winkel, W.W.R., Wijnen, W.H.F.W. (2009). Constructivist, Problem-Based Learning does work: A meta-analysis of curricular comparisons involving a single medical school. *Educational Psychologist*, 44(4): 227-249. <https://doi.org/10.1080/00461520903213592>

9 Authors

Sujuan Jia received her B.SC degree in 2005 from computer science and technology at Hebei Normal University. M.SC degree in 2008 from computer science and technology in the Hebei University of Technology. Now she is a teacher at the Department of Teaching Affairs, Hebei University of Science and Technology, Shijiazhuang 050018, China. Her main research interests include Computer Application and Management and teaching management.

Yajing Pang received her B.SC degree in 2003 from computer science and technology at Hebei University of Science and Technology. M.SC degree in 2007 from computer Application and technology in the Hebei University of Science and Technology. Now she is a teacher at the Department of Development Planning, Hebei University of Science and Technology, Shijiazhuang 050018, China. Her main research interests include: Computer Application and Management Computer Intelligence, Computer Measurement and Control and Computer Data Mining.

Article submitted 12 August 2018. Resubmitted 29 August 2018. Final acceptance 18 September 2018. Final version published as submitted by the authors.