ISSN 1407-7493 COMPUTER SCIENCE DATORZINĀTNE 2008-7493

INFORMATION TECHNOLOGY AND MANAGEMENT SCIENCE
INFORMĀCIJAS TEHNOLOĢIJA UN VADĪBAS ZINĀTNE

# PROCESSING THE RESULTS OF COMPETENCE EVOLUTION EVALUATION BY DECISION TREES

**Ieva Boļakova**, Mg.paed., Lecturer, Department of Computer Science, Daugavpils University, Parades 1, Daugavpils, LV 5400, Latvia, e-mail:ieva.bolakova@du.lv.

***Keywords: evaluation of education quality, decision trees, C4.5***

## 1. Introduction

The questionnaires are one of the possibilities to ensure student feedback in the education process. In the spring of 2008 in Daugavpils University a form of questionnaire was prepared. This form will be used to evaluate the quality of education more objectively in different study programs. The structure of the questionnaire form is based on the competence approach. The competence approach is very topical in today's education system, because the competence involves new acquired knowledge, as well as the level of understanding of this knowledge and the ability to use acquired knowledge.

The solving of the same task by different methods in separate cases may lead to contradictory results. However correlated use of the methods can increase the quality of classification, forecasting and recognition. The results of questionnaires usually are processed by different statistical methods. After the use of these methods the researchers can evaluate the comparison of opinions; they can find regularities between separate questions or problems. We can apply the machine learning methods to the questionnaire data to crystallize those attributes from whole attributes set, which have a higher rank or larger influence on another attributes. In this paper one of the knowledge formation methods is considered – the logic structures representation by decision trees.

## 2. Preparing a form of questionnaire by the competence approach

Analyzing the content of study process, learning outcomes can be separated from competences, thus indicating the different roles of academic staff and students. Desired learning outcomes of the process of learning are formulated by the academic staff. Competences are obtained or developed during the process of learning by the student [1].

Learning outcome can be determined in the description of study courses. The author of the course should point what a learner has to know, understand and/or be able to demonstrate after completion of learning. Learning outcome can be determined to the separate study course, as well as to concrete study period – for example, to each study year. Learning outcomes specify the requirements for credits acquiring.

Competences represent a dynamic combination of knowledge, understanding, skills and abilities. Fostering competences is the object of education programs. Competences are formed in various course units and obtained at different stages of education process.

Preparing the form of questionnaire for study process evaluation, the competence approach was chosen. The record of 33 competences forms the basic structure of the questionnaire form. It is offered to evaluate these competences in the 10 points scale depending on the level of development of concrete competence at the moment, when student starts studies in the university, and at the present study moment. The third evaluation, in 10 points too, determines the level of competence development, which is necessary for successful realization of professional activity. Such an approach to competence evaluation division allows one:

- To evaluate the conformity level of the study program graduates to the requirements of labor market. As a comparison basis we can use the evaluation of competence importance given by students, as well as opinion of experts – employers.
- To determine which competences
    - Are important to work, but are not developed enough;
    - Are not very important and are developed weakly;
    - Are not acknowledged as important, but are marked as developed very well;
    - Are very important and are developed very well.

All competences are grouped into three categories in the questionnaire form:

- Instrumental competences (1 - 13);
- Interpersonal competences (14 - 21);
- Systemic competences (22 - 33).

Average evaluations are counted in every competence group, and the groups of basic competences with higher rank in the appropriate group are determined.

## 3. The task statement and methods of solution

Nowadays logic rule induction can hardly be considered a new task. However the task of technology developing for correlated use of different diagnostics solutions is still topical. The proposal of such methods will be especially valuable, for example, in small size data sets, typical to areas of sociology and psychology [2].

Nowadays decision trees are one of popular methods for finding and representing regularities. Binary decision trees were used in the experiments with questionnaire data analysis – any non-leaf node is split only into two branches. The nonleaf nodes are labeled with concrete attributes, but the branches are labeled with concrete qualitative data values or - with interval of quantitative attribute values [3].

When the decision tree is constructed, it is easy to convert it into the form of "IF…THEN…" rules. The quality of the decision tree is determined by two parameters: the precision and the complexity. The precision of the tree indicates the purity of the data classification. The complexity of the tree is determined by the number of the tree nodes, number of leaf nodes, and maximal depth of the tree, i.e. distance from the root node to the decision node.

The data of the questionnaire results has to be prepared in a special form to use the machine learning methods. At first – whole data table has to be split into the learning set and the test set. This kind of data split doesn't present any problems. For example – we choose the

first third part of the data table as a learning set, but other data will be used for testing. The proportion of the split can be different.

Another requirement for the data – we have to strictly separate attributes which are some decision or conclusion, and which are analyzing to make that decision or to make conclusion. [4] For example – to find out the students satisfaction with the technical support of the program, we can be interested about their study program, about their age, the study year, participation in the lectures. In this example – the satisfaction with technical support will be the target attribute or the class. In Table 1 the class is in the last column C. The other questions are attributes, which influence the values of the class (in Table 1 - columns A1, A2, A3 and A4).

Table 1

The part of the questionnaire results data table

| No | A1 | A2 | A3 | A4 | C |
|---|---|---|---|---|---|
| 1 | 5 | 4 | 4 | 5 | 8 |
| 2 | 5 | 2 | 3 | 3 | 7 |
| 3 | 3 | 3 | 5 | 3 | 3 |
| 4 | 4 | 3 | 2 | 3 | 4 |
| 5 | 3 | 2 | 3 | 3 | 6 |
| 6 | 3 | 2 | 4 | 3 | 1 |
| 7 | 4 | 2 | 3 | 3 | 6 |
| 8 | 3 | 4 | 2 | 3 | 7 |
| 9 | 5 | 1 | 3 | 3 | 8 |
| 10 | 4 | 3 | 5 | 4 | 7 |
| ... | ... | ... | ... | ... | ... |

While composing the form of the questionnaire, some a priori indicators are offered – the set of questions, which can influence the observed branch. The expert of the branch (for example, director of the program or the council of program) offers these a priori indicators, making this choice on the basis of some theoretical considerations. We can group a priori indicators into separate subgroups (Figure 1). These subgroups reflect the factors, which the organizers of the questionnaire want to analyze. In this questionnaire as some factors we can mention, for example, the satisfaction of the students with the planning of the study process, the motivation of the education obtaining, accessibility of the literature necessary for studies etc. The audience involved in the questionnaire has not to know about these factors. The participants of the questionnaire give the answers to the separate questions or evaluate some components of the study process. When the questionnaire is performed, the summary of the results can show, which indicators practically do not participate in the branch forming or in forming the opinion about considered branch. After such experience is gained, one of the important tasks is to identify a posteriori indicators. A posteriori indicators have to be chosen so that

- they directly characterize the factors which are analyzed;
- in the case of the necessity they would be the basis for renewing the indicators as precisely as possible.

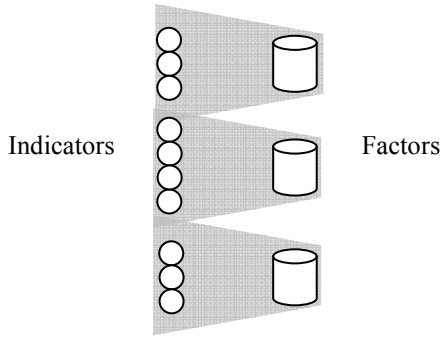


Figure 1. Summarization of separate indicators into factors

The decision tree construction methods can be used to solve the problem of reduction of the indicators space.

## 4. The analysis of questionnaire results

The technology of the decision tree construction C4.5 was used investigating the education quality in the study programs of Daugavpils University. The task of the investigation is to find out the opinion of the students and the academic staff about the organization questions of the study process and about the study program content. In the questionnaire form, the questions are included, which can help to determine the education motivation of the students. Analyzing the results of the questionnaire, we can connect the motivation of the education with the level of development of separate competences.

The 33 competences included in the questionnaire form are evaluated from three positions:

- level of competence development at the moment of studies beginning in university;
- level of the competence development at the current moment;
- the necessary level for the professional activity in the specialty.

After the questionnaire was made the mean value for the instrumental competences was calculated in the third position "necessary level for the job" (let's name it *Ideal* level, which has to be achieved to implement the professional activity on a high level).

Similarly the mean value for the instrumental competences was calculated in the other two positions "level at the moment of studies beginning" and "at the current moment achieved level" (let's name them accordingly *Start* level and *Current* level).

These values have to be calculated to somehow evaluate the necessity of education for students (i.e. – why the education is necessary for the student). Then we can evaluate this necessity of education, for example, as a difference between *Ideal* level and *Start* level.

The achieved education level can be evaluated as a difference between *Current* level and *Start* level.

If we reflect the scale of competence acquisition, then we can choose the form of display, where on the one side of the scale is *Start* level, but on other – *Ideal* level, and when the *Ideal* level is achieved, we will consider that competences are developed by 100% (see Figure 2). Then we place the level of *Current* competence development on this scale. Is this reached *Current* level somehow connected with motivation of education? – Such scale could help to answer this question. If the *Current* achieved level on the scale is close to *Ideal* level or – it is almost in the same position, then we can assume that the motivation of the student is to receive the diploma. But this decision should be based on the study year of student also. In

the fourth study year it would be logical if the appropriate competence is absolutely developed.

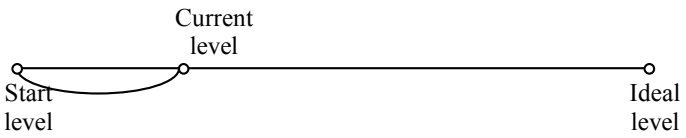


Figure 2. The scale of the competence development level

With this scale we can also survey how much (in percentage) the necessity of education is performed in study process.

What determines the necessity of education? To answer this question, correlations were estimated between the *Ideal* level and how high is development level on the *Start*.

From the results of the questionnaire, it can be seen that there are some competences, which are evaluated as important for professional activity (*Ideal* level), but these competences have no necessity of education. Maybe it is because this level is reached already.

The result of using the decision tree construction algorithm C4.5 is a graphical representation of separate regularities in the form of a decision tree, as well as the rule set (Table 2).

Table 2

The rule set received as a result of using C4.5

| | |
|---|---|
| Rule 1: | IF A1 = 3 AND A3 = 3 THEN Class 6. |
| Rule 2: | IF A1 = 1 THEN Class 7. |
| Rule 3: | IF A1 = 5 AND A2 = 2 THEN Class 7. |
| Rule 4: | IF A1 = 5 AND A2 = 1 THEN Class 8. |

The results of this concrete experiment show that the attribute A4 doesn't participate in the decision making at all. I.e. – to find out, how the level of development of competence C is connected to an estimation of attributes A1, A2, A3, and A4, further we can cancel the last question, because the algorithm has evaluated it as unimportant for formation of corresponding competence.

While processing the questionnaire data, the necessity had occurred to create any parameter, which would measure the success of the education process. The realization of the necessity of student's education increases as the study year increases. A person or a group of persons, which take the expert position (for example – director of study program, or the council of study program), evaluate the program and determine, which competences, in which study year and to which extent have to be developed. For example: first study year – 30% realization, second study year - 40% realization, etc. After that, this determination can be used to evaluate the process of education – is it successful or not, because the student opinion is considered, in what level the concrete competence or group of competences is developed the in Current position.

## 5. Conclusions

When the first questionnaire stage was performed and the acquired data has been preprocessed, it became clear that the additional tasks will appear for investigation using

decision trees. At this stage of the investigations – one of the decision trees construction algorithms advantages, which we can use in the processing of the questionnaire results, - is the ability to work with qualitative data (discrete attributes). The statistic methods, which are used to process the results of the questionnaire, such as – analysis of the correlations, regression, analysis of the factors, - these methods work with quantitative indicators (continuous attributes). However, it is quite hard to find such numerical scale for the form of questionnaire which represents the opinion of the respondent correctly/objectively/unequivocally. Even if the answers to the questions have to be evaluated by numerical mark, the respondents very differently feel zero position in the offered scale. For example, for one respondent the weakly developed competence will be evaluated with the mark from the interval [2, 4], but for another the comprehension about weakly developed competence will be in the interval [3, 5] (see Figure 3).

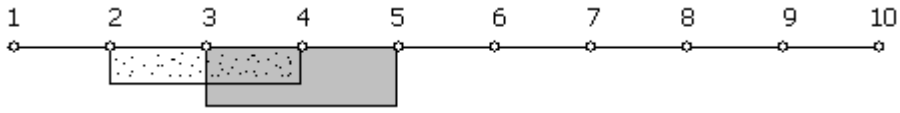


Figure 3. The scale of competence evaluating

One part of questions of the questionnaire form is included to evaluate different aspects of the material and technical support of the study program, as well as – the organization aspects of the study program. With the decision tree construction algorithm C4.5 the experiments were performed to try to crystallize out more essential indicators (attributes) from all attributes set and to observe which questions do not participate in the concrete opinion forming, for example – in the development of separate competences. The work on these experiments has to be continued because it is important to find out, why some attributes are selected, while others are ignored. Actually, at the heart of this choice is insignificance of separate indicators in formation of factors, or it is the specificity of algorithm C4.5, choosing some attributes which have a smaller number of values (number of possible answers to questionnaire questions).

## References

1. Competences, 2004. URL: http://www.tuning.unideusto.org/tuningeu/ – Visit date October 2008.
2. Берестнева О.Г., Муратова Е.А. Построение логических моделей с использованием деревьев решений. Известия Томского политехнического университета.. Т. 307. № 2, 2004. – P 154 - 160
3. Evans B., Fisher D. Overcoming Process Delays with Decision Tree Induction // IEEE Expert: Intelligent Systems and Their Applications, Vol. 9, No. 1, February 1994, P. 60 – 66.
4. Bolakova I. Classification of statistical data using inductive learning algorithms // Proceedings of the 11th International Conference on Soft Computing ‘MENDEL 2005’, Brno, Czech Republic, 2005. – P. 69–72.

**Boļakova Ieva. Kompetenču attīstības novērtēšanas rezultātu apstrāde ar lēmumu kokiem**

*Rakstā izskatītas iespējas aptaujas rezultātu analizēšanai kopā ar statistikas metodēm izmantot arī induktīvās apmācības metodes – lēmumu koku konstruēšanas algoritmus. Aptauju organizēšana ir viens no veidiem kā var nodrošināt atgriezenisko saiti ar studentiem izglītības procesā. 2008.gada pavasarī Daugavpils Universitātē tika sagatavota aptaujas anketa, kuru turpmāk izmantos, lai varētu objektīvāk novērtēt izglītības kvalitāti dažādās studiju programmās. Aptaujas anketas struktūra ir balstīta uz kompetenču pieeju. Mūsdienu izglītības sistēmā kompetenču pieeja ir ļoti aktuāla, jo kompetence sevī apvieno gan iegūtās jaunās zināšanas, gan šo zināšanu izpratnes pakāpi, spēju pielietot iegūtās zināšanas. Izmantojot algoritmu C4.5, tika veikti atsevišķi eksperimenti aptaujas rezultātu analizēšanai. Izskatīta viena no lēmumu koku konstruēšanas algoritmu pielietošanas iespējām – atribūtu kopas samazināšana. Darbs pie šiem pētījumiem tiek turpināts, jo pagaidām vēl nav pārbaudītas visas iespējas un priekšrocības, ko var dot lēmumu koku konstruēšanas metodes izmantošana aptaujas datu analizēšanā salīdzinot ar statistikas metodēm.*

**Boļakova Ieva. Processing the results of competence evolution evaluation by decision trees**

*This paper considers the use of inductive learning methods - the decision tree construction algorithms- for analyzing questionnaire results as an additional tool together with statistical methods. The questionnaires are one of the ways to ensure student feedback in the education process. In the spring of 2008 in Daugavpils University a form of questionnaire was prepared. This form will be used to more objectively evaluate the quality of education in different study programmes. The structure of the questionnaire form is based on the competence approach. The competence approach is very topical in today's education system, because the competence involves new acquired knowledge, as well as the level of understanding of this knowledge and the ability to use acquired knowledge. To analyze the results of questionnaire, a series of experiments was made using the algorithm C4.5. One of possible applications of decision tree construction algorithms is considered – reducing of the set of attributes. Investigations continue in this area, because not all possibilities and advantages of the use of decision tree construction methods in the analysis of the questionnaire data are examined.*

**Болякова Иева. Обработка результатов оценки развития компетенций с помощью деревьев решений**

*В статье рассмотрены возможности использования методов индуктивного обучения – алгоритмов построения деревьев решений для анализа результатов опроса. Организация опросов – это один из способов обеспечить обратную связь со студентами в процессе образования. Весной 2008 года в Даугавпилсском университете была подготовлена анкета, которую в будущем планируется использовать для более объективной оценки качества образования в разных программах обучения. Структура анкеты основана на компетентностном подходе. В современной системе образования компетентностный подход очень актуален, потому что компетенция объединяет в себе и полученные новые знания, и степень понимания этих знаний, и способность применять полученные знания. Используя алгоритм C4.5, было проведено несколько экспериментов для анализа результатов опроса. В статье рассмотрена одна из возможностей применения алгоритмов построения деревьев решений – уменьшение множества атрибутов. Работа над этими исследованиями продолжается, поскольку еще не изучены все возможности и преимущества, которые предоставляет использование методов построения деревьев решений для анализа данных опроса по сравнению со статистическими методами.*