10th IMEKO TC7 International Symposium
June 30-July 2, 2004, Saint-Petersburg, Russia

ESTIMATION OF KNOWLEDGE QUALITI ON A BASIS OF PROBABILITY MONITORING AND DIAGNOSING METHODS

Boris Lihtsinder[1],Lyudmila Ivanova[2]

[1]Volga State Academy of Telecommunications and Informatics, Samara, Russia
[2]Academy of Telecommunications and Informatics, Samara, Russia

Abstract – At present, the regular validation of the personnel is being practiced by various enterprises. It's made to determine the level of competency of the staff as well as their professional correspondence. At once automated methods of students' knowledge control are rapidly integrated in different educational institutions.

The knowledge control is a process taken up to determine the knowledge degree. The state of probationers, which shows their knowledge value, corresponds each level. These levels are discreted. This way the check out may be viewed as the process of diagnosing and the states characterizing the level of the man who is checked out will be viewed as diagnosing states.

Keywords: Estimation of knowledge, diagnosing, probability.

1. DIAGNOSTIC CONDITION AND INSPECTIONS

The object, which is put to the proof can be found out in the state e_i with some certainty value degree $P(e_i)$ and the higher is the said degree, the more effective should be the proof process. The proof includes fulfilling some sequence of tasks by a probationer. Each task contains one question or more. And each question should be answered only in one way while choosing from an amount of prepared ones. This way job processing may be reviewed as a range of successive verifications, while each check-up Π_k matches a question and each result of this check-up q_{jk} matches an answer.

If the probationer really possesses the level of knowledge estimated by state e_i then the a posteriori chances' value $P(e_i)$ with the correctly questions should increase after each cheek-up, while the chances' value corresponding other states should decrease.

Each check-up Π_k is uniquely characterized by a matrix of base chances with elements $P(q_{jk}/e_i)$. The matrix element determines a chance of drawing a reply q_{jk} providing that the probationer really possesses the level of knowledge corresponding the mark e_i. The values of specified base chances are being determined on the ground of an expert system of marks and they compound a base of an information system.

At the defined a priori chances' value $P(e_i)$, which took priority of regular check-up, each verification possesses some informativity.

The more informative check-up is chosen as a regular from an amount of verifications. A posteriori chances' value $P(e_i)$ are specified according to the results of the quiz and these results are taken as benchmark data at the next check-up. The process is repeated until the value of one chance will achieve the set level.

As the state e_i is a full system of events the other values of a posteriori chances should decrease. The current state of the probationer means the one which accords the maximum of a posteriori chances' values.

2. ESTIMATION OF KNOWLEDGE LEVEL WITH THE USE OF BINARY TESTS

The binary test is understood as a test question, which can be answered by the examinee in only two alternative ways, one of which is correct (true), and the other is erroneous (false).

Each testing subject (examinee) is characterized by some quantity of knowledge, which provides a particular level of knowledge - *z*. Each binary test is characterized by the complexity - σ. During the tests, the examinee can give a right answer with probability *p* and the erroneous answer - with probability *q*.

$$p = 1 - q . \qquad (1)$$

The probability of the erroneous answer *q* depends on a level of knowledge *z*. The higher is the level of knowledge, the less is the probability of erroneous answer. The probability of erroneous answer also depends on a complexity of a testing problem. The higher is the complexity of a problem, the higher is the probability of the erroneous answer at the given level of knowledge. The value $x = z/\sigma$ can be characterized as the relative knowledge level of the examinee (relating the complexity of a testing problem).

"A zero level" ($z_0 = 0$) should be selected from the amount of knowledge levels z. It is characterized by the complete lack of any examinee's knowledge, in relation to a testing problem.

The binary test can be offered the examinee in a form of a choice from several devices from which he is obliged to select the devices relevant to a right answer. At the complete lack of knowledge, the examinee selects devices a casual fashion. In that case the probability of a right answer $p(z_0) = p(0)$ depends on the parity of numbers in elements relevant to exact or erroneous answers.

2.1. Equal probability (elementary) tests

If the deriving of correct or erroneous answers at the complete lack of knowledge (random choice) is of equality probability, such binary test is called elementary.

It is obvious, that probabilities p (the exact choice) and q (an erroneous choice) are equal to each other this way $q(z_0) = p(z_0) - 0{,}5$.

The level *z* is determined by the amount of basic knowledge (Δz) of the tested problem. Each additional elementary knowledge increase the probability of a right answer on a value $\Delta p(z)$. Thus, function

$$f(z) = \frac{dp(z)}{dz} \qquad (2)$$

can be considered as a probability's distribution denseness of the right answer deriving, depending on an increment of knowledge level, while integrated function

$$p(z) = \int_{-\infty}^{z} f(x)dx \qquad (3)$$

is a probability of a right answer at the knowledge level relevant to a level z.

It is necessary to explain the sense of negative meanings of a knowledge level z. To the moment of the testing start, the examinee could have some knowledge misinforming him in giving a right way of solving the very task. For example, as a result of consultations with people, wishing to provide him the information which is certainly wrong, the examinee gains some volume of certainly false knowledge turning a level of his knowledge to the negative range of values z. Deriving of some additional elementary knowledge, which is correct, reduces disinformation and increases the level of positive knowledge. The probability of a right answer thus is augmented. Thus, the probability's distribution denseness $f(z)$ is always a positive value.

2.2. Probability's distribution for the binary elementary test

We shall accept a hypothesis that the elementary probability law of right answers submits to the normal law. The fact, that value z is an amount of basic knowledge about a subject and depends on a set of incidentals, allows the accepting of such hypothesis

$$f(z,\sigma)\frac{1}{\sigma\sqrt{2\pi}} å^{-\frac{1}{2}\left(\frac{z}{\sigma}\right)^2}, \qquad (4)$$

where σ is a root mean square deflection, defining the complexity of some binary test.

Here is the integral function of this allocation

$$F(z,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{z/\sigma} e^{-\frac{1}{2}\left(\frac{z}{\sigma}\right)^2} dz. \qquad (5)$$

If we produce replacement of a variable $x = z/\sigma$, we shall receive

$$F(z,\sigma)\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z/\sigma} e^{-\frac{õ^2}{2}} dõ. \qquad (6)$$

At the given change of variables, the parameter σ characterizes complexity of the elementary test, and the variable *x* characterizes the relative knowledge level (relatively to complexity σ of the elementary test).

The value of the given function characterize probabilities of a right answer in the elementary test.

$$F(z,\sigma) = p(z,\sigma). \qquad (7)$$

Research of function $p(z,\sigma)$ confirms validity of the made choice of elementary probability law. Really, at lack of the prior knowledge (a zero level, $x = 0$) probability of the right answer is $p(0) = 0{,}5$, and at a high level of knowledge ($x > 3$), the probability of a right answer comes nearer to unity.

Function $F(z,\sigma)$ can be noted as

$$F(z,\sigma)\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} e^{-\frac{õ^2}{2}} dõ + \frac{1}{\sqrt{2\pi}} \int_{0}^{z/\sigma} e^{-\frac{õ^2}{2}} dõ, \qquad (8)$$

where the first integral is equal 0,5, and the second is the standard function of Laplace.

$$\Phi\left(\frac{z}{\sigma}\right)\frac{1}{\sqrt{2\pi}} \int_{0}^{z/\sigma} e^{-\frac{õ^2}{2}} dõ. \qquad (9)$$

Thus,

$$p(z,\sigma) = 0{,}5 + \Phi\left(\frac{z}{\sigma}\right). \qquad (10)$$

Integral function characterizes probability of deriving a right answer at a relative knowledge level equal z/σ.

We shall review some exemplary elementary test which complexity is equal σ_0.

We converse $p(z,\sigma)$:

$$p(z;\sigma) = p\left(\frac{z}{\sigma_0};\frac{\sigma}{\sigma_0}\right) = 0,5 + \Phi\left(\frac{z/\sigma_0}{\sigma/\sigma_0}\right). \qquad (11)$$

Here the level of knowledge and complexity of the elementary test are given in the relative unities, in relation to complexity σ_0 of the exemplary test.

On fig. 1 the set of curve lines of right answer's deriving probabilities for the elementary test is submitted, it depends on the relative level of knowledge z/σ_0, for different relative values of complexity σ/σ_0 of the elementary test.

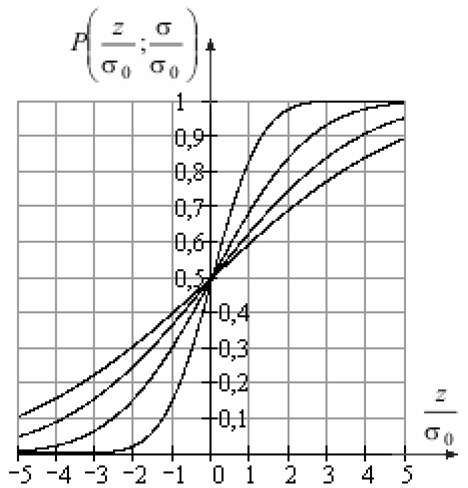

Fig. 1. Distribution of right answer's probability for elementary test

It is easy to note, that, at the change of complexity of the elementary test, in relation to exemplary, the scale of a curve on an axis z/σ_0 changes while the probability $p\left(0;\sigma/\sigma_0\right)$ does not depend on complexity of the test, and always remains equal 0,5. The probability of the answer to the same problem for the contingent of examinees concerning the second level of knowledge ($z = 2\sigma_0$) is equal 0,977. The examinees having the third level of knowledge ($z = 3\sigma_0$), will correctly answer the same test with the probability equal 0,998.

Probability of the erroneous answer

$$q(x) = 1 - p(x) = 0,5 + \Phi(x). \qquad (12)$$

2.4. Information performance of elementary binary tests

The elementary binary test has two alternate outcomes, one of which corresponds the deriving of a right answer, and the other one – corresponds the deriving of the wrong answer. Probability of the improper answer is

$$q\left(\frac{z}{\sigma_0};\frac{\sigma}{\sigma_0}\right) = 1 - p\left(\frac{z}{\sigma_0};\frac{\sigma}{\sigma_0}\right). \qquad (13)$$

The entropy of a system of two-alternate events is determined by a parity

$$H\left(\frac{z}{\sigma_0};\frac{\sigma}{\sigma_0}\right) = -p\left(\frac{z}{\sigma_0};\frac{\sigma}{\sigma_0}\right)\log_2 p\left(\frac{z}{\sigma_0};\frac{\sigma}{\sigma_0}\right) - \\ -q\left(\frac{z}{\sigma_0};\frac{\sigma}{\sigma_0}\right)\log_2 q\left(\frac{z}{\sigma_0};\frac{\sigma}{\sigma_0}\right). \qquad (14)$$

At the complete lack of knowledge, probabilities of deriving of the exact and improper answers, for the elementary binary test, are identical

$$p\left(0;\frac{\sigma}{\sigma_0}\right) = q\left(0;\frac{\sigma}{\sigma_0}\right) = 0,5, \qquad (15)$$

and the prior entropy is equal a posterior one

$$H\left(0;\frac{\sigma}{\sigma_0}\right) = -0,5\log_2 0,5 - 0,5\log_2 0,5 = 1, \qquad (16)$$

without any dependence from complexity of the elementary test.

For any knowledge level which is different from zero, a posteriori entropy $H\left(\frac{z}{\sigma_0};\frac{\sigma}{\sigma_0}\right)$ differs from the prior entropy $H\left(0;\frac{\sigma}{\sigma_0}\right)$ on an information quantity $I\left(\frac{z}{\sigma_0};\frac{\sigma}{\sigma_0}\right)$, obtained as a result of holding a testing.

$$I\left(\frac{z}{\sigma_0};\frac{\sigma}{\sigma_0}\right) = H\left(0;\frac{\sigma}{\sigma_0}\right) - H\left(\frac{z}{\sigma_0};\frac{\sigma}{\sigma_0}\right). \qquad (17)$$

Thus, for example, as a result of testing the examinees having 1-st relative level of knowledge ($z/\sigma_0 = 1$) with the help of some elementary binary test having unity complexity ($\sigma/\sigma_0 = 1$), the information quantity is gained and it is equal $I(1;1) = 0{,}368$ bits.

The examinees having 3-rd knowledge level, will receive $I(3;1) = 0{,}979$ bits as a result of testing if they were tested under the same conditions.

Accounts show that as a result of testing with the help of the elementary test of unit complexity, examinees, characterized by the first level of knowledge, will receive only 0,368 bits of the information, and will remove only small part of indeterminacy in the posed problem. The examinees characterized by the third level of knowledge, on the contrary, will receive 0,979 bits of information, and will almost completely remove the indeterminacy included in the elementary binary test.

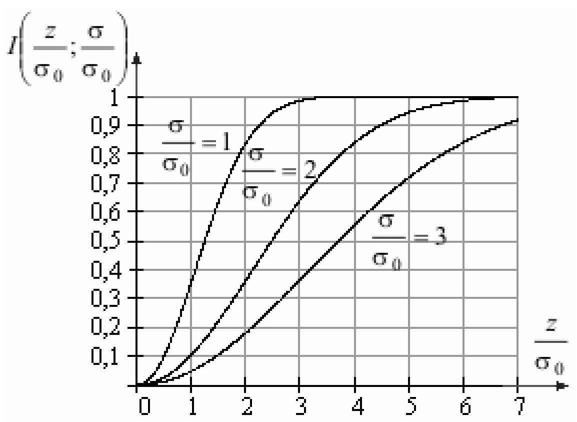

Fig. 2. Information quantity of elementary binary test with different complexities

The association of an information quantity obtained as a result of holding the elementary binary tests of different complexity by examinees, having different relative levels of knowledge is shown on fig.2.

3. PARALLEL MULTIPLE BINARY TESTS

While elementary tests contain only one choice element, the multiple binary test contains few binary elements chosen in a parallel way. For such test the probability of a right answer at the complete lack of knowledge (at a random choice), differs from 0,5.

At any answer, each of the choice elements is either in selected, or in the non-withdrawn condition. It allows to consider each of the elements, as some binary variable x_α, accepting the meanings equal to unity if the device is considered selected, and accepting zero meaning in the other way.

Then, the particular code combination of indicated binary variables $x_{i\alpha}\left(\alpha = \overline{1,n}; i = \overline{0,2^n - 1}\right)$ can be posed in correspondence to any answer X_i, where n is a number of choice elements in the test. Only 2^n different answers are possible and therefore 2^n different code combinations X_i are possible too.

$$X_i = x_{i1} \ldots x_{i\alpha} \ldots x_{in} .$$

Among the amount of code combinations there is some code combination $Y_j = y_{j1} \ldots y_{j\alpha} \ldots y_{jn}$, which depends on the requirements of the viewed task and only this one combination corresponds to the right answer.

As the examinee has a limitation of a knowledge level he can make mistakes and select some combination X_i as the exact.

The probability of such choice depends on a distinction degree of code combinations Y_j and X_i, and also depends on the probabilities of a mistake assumption at a choice of each device $x_{i\alpha}$.

At a unlimited and independent choice of elements the probability of the fact that the selected combination X_i will be accepted for the exact, is determined by a parity

$$P(Y_j / X_i) = \prod_{\alpha=1}^{n} \varphi_{ji\alpha} , \qquad (18)$$

where

$$\varphi_{ji\alpha} = \begin{cases} P_{z\alpha}, \text{ if } x_{i\alpha} = y_{j\alpha} \\ 1 - P_{z\alpha}, \text{ if } x_{i\alpha} \neq y_{j\alpha} \end{cases} , \qquad (19)$$

where

$$P_{z\alpha} = P\left(\frac{z}{\sigma_0}; \frac{\sigma_\alpha}{\sigma_0}\right). \qquad (20)$$

If the elements choice is made without limitations the probabilities $P(X_i)$ of any of code combinations X_i choice should be considered identical, and equal $1/2^n$.

If some limitations were put on a choice of elements, then a part from code combinations will be forbidden, and probabilities of their choice will be equal to zero.

The probability of a choice of all allowed combinations also should be considered identical and equal $1/m$, where m is a number of the allowed code combinations.

In other words,

$$P(X_i) = \begin{cases} \frac{1}{m}, \ i = i_a \\ 0, \ i = i_f \end{cases} , \qquad (21)$$

where i_a and i_f are accordingly the numbers of the allowed and forbidden combinations.

On the basis of the Bayes formula, we shall define the probability of the fact that the examinee will select elements of a code combination X_i while a combination Y_j corresponds the right answer

$$P(X_i/Y_j) = \frac{P(X_i)P(Y_j/X_i)}{\sum_{i \in i_p} P(X_i)P(Y_j/X_i)}. \qquad (22)$$

For all forbidden combinations ($i = i_3$) the probabilities are $P(X_{i_a}/Y_j) = 0$. For all allowed combinations $P(X_{i_a}/Y_j) = \frac{1}{m}$.

That's why

$$P(X_{i_a}/Y_j) = \frac{P(Y_j/X_{i_a})}{\sum_{i \in i_a} P(Y_j/X_{i_a})} = \frac{\prod_{\alpha=1}^{n} \varphi_{ji\alpha}}{\sum_{i \in i_a} \prod_{\alpha=1}^{n} \varphi_{ji\alpha}}. \qquad (23)$$

$P(X_{i_a}/Y_j)$ is a probability of the fact that taking in view the available limitations, the examinee will select the devices relevant to a code combination X_{i_a} if a code combination Y_j corresponds the right answer of the test .

If the selected combination X_0 in accuracy coincides with the exact code combination Y_j ($x_{0\alpha} = y_{j\alpha}; \alpha = \overline{1,n}$), then $P(X_0/X_0) = P_n\left(\frac{z}{\sigma_0}\right)$ and there is a probability of right answer deriving at the multiple binary test having limitations when the elements are chosen.

$$P_n\left(\frac{z}{\sigma_0}\right) = \frac{\prod_{\alpha=1}^{n} P\left(\frac{z}{\sigma_0}; \frac{\sigma_\alpha}{\sigma_0}\right)}{\sum_{i \in i_a} \prod_{\alpha=1}^{n} \varphi_{ji\alpha}}. \qquad (24)$$

If limitations at the elements choice are missing (the multiple binary test is without limitations), then

$$\sum_{i \in i_a} \prod_{\alpha=1}^{n} \varphi_{ji\alpha} = 1 \qquad (25)$$

and

$$P_n\left(\frac{z}{\sigma_0}\right) = \prod_{\alpha=1}^{n} P\left(\frac{z}{\sigma_0}; \frac{\sigma_\alpha}{\sigma_0}\right). \qquad (26)$$

As it was before, here the $P\left(\frac{z}{\sigma_0}; \frac{\sigma_\alpha}{\sigma_0}\right)$ probabilities of the exact choice of elementary binary tests with numbers α , generate the multiple binary test.

If complexities of all elementary binary tests are equal to each other $\left(\frac{\sigma_\alpha}{\sigma_0} = \frac{\sigma}{\sigma_0}\right)$, then

$$P_n\left(\frac{z}{\sigma_0}\right) = \left[P\left(\frac{z}{\sigma_0}; \frac{\sigma}{\sigma_0}\right)\right]^n. \qquad (27)$$

It is obvious, that the probability of a right answer at the complete lack of knowledge $\left(\frac{z}{\sigma_0} = 0\right)$ does not depend on a complexity of elementary tests

$$P_n(0) = 0{,}5^n. \qquad (28)$$

The bigger is the number of devices - n in the multiple test, the less is the probability of a right answer deriving at the complete lack of knowledge, and, hence, the higher is the efficiency of the test.

For example we shall consider a widely meeting special case of the multiple binary test which in fact is a choice of one element from the suggested n elements. It is admissible that all n of elementary tests, generating choice elements, have identical complexities

$$P\left(\frac{z}{\sigma_0}; \frac{\sigma_\alpha}{\sigma_0}\right) = P\left(\frac{z}{\sigma_0}; \frac{\sigma}{\sigma_0}\right) = P(z, \alpha), \qquad (29)$$
$$\left(\sigma_\alpha = \sigma; \alpha = \overline{1,n}\right).$$

The choice of one of n elements superimposes the limitations on the chosen code combinations. These limitations mean that n combinations from all the set of 2^n combinations appear allowed. It is uneasy to be convinced, that for the indicated limitations

$$\prod_{\alpha=1}^{n} P\left(\frac{z}{\sigma_0}; \frac{\sigma_\alpha}{\sigma_0}\right) = P^n(z, \sigma), \qquad (30)$$

and the total is

$$\sum_{i \in i_a} \prod_{\alpha=1}^{n} \varphi_{ji\alpha} = P^n(z,\sigma) + \qquad (31)$$
$$+(n-1)\left[1-P(z,\sigma)\right]^2 P^{n-2}(z,\sigma).$$

After substitution and the relevant transformations, we shall receive

$$P_n\left(\frac{z}{\sigma_0}\right) = \frac{1}{1+(n-1)\left[\frac{1}{P(z,\sigma)}-1\right]^2}. \qquad (32)$$

The probability of a right answer in case, when examinees with the complete lack of knowledge $\left(\frac{z}{\sigma_0}=0\right)$ are being tested, is defined if $P(0,\sigma)=0{,}5$.

$$P_n(0) = \frac{1}{1+(n-1)\left[\frac{1}{0{,}5}-1\right]^2} = \frac{1}{n}, \qquad (33)$$

it also does not depend on complexity σ of the used elementary binary tests.

On fig. 3 associations $P_n\left(\frac{z}{\sigma_0}\right)$ for different meanings of the relative complexity $\left(\sigma/\sigma_0\right)$ of used elementary binary tests are shown.

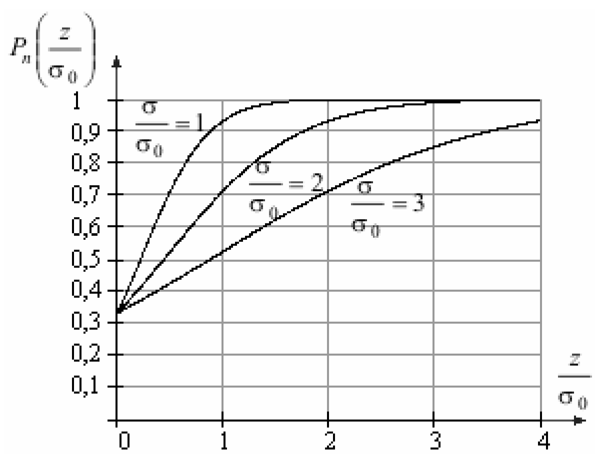

Fig. 3. Right answer probability's distribution for multiple binary test with different complexities

In an often case when only two devices of a choice $n=2$ (the elementary binary test) are present, we shall receive $P_2(0)=0{,}5=P(0,\sigma)$, that completely coincides with the outcome obtained earlier for the elementary binary test.

Thus, application of methods of the diagnosing, grounded on probability criteria, provides a solution of a problem of an objective estimation of the knowledge quality.

4. CONCLUSION

The viewed algorithms, using probability methods of diagnosing, happened to be a basis of the created program environment of a knowledge estimation system. The system has the following features:

- The adjustability under different procedures and for monitoring in different subject domains is provided;
- Procedure of testing and the results' check-out is based on probability performances. It allows to increase objectivity and reliability of inspection by removing the consequences of the unbiased errors, which are the result of incorrectly posed problems or deriving uncertain answers;
- Depending on the requirements, the change of testing complexity without changing the system database is possible. The system environment was created as an application for operating in MS Windows and can be modified to solve monitoring and tutoring problems.

REFERENCES

[1] B.J. Lihtsinder, "The estimation of knowledge with the use of probability methods of diagnosing", the collection of works "Computer Science, the Radio Engineering, Communication", release № 4, Samara, 1999.

[2] B.J. Lihtsinder, L.B. Ivanova, "Probability methods of a quality monitoring and diagnosing of multipleparameter objects", "Informpribor", Moscow, № 3682 from 12.02.87.

Authors: Lihtsinder Boris, professor, academician of IAI and ATI, Vice-president of the Academy of Telecommunication and Informatics (ATI), The Head of Chair of Discrete Message Transmission of VSATI.
Ivanonva Lyudmila, doctor of science, the corresponding member of the ATI, Chief Scientific Secretary of the ATI.
Adress: 443110, L.Tolstogo st., 23, of. 236, Samara, Russia
Tel./Fax: +7–8462–33–47–69
e-mail: lixt@samtel.ru