Decision Trees-based Anomaly Detection in Computer Assessment Results

V L Panasov¹ and N M Nechitaylo²
¹ Rostov State Transport University (RSTU), 2, Rostovskogo Strelkovogo Polka Narodnogo Opolcheniya sq., Rostov-on-Don, 344038, Russian Federation
² Federal State Institution of Higher Education "Russian University of Transport" (RUT), 9, bld. 9, Obraztsova st., Moscow, 127994, Russian Federation

E-mail: vitekp@donpac.ru

Abstract. A survey of the current software in the area of computer assessments of students at a university is done. This work shows ways to improve quality of computer assessments and educational management. It suggests options to build a computer system for detecting anomalous assessment results of individual students and entire disciplines. This information could then be used to amend lectures and instructional material or management and better intercommunication to students. This work suggests feature sets for the analysis of student assessment results and demonstrates their availability in the current university software. After surveying known artificial intelligence systems this work proves the choice of decision trees for this problem. It suggests methods and algorithms to improve positive prediction rate of decision trees based on ideas of bootstrap method. A software tool was developed that implements the suggested algorithms.

1. Introduction

Computer assessments are widely used at various stages of education. At Rostov State Transport University (RSTU) computer assessments have been used for several years.

Since computer assessments have been used for a long time and due to high number of students at our university we have been able to collect enough statistics of student performance and utilize modern methods of mathematical statistics and machine property analysis in order to improve quality of education. Improvement can be achieved in the technology of computer assessments as well as in finding best approaches and practices of intercommunication with students.

One possible approach to improve quality of computer assessments is by means of machine analysis using decision trees [1]. In this approach decision trees help to detect semantic correlations between various tests in an assessment and compute weights to amend final score to be more fair.

2. Materials and methods

Currently student test answers are analyzed at the department of student performance monitoring at RSTU in order to:

- Detect excessively sophisticated tests or tests with poorly composed question or answer options that produce anomalously low results scored by students;
Detect excessively trivial tests that produce anomalously high results scored by students.

Detection of such anomalous tests is done by means of empirical borderlines. An expert teacher would then amend these tests in order to eliminate such anomalies [2].

Such detection technique have been computerized due to its algorithmic simplicity, but it is currently far from being optimal. It does not take into account many factors affecting the number of correct answers given by a student. Also it provides no possible explanations of what caused such anomalies.

The term of anomaly in computer science was introduced some time ago. Anomaly is defined as a point in a dataset that significantly differs from the remaining data [3]. Such points are sometimes called as surprises, outliers, exceptions or garbage depending on application [4-5].

When detected an anomaly is subject to a more detailed analysis. Additional data can be helpful that could explain why a point was marked anomalous. Applied to computer assessments possible causes of anomalous points include:

- Poor preparation of a student in this course;
- Poor preparation of a student in another preceding affiliated course;
- Missing classes by a student or a whole group because of fair or unfair reasons;
- Flaws in instructional material or lectures in some topics;
- Insufficient number of tests in the topic;
- Excessively sophisticated test question or answer options (distractors);
- Excessively trivial test question or excessively weak distractors;
- Insufficient adjustment of lectures to a given profession;
- Low correlation of lectures to preceding affiliated courses;
- Prompts to a student during testing.

A computer system by means of analysis of computer assessment results in various courses and some other features (profession, gender etc.) can improve positive rate of anomaly detection as well as give keys to understanding its causes.

Generally such intelligent analysis can be done in two aspects:

- Detect anomalous assessment results of an individual student;
- Detect anomalous assessment results of an entire discipline.

In the former a machine analysis system would look at features affecting results of an individual student. For example, if a student had bad results in mathematics it would be reasonable to have bad results in a topic involving many complex calculations. On the contrary if a student had good results in engineering graphics then bad results in drawing of train time charts should be considered an anomaly. Also there may be some correlation between excessively poor or excessively good results in a topic and gender or profession.

In the latter a machine analysis system would look at some peculiarities of a given profession. For example, bad results in understanding computer algorithms would be expected in the profession of Customer Service and Tourism. However in the profession of Railway Dataware it should be considered an anomaly. Overall assessment results of a course can be affected by the amount of class hours (which can deviate in various educational standards) or the amount of missed classes because of holidays for example. Bad results in a topic can be caused by bad results in an affiliated preceding discipline. Grade should also be taken into account because usually higher grades show greater eagerness to education.

All these correlations could be discovered by the proposed intelligent analysis system of assessment results which could make use of the positive answer rate in tests as well as some additional
features. It should be noted that current software of a university such as RSTU could provide some additional features such as:

- Which topic or competence a test belongs to? In the Assessment Fund (AF) developed by the Department of Student Performance Monitoring (DSPM) of RSTU each test is categorized by one or several competences learned by a profession;
- Which disciplines and topics teach each competence? This information is available in the Sub-system of Discipline Curriculum (SoDC) developed and used in RSTU. It allows to put together various disciplines and their topics that teach the same competence;
- How preceding and following disciplines are interconnected? This information is also available in the SoDC. It allows to take into account the effect of preceding disciplines and student performance in them on the student performance in a given discipline;
- What scores had a student in preceding disciplines learned at school? Links to preceding school disciplines are also available in SoDC. Scores gained in the Unified State Examination (USE) are also available in the corresponding software;
- How many class hours, lectures, lab instructions in electronic form are assigned to a given discipline? This information is also available in SoDC;
- The amount of missed classes by individual students and entire groups (because of holidays) are also available in the Sub-system of Class Attendance and Learning Tasks Fulfillment (SoCAaLTF) available in the Electronic Educational Environment at RSTU;
- Gender, age, profession, grade, number of tests in a topic of a discipline and other information are also available in the corresponding software of the personal department or DSPM.

There are many approaches to intelligent data analysis, but most of them assume a preliminary stage of knowledge acquisition. At the knowledge acquisition stage (or just training stage) the class of a training object is known. At the working stage the aim of an intelligent system is to detect a proper class.

In the domain of Artificial Intelligence such problems are qualified as Classification. After an intelligent system has been trained it can analyze features of a new object (in our case: student or group) and classify it i.e. predict its class (in our case: predict the percentage of correct answers in a competence) [6-9].

To know how anomalous an object is (in our case: assessment results of a student or a group) one should compute its mahalanobis distance to $k$-nearest neighbors. Applied to our problem this algorithm has the following shortcomings:

- Given a high number of features one should use additional weights to level features of different nature;
- No interpretation could be deduced to understand its decision, because prediction logic is hidden behind formulas and weights.

To overcome these shortcomings we suggest to use algorithms based on Decision Trees [1]. The classification rule of such intelligent systems is a hierarchy where each feature is tested in a node. At training stage an algorithm of building decision tree would look for a feature that would better split the training set of object cases in two or more parts.

3. Results
Focusing on the task of the analysis of an individual student assessment results given the information available in the software of a modern university we suggest a dataset with the following features in (table 1).

We suggest to analyze assessment results separately for each competence: $C_1, C_2, ..., C_N$, where $N$ is the number of competences taught by a given discipline. $PD_1, PD_2, ..., PD_M$ denote preceding
affiliated disciplines teaching a given competence where $M$ is the number of preceding disciplines. $SD_1$, $SD_2$, ..., $SD_K$ denote preceding affiliated school disciplines where $K$ is the number of school disciplines.

**Table 1.** Feature set for the analysis of $C_1$ competence for a student.

| Feature no. | Feature description |
|-------------|---------------------|
| 1           | Percentage of correct answers in the competence $C_1$ of the discipline $PD_1$ |
| 2           | Percentage of correct answers in the competence $C_1$ of the discipline $PD_1$ |
| ...         | ...                 |
| $M$         | Percentage of correct answers in the competence $C_1$ of the discipline $PD_M$ |
| $M+1$       | USE score in the discipline $SD_1$ |
| $M+2$       | USE score in the discipline $SD_2$ |
| ...         | ...                 |
| $M+K$       | USE score in the discipline $SD_K$ |
| $M+K+1$     | Number of tests in the competence $C_1$ |
| $M+K+2$     | Percentage of missed classes by a student |
| $M+K+3$     | Gender               |
| $M+K+4$     | Age                  |
| $M+K+5$     | Grade                |
| $M+K+6$     | Profession            |

The percentage of positive answers given by an individual student in a given competence of a given discipline $C_1$ can be marked as a class. Initially one should build a knowledge base using recorded assessment results of former students. It would then allow to predict assessment results of new students. If predicted results greatly differ from the actual ones then it should be considered an anomaly that should be looked at by the teacher of the discipline. Depending on a chosen prediction model additional information explaining this anomaly could be provided by an intelligent system to the discipline teacher that would help him. In this case possible causes of an anomaly include: missed classes, prompts during exam, personal circumstances.

Also note that separate knowledge base should be built for each competence of a given discipline.

Focusing the second task – analysis of the assessment results of an entire discipline we suggest to use average positive answer rate of each group thus looking at a group as a whole. Such approach allows to smooth individual peculiarities of a student. Averaging assessment results in a group would allow to shift focus to the used educational methodology based on peculiarities and demands of a given profession, evaluation of interconnections to other disciplines and quality of AF.

Applied to the prediction of assessment results of an individual student a decision tree would look as in (figure 1).

As seen in (figure 1) one can easily interpret the decision tree because it is understandable for a human even though it is built by a machine using one of the known algorithms: ID3, C4.5, CART etc. [10]. After prediction has been made it is possible to observe which way – nodes the prediction was made through. It would help a teacher to analyze assessment results of an individual student or a group.

We should note that modern algorithms for building decision trees use methods of Pruning decision trees that increase positive prediction rate. Naturally it is based on the idea of cutting-off cases that don’t feet into the entire case set. It is the same idea which anomaly detection is based on. Obviously known algorithms of building and pruning decision trees could be naturally used in the problem of anomaly detection.
In our work we suggest to use a new improved algorithms for building decision trees and ensembles of decision trees. The methods described below have been implemented and tested. Their positive effect have been proved on test statistical databases.

4. Discussion

Building decision trees in all algorithms is recursive. Selection of a feature $A$ and its outcomes $v_u$ for each node $t$ of a decision tree $L$ is done based on ambiguity measure – entropy of the training set assigned to the current node which is the diversity of classes fallen into the training sub-set:

$$Info(T) = \sum_{j=1}^{k} \frac{freq(C_j, T)}{|T|} \cdot \log_2 \frac{freq(C_j, T)}{|T|}$$  \hspace{1cm} (1)$$

where $T$ – training set; $C_j$ – a class in this set; $freq(C_j, T)$ – the number of cases in the sub-set $T$ marked by class $C_j$.

Decision trees built with this method would allow for easy detection of anomalies. Anomalies are usually placed in the leaf nodes at small depths $z$.

One of the shortcomings of such entropy measures is the increase of error rate at building lower nodes because of the decrease of the training set size. In order to overcome this problem it is suggested to use the ideas of the bootstrap method [11].

One of the suggested applications of the bootstrap method is the algorithm of recursive building of decision trees $P.Tree$ which is described verbally below:

1. Make $B$ bootstrap selections $\{T_{t*i}\}$ out of the current case set under the following conditions:
   - The size of the bootstrap selection $T_{t*i}$ must be lower than the original $T_t$;
   - All cases in the original selection have equal probability of being selected in the new bootstrap selection of $1/N$, where $N$ – the number of cases in the current set $T_t$;
   - Original cases could be doubled in the new bootstrap selection.

2. Build normal tree $L^*_i$ for each selection using criteria (1).

3. Evaluate the error rate of prediction of the tree on the independent test selection $T_{t^{*}i}^{(*)} = (T_t - T_{t*i})$ using formula (2):

$$R_t(L^*_t) = \frac{1}{N_{t^{*}i}} \cdot \sum_{j} \chi(d_{L^*_t}(x_j) \neq C_j)$$  \hspace{1cm} (2)$$

where $d_{L^*_t}(x)$ is the prediction rule of the sub-tree $L^*_i$ which root is in the node $t^{*}_i$. 

**Figure 1.** Decision Tree fragment for the prediction of assessment results in the competence $C_1$. 

Figure of a decision tree fragment for the prediction of assessment results in the competence $C_1$. The tree includes nodes for gender, percentage in mathematics, boolean algebra, engineering drawing, and profession A. The tree structure shows the decision-making process for predicting assessment results.
4. Out of all candidate sub-trees $L_i^*$ choose the one $L_i'^*$ that has the lowest prediction error rate.
5. Prune all branches of the tree $L_i'^*$ leaving just the root node $t^*$.
6. Rebuild all branches of the node $t^*$ with this algorithm recursively to all underlying nodes.

Decision trees built by the algorithms described could be pruned by the following method: Compute bootstrap evaluation of the error prediction rate of each node. If total error rate of all outcomes of a node on a test selection $T_{t^*}$ is greater than the error rate of the node itself, then prune all outcomes and mark this node as a leaf with the most frequent class found in the selection $T_{t^*}$. 

5. Conclusion
The suggested algorithm has the following advantages over the know ones:

- It detects poor decision trees with high prediction error rate;
- It estimates new tree coherence to the previously accumulated collection and detects bad combinations of decision trees;
- It automatically detects optimal size of collection;
- It optimizes using a pseudo-independent prediction error estimation which incorporates some information about the decision tree generalizing ability.

In the end we should state that current level of data awareness of educational process at higher educational institutions allows for a deeper analysis of student performance, detect anomalous performance deviations of individual students and entire disciplines automatically, and help in finding ways to eliminate them. It seems promising to use methods of artificial intelligence based on decision trees and bootstrap ideas for this task.

References
[1] Quinlan J R 1999 Simplifying decision trees. *International Journal of Human-Computer Studies* 51(2) 497-510
[2] Agrawal S and Agrawal J 2015 Survey on anomaly detection using data mining techniques. *Procedia Computer Science* 60 708-713
[3] Kovalev S M, Sukhanov A V, Sukhanova M V and Sokolov S V 2018 Adaptive Approach for Anomaly Detection in Temporal Data Based on Immune Double-Plasticity Principle. *Advances in Intelligent Systems and Computing* 679 234-243
[4] Butakova M A, Chernov A V, Kovalev S M, Sukhanov A V and Zajaczek S 2020 Network traffic anomaly detection in railway intelligent control systems using nonlinear dynamics approach. *Lecture Notes in Electrical Engineering* 554 475-483
[5] Butakova M A, Chernov A V, Shevchuk P S and Vereskun V D 2017 Complex event processing for network anomaly detection in digital railway communication services. *Materials of the 25th Telecommunications Forum (TELFOR) Belgrade* 1-4
[6] Kostoglotov A A, Pugachev I V, Penkov A S and Lazarenko S V 2020 Synthesis of Multimodel Algorithms for Intelligent Estimation of Motion Parameters Under Conditions of Uncertainty Using Condition of Generalized Power Function Maximum and Fuzzy Logic. *Advances in Intelligent Systems and Computing* 1156 541-547
[7] Lyabakh N N and Gibner Y M 2020 Analytical Analogies Calculus. *Advances in Intelligent Systems and Computing* 1156 597-606
[8] Lyabakh N N, Saryan A S, Dergacheva I V, Nebaba A N, Lindenbaum T M and Panasov V L 2020 Development of the pattern recognition theory for solving the tasks of object classification and yard processes. *Advances in Intelligent Systems and Computing* 1226 59-68
[9] Savvas I K, Chernov A V, Butakova M A and Chaikalis C 2018 Increasing the Quality and Performance of N-Dimensional Point Anomaly Detection in Traffic Using PCA and DBSCAN. *Materials of the 26th Telecommunications Forum (TELFOR) Belgrade* 1-4
[10] Wanga X, Zhoua C and Xu X 2019 Application of C4.5 decision tree for scholarship evaluations. *Procedia Computer Science* **151** 179-184

[11] Barsacchi M, Bechini A and Marcelloni F 2020 An analysis of boosted ensembles of binary fuzzy decision trees. *Procedia Expert Systems with Applications* **154** 113436