Journal of Engineering and Applied Sciences 12 (Special Issue 10): 8780-8784, 2017
ISSN: 1816-949X

Intelligent Methods for Evaluation of Student Written Works

Uzunova-Dimitrova Boryana, Zhelezov Stanimir and Paraskevov Hristo
Department of Computer Systems and Technologies, Faculty of Mathematics and Computer Science,
Konstantin Preslavsky University of Shumen, Shumen, Bulgaria

Abstract: A formalized approach is proposed for assessing written works of students through unconventional application of the dichotomous Rasch Model with the use of fuzzy assessments of the criteria. The model allows program realization and can be embedded into e-Learning platforms.

Key words: e-Learning in higher education, evaluation of written works of students, one-parameter Rasch Model, Rasch measurement, formalization of the evaluation process, students

INTRODUCTION

To obtain an objective assessment of learner's knowledge various measuring tools are used: tests, essays, reviews, term papers, course projects, reports for the internship period, solving cases, tasks from the subject area, diploma theses and others.

Item Response Theory (IRT) develops formal models of assessment for test control of student's knowledge some of which have been software realized (Lili, 2016). Winsteps (dichotomous Rasch Model, partial credit model, rating scale model) bilog, multilog, parscale (birnbaum models and their extensions for polytomic tasks); facets (many-facet Rasch Model); RUMM (Rasch Unidimensional Measurement Models) conquest (multidimensional models), etc.

Test systems allow rapid and objective assessment of student's knowledge in a particular subject area but they also have major disadvantages for example, the possibility of random answer selection, the complexity of developing high-quality tests, the lack of specialists in the field of test theory, the impossibility to assess abstract, creative knowledge and more.

Striving to avoid the disadvantages of test systems increasingly directs the attention of experts to the use of written works as a tool to measure the knowledge or skills that students acquire in certain subject areas (Veldkamp and Davier, 2016). Development of essays, term papers, review papers, course projects, etc., develops creative thinking, ability for structuring, generalization and analysis of information, skills to use acquired knowledge in practice and much more.

Creative works have greater informativeness than formalized tests. Their assessment is an intellectual process its automation is connected with the realization of high technologies for text analysis. Models use artificial intelligence methods which allow an adequate modelling of the decision-making process of assessing learner's knowledge. For formalization of knowledge graph models, semantic networks, neural networks, fuzzy sets, fuzzy logic and other formalisms for modelling under conditions of uncertainty are used. Those works still have theoretical and research nature but will undoubtedly lead to obtaining substantial results and software methods for their realization.

MATERIALS AND METHODS

Research objective: The student's works such as essays, papers, term papers, reviews, course projects, course works and others are presentations in free form which are assessed by a lecturer. Here, we exclude the assessment of diploma works and computer programs which have different models.

The problem of assessing is non-formalized and is solved under conditions of uncertainty of various types determined by criteria and linguistic rules for decision making.

The objective of this research is to propose and give reasons for a formalized approach to assessing student's written works by means of unconventional application of the one-parameter Rasch Model. The model allows software realization and can be used in all forms of education but it is especially useful in distance learning where there is no direct contact between the assessor and the assessed.

Applicability of the Rasch Model: The possibility of using the one-parameter Rash Model (Bond and Fox, 2015; Maslak *et al*., 2005) to formalize the assessment of

Corresponding Author: Uzunova-Dimitrova Boryana, Department of Computer Systems and Technologies, Faculty of Mathematics and Computer Science, Konstantin Preslavsky University of Shumen, Shumen, Bulgaria

written works of students derives from the following considerations which are not inconsistent with the actual practice and the nature of the thinking process of the lecturer.

Learner's knowledge and difficulty of written work are assessment parameters which allow an objective assessment, regardless of the assessor and the used measuring tool. Written work is a tool to measure student's knowledge in a particular subject area. Subject to assessment is the latent variable <quality of the student's written work> that does not allow direct measurement but can be measured objectively.

The lecturer is able to assess objectively the quality of the work, moreover, higher quality work will be rated higher. Scores of different, equally competent lecturers on one and the same work can differ from one another insignificantly because of unavoidable measurement errors but not due to the differences in competencies.

Thorough analysis of lecturer's thinking in the decision-making process for evaluation formation shows that lecturers have a subjective inclination to encourage knowledgeable students and negative expectations of unknowing ones which affects the assessment. Likewise, because of non-linearity, the Rasch Model favours knowledgeable students and is unfriendly towards unknowing ones.

Formalization of the problem of assessing written works: The latent variable <quality of the student's written work> does not allow direct measurement but can be decomposed into indicator variables (or criteria) that can be evaluated more easily. We introduce the following designations:

$S = \{s_1, s_2, \ldots, s_m)\ )$ is a finite, discrete set of written works which are subject to evaluation; D: a finite, discrete set of diagnoses; $C = \{c_1, c_2, \ldots, c_n\}$ is a finite, discrete set of given criteria (indicator variables); $A = \|a_{ij}\|$, $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n$ is a matrix containing the results of the evaluation; $a_{ij} \in L$ is the rating of the work s_i on the c_j criterion; L is a discrete scale of assessment's values.

Using the linguistic model of decision making for evaluation of written works, the greatest semantic proximity can be achieved if it is regarded as a task for a diagnosis of a type (Eq. 1):

$$<S, C, L, A, D> \quad (1)$$

with the following formulation: For any written work $s_i \in S$ determine the diagnosis $d \in D$ on the basis of results A from the estimations based on criteria C given in the scale L. Formally, this means to find an injective image (Eq. 2):

$$\Omega: S \rightarrow A \rightarrow D \quad (2)$$

of quantitatively measured opinion of the quality of the written works A in the set of diagnoses D. The image $S \rightarrow A$ is obtained as a result of evaluation of works $s_i \in S$ on the basis of given criteria $c_j \in C$ by the lecturer. In order to obtain $A \rightarrow D$ we apply the Rasch Model for dichotomous data (Maslak *et al*., 2005).

Unidimensional Rasch Model: Paraphrasing the Rasch Model (Bond and Fox, 2015; Maslak *et al*., 2005), we can assume that the probability P for a student with proficiency S to develop a work of high-quality with difficulty T is defined by the following equation:

$$P(S, T) = \frac{S}{S+T}$$

The function P(S, T) is called a function of success, and the variables S and T are latent variables. If we introduce the following designations:

$$A = LN(S),\ S = \exp(A)$$
$$B = LN(T),\ T = \exp(B)$$

for P, we get:

$$P(S, T) = \frac{\exp(A)}{\exp(A)+\exp(B)} = \frac{1}{1+\exp(B-A)}$$

The resulting ratio is called the basic logistic model of Rasch. From the last formula it is seen that the probability of success depends only on the difference B-A which is why the Rasch Model is one-parameter.

Selection of a scale: The Rasch Model is applied with the dichotomical scale {yes, no} = {0, 1} which is a little informative. We select the scale L = {bad, good, excellent} = {1, 0.5, 1}. It is applicable with minor modification to the Rasch Model. This scale is convenient for lecturers and gives greater opportunity for the formation of unambiguous assessment compared to the scales in which the number of terms k>3.

Course project assessment criteria: The choice of criteria for assessing the quality of the project has significant relevance to the objectivity of the evaluation. Criteria are described linguistically.

The lecturer follows the non-formalized rules that lead to ambiguous assessment due to the subjectivity of experts and linguistic uncertainty of the terms used. Evaluation is formed as a result of operations that are difficult to formalize.

For the purpose of the experiment there were selected 11 criteria for assessing the quality of the scientific review papers. They were selected on the basis of the analysis of internet sources and consultations with lecturers.

Criteria:

- Compliance with the requirements of the work layout
- Clearly defined objective
- Compliance of the content with the objective of the work
- An overview of world achievements on the topic
- Relevance of the overview to the topic
- Adequate use of scientific terminology
- Logically organized text
- Completeness of the topic
- The deep analysis of the problem
- The credibility of conclusions
- Independence of the work

The last criterion cannot be assessed objectively by the lecturer. For this purpose an anti-plagiarism system is used (Advego Plagiatus, http://advego.ru/plagiatus/), which checks the text on the uniqueness and gives the final result as a percentage. In the presence of borrowings greater than a preselected threshold of acceptability (for example, 40%), the work is disqualified and a bad grade is given.

Research design: The assessment is performed in the following sequence:

1. The teacher evaluates the work on pre-selected criteria in the scale L. As a result of the expert evaluation, we obtain the A matrix with dimensions m×n, where m is the number of verified works, n is the number of the criteria.

2. We calculate the primary score b_i, i = 1, 2, ..., m of the students as the sum of the ratings on the lines:

$$b_i = \sum_{j=1}^{n} a_{ij}$$

3. We calculate the parameters p_i , i = 1, 2, ..., m by the equation:

$$p_i = \frac{b_i}{n}$$

We ignore the extremal scores as follows: if $b_i = 0$, we make $p_i = \varepsilon$; if b_i is equal to the maximum score, we make $p_i = 1-\varepsilon$ where ε is quite a small number for example, $\varepsilon = 0.001$.

4. The initial approximation of the evaluation of the i-s work is calculated by the equation:

$$A_i = LN\left(\frac{p_i}{1-p_i}\right), i = 1, 2, ..., m$$

5. We calculate the primary score c_j, j = 1, 2, ..., m of the criteria, obtained by adding the ratings on the columns:

$$c_j = \sum_{i=1}^{m} a_{ij}$$

6. We calculate the parameters p_j, j = 1, 2, ..., n by the equation:

$$p_j = \frac{c_j}{m}$$

Similarly to 3 if $c_j = 0$, make $p_j = \varepsilon$; if c_j is equal to the maximum score, we make $p_j = 1-\varepsilon$.

7. We calculate the initial values of the criteria difficulty by the equation:

$$B_j = LN\left(\frac{1-p_j}{p_j}\right), j = 1, 2, ..., n$$

The final assessment of the written work is obtained by linear transformation $A_i \in [\min(A_i), \max(A_i)]$, i = 1, 2, ..., n of the scale {2, 3, 4, 5, 6} = {bad, fair, good, very good, excellent}.

RESULTS AND DISCUSSION

To test the possibility of using the described method an experiment was carried out. The developed review papers of students from Shumen University, studying the discipline "Programming for office systems" were assessed with the use of the Rasch Model and by proof test on the same material studied. Part of all 790 conducted experiments are presented in Table 1.

Analysing the results, we can draw the following conclusions: the proposed model realizes a multi-criteria approach for measuring the quality of written works of students.

Convolution of linguistically defined vector score in scalar using the Rasch Model helps the lecturers in their research. The criteria should allow an objective assessment. Therefore, they must be accurate, clear, unambiguous without use of logical connections (and or not).

Table 1: Experimental results of evalution of review papers of students

Citeria/ Students	1	2	3	4	5	6	7	8	9	10	11%	11 {0; 1}	Primary ball bi	Pi	Ai = Ln (Pii))1. in logits	Evalution of the proposed method	Evalution test
S1	1	1	0	1	1	1	1	1	1	1	16	1	10	0.999	4.509	6	6
S2	0.5	1	0.5	1	1	1	1	1	1	1	22	1	10	0.999	4.509	6	6
S3	1	1	0	1	0	1	0	1	1	0	30	1	7	0.727	0.981	5	6
S4	0	0.5	0	0	0.5	0	1	0	1	0	27	1	4	0.455	-0.182	4	4
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	
S72	0	0	0	0	0	0	1	0	1	0	74	0	2	0.182	-1.504	2	3
S73	0	0	0.5	0	0	0	1	0.5	1	1	53	0	4	0.364	-0.560	3	3
S74	1	0	0	0	1	0	0	0	0	1	35	1	4	0.455	-0.182	4	4
S75	1	1	0.5	0	0	0.5	0	1	0	0	79	0	4	0.364	-0.560	3	4
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	
S111	0	1	1	1	0	1	0	0	1	0	20	1	6	0.636	0.560	5	6
S112	1	1	0.5	0	1	0.5	1	0	1	0	36	1	7	0.727	0.981	5	5
S113	1	0	0	0	0.5	1	0	0	0	0.5	93	0	3	0.273	-0.981	3	3
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	
S510	0	0	0.5	0	0	0.5	0	0	0	0	98	0	1	0.091	-2.303	2	2
S511	0	0	0	1	1	0	0.5	0	0	0.5	62	0	3	0.273	9.813	3	
S512	0.5	1	1	0.5	1	0	1	1	1	1	41	0	8	0.727	0.981	5	6
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	
S787	1	0	0.5	0.5	0	0.5	0	0	0.5	0	37	1	4	0.455	-0.182	4	4
S788	0.5	1	0	1	0.5	0	1	1	0	1	38	1	7	0.727	0.981	5	5
S789	0	0	0.5	0.5	0.5	1	0	0	0.5	1	11	1	5	0.545	0.182	4	4
S790	1	0.5	0	0	0	1	0.5	0	0	0	45	0	3	0.273	-0.981	3	3
Primary ball Ci	9.5	9	5.5	7.5	8	9	9	6.5		10							
Pi	0.317	0.300	0.183	0.250	0.267	0.300	0.300	0.217	0.333	0.267		0.5556					
Bi = Ln ((1-pi) pi) in logits	0.769	0.847	1.494	1.099	1.012	0.847	0.847	1.285	0.693	1.012		0.2231					

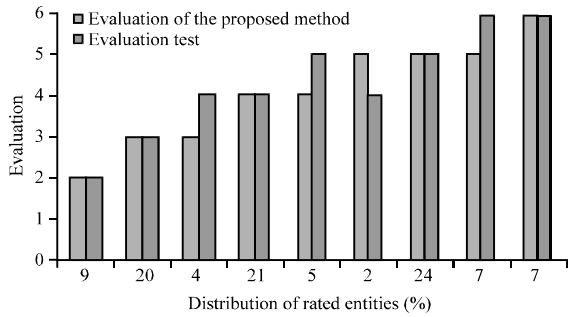

Fig. 1: Comparing the results of the assignments to the proposed model with those of the test

From the presented experimental results in Fig. 1. It is seen that there is over 80% duplication of the results. The differences in the results of the evaluation of course projects compared to the newly proposed modified Rasch Model and the test scores are insignificant. Differences in assessments occur mainly because of the 11th criteria which is Plagiarism. This means that the proposed method of assessing knowledge is objective and reliable. The criteria are of unequal weight. Artificial determining of weights (for example when processing an expert opinion) can distort the information on the proficiency of the students.

In repeated real experiments, statistical data can be used to "fit" ("Within population item-fit") the input parameters (criteria) to the requirements of the model. The simple criteria (which are met by all) and the complex ones (which none of the works meets) can be excluded.

Plagiarism compromises the assessment. All written works must be checked for uniqueness by means of an anti-plagiarism system. Non-unique works (for example, plagiarism >40%), get a poor mark. If the percentage of plagiarism is low or missing, it is impossible to formally determine the authorship of the texts. Authorship of the student's work can only be ascertained in a direct contact between the assessor and the assessed.

CONCLUSION

A software implementation of the proposed method is realized as part of the developed "Web-based intelligent control system of student's knowledge". Using sensor networks has its place in the evaluation of knowledge in e-Learning (Wei *et al*., 2012).

The research is being carried out in the following areas. Development of a formalized method for determining the criteria. Development of methods for assessing student's written works. In addition to the Rasch Model, a fuzzy model based on the theory of fuzzy sets is being developed. Selection of scales and methods to correctly convert assessments from one scale to another.

REFERENCES

Bond, T. and C.M. Fox, 2015. Applying the Rasch Model: Fundamental Measurement in the Human Sciences. 3rd Edn., Routledge, Abingdon, England, ISBN:978-1-315-81469-8, Pages: 383.

Lili, T., 2016. System quality assessment of knowledge of students distance learning. Altai State Technical University, Barnaul, Russia. https://translate.google.com.pk/translate?hl=en&sl=ru&u=http://elib.altstu.ru/elib/books/Files/pa2010_2/pdf/196tomasheva.pdf&prev=search.

Maslak, A.A., G. Karabatsos, T.S. Anisimova and S.A. Osipov, 2005. Measuring and comparing higher education quality between countries worldwide. J. Appl. Meas., 6: 432-442.

Veldkamp, B. and V. Davier, 2016. Methodology of Educational Measurement and Assessment. Springer, Berlin, Germany,.

Wei, W., X.L. Yang, P.Y. Shen and B. Zhou, 2012. Holes detection in anisotropic sensornets: Topological methods. Int. J. Distrib. Sens. Netw., 2012: 1-9.