

CORRELATION – REGRESSION ANALYSIS OF THE KNOWLEDGE AND SKILLS COMPLEX EXPERT SYSTEM

Irena Kulisha^{1 a}, Ravil I. Muhamedyev², Alexander Mrochko², Abduhali N. Shynybekov³

¹Latvian University of Agriculture, J. Čakstes 5, Jelgava, LV- 3001, Latvia

²Information Systems Institute, Lomonosova 1, Riga, LV-1019, Latvia

³Kazakhstan National Agrarian University, Almaty, Kazakhstan

Received 3 March 2009, accepted 4 April 2010

Abstract. In order to define development indicators of future HM service culture, the complex experts' system (CES) named *Evaluation of Hospitality Service Competence* ("EHSC") v 2.0 created by one of the authors based on software "Mini Experts' System" was applied. The program operates as a cover of a simple ES based on Bayesian system of conclusions. The experimental diagnostics detected few faults in question formulations of "0 - EHSC". This resulted in the effects not being credible. During the experimental probation of CES "EHSC" the following was ascertained: the diagnostics process of students turned out to be labor and time-consuming (up to 3 hours); students did not understand some of ES questions because of the incorrect formulation. Due to this the authors made a decision to optimize CES "0 - EHSC" performing a correlation - regression analysis, by reducing the ES number and correcting the questions. After the optimization the CES "EHSC" was developed CES "1 - EHSC". The significant number of questions increases the time of knowledge evaluation (about 3 hours). Correlation analysis methods were used to reduce the number of expert systems. The number of expert systems was reduced to 20 and overall time of testing decreased to 1,5 hour.

Keywords: Diagnostics, expert system, correlation – regression analysis.

Short title: Correlation - regression analysis

Introduction

A situation analysis of today's education allows one to state that one of the main problems of pedagogic diagnostics is the lack of diagnosing "set of instruments", which enables: i) to define peculiarities of a students' acquirement of knowledge, abilities and skills in different study subjects; ii) to determine the characteristics of student's difficulties and reasons of their emergence; iii) to assess the changes taking place in students' development.

In order to perform diagnostics of future hospitality managers' development of HM service competence, the author offers the application of the diagnosing method - *Expert System* (ES) as a modern and contemporary method of research.

1. Idea of Expert System (ES)

Intellectual systems and technologies, expert systems and artificial intellects are a leading and perspective direction of the development of the 21st century, opening new possibilities in optimization of various processes, optimal control, compliance control, testing and other processes of intellectual knowledge acquirement, applying the achievements of information technologies [1]. Expert system – a programmable system, combines the abilities of a computer and man – expert's experience and knowledge. Expert system can offer a reasonable council or find a reasonable decision to the formulated task [2]. ES are applied systems of artificial intellect, in which the knowledge base is formalized empiric knowledge of highly qualified specialists (experts) within a confined area of study subjects.

^aCorresponding author, email: irena.kulish@inbox.lv

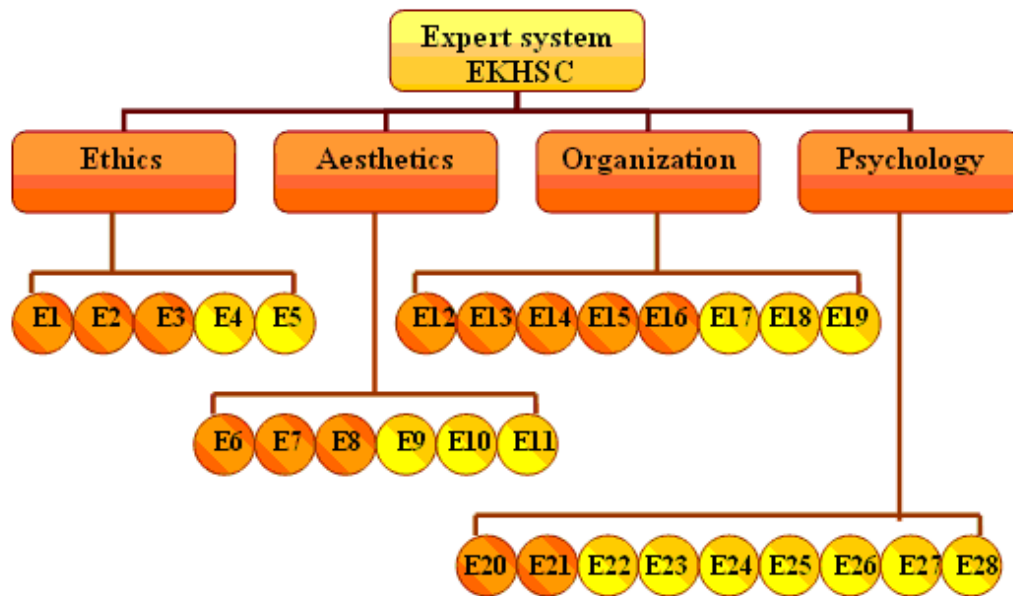


Fig. 1. The structural scheme of the complex experts' system "EHSC" for competency assessment in service culture.

Bielawski and Lewand explain this term like this: "Experts' systems are computer programs, in which for the modeling of human-expert activity within a narrow area of knowledge, the knowledge and experience is applied" [3].

One of the approaches is the application of ES in the study process, which in action is based on the application of the uncertain logics [4]. Application of uncertain logics helps to complete tasks successfully when their output data is formalized weakly. Service culture in a hospitality field is such a field of human activity, where it is difficult to formalize the knowledge of managers.

In order to define development indicators of future HM service culture, the *complex experts' system* (CES) titled *Evaluation of Hospitality Service Competence* ("EHSC") v 2.0 [5,6] was created by the author. This system is based on software "Small Experts' System" was applied [7]. The program operates as a cover of a simple ES, based on Bayesian system of conclusions [8].

The structure of CES "EHSC" is showed in Fig. 1. The system is formed of 28 local experts' systems, depicted in circles; they are divided into four groups. Each group assesses one aspect of service culture: ethical, aesthetic, organizational, psychological. The colored circles are applied to technological culture, while the uncolored - to humanitarian culture.

The diagnostics of future HM service competence's development are carried out based on the answers given to 374 questions, covering the most significant criteria of managers' service culture to the maximum. Formulated questions consider the specifics of activities within the hospitality field.

The program operates as a cover of a simple ES, based on Bayesian system of conclusions [8], which foresees that most of the time the information is not absolutely precise, it possesses the nature of probability. Therefore, after making a list of questions the expert determines the probability of a positive or negative assessment, the answers 'yes' or 'no', depending on the formulation of each question.

The same way the approximate probability is determined for each mentioned conclusion obtained as a result of the respective analysis. After processing the answers, obtained during the analysis of service culture's criteria (aesthetical, ethical, organization technological and psychological), the system of experts logically chooses one of the following levels (conclusions): level 1- high; level 2- average; level 3 - low.

The concluding stage of the CES formation is the experimental probation, which was carried out while the user (student) was working with the experts' system. The dialogue of the user and the system is formed in the format of questions and answers. The value scale of answer variants is from -5 to 5. According to the scale, the answer "-5" mean the answer "no", "0" - the student cannot formulate the answer clearly, "5" - the answer is "yes".

In order to try the test the CES "EHSC" author has performed an experimental diagnostics of respondents (testing No.0), applying CES "0 - EHSC", in Information System Management Institute (ISMI). During the process of diagnostics 33 first year students of ISMI who acquired knowledge in study programs "Manager of Tourism Business Activities" and "Hospitality Management" were involved.

The experimental diagnostics detected few faults in question formulations of "0 - EHSC".

1. Conclusions made by the system after processing the received answers, rather clearly define the situation in each of the reviewed structure components of the service competence. After processing the answers obtained by performing the analysis of future HM's service competence's structure components, ES logically chose one of the conclusions (high level, average level, low level).

2. During the probation no discontinuities were ascertained in the operation of the system. The feasibility of the system to determine a development level of students - the future HM service competence during the dialogue, confirmed the practical significance of CES "EHSC"'s application.

3. During the experimental probation of CES “EHSC” the following was ascertained: the diagnostics process of students turned out to be labor and time-consuming (up to 3 hours).
4. Students did not understand some of ES questions because of the incorrect formulation.

Due to this the author made a decision to optimize CES “0 - EHSC” performing a correlation – regression analysis, by reducing the ES number and correcting the questions.

2. Correlation - regression analysis

It is possible to optimize CES “EKHSC” by the number of local expert systems, determining the mutual connection of evaluations of these systems. All connections which can be expressed and measured in numbers can be defined as statistical, including the functional connections. A connection between two or more accidental quantities can be calculated applying a method of mathematical statistics - the analysis of correlations and regressions.

A correlation connection is the substantial case of statistic connection, when various significances of one variable quantity correspond with different average significances of another quantity. A correlation connection among the features can be caused by several reason.

1. Dependence of outcoming features (its variations) from the variation of the factor feature (cause-effect relationship); a case when it is logically clear which feature is the independent variable (factor) x , which – the dependent variable (result) y .
2. Interdependency of two results of a common cause.
3. Interconnection of features, where each is both the cause and the effect. The case when each feature can be both the dependent variable and the independent variable.

According to the essence of correlation connection, there are two tasks for its research: i) determination of the connection type and evaluation of the equation parameters, which displays the connection between the average quantities of the dependent variable and the quantities of the independent variable (regression analysis); ii) measurements of connection closeness of two or more features [5]. The first task is general and covers functional connections, but the second task is characteristic to statistical connections.

In order to determine the closeness of the connection several indicators are applied.

Firstly, it is the determination ratio R^2 and the correlation relation $\eta_{y,x}$:

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (1)$$

where n – number of units in the set; \hat{y}_i individual quantity after equation of connection; \bar{y} – total average quantity of the resultative feature; y_i – total average quantity of the resultative feature:

$$\eta_{y,x} = \sqrt{R^2}. \quad (2)$$

The simplest system of correlation connection is the linear connection between two features. The increased attention paid to linear connection can be explained by the variation confinement of the variable quantity, as well as by the fact that in most

cases applying calculations to the non-linear forms of connection they transform into linear forms.

Secondly, the closeness of connection pairs' linear correlation can be determined by correlation relation $\eta_{y,x}$. Besides, in the linear form of connection equation another closeness indicator of interconnection is applied – correlation ratio $r_{x,y}$. This indicator is a ratio, which is expressed not as absolute measurement units of features, but as a specific weight of the quadratic mean deflections:

$$r_{x,y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}, \quad (3)$$

where n – number of units in the set; $\bar{y} \cdot \bar{x}$ – common average quantities of the features; $y_i \cdot x_i$ – individual quantities of the features.

The correlation ratio does not depend on the conventional measurement units of the features, thus it is comparable to any feature. Usually [6] a significantly close connection is considered in cases when $r_{x,y} > 0.7$; average close – $0.5 < r_{x,y} < 0.7$; weak close – $r_{x,y} < 0.5$.

Applying this method of correlation and regression analysis, one must consider that correlation indicators can be interpreted only within the terms of deflection variations from the average quantity (in space of differences) and the correlation of features should not be interpreted as the connection of its levels.

The indicators of correlation connections, determined by a limited set (size n), actually are only evaluations of feature - selection of statistical connection. Therefore an assessment of the preciseness and extent of credibility of the correlation's parameters is required. With credibility one must understand a probability that the quantity of the parameter under testing will not be equal to zero. The credibility probability of correlation parameters is determined according to testing rules of statistical hypotheses developed by mathematic statistic, i.e. comparing the evaluation quantity with the mean random error of evaluation [5, 9].

The mean error of correlation ratio's $r_{x,y}$ evaluation is calculated according to the following formula:

$$S_r = \frac{1 - r_{x,y}^2}{\sqrt{n - 2}}. \quad (4)$$

Knowing the mean error of the correlation ratio's evaluation, γ – can be calculated, the probability that the zero value of the ratio falls within the credibility interval, taking into consideration the errors of quantities (probation of correlation ratio's significance). In this connection the ratio is related to its mean error, i.e. Student's t – criterion:

$$t = \frac{r_{x,y} \sqrt{n - 2}}{1 - r_{x,y}^2}. \quad (5)$$

Critical quantity is determined by tables of Student's division, taking into account the condition that:

$$P = (|t| > t_{\gamma,k}) = r_{x,y}. \quad (6)$$

Fig. 2. Table of correlation ratios of local expert systems

In that case $k=(n-2)$ – a figure of degrees of freedom; γ - level of significance.

For the local experts’ systems, whose $r_{x,y}$ came near to value (1), the assessment credibility of the correlation ratio was examined (the value level was calculated).

$$r_{x,y} \rightarrow 1 \tag{7}$$

The values of Student’s criteria were calculated and for the level $\gamma=0.95$ the value of correlation value was defined.

If the correlation ratio is close to one, then the division of its evaluation differs from the average or Student’s division, because it is limited by the quantity 1. In cases like this R. Fisher suggested to re-form evaluation of ratio’s credibility into a form which does not possess such a limit:

$$z = 0.5 \ln \frac{1 + r_{x,y}}{1 - r_{x,y}} \tag{8}$$

The mean error of quantity z is calculated by the formula:

$$S_z = \frac{1}{\sqrt{n-3}}, \tag{9}$$

but the quantity of Student’s criteria is determined as:

$$t = \frac{z}{S_z} \tag{10}$$

In credibility examination of pair correlation’s parameter assessment one must consider that in cases of small selection fields only close connections can be determined precisely, but in cases of big sets the connections of weak closeness are also credibly determined.

3. Results of the complex expert system’s “EHSC”: correlation - regression analysis

The task of correlation analysis was to determine correlation dependence among separate local expert systems, in which 28 ser-

vice competences of future hospitality managers developed by the author were reflected upon.

Clarifying correlation dependence among various evaluations of competence, provided by different local expert systems, it can be assumed that such a dependency can be caused and it can exist due to dependence between two effects of a common cause.

In order to optimize the number of the local experts’ systems, 28 local experts’ systems were subjected to the correlation analysis. For each pair of the local experts’ system by selection in amount of 33 units (number of students) the pair correlation ratios were obtained $r_{x,y}$ (Fig. 2). The calculations of correlation ratios were performed applying the package of *MS Excel* data analysis [10].

For local expert systems whose $r_{x,y}$ came close to significance 1, the evaluation credibility of correlation ratio was examined (the level of significance was calculated). Significances of Student’s t-criterion were calculated according to formulas (8, 9, 10) and significance of correlation ratio was determined for the level (see Table 1).

The results of data correlation analysis enabled the optimization of CES “EKHSC. The number of ES was cut down to 20, but the number of questions - down to 190. After the optimization the CES “EKHSC” was developed in both Latvian and Russian and named CES “1 – EHSC”.

The initial testing at ISMI was conducted twice, at the beginning performing the testing with “0 - EHCS”, but after a week with “1 - EHCS” (optimized CES). Comparing the results of both testings, a conclusion can be made that the results of both testings’ differ only slightly (see Table 2). Thus, in further researches it would be rational to use the optimized ES- “1 - EHCS”.

The initial diagnostics, applying the complex expert system “1 - EHSC” was performed in the three education establishments. Firstly, in the Information System Management Institute (ISMI) in Riga (Latvia). Then the research was continued in: i) Klaipėda Business and Technology College (KBTC), Klaipėda (Lithuania), ii) College of Hotel Management (CHM), Belgrade (Serbia).

Table 1. Significance of correlation coefficient. $\gamma=0.95$, $k=31$.

x, y	$r_{x,y}$	t	$t_{\gamma,k}$	Significance
3, 5	0.98	13.6	2.04	significant
7, 24	0.99	13.6	2.04	significant
8, 24	0.73	13.1	2.04	significant
17, 4	0.99	13.6	2.04	significant
19, 23	0.96	13.5	2.04	significant
20, 13	0.93	13.5	2.04	significant
26, 1	0.89	13.4	2.04	significant
28, 1	0.94	13.5	2.04	significant

The first testing was performed applying the developed CES "1 - EKHSC" after correlation and reduction. This analysis involved 124 first year students of ISMI, 22 first year students of KBTC and 43 first year students of CHM. In total 189 first year students were involved in this research.

A completion of full analysis took 1 hour 15 minutes for ISMI students, for KBTC students it took 1 hour 40 minutes, for CHM students – 1 hour 50 minutes.

Correlation analysis methods were used to reduce the number of expert systems. The number of expert systems was reduced to 20 and overall time of testing decreased to 1,5 hour.

Conclusions

Based on previously mentioned researches, it can be concluded as follows.

1. In this work the system of Bayes' decision making was applied. Experts' system for assessing hospitality culture knowledge was developed - formed by 28 local subsystems, which assess the knowledge of becoming specialists in 4 areas. Testing of the system was performed, which approved a self-evident improvement of results depending on the training time.
2. After a correlation - regression analysis the number of expert systems lowered to 20. Testing time reduced to 1,5 hours.

Table 2. Comparison of diagnostic results ($r_{x,y}$ distribution) before and after optimization of CES "0 - EHSC". Testing was provided at Information Systems Management Institute, Riga.

High level		Average level		Low level	
Before optimization	After optimization	Before optimization	After optimization	Before optimization	After optimization
testing with "0 - EHCS"	testing with "1 - EHCS"	testing with "0 - EHCS"	testing with "1 - EHCS"	testing with "0 - EHCS"	testing with "1 - EHCS"
0.77617	0.70197	0.64624	0.51330	0.16393	0.23050

References

1. Bielawski, L., Lewand, R. (1991) Intelligent systems defined. In: Bielawski, L., Lewand, R. Intelligent Systems Design: Integrating Expert Systems, Hypermedia, and Database Technologies. New York ... [etc.] : John Wiley & Sons, p. 17-62.
2. Kulisha, I. Muhamedijev, R. (2009) Experts' System for Assessing Students' Knowledge in the Area of Hospitality Technology. "Computer Modeling and New Technologies". Rīga : TSI, Vol.13, No.2, 31-37 lpp.
3. Moskvins, G. (2008) Intelektuālās sistēmas un tehnoloģijas: mācību līdz. Jelgava : LLU, 5. lpp.
4. Rivža, P., Arhipova, I. (1995) Ievads ekspertu sistēmās: lekciju konspekts LLU studentiem. – Jelgava, 9 lpp.
5. Buchnin A. (2001) "Malaja ekspertnaja sistema" v2.0. [http://bukhnin.chat.ru/], retrieved 2008 12 15.
6. Eliseeva, I.I., Juzbashev M.M. (2006). Obshchaja teorija statistiki: uchebnik (in rus.) Ed. Eliseeva, I.I. 5th issue. - Moscow: Finansy i statistika - P. 327-349.
7. Kozlov A. Ju.; Shishov V. F. (2003) Paket analiza MS EXCEL v ekonomiko-statisticheskich raschetach (in rus.) Ed. Mchitarian V.C. - Moscow: ONITI-DANA - P. 117-126.
8. Pustynnikova I.I. (2001) Technologija ispolzovanija ekspertnych sistem dlja diagnostiki znaniy i umenij // Education Technology & Society, Nr.4(4), c. 80-82
9. Joseph F. Healey. Statistics: A Tool for Social Research // Amazon.com Product, ISBN 0534627943.
10. Chastnikov A. P., Belov D.L., Gavrilova T. A. (2003) Razrabotka ekspertnych sistem (in rus.) - Sanct-Peterburg: BCHB-Peterburg - P. 12-33.