

Quantitative Forecasting and Assessment Models in the State Education System

Vydūnas ŠALTENIS

*Institute of Mathematics and Informatics
Akademijos 4, LT-2021 Vilnius, Lithuania
e-mail: saltenis@ktl.mii.lt*

Gintautas DZEMYDA, Vytautas TIEŠIS

*Vilnius Pedagogical University
Studentų 39, LT-2004 Vilnius, Lithuania
Institute of Mathematics and Informatics
Akademijos 4, LT-2021 Vilnius, Lithuania
e-mail: {dzemyda, tiesis}@ktl.mii.lt*

Received: June 2002

Abstract. This paper presents model-based forecasting of the Lithuanian education system in the period of 2001–2010. In order to obtain satisfactory forecasting results, development of models used for these aims should be grounded on some interactive data mining. The process of the development is usually accompanied by the formulation of some assumptions to background methods or models. The accessibility and reliability of data sources should be verified. Special data mining of data sources may verify the assumptions. Interactive data mining of the data, stored in the system of the Lithuanian teachers' database, and that of other sources representing the state of the education system and demographic changes in Lithuania was used. The models cover the estimation of data quality in the databases, analysis of the flow of teachers and pupils, clustering of schools, the model of dynamics of the pedagogical staff and pupils, and the quality analysis of teachers. The main results of forecasting and integrated analysis of the Lithuanian teachers' database with other data reflecting the state of the education system and demographic changes in Lithuania are presented.

Key words: modelling, data mining, neural networks, clustering, forecasting, education system.

1. Introduction

Forecasting models of complex social or state systems are usually based on the analysis of a great amount of information by using suitably adjusted methods. Selection of the methods and parameters of models is usually accompanied by the formulation of some hypotheses. Another origin of hypotheses is a limited possibility of accessing the necessary data as well as insufficient reliability of data.

Usually data mining tasks involve forecasting of some variables (Han and Kamber, 2000). In this paper, we discuss a general scheme of interactive data mining in the data

sources that may verify the assumptions and ensure the construction of well-founded forecasting models. Moreover, we apply the general scheme of special data mining of the data stored in the system of the Lithuanian teachers' database, pupils' database, and that of other data sources representing the state of the education system. The data on demographic changes in Lithuania were also used.

The models developed cover the estimation of data quality in the databases, analysis of the flow of teachers and pupils, clustering of schools, the model of dynamics of the pedagogical staff and pupils, and the quality analysis of teachers. The paper presents the items of forecasting and quality analysis more in detail.

As a result, model-based forecasting of the Lithuanian education system in the period of 2001–2010 has been performed. The results of the investigation discovered significant problems that may be faced in the future by the state education system: redundancy of the pedagogical staff for some subjects and for some regions (especially, in primary schools).

This paper is organised in the following way. In Section 2, the quantitative forecasting and assessment models of education system are presented. In Section 3, a general scheme of the interactive data mining from different sources is discussed. In Section 4, some results are presented for the case of the Lithuanian education system.

2. Models of the Education System

Let us consider the process of model constructing by the example of the Lithuanian education system, a schematic model of which is presented in Fig. 1.

The model is complicated, therefore we will expose only a small part of its features in this paper.

The aim of forecasting models was to evaluate the evolution of the state education system in Lithuania in the nearest future. A continuous reduction of birth rate in the state requires to be ready for future problems.

The main parts of forecasting models are as follows:

- pupils' number forecasting models;
- teachers' number forecasting models;
- models for clustering of schools;
- models that evaluate the teachers' qualification and pupil's progress.

All these models are interrelated: e.g., the teachers' number depends on the pupils' number on the types of schools; the pupils' progress depends on the quality of teachers and so on.

Fig. 1 illustrates the pupils' flow. The majority of children from the birth flow comes to the 1-st forms. During the school year a part of them drops out of the school.

The flow of teachers comes to a school from universities and colleges. They may leave a school temporarily or permanently.

Schools, teachers, and pupils are characterised by some quality indices (see Fig. 1), therefore their values may be included in the forecasting models.

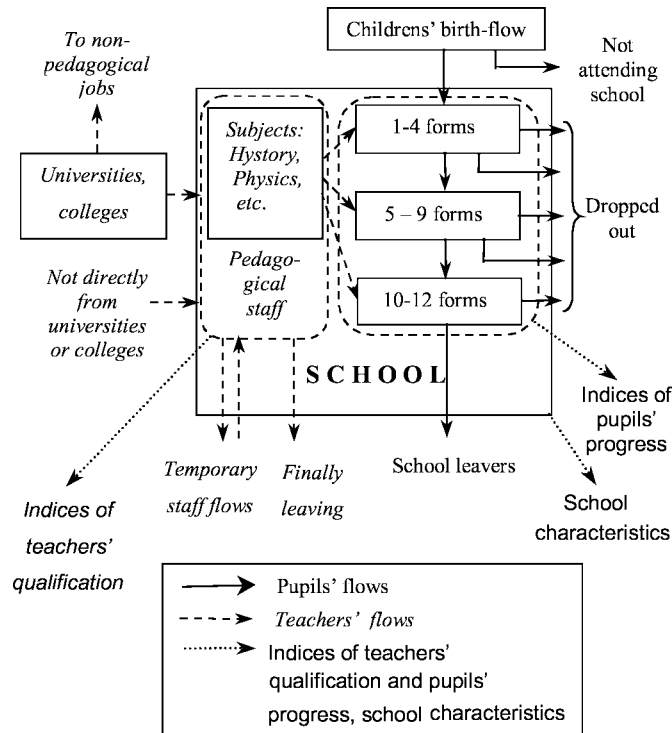


Fig. 1. The flows of pupils, teachers, indices of teachers' qualification and pupils' progress, and school characteristics.

2.1. Modelling of Pupils' Number

The pupils' number models include:

- forecasting of pupils' input flows into the first forms on the base of analysis of birth rate data;
- forecasting of dropped out pupils' flows;
- forecasting of the distribution of these flows in various administrative regions.

Let us introduce the notation:

y_{ik} is the number of pupils in the i -th region ($i = 1, \dots, m$) and in the k -th form ($k = 1, \dots, 12$), where m is the number of regions;

y_{ik}^+ is the number of pupils that replenished the corresponding form;

y_{ik}^- is the number of pupils that left the corresponding form.

The values y_{ik}, y_{ik}^+ and y_{ik}^- are time dependent, therefore they are denoted as the functions of time $t(t = t_1, \dots, t_p)$, where p is the number of years in the model.

Then the balance equations for the i -th region and the k -th form are:

$$\Delta y_{ik}(t) = y_{ik}(t) - y_{i,k-1}(t-1) = y_{ik}^+(t) - y_{ik}^-(t),$$

where $\Delta y_{ik}(t)$ is the change of the pupils' number in the t -th year.

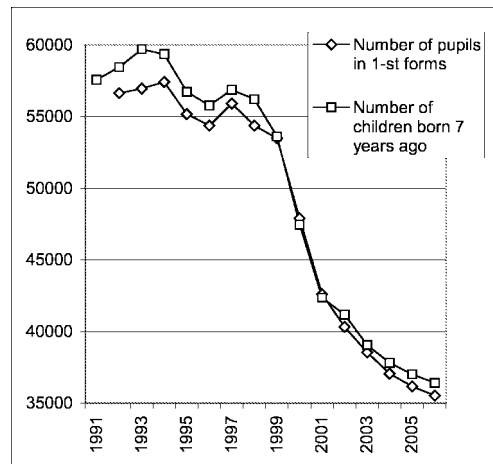


Fig. 2. The number of pupils coming to the 1-st forms, and the number of children born 7 years ago.

The number of pupils in the first forms $y_{i1}(t)$ mainly depends on the birth number $b_i(t - t_b)$ before t_b years, therefore:

$$y_{i1}(t) = k_i^b b_i(t - t_b),$$

where the coefficient k_i^b describes the part of new-borns that come to school after t_b years.

The corresponding equations for all state numbers may be obtained if we use the summary values:

$$y_k(t) = \sum_{i=1}^m y_{ik}(t), \quad b(t) = \sum_{i=1}^m b_i(t). \quad (1)$$

Fig. 2 illustrates the main idea of the pupils' forecasting model: the input flow of pupils may be evaluated using demographic birth-rate data sources.

Data mining discovered that, in the whole state the values of coefficients may be evaluated as equal to $k^b = 0.971$ and $t_b = 7.3$. In other words, 97.1% of all new-borns, that were born 7.3 years ago, attend the first forms (in average). According to the analysis of corresponding data these characteristics are stable during a long time period.

In an analogous way, changes in the pupils' number in the t -th year $\Delta y_{ik}(t)$ were evaluated.

2.2. Modelling of the Teachers' Number

The forecast is usually mined by two contrasting general approaches: the formal extrapolating trends or the functional modelling and simulation based on the nature of data (Brauer and Castillo-Chavez, 2001). The first one treats data formally as statistical numbers, so it is impossible to take into account the known, evolving in time, factors that impact on the data. On the other hand, the trends may be simply implemented.

The second approach is based on the analytic model, which usually uses differential equations or statistical simulation, and the data are used to mine parameters of the model or to verify assumptions on the model. However, sometimes it is difficult to construct a complex model, to evaluate a great number of parameters, or to verify assumptions due to the insufficiency of data. In such cases, we have introduced a mixed approach in which functional models involve factor trends as parameters. Let $x(t)$ be the vector-function of forecasted values at the moment t , k_{-0} be the previously known values of parameters whose forthcoming values will be forecasted as trends $k(t, k_{-0})$, c be the vector of stable parameters and $y(t)$ be the vector of factors to be forecasted by another separate model. Then the model is as follows:

$$x(t) = F(t, X(t-1), k(t, k_{-0}), c, y(t)), \text{ where } X(t-1) = (x(t-1), \dots, x(0), x_{-0}).$$

Therefore we have three types of parameters:

- $k(t, k_{-0})$ are forecasted by trends;
- $y(t)$ are evaluated by modelling or simulation;
- c values are supposed to be stable.

The examples of such an approach in the construction of teachers' job market forecast are presented in the subsection below.

The necessary teachers' number forecast

The simple model (2) presented in this sub-subsection is built on the basis of teachers' job market. It is suitable to forecast the number of necessary employees in some other job markets.

It is important for education officials to forecast the amount of teachers $P(t)$ necessary to serve pupils. The trends of $P(t)$ are not suitable for forecasting because the amount of teachers obviously depends on the number of pupils as well as on many other education policy factors and social circumstances. The number of pupils $y(t)$ may undergo essential changes and it is impossible to forecast $y(t)$ by trends. Therefore the number $y(t)$ is forecasted by model (1) described in Section 2.1. The education policy factors and social circumstances are hardly to be described and evaluated. Therefore we have applied the assumption that the number of teachers is proportional to the number of pupils, and the ratio pupils/teachers $k(t)$ that slowly changes for most education subjects. The ratio accumulates all the social and educational policy factors and may be often forecasted by the trends.

So we have proposed a simple model:

$$P(t) = y(t)/k(t), \tag{2}$$

where $P(t)$ is the necessary number of teachers for some education subject, $y(t)$ is the number of pupils in the forms, in which the subject is taught, $k(t)$ is the ratio y/P for the subject. The ratio $k(t)$ is forecasted by the regression trends based on former data.

Model (2) lies between two traditional approaches: the forecast by trends and by modelling. Some trends of the ratio pupils/teachers, mined from the Lithuanian teachers' database, are presented in Fig. 3.

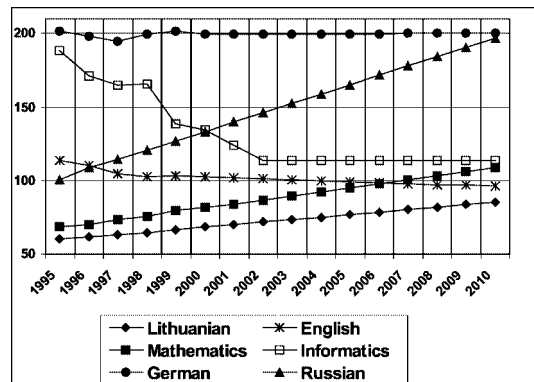


Fig. 3. The mined and forecasted ratio pupils/teachers for some education subjects.

The trends in Fig. 3 are smooth except for the ratio pupils/(teachers of informatics). So, the proposed assumption of the simple model does not apply to the informatics subject due to a fast development of teaching this subject. The ratios found were used to forecast the necessary number of teachers.

The unsatisfied teachers' demand/surplus forecast

Like model (2), model (3), presented in this subsection, is built on the basis of teachers' job market. It is universal, too.

The future supply and demand in the teachers' job market depends on teachers' flows, presented in Fig. 1, as well as on the necessary teachers' numbers evaluated in the previous subsection. An assumption has been made that the main factor to leave a teacher's job is age. The incoming flow into the market consists of mainly young teachers, and we have supposed to forecast this flow by the trend.

Let us describe a model of number D of teachers working in schools or looking for a teacher's job. In order to average variations of different year numbers, the teachers of contiguous age are grouped: x_0 is the number of teachers under 24; x_1 is the number of teachers of age 24 to 27 inclusive; $x_2 - x_{11}$ is the numbers of teachers in the successive groups, spanning age ranges of four years; x_{12} is the number of teachers above 67. The data show that the ratio $x_i(t)/x_{i-1}(t-4)$ is relatively stable, with an exception of two first groups x_0 and x_1 because their size is determined by the income flow and may be forecasted by extrapolating trends.

So we have such a model:

$$\begin{aligned}
 x_j(t) &= f(t, x_{-0j}), \quad j = 0, 1; \\
 x_i(t) &= c_i x_{i-1}(t-4), \quad i = \overline{2, 12}; \\
 D(t) &= \sum_{i=0}^{12} x_i(t).
 \end{aligned} \tag{3}$$

Here t is the year, f is the function that extrapolates the size of two youngest groups on the basis of known former data x_{-0} , c_i is the coefficient of teachers' transition from the

$(i - 1)$ -th group to the i -th group. The coefficient c_i is evaluated from the previous sizes of teachers' groups. It depends on incoming and outgoing teachers' flows and may be slightly more than 1 for some groups. So, the recurrent equations (3) evaluate the number $D(t)$ of teachers, working or looking for a job, under the assumption that tendencies of teachers' training and flows will remain.

The number of active teachers of some speciality or qualification may be forecasted by model (3), depending on the nature of x_{-0} data used.

If the difference between the necessary and active teachers' number $d(t) = P(t) - D(t)$ is positive, then there is an unsatisfied demand, otherwise, there is a surplus in teachers of the considered speciality or qualification.

2.3. Models of School Quality Evaluation

The goal of this section is to develop a method that allows us to qualitatively compare the schools from the standpoint of "city – rural district" or "gymnasium – secondary school". In most cases, education in gymnasias is of a higher quality as compared with that in usual secondary schools. This may often be concluded when comparing education in city schools with the rural district ones. However, how great are these differences? In addition, it would be useful to get some knowledge of the influence of dynamics of the qualification, age, and number of teachers on the state of school. Such an analysis, taking into account the time factor, gives a possibility to observe the changes of differences of schools and to look for the reasons of such a change.

The research is based on 19 schools from the Panevėžys city and 9 schools from the Panevėžys district. In the tables and figures of this section, the schools from the city are labelled by numbers from 1 to 19, and that of the district are labelled by the numbers from 20 to 28. There are two gymnasias (numbers 1 and 2). The remaining schools are the secondary ones.

The analysis covers teachers that work with pupils of 5–12 forms. 1997/1998 and 1999/2000 school year data were analysed. Here the school year is defined by two numbers of years: the school year begins on September 1 of the first year, and ends in June of the second year.

The following indicators were selected to describe a school:

- x_1 – percent of teachers of the highest qualification (degree of a methodologist or expert);
- x_2 – percent of teachers that have not a desired qualification (i.e., who don't do the job they were trained for);
- x_3 – percent of teachers whose age is over 55 years;
- x_4 – percent of teachers who are younger than 35 years;
- x_5 – percent of the annual increase in the number of teachers.

In fact, the indicators describe a school from three points of view: x_1 and x_2 show the qualification of teachers, x_3 and x_4 indicate their age, and x_5 characterises the dynamics of the number of teachers.

For each school, the values of the n -dimensional vector $X_i = (x_{i1}, x_{i2}, \dots, x_{in})$, $i = \overline{1, v}$, may be computed from the databases. In our case, $v = 28$ and $n = 5$, because each school among 28 is described by five indicators. The actual scales of indicators are different. Therefore, it is necessary to normalise them before a further analysis: evaluation of the mean value \bar{x}_j and variation σ_j^2 of each indicator x_j using v its values; normalisation of the values x_{ij} of an indicator using the formula $(x_{ij} - \bar{x}_j)/\sigma_j$.

We suggest analysing visually the normalised vectors X_1, \dots, X_v by means of the mapping methods and artificial neural networks. It is necessary to make some compression of data dimension from n to two. We apply below the self-organising map (SOM), proposed by Kohonen (2001), and its combination with Sammon's (1969) mapping. Additional details on the SOM, Sammon's mapping and their combination are presented in the papers by Dzemyda (2001) and Dzemyda and Kurasova (2002).

Any neurone in the rectangular SOM is entirely defined by its location on the grid (number of row i and column j) and by the codebook vector $m_{ij} = (m_{ij}^1, m_{ij}^2, \dots, m_{ij}^n) \in R^n$. After the learning process, the SOM is self-organised and n -dimensional input vectors X_1, \dots, X_v are mapped – each input vector is related to the nearest neuron (so-called neuron-winner), i.e., the vectors are distributed among the elements of the map. Some elements of the map may remain unrelated with any vector from X_1, \dots, X_v , but there may occur an element related with some input vectors. In fact, using the SOM-based approach, we can draw a two-dimensional table with cells corresponding to the neurons. The cells corresponding to the neurons-winners are filled with the order numbers of vectors X_1, \dots, X_v . Some cells may remain empty. One can visually evaluate the distribution of vectors X_1, \dots, X_v in the n -dimensional space R^n in accordance with their distribution among the cells of the table.

The table, filled according to the distribution of vectors X_1, \dots, X_v among the cells of the SOM, does not answer the question, how much the vectors of the neighbouring cells are close in the n -dimensional space. Therefore, the next stage is to analyse the codebook vectors corresponding to non-empty cells of the table (SOM) by using Sammon's mapping, i.e., to visualize the relative distances between the codebook vectors corresponding to the neurons-winners. In the case of such combined mapping, the neural network performs some sorting (clustering) of data, and Sammon's algorithm presents the results visually to gain an additional insight.

The results of analysis are presented in Tables 1 and 2 and Figs. 4 and 5: Table 1 and Fig. 4 provide the results of the 1997/1998 school year, and Table 2 and Fig. 5 provide those of the 1999/2000 school year. Tables 1 and 2 contain the results of application of the SOM. In both cases there are square tables (four rows and four columns) containing order numbers of schools. The schools in the nearer cells of tables are more similar, and those in the farther ones are more different. The results in Tables 1 and 2 are easier perceived if we analyse them jointly with those in Figures 4 and 5 where Sammon's mapping is applied to the results of the SOM.

Let us note (see Figs. 4 and 5) that the schools of the Panevėžys city and district form several clusters in both school years.

1997/1998 school year. The upper side of Fig. 4 reveals the cluster of city schools with an exception of one district school (23) falling into this cluster: (1, 2, 4–8, 10–12, 14–17,

Table 1

Distribution of the Panevėžys city and district schools on the 4×4 SOM: 1997/1998 school year

3, 27	20, 24, 26		21, 25
18, 28			9, 13, 22
19, 23	14		10, 12
	2, 4, 8, 15, 16, 17	6, 11	1, 5, 7

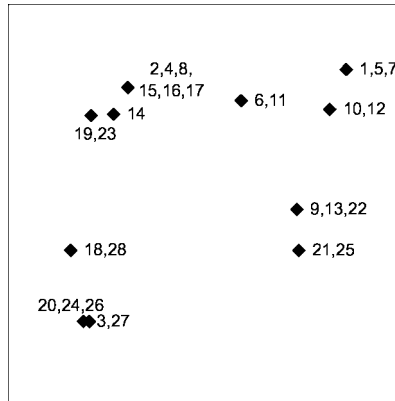


Fig. 4. Results of the combined mapping: distribution of the Panevėžys city and district schools in 1997/1998 school year.

Table 2

Distribution of the Panevėžys city and district schools on the 4×4 SOM: 1999/2000 school year

20	3, 28	18	21, 25, 27
22, 24	13	23	19
7, 10	4, 9, 26		8, 14
1	5, 12	6, 11, 15	2, 16, 17

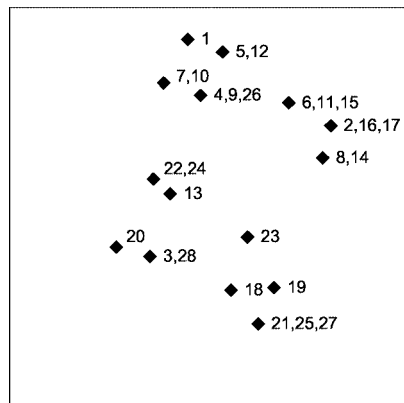


Fig. 5. Results of the combined mapping: distribution of the Panevėžys city and district schools in 1999/2000 school year.

19, 23). This cluster may be visually divided into two subclusters (2, 4, 8, 14–17, 19, 23) and (1, 5, 7, 10, 12): secondary schools of the first subcluster belong to the “attraction zone” of the first gymnasium (1) and the secondary schools of the second one are related with the second gymnasium (2). Two city schools (6, 11) here are in the junction of these two subclusters. In the second and third clusters (3, 18, 20, 24, 26–28) and (9, 13, 21, 22, 25), district schools dominate, however these clusters contain city schools, too.

1999/2000 school year. In Fig. 5 we observe two clusters of schools. In the upper side of Fig. 5 we observe a cluster of city schools where there are both gymnasia. This cluster contains only one district school (26). In comparison with the 1997/1998 school year, relations of the first gymnasium (1) with other schools became weak: in Table 2 as compared with Table 1, the first gymnasium is located in a separate cell. Like in the 1997/1998 school year, this cluster may be divided into two subclusters that are zones of attraction of two gymnasia: (1, 4, 5, 7, 9, 10, 12, 26) and (2, 6, 8, 11, 14–17).

The second cluster (lower side of Fig. 5) contains both city and district schools, but district schools dominate here. It also may be visually divided into two subclusters: (3, 13, 20, 22, 24, 28) and (18, 19, 21, 23, 25, 27). However, these clusters are closer as compared with the second and third clusters in the 1997/1998 school year.

When comparing 1997/1998 and 1999/2000 school years, we observe that the conclusions on the differentiation of schools according to their location (city or rural district) or to their status (gymnasium or secondary school) remain. However, if we consider a separate school, we notice changes in its situation among other schools, and these changes are sometimes essential.

The results above allow suggest a conclusion that the analysis of schools by using artificial neural networks gives a possibility to get quantitative estimates of differences among schools, characterised by the set of five indicators. The experiments on the basis of data about Panevėžys schools proved the possibility of a qualitative comparison of schools from the standpoint “city – rural district” or “gymnasium – secondary school”. Moreover, bearing in mind the results of such an analysis, it is expedient to solve an inverse problem: to determine a preferable set of indicator values that may try to reach a not leading school.

The method, proposed in this section, is universal. It is independent of the selected set of indicators. This opens wide possibilities for improving the existing system of indicators or developing a new one.

2.4. *Model of Staff Quality Evaluation*

We present here an example of data mining from the database in order to gain knowledge on a related phenomenon. The approach is often used in statistics, e.g., the level of purchase of consumer goods is an informative indicator of income.

It is common to evaluate the staff quality by formal indicators, such as educational qualifications, professional ranks, and the previous experience. Such data are in the teachers’ database. However, objective indicators of the staff quality (e.g., examination results of their pupils) seem to be more informative.

In the model of staff quality evaluation, the results of state school-leaving exams were used as the base of objective indices. The object of evaluation was the staff that teaches some subject in the last two forms. Integral results of the exams of each municipality have been analysed.

The municipality was regarded as an independent statistical variable, and the results of exams in a considered subject as well as professional ranks of corresponding staff were regarded as dependent random values.

The results of the state exam in mathematics (Zabulionis, 2000) are presented in Fig. 6 and Fig. 7. The axis y represents the percent of pupils that score more than 74. In Fig. 6, the axis x represents the percent of teachers that have the highest professional ranks (methodologists and experts). In Fig. 7, the axis x represents the percent of mathematicians above 45 in two senior forms. A marker in the chart corresponds to one municipality. The statistical regression trendlines are created by *MS Excel*. Here the interactive actions may be useful because the formal automatic criteria of the best fitness of trendlines sometimes fail in the choice of the trend type.

In the case of the reasonable teachers' assessment, the professional ranks should be correlated with the objective indicator used (integral exam results in different municipalities). However, a confusing result was obtained: the weak relation between the formal teachers' indicators and the objective indicator. R^2 is the coefficient of determination. In Fig. 6 and Fig. 7, the coefficient R^2 is rather low to conclude that the trends are statistically reliable.

Two conclusions may be done:

- the assessment system of teachers is not proper; the significant increase of the higher professional ranks in recent years (Dzemyda, Gudynas *et al.*, 2001) may indicate the adaptation of teachers to formal requirements;
- the pupils' exam results is not a proper measure of teachers' quality.

The phenomenon (see Fig. 6 and Fig. 7) needs an intense study.

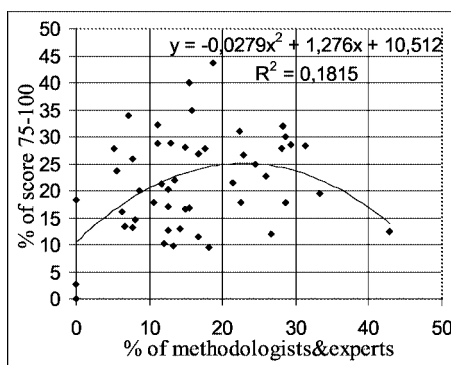


Fig. 6. The impact of professional ranks.

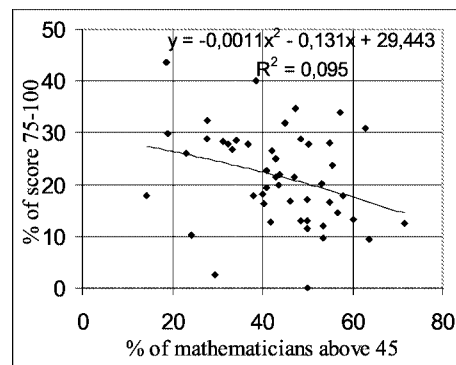


Fig. 7. The impact of age.

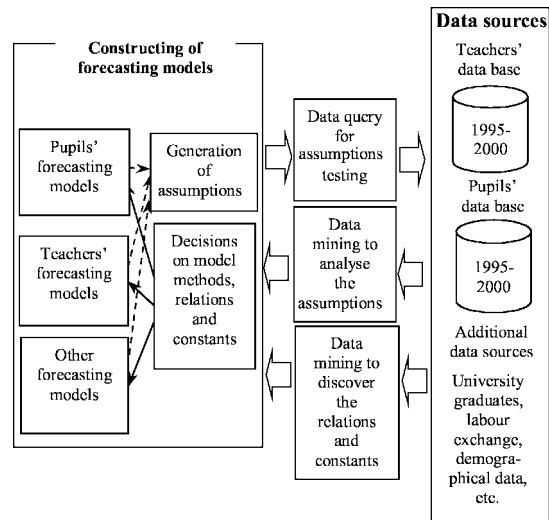


Fig. 8. The flow-chart of the interactive data mining for constructing the forecasting models of the state education system.

3. Interactive Data Mining for Constructing the Models

The interactive data mining scheme for the forecasting models of the Lithuanian state education system is illustrated in Fig. 8. The data sources here consist of a teachers' database, a pupils' database, and some additional data sources containing information about the demographic situation, university graduates, etc.

During the construction of models, the researcher forms models interactively, suggests assumptions, and finally decides on the data mining results.

The main assumptions used in the interactive data mining may be divided into the following groups:

- adequacy of the model to the data structure;
- accessibility, quality and reliability of the data used;
- stability of some values or ratios of values (for example, see Section 2.1).

4. Performance Results: the Case of the Lithuanian Education System

The forecasting scheme used for the Lithuanian education system is presented in Fig. 9.

The total pupils' number forecast is presented in Fig. 10. Here we see that the number of pupils in primary schools (1–4 forms) dramatically decreases. As we can see from the figure the wave of low birth rate will influence the pupils' number of higher forms in the near future.

The forecast of necessary teachers' numbers $P(t)$ is gained, using the discovered pupils' numbers $y(t)$ and mined pupils/teachers ratios $k(t)$ (see model (2)). Obviously, the forecast of model parameters is not exact. Thus, we use two forecasting scenarios:

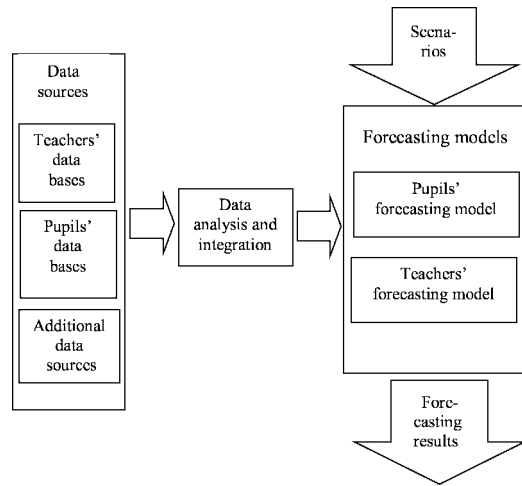


Fig. 9. The forecasting scheme used for the Lithuanian education system.

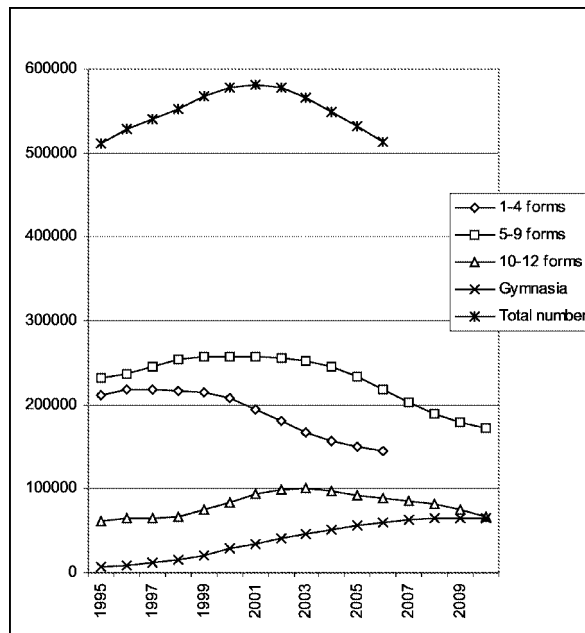


Fig. 10. Pupils' number forecast.

the first one uses the forecasted parameters (see Fig. 3), the second one assumes that the parameters remain constant after the 2002/2003 school year. In Fig. 11, the first forecast is presented by marked continuous lines, and the second one is presented by dashed lines.

Model (3) was used for the forecast number $D(t)$ of working and looking for a job (active) teachers. The transition coefficients c_j , $j = \overline{2, 12}$ and trends of size for two

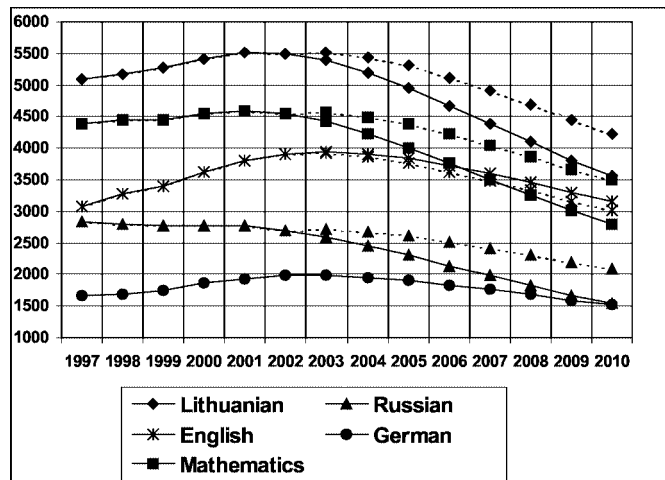


Fig. 11. Forecast of the necessary numbers of teachers.

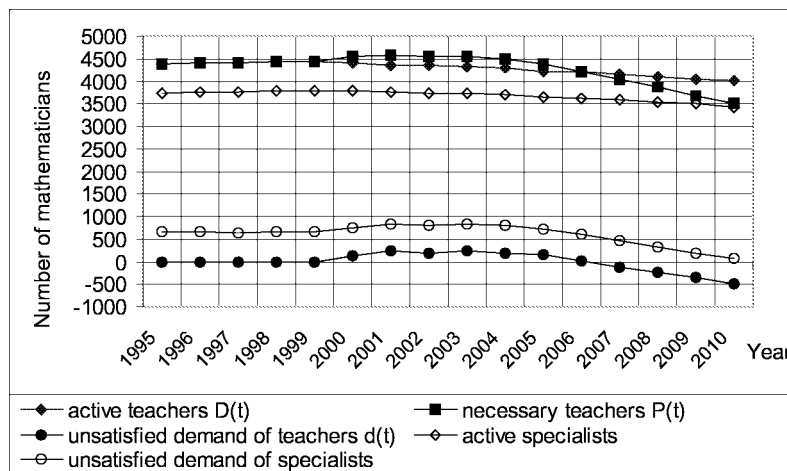


Fig. 12. Forecast of the number and demand for mathematicians.

youngest teachers' groups $x_j(t) = f(t, x_{-0j}), j = 0, 1$ were mined from the Lithuanian teachers' database. The unsatisfied demand/surplus of teachers $d(t) = P(t) - D(t)$ is also important in education policy. The number $D_s(t)$ of professional specialists was forecasted by the same procedure (3). The corresponding unsatisfied demand/surplus of specialists $d_s(t) = P(t) - D_s(t)$ was also forecasted.

E.g., in Fig. 12, the forecast number of mathematicians and their demand is presented for instance. It is interesting to note that only professional mathematicians may be engaged as teachers in 2010. The forecast for all the rest subjects is presented in (Dzemyda, Gudynas *et al.*, 2001).

5. Conclusions

The paper proposes an approach of data mining in the construction of quantitative forecasting and assessment models in the state education system.

In order to obtain satisfactory forecasting results, the models used for these aims should be grounded on some interactive data mining. Process of constructing is usually accompanied by the formulation of some assumptions to background the methods or models.

The main quantitative models developed are as follows: pupils' and teachers' number forecast, teachers' demand forecast, school and its staff quality evaluation.

The models are applied to analyse the Lithuanian teachers' database and data from other sources representing the state of the education system and demographic changes. The analysis enables us to forecast a significant reduction in the teachers' demand. The main reason for this phenomenon is that the number of pupils dramatically decreases in primary schools at present. The wave of low birth rate will influence the pupils' number of higher forms in the near future. The method allowed us to get quantitative estimates of teachers' demand or surplus for various subjects. The discovered knowledge serves as a basis for evidence-based policy of the Lithuanian education system.

6. Acknowledgements

This research is part of the project supported by the Lithuanian Ministry of Education and the Open Society Fund-Lithuania.

References

- Brauer, F., and C. Castillo-Chavez (2001). *Mathematical Models in Population Biology and Epidemiology. Texts in Applied Mathematics*, Vol. 40. Springer-Verlag, New York Berlin Heidelberg.
- Dzemyda, G. (2001). Visualization of a set of parameters characterized by their correlation matrix. *Computational Statistics and Data Analysis*, **36** (1), 15–30.
- Dzemyda, G., P. Gudynas, V. Šaltenis and V. Tiešis (2001). *Teachers and Pupils in Lithuania: Analysis and Prognosis*. Mokslo Aidai, Vilnius (in Lithuanian).
- Dzemyda, G., and O. Kurasova (2002). Comparative analysis of the graphical result presentation in the SOM software resources. *Informatica*, **13** (3), 275–286.
- Han, J., and M. Kamber (2000). *Data Mining: Concepts and Techniques*. Morgan Kaufmann.
- Kohonen, T. (2001). *Self-Organizing Maps, 3rd ed. Springer Series in Information Sciences*, Vol. 30. Springer-Verlag.
- Zabulionis, A. (2000). *The Results of State School-Leaving Exams in Mathematics and History on 2000 in Lithuanian Schools*. National Centre of Examination, Vilnius (in Lithuanian).

V. Šaltenis graduated from the Kaunas Technological Institute, Lithuania, in 1959. He received a Ph.D. degree from the Moscow Energy Institute of the USSR Academy of Sciences in 1966 and the degree of Doctor Habilius from the Institute of Mathematics and Informatics, Vilnius in 1998. He is a principle researcher of the Optimization Department at the Institute of Mathematics and Informatics, Lithuania. His present research interests include both theory and applications of the structure of optimization problems, modelling, multicriteria decision support systems.

G. Dzemyda graduated from Kaunas University of Technology, Lithuania, in 1980, and in 1984 received there the doctoral degree in technical sciences (Ph.D.) after post-graduate studies at the Institute of Mathematics and Informatics, Vilnius, Lithuania. In 1997 he received the degree of Doctor Habilius from Kaunas University of Technology. He was conferred the title of Professor (1998) at Kaunas University of Technology. He is a Deputy Director of the Institute of Mathematics and Informatics and a principal researcher at the Optimization Department of the institute. The main field of scientific interests is the interaction of optimization and data analysis. The interests include optimization theory and applications, multiple criteria decisions, neural networks, and data analysis.

V. Tiešis graduated from the Kaunas University of Technology. He occupies the position of a research fellow at the Institute of Mathematics and Informatics. Alone or with co-authors he has published more than 80 scientific papers and a monograph. His research interests include non-linear and integer optimisation algorithms, conditions of convergence, and efficiency of rough algorithms. His present interests include modelling and control in sociology, epidemiology, economics, military operation, and technology.

Kiekybiniai prognozavimo ir vertinimo modeliai švietimo sistemoje

Vydūnas ŠALTENIS, Gintautas DZEMYDA, Vytautas TIEŠIS

Straipsnyje pasiūlyti kiekybiniai prognozavimo bei kokybės vertinimo modeliai švietimo sistemoje. Modeliai panaudoti Lietuvos švietimo sistemos analizei ir prognozavimui iki 2010 metų. Konstruojant modelius formuluojamos prielaidos apie šių modelių savybes bei parametrus. Šios prielaidos patikrinamos interaktyviu būdu analizuojant žinias, išskirtas iš Lietuvos mokytojų bei mokinių duomenų bazių ir kitų šaltinių, turinčių duomenų apie švietimo sistemą bei demografinius reiškinius. Modeliai apima mokytojų bei mokinių srautų analizę ir prognozę, mokytojų personalo kokybės įverčius ir mokyklų klasifikavimą. Pateikiami minėtų duomenų šaltinių integruotos analizės bei prognozės duomenys, atspindintys švietimo sistemos būklę bei demografinių pokyčių įtaką.