Foundations of Management, Vol. 5, No. 2 (2013), ISSN 2080-7279
DOI: 10.2478/fman-2014-0012

PROJECT PARAMETER ESTIMATION ON THE BASIS OF AN ERP DATABASE

Marcin RELICH

Faculty of Economics and Management, University of Zielona Góra, Poland
e-mail: m.relich@wez.uz.zgora.pl

Abstract: Nowadays, more and more enterprises are using Enterprise Resource Planning (EPR) systems that can also be used to plan and control the development of new products. In order to obtain a project schedule, certain parameters (e.g. duration) have to be specified in an ERP system. These parameters can be defined by the employees according to their knowledge, or can be estimated on the basis of data from previously completed projects. This paper investigates using an ERP database to identify those variables that have a significant influence on the duration of a project phase. In the paper, a model of knowledge discovery from an ERP database is proposed. The presented method contains four stages of the knowledge discovery process such as data selection, data transformation, data mining and interpretation of patterns in the context of new product development. Among data mining techniques, a fuzzy neural system is chosen to seek relationships on the basis of data from completed projects stored in an ERP system.

Key words: knowledge management, new product development, knowledge discovery in databases, data mining, ERP system.

1 Introduction

In recent years, the advancement of information technology in business management processes has placed Enterprise Resource Planning (ERP) system as one of the most widely implemented business software in various enterprises. The use of an ERP system is especially significant in the production enterprises, in which a number of operational processes is enormous. ERP is a system for the seamless integration of all the information flowing through the company such as finances, accounting, human resources, supply chain and customer information [8]. The primary task of an integrated system is to maintain the data flow of an organisation and to reduce the redundancy [12, 16].

The present information and communication technologies have become one of the most important factors, conditions and chances of the company development. These technologies enable the collection, presentation, transfer, access and using of enormous amount of data. The data are a potential source of information that in connection with manager skills and experience may influence on the choice of the correct decision. ERP systems help to collect, operate and store data concerning daily activities of an enterprise (e.g. client orders), as well as the results of previous projects (development of products) [14, 24].

One of the functionalities of an ERP system concerns project management that a company can use to develop new products. Project management can be supported by knowledge management [9], for instance in the context of information acquisition for project financing [13]. To obtain a project schedule, there is required data specification concerning resources and activities, including their sequence and duration. Project parameters can be specified by the experts or estimated with the use of an ERP database. First approach is suitable for the projects that have very unique form, e.g. for the construction projects. In turn, if an enterprise develops new products and a new version of product is connected with the superficial modification of a product specification, then it is possible to acquire the knowledge from the ERP database and to use it for the improvement of estimation quality of project parameters.

The goal of this paper is to present the possibility of the use of the ERP database for seeking the relationships between the ERP attributes (e.g. delay of material delivery by suppliers, number of subcontractors, team members) and the project parameters (e.g. project duration).

The sought relationships can support the user in the assessment of project parameters, and as a consequence to obtain the more relevant estimates. It is unrealistic to expect very accurate estimates of project effort because of the inherent uncertainty in development projects, and the complex and dynamic interaction of factors that influence on its development.

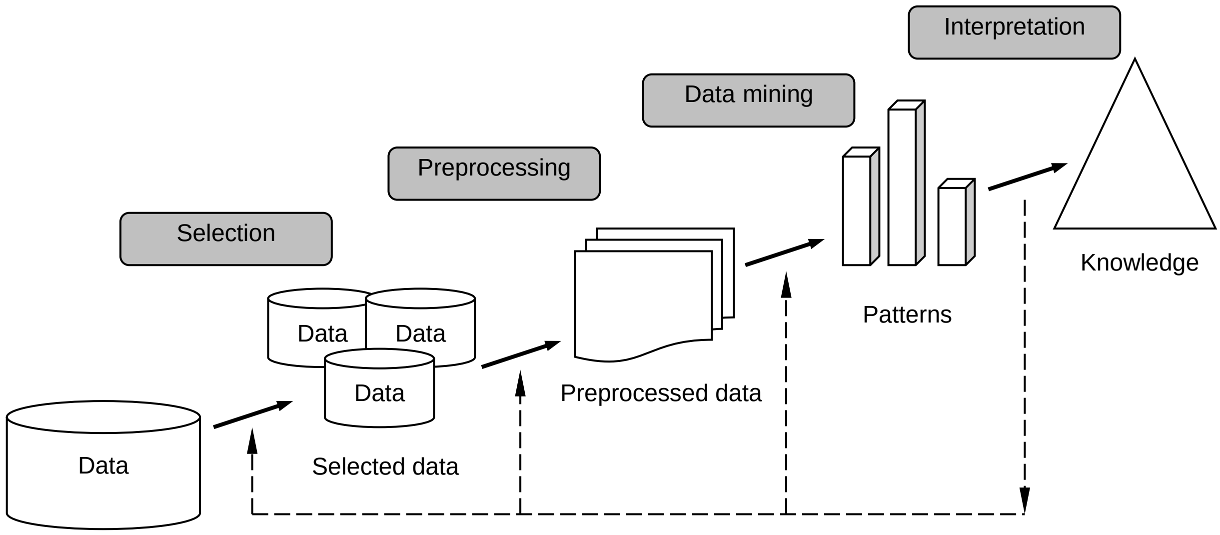

Figure 1. Model of knowledge discovery
(*source: self study base on [10]*)

However, even a small improvement in the estimation quality can influence positively on planning and monitoring the project, for instance, in project cost, resource allocation and schedule arrangement. This is the motivation to try to use the ERP database in order to discover the useful patterns.

2 Model of a knowledge discovery in databases

The knowledge discovery in databases (KDD) is concerned with the development of methods and techniques for making sense of data. KDD is the nontrivial process of identifying valid, novel, potentially useful and ultimately understandable patterns in data. The basic problem addressed by the KDD process is one of mapping low-level data (which are typically too voluminous to understand and digest easily) into other forms that might be more compact (e.g. a short report), more abstract (e.g. a descriptive approximation or model of the process that generated the data), or more useful (e.g. a predictive model for estimating the value of future cases). At the core of the process is the application of specific data-mining methods for pattern discovery and extraction [10].

One of the most used methodologies for developing data mining and knowledge discovery projects is CRoss-Industry Standard Process for Data Mining (CRISP-DM) and KDD process proposed by Fayyad et al. CRISP-DM states which tasks have to be carried out to complete a data mining project.

This methodology consists of the following stages: business understanding, data understanding, data preparation, modelling, evaluation and deployment [15]. In turn, KDD process includes nine steps: developing and understanding of the application domain, creating a target data set, data cleaning and preprocessing, data reduction and projection, choosing the DM task, choosing the DM algorithm, DM, interpreting mined patterns and consolidating discovered knowledge [10].

The steps of the KDD process in the above-presented models can generally be grouped into four main tasks:

- data selection,
- data preprocessing,
- data mining,
- interpretation of patterns.

Figure 1 presents the model of KDD process that is further considered in the context of knowledge acquisition from an ERP database.

The data selection step is connected with an understanding of the application domain and the relevant prior knowledge and identification of the goal of the KDD process from the customer's viewpoint. This step also concerns creating a target data set: selecting a data set, or focusing on a subset of variables or data samples for the knowledge discovery.

Second step – data preprocessing consists of data cleaning, preprocessing and reduction. Basic operations include removing noise if appropriate, collecting the necessary information to model or account for noise, deciding on strategies for handling missing data fields, accounting for time-sequence information and known changes, finding useful features to represent the data depending on the goal of the task.

With dimensionality reduction or transformation methods, the effective number of variables under consideration can be reduced, or invariant representations for the data can be found [10].

Data mining step includes choosing the data mining algorithm(s), selecting method(s) to be used for searching for data patterns, and searching for patterns of interest in a particular representational form or a set of such representations, including classification rules or trees, regression and clustering. The user can significantly aid the data mining method by correctly performing the preceding steps.

Fourth step consists of interpreting mined patterns and sometimes the visualisation of the extracted patterns and models or visualisation of the data given the extracted models. The discovered knowledge can be used directly, incorporating the knowledge into another system for further action, or simply documenting it and reporting it to interested parties. This step also includes checking for and resolving potential conflicts with previously believed (or extracted) knowledge. It is noteworthy that the KDD process can involve iterations and can contain loops between any two steps [10].

The knowledge acquisition from an ERP system requires using the KDD methods that are suitable for the characteristics of an ERP database. The proposed method for seeking relationships in the order of project parameter estimation is presented in the next section.

3 Method for project parameter estimation in the context of ERP systems

The presented method is dedicated for the production enterprises that use an ERP system also to develop the new products. The phases of product development depend on the characteristics of product and company, in which it is designed. However, some common phases can be distinguished, for example [1]:

- plan and define programme,
- product design and development verification,
- process design and development verification,
- product and process validation,
- production.

These phases can also be considered in the context of concept initiation, programme approval, prototype, pilot and launch. Each phase requires the specification of duration and cost. In each of these phases, the critical factors (parameters of an ERP database) that significantly influence on new product development are sought. The estimation of these parameters is especially desired in the medium and large enterprises that develop a few new products simultaneously. In the case of significant variance of a project parameter, the use of the average or time series models can result in the inaccurate estimates. Thus, the search of conditional rules using an ERP database is proposed. The sought relationships can improve the quality of estimates that are input into an ERP system, in a project management module.

Among the KDD steps, two steps seem to be especially important in the context of knowledge acquisition from an ERP database, i.e. data selection and data mining. An ERP database contains dozens of attributes that can be irrelevant to the mining task or redundant. To reduce the data set size, attribute (feature) subset selection can be used that removes irrelevant or redundant attributes (or dimensions). The goal of attribute subset selection is to find a minimum set of attributes such that the resulting probability distribution of the data classes is as close as possible to the original distribution obtained using all attributes. Mining on a reduced set of attributes has an additional benefit. It reduces the number of attributes appearing in the discovered patterns, helping to make the patterns easier to understand [11]. One of the variable reduction methods is principal component analysis that reduces the dimension for linearly mapping high dimensional data onto a lower dimension with minimal loss of information.

Knowledge acquisition requires some data mining techniques that cope with the description of relationships among data and that solve the problems connected with e.g. classification, regression and clustering. These techniques include neural networks, fuzzy sets, rough sets, time series analysis, Bayesian networks, decision trees, evolutionary programming and genetic algorithms, Markov modelling, etc. Data mining should be connected with matching the goals of the KDD process from the user's viewpoint to a particular methods,

Table 1. Input variables for product prototype phase
(*source: self study*)

Suppliers	**Materials Management**	**Production**	**Project Management**
Value of material purchase	Number of materials in warehouses	Productive capacity (actual/maximal)	Number of standard tasks in the project phase
Number of suppliers selling required materials	Number of warehouse transfers	Number of resource overloads	Number of unique tasks in the project phase
Number of delivery reminder documents		Time of machine inspection	Number of changes in the project phase specification
Delivery duration		Number of machines	Number of subcontractors
Delay of delivery		Number of work orders	Number of team members
			Number of materials needed in the project phase

for example, summarisation, classification, regression, clustering. Blind application of data mining methods can lead to the discovery of meaningless and invalid patterns [10].

Database of an ERP system comprises an enormous number of parameters that can be considered as potential variables to identify the project parameters. One of the data mining techniques is fuzzy neural system that can take into account the imprecise character of data, cope with enormous amount of data, and identify the relationships among data.

Fuzzy logic and artificial neural networks are complementary technologies and powerful design techniques that can be used in the identification of patterns from among a large database and noisy data.

The fuzzy neural system has the advantages of both neural networks (e.g. learning abilities, optimisation abilities and connectionist structures) and fuzzy systems (simplicity of incorporating expert knowledge). As a result, it is possible to bring the low-level learning and computational power of neural networks into fuzzy systems and also high-level human like IF-THEN thinking and reasoning of fuzzy systems into neural networks.

The fuzzy neural method is rather a way to create a fuzzy model from data by some kind of learning method that is motivated by learning procedures used in neural networks. This substantially reduces development time and cost while improving the accuracy of the resulting fuzzy model. Being able to utilise a neural learning algorithm implies that a fuzzy system with linguistic information in its rule base can be updated or adapted using numerical information to gain an even greater advantage over a neural network that cannot make use of linguistic information and behaves as a black box [2].

The behaviour of a fuzzy neural system can be represented by a set of humanly understandable rules or by a combination of localised basis functions associated with local models, making them an ideal framework to perform nonlinear predictive modelling. Nevertheless, one important consequence of this hybridisation between the representational aspect of fuzzy models and the learning mechanism of neural networks is the contrast between readability and performance of the resulting model [2].

The combination of fuzzy systems and neural networks has recently become a popular approach in engineering fields for solving problems in control, identification, prediction, pattern recognition, etc. [6–7, 26]. One well-known structure is the adaptive neuro-fuzzy inference system (ANFIS) that is a universal approximator and enables the non-linear modelling and forecasting [17].

4 Illustrative example

The following example refers to four steps of knowledge discovery from an ERP database in the context of new product development. The output variable is the duration (in days) of the j-th phase in project i. In turn, the input variables of the j-th phase in project i are presented in Table 1.

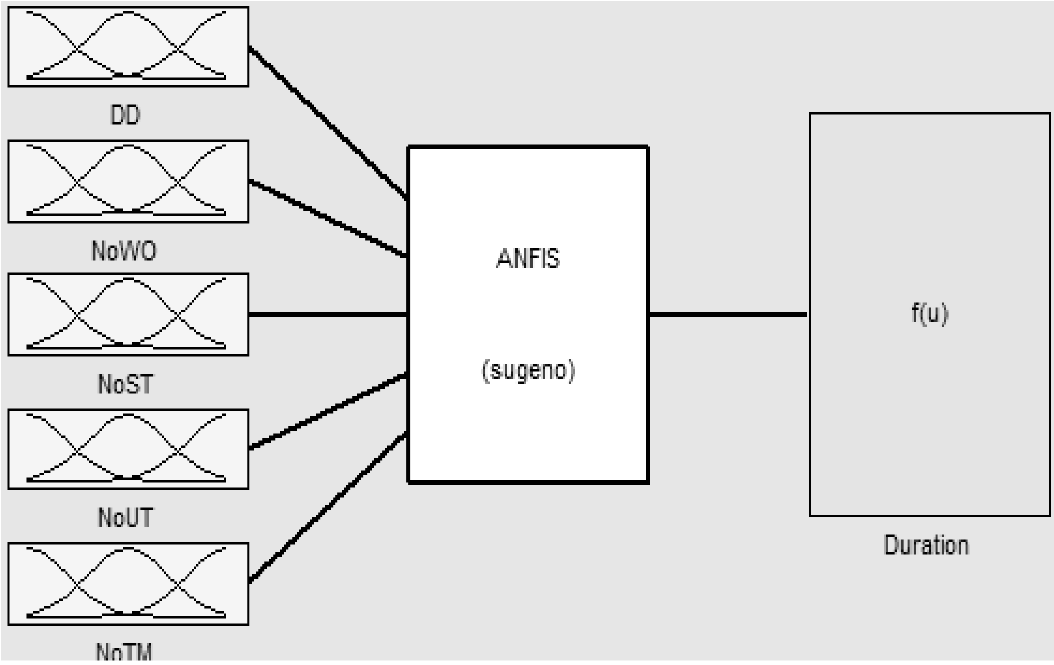

Figure 2. Structure of adaptive neuro-fuzzy inference system
(*source: self study*)

These variables are derived from the ERP system modules that are connected with new product development (project management). The development of a product prototype requires the purchase of materials from the suppliers, storage of materials and usage of materials in production. The first step of the knowledge discovery process concerns data selection and it can be divided into two approaches: the expert chooses the data according to his/her experience and the use of one of the variable reduction methods. The input variables presented in Table 1 have been chosen according to the expert approach. All these variables have the numerical format, but different units, for example, purchasing is in monetary unit, delivery duration in days and productive capacity in percent.

Therefore, the data requires transformation before one of the variable reduction methods is used. The principal component analysis has been chosen as the variable reduction method. This analysis transforms the input data so that the elements of the input vectors will be uncorrelated. In addition, the size of the input vectors may be reduced by retaining only those components that contribute more than a specified fraction of the total variation in the data set. After the use of principal component analysis, the data set has been reduced from 18 input variables to 5 (Delivery duration – DD, Number of work orders – NoWO, Number of standard tasks in the project phase – NoST, Number of unique tasks in the project phase – NoUT, Number of team members – NoTM). Moreover, the data set has been normalised so that it has a zero mean. Data preprocessing (transformation) is the second step of the knowledge discovery process and helps a fuzzy neural system obtain more accurate results.

The third step of the knowledge discovery process regards to the use of data mining techniques/tools. In the order to seek relationships, the adaptive neuro-fuzzy inference system (ANFIS), which is the tool of Matlab® software, has been applied. Figure 2 presents the structure of ANFIS for the duration of project prototype phase.

In order to eliminate the overtraining of ANFIS (too strictly function adjustment to data) and to increase the estimation quality, the data set is divided into learning (P1-P12) and testing set (P13-P15). The learning phase requires the declaration of a membership function type of fuzzy sets (e.g. triangular, Gaussian function), defuzzification method, method of weights optimisation and stop criterion (e.g. error value or the number of iteration). ANFIS tool proposes two methods concerning the identification of the shape of membership functions: grid partition and subtractive clustering method. The shape of membership functions for input variables is presented in Fig. 3.

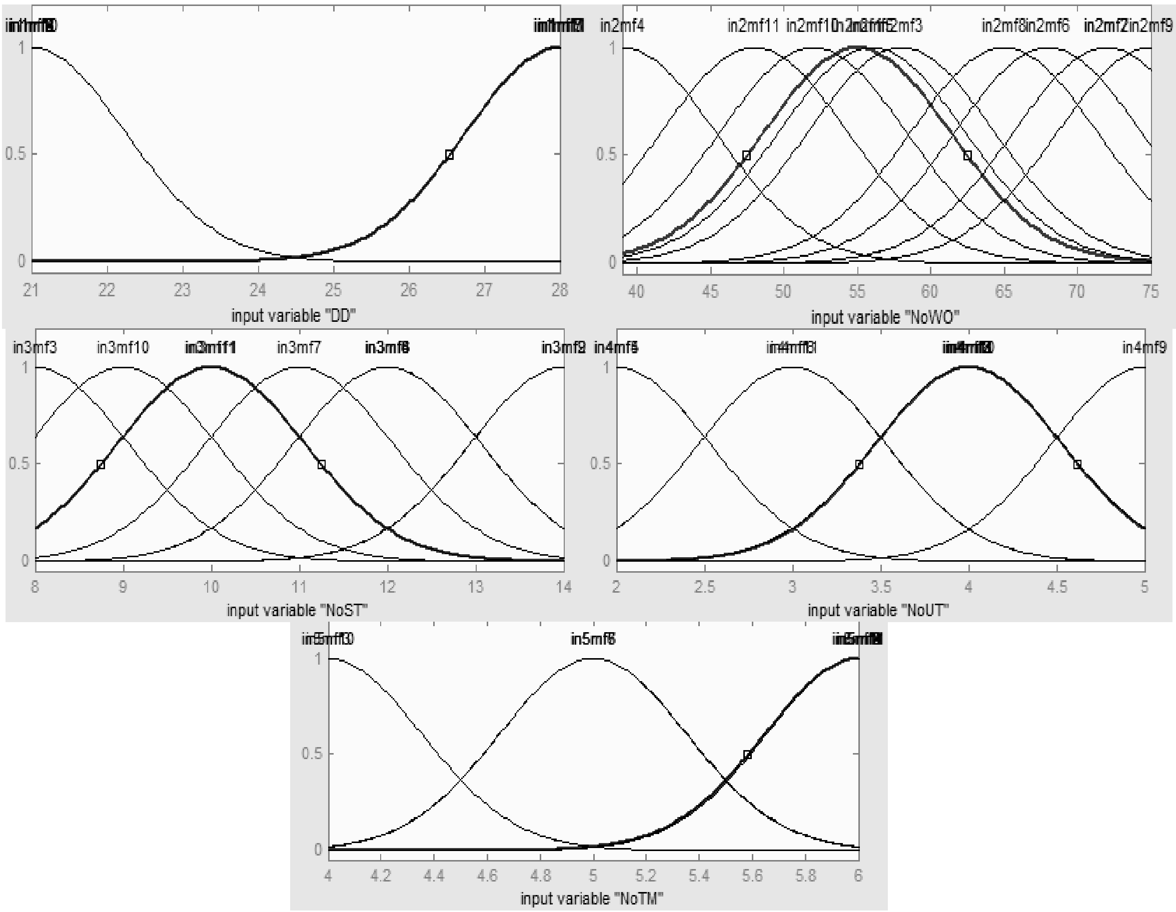

Figure 3. Membership function for input variables
(*source: self study*)

1. If (DD is in1mf1) and (NoWO is in2mf1) and (NoST is in3mf1) and (NoUT is in4mf1) and (NoTM is in5mf1) then (Duration is out1mf1) (1)
2. If (DD is in1mf2) and (NoWO is in2mf2) and (NoST is in3mf2) and (NoUT is in4mf2) and (NoTM is in5mf2) then (Duration is out1mf2) (1)
3. If (DD is in1mf3) and (NoWO is in2mf3) and (NoST is in3mf3) and (NoUT is in4mf3) and (NoTM is in5mf3) then (Duration is out1mf3) (1)
4. If (DD is in1mf4) and (NoWO is in2mf4) and (NoST is in3mf4) and (NoUT is in4mf4) and (NoTM is in5mf4) then (Duration is out1mf4) (1)
5. If (DD is in1mf5) and (NoWO is in2mf5) and (NoST is in3mf5) and (NoUT is in4mf5) and (NoTM is in5mf5) then (Duration is out1mf5) (1)
6. If (DD is in1mf6) and (NoWO is in2mf6) and (NoST is in3mf6) and (NoUT is in4mf6) and (NoTM is in5mf6) then (Duration is out1mf6) (1)
7. If (DD is in1mf7) and (NoWO is in2mf7) and (NoST is in3mf7) and (NoUT is in4mf7) and (NoTM is in5mf7) then (Duration is out1mf7) (1)
8. If (DD is in1mf8) and (NoWO is in2mf8) and (NoST is in3mf8) and (NoUT is in4mf8) and (NoTM is in5mf8) then (Duration is out1mf8) (1)
9. If (DD is in1mf9) and (NoWO is in2mf9) and (NoST is in3mf9) and (NoUT is in4mf9) and (NoTM is in5mf9) then (Duration is out1mf9) (1)
10. If (DD is in1mf10) and (NoWO is in2mf10) and (NoST is in3mf10) and (NoUT is in4mf10) and (NoTM is in5mf10) then (Duration is out1mf10) (1)
11. If (DD is in1mf11) and (NoWO is in2mf11) and (NoST is in3mf11) and (NoUT is in4mf11) and (NoTM is in5mf11) then (Duration is out1mf11) (1)

Figure 4. Fuzzy rules for duration assessment
(*source: self study*)

According to the shape of membership functions, the rules are built. The example of fuzzy rules for the duration is presented in Fig. 4.

After learning phase, the testing data are led to input of system to compare the error between different models. Root mean square errors (RMSE) for various models are presented in Table 2.

Table 2. Comparison of RMSE for different models
(*source: self study*)

Model	**RMSE**
Average	13.06
Linear model	12.65
ANFIS – grid partition	5.88
ANFIS – subtractive clustering	3.74

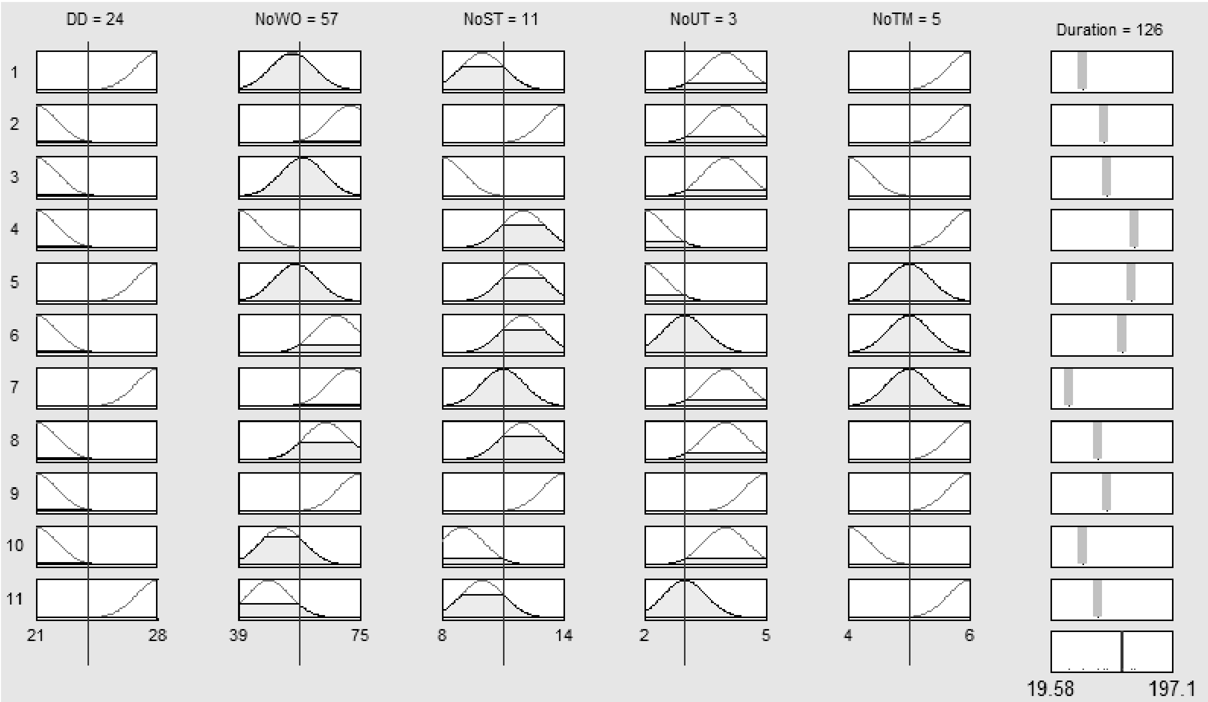

Figure 5. Estimation of project duration
(*source: self study)*

The least error in testing set for the duration of project prototype phase has been generated with the use of ANFIS with subtractive clustering method. The comparison of different models is especially recommended in the case of low level of variance for a dependant variable (in the considered case for the duration of a project phase). The membership functions and rules are a basis to evaluate the duration of an actual project. Let us assume that for the actual project are considered the following values: delivery duration – 24 days, number of work orders – 57, number of standard tasks in the project phase – 11, number of unique tasks in the project phase – 3, and number of team members – 5.

Figure 5 presents the memberships functions for 8 rules that determine the planned duration of project phase at 126 days.

The presented analysis can be broadened into multidimensional analysis to support the decision-maker in determining optimal values of some parameters. For instance, the decision-maker would like to know the number of team members (from 5 to 6) and the number of work orders (from 55 to 65), for which the planned duration of project phase is the least. Table 3 presents the results for the above-mentioned constraints. The results indicate that the minimal duration is for 6 members of project team and 55 number of work orders. Moreover, the increase of project duration at 5 team members is proportionally larger than at 6 members

Table 3. Example of what-if analysis for project duration
(*source: self study*)

Number of work orders	**Number of team members**	
	5	**6**
55	122	88.0
56	124	88.5
57	126	89.4
58	127	90.4
59	129	92.2
60	131	93.8
61	133	95.6
62	134	97.4
63	136	99.2
64	138	101.0
65	139	103.0

For instance, the reduction of work orders from 60 to 55 for 5 team members decreases project duration by 9 days (131 days – 122 days), whereas for 6 team members, the decline equals about 6 days (93.8 days – 88 days). In the case of extensive search space, the time of obtained solution can be significantly reduced, e.g. using constraints programming languages [4–5, 18–21, 25]. The above-presented analysis is conducted for each phase of project and the obtained estimates can be used for further evaluation, e.g. in the planning of cash flow, working capital and financial reserves.

5 Closing remarks

The knowledge discovery process in the context of an ERP database can be considered as four steps: data selection, data transformation, data mining and pattern interpretation. An enormous number of data and attributes in an ERP database require paying attention to the proper choice of variable reduction and data mining methods. In the paper, the principal component analysis has been chosen as the variable reduction method and the fuzzy neural system as data mining technique.

The fuzzy neural system has the advantages of both neural networks (e.g. learning abilities, optimisation abilities and connectionist structures) and fuzzy systems (simplicity of incorporating expert knowledge). As a result, it is possible to bring the learning and computational power of neural networks into fuzzy systems and also human like if-then thinking and reasoning of fuzzy systems into neural networks.

The advantages of proposed approach include the search of conditional rules into an ERP database and using them for the project parameter estimation. This is especially important in the case of significant variance of a project parameter, for which the average and time series models result in the inaccurate estimates. More exact identification of project duration and cost enables more precision of project planning and control, as well as the improvement of cash flow planning. The proper choice of a set of new products can lead to increasing the market share, profitability and liquidity of an enterprise. The proposed approach is especially valuable for an enterprise that has a database of past projects, because there is the possibility to gather additional information in the form of conditional rules. The application of the proposed approach encounters some difficulties, among other things, by the collecting enough amounts of data of the past similar projects. Moreover, the lack of uniform rules that concern the development of fuzzy neural systems may cause an acceptance problem for the decision-makers. However, the presented approach seems to have the promising properties for acquiring information from an ERP system.

Further research focuses on the development of the proposed method for project parameter estimation in the context of ERP systems, for instance, towards the choice of an optimal set of new products with the use of constraint programming techniques. The value of discovered patterns obviously depends on the quality of database of an ERP system.

The data quality can be improved through using standards or methodology (according to the certification process) for managing of the project [22], proper project communication [23] and the education approach in an organisation [3].Hence, future research can be aimed at adjusting the proposed approach in the aspect of other project factors, including the improvement of training system in the company and communication in the project team.

6 References

[1] *Advanced product quality planning and control plan. Reference Model.* Carwin Continuous Ltd., Essex 1994.

[2] Azar A.T. - *Adaptive neuro-fuzzy systems* [in] Fuzzy systems. InTech 2010.

[3] Bazsova B. - *Use of Attis software tool in education* [in] Proceedings of the 10th International Conference on Strategic Management and its Support by Information Systems, pp. 23-34, VSB - Technical University of Ostrava 2013.

[4] Banaszak Z., Zaremba M., Muszyński W. - *Constraint programming for project-driven manufacturing* [in] Int. J. Production Economics, pp. 463-475, Vol. 120, 2009.

[5] Bocewicz G., Banaszak Z. - *Declarative approach to cyclic steady state space refinement: periodic process scheduling* [in] The International Journal of Advanced Manufacturing Technology, Springer, pp. 137-155, Vol. 67, Issue 1-4, 2013.

[6] Cheng M.Y., Tsai H.C., Sudjono E. - *Evolutionary fuzzy hybrid neural network for project cash flow control* [in] Engineering Applications of Artificial Intelligence, pp. 604-613, Vol. 23, 2010.

[7] Chien S.C., Wang T.Y., Lin S.L. - *Application of neuro-fuzzy networks to forecast innovation performance* [in] Expert Systems with Applications, pp. 1086-1095, Vol. 37, 2010.

[8] Davenport T. - *Putting the enterprise into the enterprise system* [in] Harvard Business Review, pp. 121-131, July-August 1998.

[9] Doskočil R., Smolíková L. - *Knowledge management as a support of project management* [in] Knowledge for Market Use 2012, Societas Scientiarum Olomucensis II, pp. 40-48, Olomouc 2012.

[10] Fayyad U., Piatetsky-Shapiro G., Smith P. - *From Data Mining to Knowledge Discovery in Databases*. [in] American Association for Artificial Intelligence, pp. 37-54, Fall 1996.

[11] Han J., Kamber M. - *Data Mining. Concepts and Techniques*. 2nd edition, Morgan Kaufmann Publishers, San Francisco 2006.

[12] Imtiaz A., Kibria M.G. - *Modules to optimize the performance of an ERP based integrated information system*. IEEE International Conference on Informatics, Electronics & Vision, pp. 598-601, 2012.

[13] Kormancová G. - *Project financing in Slovak companies* [in] European Scientific Journal, University of the Azores, Portugal, Special edition N. 1, pp. 367-370, June 2013.

[14] Lis M. - *Zarządzanie portfelem projektów jako element strategii oraz sposób zarządzania w organizacji*. Wydawnictwo Akademii Techniczno-Humanistycznej, pp. 126-139, Bielsko Biała 2010.

[15] Marban O., Mariscal G., Segovia J. - *A data mining & knowledge discovery process model* [in] Data Mining and Knowledge Discovery in Real Life Applications, I-Tech,Vienna 2009.

[16] May J., Dhillon G., Caldeira M. - *Defining value-based objectives for ERP systems planning* [in] Decision Support Systems, pp. 98-109, Vol. 55, 2013.

[17] Relich M. - *A decision support system for alternative project choice based on fuzzy neural networks* [in] Management and Production Engineering Review, pp. 46-54, Vol. 1, No. 4, 2010.

[18] Relich M. - *Project prototyping with application of CP-based approach* [in] Management, pp. 364-377, Vol. 15, No. 2, 2011.

[19] Relich M., Banaszak Z. - *Reference model of project prototyping problem* [in] Foundations of Management - International Journal, pp. 33-46, Vol. 3, No. 1, 2011.

[20] Relich M. - *Fuzzy project scheduling using constraint programming* [in] Applied Computer Science, pp. 3-16, Vol. 9, No. 1, 2013.

[21] Rostański M. - *Business process analysis with the higher education institution example* [in] Proceedings of the 10th International Conference on Strategic Management and its Support by Information Systems, pp. 173-182, VSB - Technical University of Ostrava 2013.

[22] Samáková J., Šujanová J. - *Project management certification approaches in Slovak industry enterprises* [in] Efficiency and Responsibility in Education 2013: proceedings of the 10th International Conference, pp. 542-549, Czech University of Life Sciences Prague 2013.

[23] Samáková J., Šujanová J., Koltnerová K. - *Project communication management in industrial enterprises* [at] Proceedings of the 7th European Conference on Information Systems Management and Evaluation, pp. 155-163, University of Gdańsk 2013.

[24] Sikora M., Borowski Z., Majchrzak M. - *Zintegrowane systemy informatyczne w organizacji gospodarki magazynowej* [in] Logistyka No. 2, 2012.

[25] Sitek P., Wikarek J. - *A declarative framework for constrained search problems* [in] New Frontiers in Applied Artificial Intelligence, Lecture Notes in Artificial Intelligence, Nguyen, N.T., et al. (Eds.), Vol. 5027, Springer-Verlag, Berlin-Heidelberg, pp. 728-737, 2008.

[26] Zeng J., An M., Smith N.J. - *Application of a fuzzy based decision making methodology to construction project risk assessment* [in] International Journal of Project Management, pp. 589-600, Vol. 25, 2007.