STUDYING TESTING EFFECTIVENESS DYNAMICS IN TRAINING OPERATORS OF COMPLEX TECHNICAL SYSTEMS

P.N. Dumin
PhD (Physics and Mathematics), Head of Scientific Laboratory,
Moscow State University of Psychology and Education,
Computer Science Faculty,
29 Sretenka Street, 127051 Moscow, Russia

L.S. Kuravsky
DSc (Engineering), Professor, Dean of Computer Science Faculty,
Moscow State University of Psychology and Education,
Computer Science Faculty,
29 Sretenka Street, 127051 Moscow, Russia

ABSTRACT
Under consideration is an approach for studying the process of evaluating the intellectual abilities of operators of complex technical systems. The procedure involves both analyzing dynamics of test score accumulation and classifying a test subject to one of the diagnosed classes. The results obtained can be used to correct test contents and optimize testing the operator intellectual abilities. An example of practical application for the proposed approach is given.

Keywords: operators of complex technical systems, testing procedure, skill class recognition, intelligence testing


1. INTRODUCTION
Computer testing is widely used in medicine, psychology and education to diagnose, determine the level of competence and suitability of the subjects to perform certain functions, including the teaching quality monitoring. Testing quality and reliability of its results largely depend on the testing technology, which in recent decades have become the subject of active scientific research. Currently, the dominant approach in this area is the use of technologies set on the so-called Item Response Theory (IRT) based on latent structure analysis [1, 9].
basic concept of this approach, proposed by G. Rasch in 1960, assumes that the probability of a correct response to the test item task is determined by the difference ability level (or knowledge) and the test difficulty. In practice, depending on the conditions of the applied problem, there are other, more complex models based on this concept [1].

Using the IRT technology leads to the a number of problems, among which are ignoring the fact that the result of the test can, in general, significantly change over time due to fatigue of the subjects and other factors, taking different values during the test session; the necessity of performing a sufficiently large number of jobs test items to provide estimates with acceptable accuracy, etc. There are approaches to avoid such problems [3, 4]. Such approaches based on the use of learning structures in the form of the Markov models with discrete and continuous time. Its features, which provide advantages over similar methods of testing, are:

- the identification and use of the temporal dynamics of changes of the ability to solve the tasks of the test, test items during the construction of estimations;
- the possibility of taking time spent on the solving of test tasks into account, during the construction of estimations;
- the possibility of researching the temporal dynamics of knowledge or abilities skills both in discrete and in continuous time scale;
- smaller, in comparison with other approaches, quantity of tasks that must be presented to a subject to provide estimations of the knowledge or abilities skills with given precision, which speeds up the testing process;
- obtaining of a probability distribution of possible test results as the final result;
- the advanced technique of identification of the model parameters.

There are systems for decision support based on the foregoing testing techniques, which are using Markov and factor models [3-5], which accelerate the testing process, optimizing the presentation of test items. Diagnostic conclusions are based on the probabilistic estimations of subject belonging to different categories, specified in the testing process. These tools give additional information to the practicing specialist for analysis and, in the case of Markov models, provide recommendations on the choice of the next test, which has, compared with other, most discriminating fineness ability for a given subject.

Training operators of complex technical systems is a complex multi-stage procedure involving the measurement of many different personality characteristics [2]. Measurements may relate to the psychophysiological characteristics of the operator during the management of complex systems, stress tolerance during critical situations, etc. One of the stages of the study is the intellectual abilities assessment of operators using standardized test procedures. Given the complex nature of the study of the level of preparedness of operators, the proposed approach allows us to reduce the time spent on measuring intellectual abilities. An analysis of the results of the dynamics of the test and the test points accumulation allows you to adjust the test materials and choose among them only the part that provides the researcher with information about the actual level of abilities of the subject. The proposed approach allows us to better assess the contents of test materials and optimize them to more quickly obtain the necessary data on the level of intellectual abilities of the operator. At the same time, the proposed procedure makes it possible to quantify the degree to which a subject belongs to one of the diagnosed classes. A continuous assessment of this kind allows you to control the reliability degree of the decision. The algorithm is simple to use and does not require the configuration of additional parameters inherent in the complex diagnostic models used in computer testing.
2. ABILITY LEVEL DIAGNOSTICS: PRACTICAL APPLICATION

The testing procedure was carried out in two stages. At the first stage, the subjects were asked to complete 12 tasks from the Raven progressive matrix test. The second stage consisted of presenting the tasks of the first three subtests of the Amthauer test. The Amthauer test typically consists of 9 subtests aimed at assessment of several subjects’ abilities. The brief description of subtests is given below.

- “Logical selection”. Subtest aimed at assessment inductive reasoning and linguistic skills of subject. Main goal of this subtest is complete the sentence with one of the proposed words.
- The second subtest aimed at study subject’s abilities to abstraction and manipulation of verbal concepts. The main goal is to find an extra word among the five proposed options.
- The third subtest is called “Analogy” and aimed at assessment of combinatorial abilities of subject. Each test item consists of three words, first and second words have semantic connection to each other. Subject should choose one word among the five proposed, which will have the same semantic connection with the third word as the first two.
- The fourth subtest is called “Classification”. The subtest is aimed to assessment subject ability level to formulate a point of view. The main goal is to combine two words with a general representation.
- The fifth subtest called "Computational task" and aimed at assessment of subject abilities to solve arithmetic tasks.
- The subtest called "Series of numbers" aimed at study of subject abilities to numbers manipulation. The subject should find a pattern in the number series and continue it.
- "Figures choosing". The subtest is aimed to assessment of combinatorial abilities of subject. Subjects are offered cards with geometric figures partitioned into several parts. The subject needs to find a card with a figure that matches the divided into parts figure.
- The eighth subtest is similar to the seventh, but the basic figure is a cube.
- The ninth subtest contains tasks to assess the ability to focus attention and remember learned information. The subject must remember several certain words and find them among others suggested in the task. Words intended for memorization are listed in the table by certain groups.

The experimental treatment was subjected to the results of adapting versions of these tests [9]. The results of the testing procedure were presented as matrix of responses. Further analysis was carried out using the algorithm proposed in [5]. The main steps of approach are presented in Figure 1.
Prior to the following computations, time series representing the histories of test item performance are brought to the scales, in which each element is cumulative sum of difficulties, which are measured in logits for successfully performed items (from the beginning to a current time point), with scale zeros being moved to minimal difficulty values. Distance matrices can be calculated using two metrics: Euclidean metric and Kohonen metric [6]. Bar charts Figure 2 (a, b) represents the distance distribution between current subject and other subjects participating in the experiment. As we can see the distances between the subjects have a certain pattern that allows you to conclude that there is a structure in the data under study. In the right part of figure 1 (a, b) the pairwise distance matrices is represented in color and correspond to the two used metrics: Euclidean metric (2.a) and Kohonen metric (2.b). Both distance matrices show a similar structure for the relative position of subjects in space.

Figure 2 Distance matrices calculated using the Euclidean metric (a) and Kohonen metric (b). Bar chart shows the distance distribution from current subject to all others.
Figure 3 shows graphs of the dynamics of scoring for specific subjects from two different groups: a low ability group and a high ability group. As can be seen, a group of strong subjects more evenly scores test points during the testing process for all three tests (green, white and red lines). Subjects from the low ability group during the test is in a static state longer, corresponding to a constant number of test points. Data on the dynamics of the set of test scores can serve for a more detailed study of both the abilities of the subjects and the effectiveness of specific test procedures used in measuring the abilities of subjects.

Figure 3 The dynamics of scoring for two different groups: high ability group (a, b) and low ability group (c, d).
It is possible to correct the test tasks composition or change their sequence by examining
time periods in which the dynamics of points accumulation by the subjects had certain
regularities. This approach can be applied in the development of test tasks and in adaptive
testing.

Using mutual distances matrices as initial data for the multidimensional scaling procedure,
one can visualize the results in a space of acceptable dimension. This embedding can be used
for cluster analysis, which allows to classify the test subject based not only on the number of
raw test results, but also considering the dynamics of accumulation of test points over time.
The Figure 4, 5 shows clustering and classification results for several subjects. For each case,
the probability of the subject belonging to a certain group of subjects was calculated for two
cases: Euclidean (Figure 4) and Kohonen metrics (Figure 5).

**Figure 4** Classification results for two subjects from different groups (a – low ability group, b – high
ability group). Marker with a blue border corresponds to a current classified item. Red squares
correspond to cluster centroids. The left scatterplot contains only the objects of the target cluster. Bar
charts show probability estimates of a selected subject belonging to different clusters.
Figure 5 Classification results for two subjects from different groups (a – low ability group, b – high ability group). Marker with a blue border corresponds to a current classified item. Red squares correspond to cluster centroids. The left scatterplot contains only the objects of the target cluster. Bar charts show probability estimates of a selected subject belonging to different clusters.

The proposed approach can be applied in adaptive testing processes. During testing, it is possible to set a threshold for evaluating the probability that the subject belongs to a cluster that corresponds to the ability level. When the threshold is reached, the testing procedure stops. This approach allows you to reduce the testing time and get the result of the specified reliability.
3. PRINCIPAL RESULTS AND CONCLUSIONS

- The proposed approach allows to solve diagnostic problems with limited observation results.
- Probability estimates of belonging to different clusters can act as criteria in the process of adaptive testing procedures.
- The proposed algorithm does not require additional parameter settings and can be used on the raw data.
- Analysis of the dynamics of accumulation of test points allows one to examine the test tasks composition: to detect time periods in which the dynamics of accumulation of points had certain regularities.

ACKNOWLEDGEMENTS

This work has been supported by the Russian Foundation for Basic Research (Project 17-29-07034).

REFERENCES


http://www.iaeme.com/IJARET/index.asp
editor@iaeme.com