Knowledge Discovery in Enterprise Databases for Forecasting New Product Success

Marcin Relich[1(✉)] and Krzysztof Bzdyra[2]

[1] Faculty of Economics and Management, University of Zielona Gora, Zielona Gora, Poland
`m.relich@wez.uz.zgora.pl`
[2] Faculty of Electronic and Computer Engineering, Koszalin University of Technology,
Koszalin, Poland
`krzysztof.bzdyra@tu.koszalin.pl`

Abstract. This paper presents the knowledge discovery process that aims to improve the forecast quality of the success of new product development projects. The critical success factors for new product development are identified on the basis of information acquired from an enterprise system, including the fields of sales and marketing, research and development, production, and project management. The proposed knowledge discovery process consists of stages such as data selection from enterprise databases, data preprocessing, data mining, and the use of the discovered patterns for forecasting new product success. The illustrative example presents the use of fuzzy neural networks for forecasting net profit from new products.

Keywords: Knowledge retrieval · Data mining · Fuzzy neural systems · New product development · Rule-based systems

1 Introduction

Popularity of data mining and knowledge discovery is caused by an increasing demand for tools that help reveal and comprehend information hidden in huge amount of data [1]. Such data is routinely generated in scientific domains (e.g. astronomical, biological data) as well as through business transactions in banks, credit card companies, supermarkets, and other organizations. Data reflecting business transactions is stored increasingly in an integrated enterprise information system such as enterprise resource planning or customer relationship management system. As the amount of available data in companies becomes greater and greater, companies have become aware of an opportunity to derive valuable information from their databases, which can be further used to improve their business [2]. One of the most important aspects of continuing business is the successful new product development (NPD) and launch.

The NPD process allows the company to maintain its competitive position and continue business success. This process consists of the stages such as idea generation, idea screening, concept testing, development of the selected concepts, testing and commercialisation of new products [3]. As the success of the NPD projects is closely connected with the success of the entire company, the NPD projects play an important

role in an organisation's growth and development [4]. In turn, the unsuccessful products may decrease the volume sales, profit, and finally, lead the company to bankruptcy. To survive and succeed in the dynamic business environment, companies usually focus on several areas to improve their new product development, such as identifying customer needs for continuous new product development, improving product quality, and accelerating the process of commercialization [5].

The new product development literature focuses on identifying success factors of new products with the use of multi-sectors surveys that often base on evaluating NPD process, organization, culture, and company strategy [6–8]. These surveys are useful to identify the key success factors, which bring positive influence on NPD (e.g. clear definition of product before development begins, high-quality preparatory work), and trends in the context of the various industries. However, the usefulness of multi-sectors surveys is limited for forecasting the success of a new product in the specific company. Therefore, the proposed approach bases on the success factors of NPD that are identified with the use of an enterprise system. The advancement of information technology helps today's organisations in business management processes and collecting data that is potential source of information [9]. A main task faced by NPD projects is how to acquire knowledge and sustain success rate among the products.

The previous research has focused mainly on the use of the econometric models for forecasting the success of a new product [10, 11]. This paper aims to present the use of fuzzy neural network as data mining techniques to forecast the success of new products and compare the forecast quality with the econometric models. The proposed approach takes into account data of the previous projects that is stored into an enterprise system. The remaining sections of this paper are organised as follows. Section 2 presents the knowledge discovery and data mining process. The proposed method of forecasting new product success is presented in Sect. 3. An example of the proposed approach is illustrated in Sect. 4. Finally, some concluding remarks are contained in Sect. 5.

2   Knowledge Discovery and Data Mining Process

Knowledge discovery in databases is a nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data [12]. Knowledge discovery (KD) is concerned with the development of methods and techniques for making sense of data. The basic problem addressed by KD is mapping low-level data (which is 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) [12].

A knowledge discovery process can be described as a series of steps that differ in their number and range. Fayyad et al. presented the nine-step model that includes developing and understanding of the application domain, creating a target data set, data cleaning and preprocessing, data reduction and projection, choosing the data mining task, choosing the data mining algorithm, data mining, interpreting mined patterns, and finally consolidating discovered knowledge [12]. Cabena et al. described five steps of

the KD process: business objectives determination, data preparation, data mining, domain knowledge elicitation, and assimilation of knowledge [13]. Cross-Industry Standard Process for Data Mining (CRISP-DM) states which tasks have to be carried out to complete a data mining project. These tasks contain business understanding, data understanding, data preparation, modeling, evaluation, and deployment [14]. Cios et al. adapted the CRISP-DM towards providing a more general, research-oriented description of the KD model through the following steps: understanding the domain, understanding the data, preparation of the data, data mining, evaluation of the discovered knowledge, and using the discovered knowledge [1, 14]. In general, the presented KD models include the stages such as data understanding and selection, data preprocessing, data mining, and implementation of the discovered patterns [15]. This framework of the KD process is further considered in this study (see Fig. 1).
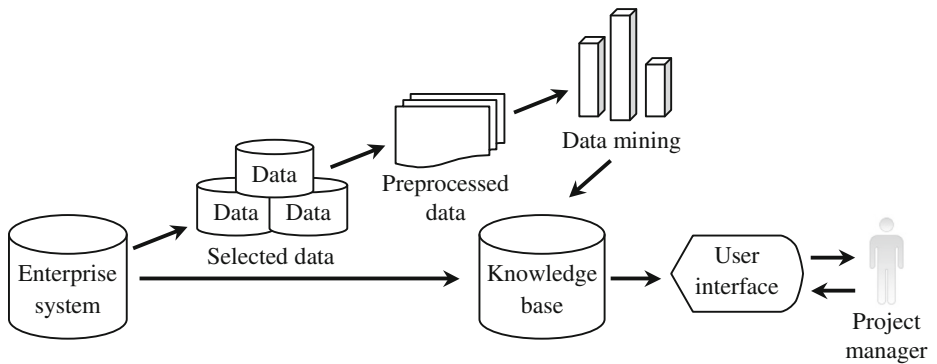

Fig. 1. Framework of decision support system for forecasting new product success

At the core of the KD process is the application of specific data mining methods for pattern discovery and extraction [12]. Data mining tasks can be classified as descriptive and predictive [15]. While the descriptive techniques provide a summary of the data, the predictive techniques learn from the current data in order to make predictions about the behaviour of new data sets. The most commonly used tasks in data mining include classification, clustering, associations, visualization, and forecasting that is further considered. Among data mining techniques, a fuzzy neural system has been chosen for identifying patterns and forecasting new product success.

The fuzzy neural system has the advantages of both neural networks (e.g. learning abilities, optimization abilities and connectionist structures) and fuzzy systems (e.g. if-then reasoning, simplicity of incorporating expert knowledge). The outcome of a fuzzy neural system is a set of if-then rules that can be used to perform nonlinear predictive modelling, classification, simulation, and forecasting [16, 17]. One well-known structure of fuzzy neural networks is the adaptive neuro-fuzzy inference system (ANFIS) that has been used in this study.

The methodology of the proposed approach is based on CRISP-DM model in which understanding the business and data is provided by a domain expert, the variables to analysis are preprocessed according to principal component analysis, data mining is

implemented in the ANFIS, and discovered knowledge is used for forecasting the success of a new product in order to support the project manager.

3   The Proposed Method of Forecasting New Product Success

The success of a new product depends on several external (e.g. customer demand, changes in technology, environmental regulations) and internal factors (resources and processes used during new product development). The challenge is to increase the forecast quality of product success on the basis of an accessible enterprise database. The forecasts of the product success can help the project manager in taking decision of continuing or deleting NPD projects.

Nowadays, more and more enterprises support their business processes using an enterprise information system such as enterprise resource planning (ERP), customer relationship management (CRM), or computer-aided design (CAD) [18]. Databases of these systems store specifications of the previous NPD projects, customer requirements, and design parameters. Consequently, an enterprise database is a potential source of information that can be revealed. Figure 1 illustrates the framework of knowledge discovery from an enterprise system and the use of knowledge base for forecasting new product success.

The presented approach includes the stages of the knowledge discovery process such as data understanding and selection, data preprocessing, data mining, addition of the discovered patterns to knowledge base, and the use of knowledge base for forecasting new product success. A domain expert selects from an enterprise system the data that is suspected of the impact on the product success. The expert can also add some facts and constraints (e.g. amounts of resources, the number of activities in a NPD project) directly to knowledge base. In the next step, the chosen variables are reduced and preprocessed according to principal component analysis. Taking into account good forecasting properties and the possibility of obtaining if-then rules that can be stored in knowledge base, among the data mining techniques, the ANFIS has been chosen in this study. The identified patterns are stored in the form of rules in knowledge base that helps the project manager forecast the success of a new product. The forecasts can be further used for selecting a set of the most promising new products [19].

Taking into account the large number of potential new products, there is a need to use some techniques that reduce a search space. The processing time of calculations can be significantly reduced with the use of constraints programming techniques [20]. Constraints programming has embedded ways to solve constraints satisfaction problems with greatly reduction of the amount of search needed [21]. This is sufficient to solve many practical problems such as supply chain problem [22, 23] or scheduling problem [24].

Selection of new products for further development usually bases on metrics of the product success, and it should also take into account the company's resources. A successful NPD project can be considered as the achievement of an acceptable level of performance such as sales, market share, return on investment, profit, or customer satisfaction [8]. Taking into account the product lifetime and return on product

development expense, the average net profit from a product per month is considered as a metric of the product success in this study. This metric is estimated on the basis of the previous developed products which specification can be retrieved from an enterprise system. The next section presents an example of using the proposed decision support system for forecasting the success of a new product.

4   Illustrative Example

The example refers to the above-described KD process and includes the steps such as data selection, data preprocessing, data mining, and the use of the retrieved patterns for forecasting the success of a new product. Table 1 presents the input variables of four fields that are suspected of the impact on the success of the NPD project.

Table 1. Input variables

Marketing and sales	R&D	Production	Project management
Number of customer requirements for the NPD project	Percentage of existing parts used in the NPD project	Number of resource overloads in the production phase of NPD	Number of project team members
Number of customer complaints for the NPD project	Percentage of customer requirements translated into speci-fication	Number of work orders	Average time spent on communication in project team members
Duration of marketing campaign of the product	Number of standard activities in the NPD project	Number of subcontractors for the NPD project	Duration of the NPD project
Cost of marketing campaign of the product	Number of activities in the NPD project	Unit cost of production for the product	Cost of the NPD project

Data selection can be divided into two phases. First, the expert selects the variables from the enterprise database according to his/her experience (Table 1). In the second phase, the optimal number of variables is sought with the use of a variable selection method. In this study, the stepwise procedure has been used, in which 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 using the stepwise procedure, the data set has been reduced from 16 input variables to 5 following variables: percentage of customer requirements translated into technical specification (CR), duration of the NPD project (D), cost of the NPD project (C), unit cost of production for the product (UC), and cost of marketing campaign for the product (MC).

The considered example includes data from 33 previous product development projects (P1-P33) that derive from 3 separate product lines. The collected data has a

numerical format but different units, for example, cost of the NPD project is measured in monetary unit, duration of the NPD project in days, whereas the share of customer requirements translated into technical specification is measured in percent. As a result, the data requires transformation in order to use data mining techniques more effective. Data preprocessing facilitates the improvement of modeling accuracy of a fuzzy neural system (ANFIS). In this study, the principal component analysis has been used to transform data before the learning phase of ANFIS.

Data mining stage is connected with revealing the hidden relationships with the use of ANFIS. In order to reduce the overtraining of ANFIS and to increase the forecast quality, the data set has been divided into learning (P1-P27) and testing sets (P28-P33). Fuzzy inference system has been generated with the use of two methods: grid partition and subtractive clustering. The grid partition (GP) method has used three triangular membership functions for each input variables, and constant and linear type of output membership function (MF). In turn, the subtractive clustering (SC) method has used Gaussian membership functions and different values of the parameter concerning the range of influence (RI). After learning phase, the testing data has been led to input of the fuzzy neural system to compare the RMSE for ANFIS and other models. The results have been obtained in the Matlab® software and presented in Table 2 as the root mean square errors (RMSE) for the learning set (LS) and testing set (TS), as well as the number of rules generated. The comparison also includes the average and linear regression model.

Table 2. Comparison of forecasting models

Model	RMSE for LS	RMSE for TS	Number of rules
ANFIS (SC, RI = 0.3)	0.001	68.78	15
ANFIS (SC, RI = 0.5)	0.134	21.35	6
ANFIS (SC, RI = 0.7)	1.58	6.97	3
ANFIS (GP, MF = constant)	0.005	29.18	243
ANFIS (GP, MF = linear)	0.0001	29.25	243
Linear regression	20.57	20.88	1
Average	69.85	74.32	1

The results presented in Table 2 indicate the importance of parameter adjustment of specific data mining techniques. All the RMSE generated by ANFIS are less than the average, but only one case (ANFIS trained with the use of subtractive clustering method and range of influence equals 0.7) has obtained the RMSE less than linear regression model. The presented forecasting models also differ in the number of rules that influence the performance of decision support system. The identified relationships can be described as if-then rules and used for forecasting net profit from a new product. Let us assume that for the actual NPD project the following values are considered: customer requirements translated into technical specification – 70 %, duration of the NPD project – 30 weeks, cost

of the NPD project – 65 monetary units, unit cost of production for the product – 0.35 monetary unit, and cost of marketing campaign for the product – 90 monetary units. Figure 2 presents three rules generated by ANFIS (case: SC, RI = 0.7) that have been used for forecasting net profit which reaches 227 monetary units for the considered product. The forecasts of net profits from the NDP projects can be further used for seeking the NPD project portfolio that ensures the maximal total net profit from all products.
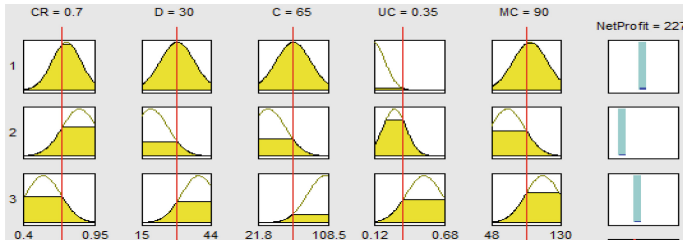

Fig. 2. Forecast of net profit with the use of ANFIS

5   Conclusions

The characteristics of the presented approach includes the use of an enterprise system database to knowledge retrieval and fuzzy neural networks to seek the patterns in the context of forecasting new product success, as well as the description of the identified patterns in the form of if-then rules that are suitable for human-like reasoning. The knowledge base includes the rules identified by fuzzy neural network or/and an expert, facts (including company's resources), and it allows the project managers to obtain the forecast of net profit from the NPD projects.

This research includes the comparison of the different methods in the context of the forecast quality and the number of rules generated. The results indicate that the forecast quality with the use of fuzzy neural networks can outperform the forecasting models that base on the average and linear regression. However, there is needed the adjustment of learning parameters of a fuzzy neural network what can be treated as a disadvantage. Nevertheless, fuzzy neural networks are able to identify the nonlinear and complex patterns (if there are any), develop the knowledge base, and finally, use these patterns to perform nonlinear predictive modelling, simulation, and forecasting.

The identified patterns are used in the decision support system to help the managers in forecasting the success of the NPD projects, selecting the most promising products, and reducing the risk of unsuccessful product development. On the other hand, the application of the proposed approach encounters some difficulties by collecting enough amount of data of the past similar NPD projects and ambiguous principles to build the structure of a fuzzy neural network. Nevertheless, the presented approach seems to have the promising properties for acquiring additional information from an enterprise system and using them to forecasting new product success. The further research aims to develop an interactive decision support system for identifying the most promising product portfolio according to the user's preferences.

References

1. Cios, K.J., Kurgan, L.A.: Trends in data mining and knowledge discovery. In: Pal, N., Jain, L. (eds.) Advanced Techniques in Knowledge Discovery, pp. 1–26. Springer, London (2005)
2. Li, T., Ruan, D.: An extended process model of knowledge discovery in database. J. Enterp. Inf. Manage. **20**(2), 169–177 (2007)
3. Trott, P.: Innovation Management and New Product Development. Prentice Hall, Essex (2005)
4. Spalek, S.: Does investment in project management pay off? Ind. Manage. Data Syst. **114**(5), 832–856 (2014)
5. Chan, S.L., Ip, W.H.: A dynamic decision support system to predict the value of customer for new product development. Decis. Support Syst. **52**, 178–188 (2011)
6. Mishra, S., Kim, D., Lee, D.: Factors affecting new product success: cross-country comparisons. J. Prod. Innov. Manage **13**(6), 530–550 (1996)
7. Lynn, G., Schnaars, S., Skov, R.: A survey of new product forecasting practices in industrial high technology and low technology businesses. Ind. Mark. Manage. **28**(6), 565–571 (1999)
8. Ernst, H.: Success factors of new product development: a review of the empirical literature. Int. J. Manage. Rev. **4**(1), 1–40 (2002)
9. Relich, M.: Knowledge acquisition for new product development with the use of an ERP database. In: Federated Conference on Computer Science and Information Systems (FedCSIS), pp. 1285–1290 (2013)
10. Hardie, B., Fader, P., Wisniewski, M.: An empirical comparison of new product trial forecasting models. J. Forecast. **17**, 209–229 (1998)
11. Kahn, K.: An exploratory investigation of new product forecasting practices. J. Prod. Innov. Manage **19**, 133–143 (2002)
12. Fayyad, U., Piatetsky-Shapiro, G., Smith, P.: From data mining to knowledge discovery in databases. Am. Assoc. Artif. Intell. 37–54 (1996). Fall
13. Cabena, P., Hadjinian, P., Stadler, R., Verhees, J., Zanasi, A.: Discovering Data Mining: From Concepts to Implementation. Prentice Hall, Saddle River (1998)
14. 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 (2009)
15. Han, J., Kamber, M.: Data Mining. Concepts and Techniques. Morgan Kaufmann Publishers, San Francisco (2006)
16. Hudec, M., Vujosevic, M.: Integration of data selection and classification by fuzzy logic. Expert Syst. Appl. **39**, 8817–8823 (2012)
17. Relich, M., Muszynski, W.: The use of intelligent systems for planning and scheduling of product development projects. Procedia Comput. Sci. **35**, 1586–1595 (2014)
18. Gola, A., Świć, A.: Computer-aided machine tool selection for focused flexibility manufacturing systems using economical criteria. Actual Probl. Econ. **10**(124), 383–389 (2011)
19. Relich, M., Pawlewski, P.: A multi-agent system for selecting portfolio of new product development projects. In: Bajo, J., Hallenborg, K., Pawlewski, P., Botti, V., Sánchez-Pi, N., Duque Méndez, N.D., Lopes, F., Vicente, J. (eds.) PAAMS 2015 Workshops. CCIS, vol. 524, pp. 102–114. Springer, Heidelberg (2015)
20. Sitek, P.: A hybrid CP/MP approach to supply chain modelling, optimization and analysis. In: Federated Conference on Computer Science and Information Systems, pp. 1345–1352 (2014)

21. Van Roy, P., Haridi, S.: Concepts, Techniques and Models of Computer Programming. Massachusetts Institute of Technology, Cambridge (2004)
22. Grzybowska, K., Kovács, G.: Logistics process modelling in supply chain – algorithm of coordination in the supply chain – contracting. In: de la Puerta, J.G., Ferreira, I.G., Bringas, P.G., Klett, F., Abraham, A., de Carvalho, A.C.P.L.F., Herrero, Á., Baruque, B., Quintián, H., Corchado, E. (eds.) International Joint Conference SOCO'14-CISIS'14-ICEUTE'14. AISC, vol. 299, pp. 311–320. Springer, Heidelberg (2014)
23. Grzybowska, K.: Selected activity coordination mechanisms in complex systems. In: Bajo, J., Hallenborg, K., Pawlewski, P., Botti, V., Sánchez-Pi, N., Duque Méndez, N.D., Lopes, F., Vicente, J. (eds.) PAAMS 2015 Workshops. CCIS, vol. 524, pp. 69–79. Springer, Heidelberg (2015)
24. Bzdyra, K., Banaszak, Z., Bocewicz, G.: Multiple project portfolio scheduling subject to mass customized service. In: Szewczyk, R., Zieliński, C., Kaliczyńska, M. (eds.) Progress in Automation, Robotics and Measuring Techniques. AISC, vol. 350, pp. 11–22. Springer, Heidelberg (2015)