DATA MINING DRIVEN DECISION SUPPORT

Serge Parshutin[1] and Arkady Borisov[2]

serge.parshutin@rtu.lv, arkadijs.borisovs@cs.rtu.lv

Abstract

Production planning is the main aspect for a manufacturer affecting an income of a company. Correct production planning policy, chosen for the right product at the right moment in the product life cycle (PLC), lessens production, storing and other related costs. This arises such problems to be solved as defining the present a PLC phase of a product as also determining a transition point - a moment of time (period), when the PLC phase is changed.

The paper presents a structure of the Data Mining driven Decision Support system, meant for supporting a production manager in his/her production planning decisions. The developed system is based on the analysis of historical demand for products and on the information about transitions between phases in life cycles of those products. The architecture of the developed system is presented as also an analysis of testing on the real-world data results is given.

Keywords: Data Mining, Self-Organizing neural map, Forecasting, Decision Module

1. Introduction

Constantly evolving computer technologies are more and more becoming an inherent part of successful enterprises management and keeping its activity at a high level. Different statistical methods are employed as well, though an increasing interest in computational intelligence technologies and their practical application can be observed ever more.

For quite a long time neural networks have been one of the most popular areas in the sphere of various processes forecasting including non-linear ones. The number of publications, books and monographs published within the last few years gives apparent evidence of that. A special place among neural networks is occupied by self-organising maps whose primary goal is to transform the incoming vectors of signals that are of deliberate dimensionality into single- or two-dimensional discrete map.

This paper focuses on studying the ability of application of the Data Mining driven Decision Support to the real-world problem. A task of product life cycle phase transition point

[1] PhD Student of Riga Technical University, Institute of Information Technology

[2] Dr.habil.sc.comp., Professor of Riga Technical University

forecasting can serve as an example of such problem. From the viewpoint of the management it is important to know, in which particular phase the product is. One of applications of that knowledge is selection of the production planning policy for the particular phase [1]. For example, for the maturity phase in case of determined demand changing boundaries it is possible to apply cyclic planning [2], whereas for the introduction and decline phase an individual planning is usually employed. As the technologies are evolving, the variability of products grows, making manual monitoring of PLCs a difficult and costly task for companies. Thus having an autonomous System that monitors products, creates and automatically updates lists of products for what it is reasonable to consider a production planning policy replacement, is one valuable alternative. This paper proposes a model of modular multi-network system that ensures the solving of the aforementioned task as well as provides an analysis of system testing results.

2. Problem statement

Any created product has a certain life cycle. The term "life cycle" is used to describe a period of product life from its introduction on the market to its withdrawal from the market. Life cycle can be described by different phases: traditional division assumes such phases like introduction, growth, maturity and decline [3]. For products with conditionally long life cycle, it is possible to make some simplification, merging introduction and growth phases into one phase - introduction. An assumption that three different phases, namely, introduction, maturity and end-of-life are possible in the product life cycle, gives us two possible transitions. The first transition is between introduction and maturity phases and the second - between maturity and product's end-of-life.

From the side of data mining [4,5,6], information about the demand for a particular product is a discrete time series, in which demand value is, as a rule, represented by the month. A task of forecasting a transition points between life cycle phases may be formulated as follows. Assume that $D = \{d_1, \ldots, d_i, \ldots, d_n\}$ is a dataset and $d = \{a_1, \ldots, a_j, \ldots, a_l\}$ is a discrete time series whose duration equals to l periods, where $l \in L = \{l_1, \ldots, l_h, \ldots, l_s\}$ and varies from record to record in the dataset D. For simplification, the index of d is omitted. Time series d represents a particular phase of a product life cycle, say introduction. Assume that for a particular transition, like introduction to maturity, a set of possible transition points $P = \{p_1, \ldots, p_k, \ldots, p_m\}$ is available. Having such assumptions, the forecasting of a transition point for a new product, represented by a time series $d' \notin D$, will start with finding an implication between historical data sets D and P, $f: D \rightarrow P$. As this step is done, the found model is then applied to determine possible transition points for a new product. Having information about transitions the Decision Support Module autonomously marks products, for which it would be reasonable to reconsider a production planning policy.

3. Structure of the System

The developed system consists of three main elements - Data Management Module (DMM), Neural Block (NB) and Decision Support Module (DSM), shown in Figure 1.

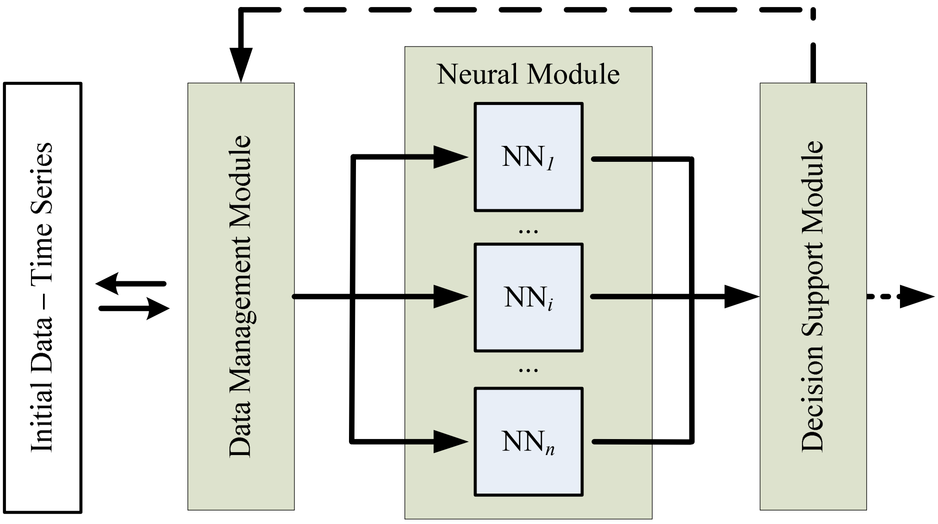

Figure 1. Structure of the system.

The description of each module is given, as also the system functioning algorithm is fully described in source [7]. As the system is trained, the application of it begins. In this stage the DSM starts its inner processes. With regard to the task stated in section 2, the DSM will take a number of duties.

The DSM monitors a number of products a company manufactures. At a specific moment of time, say at the end of each period, the DSM starts the process, displayed in Figure 2. The depicted process includes three main steps - Determination of the Best Matching Cluster (BMC) for each of the products that are monitored by the DSM; Formation of the List of Interest (LOI), the list of products for which it would be reasonable to reconsider the planning policy; Evaluation of a cyclic and non-cyclic planning policy for each product in the List of Interest. And finishes with the fourth step - Reporting the results of the evaluation to a manager. Let us describe the processes hidden behind each of the main steps.

Step 1: Determination of the BMC for each of the monitored products. For each of the monitored products DSM collects a demand data from a specific period - the first period of the Introduction phase for managing the $M1$ transition point or the first period of the Maturity phase for managing the $M2$ transition point; till last finished period. The collected data is formed as a discrete time series and is included in a list which being sent to the Data Managing Module (dashed arrow in Figure 1) with a command to find a BMC in a corresponding neural network for each of time series in the list.

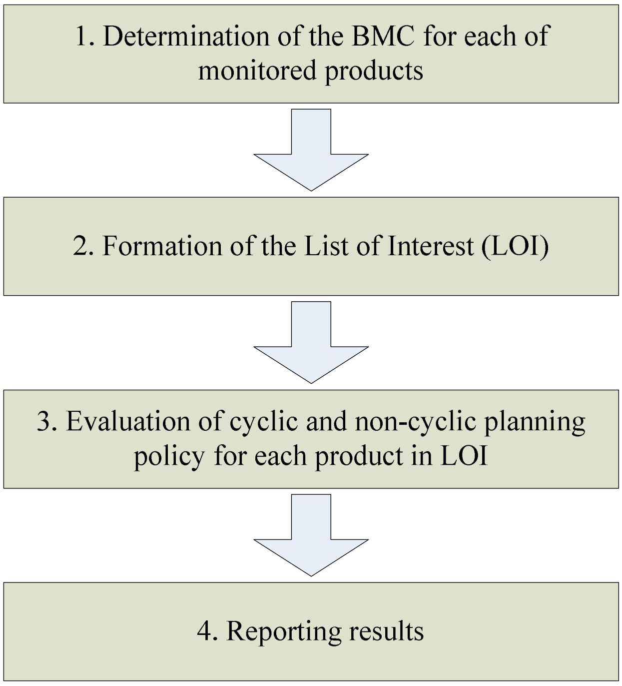

Figure 2. DSM application process.

The DMM receives the list and calculates the duration (l) of each of the product demand time series. Knowing the duration of each demand time series, the DMM sends each record to the corresponding neural network n_l, that processes time series with duration l, in the Neural Module.

The Neural Module finds the Best Matching Cluster for each of the product demand time series and returns a list of found cluster to the Decision Support Module. The information from the BMC contains a list of possible transition points for each analysed product. This is where the Formation of the LOI begins.

Step 2: Formation of the List of Interest. The BMC may contain different information for products and several cases are possible:

1. The simplest case (*C1*) when the Best Matching Cluster contains only one possible transition point. In this case the DSM assumes this transition point as preferable one and follows a solution (*S1*) containing three major rules:
   a. If $l < p$ and $p - l > \theta$ Then: Product remains monitored and is not included in the List of Interest;
   b. If $l < p$ and $p - l \leq \theta$ Then: Product is included in the List of Interest;
   c. If $l \geq p$ Then: Product is included in the List of Interest.

   Where l is the duration of the demand time series in periods; p - a possible transition point (period) for a demand time series; and variable θ stores the minimal threshold of interest for either including the product in the List of Interest or not.
2. The case (*C2*) when the BMC contains more than one possible transition points for a demand time series, but one of transition points have an expressed appearance frequency f. An appearance frequency may be stated as expressed if it exceeds some threshold, like 50%. In such case the solution (*S2*) will be that the Decision Support Module accepts the transition point with an expressed f as preferable one and follows rules from the (*S1*) solution.
3. The third possible case (*C3*) is that the BMC contains several possible transition points, but there is no one with an expressed appearance frequency present. For this case several solutions are possible:
   a. Solution *S3*. The Decision Support Module checks if the neural network n_i where the BMC for a current product was found, is the only one network in the Neural Module or if it is the last network in the NM. If so, then the DSM follows the next two rules:
      i. If only one transition point has the highest (not expressed) f, then select it as a preferable one and follow rules from solution *S1*;
      ii. If several transition points has the highest f, then select a transition point with a minimal value as preferable one and follow rules from solution *S1*. Thus if transition points with the highest f are 4th, 5th and 8th period then the DSM will choose the 4th period.

If at the end of formation of the List of Interest the list is empty then the Decision Support Module bypasses the third step and reports that products, for which it would be reasonable to reconsider the planning policy, were not found. In case when LOI contains at least one product the DSM starts processes in the third step.

Step 3: Evaluation of cyclic and non-cyclic planning policy for each product in LOI. At this step The Decision Support Module measures a gap between expenses of using cyclic or non-cyclic planning policy for each product in the List of Interest. As stated in [2] the measure of Additional Cost of a Cyclic Schedule (*ACCS*) may be used for those purposes. The *ACCS* is calculated by formula (1).

$$ACCS = \frac{CPPC - NCPPC}{NCPPC}, \tag{1}$$

where *CPPC* is the Cyclic Planning Policy Cost and *NCPPC* - the Non-Cyclic Planning Policy Cost.

As the third step is finished the manager receives a results of analysis of the products from the List of Interest.

4. Gathered results

The fact that the data describes real life process and marks of transitions were putted by experts implies that some noisiness in data is present.

The obtained dataset contains 199 real product demand time series with minimal duration equals to 4 and maximal - to 24 periods. Each time series contains the demand during the introduction phase of a specific product plus one period of the maturity phase, and is marked with *M1* marker. To normalize the data, the Z-score with standard deviation normalization method was applied. As the true bounds of the demand data in the dataset are unknown and the difference between values of various time series is high, the chosen normalization method is one of the most suitable ones.

The main target of the performed experiments was to analyse comparatively the precision of forecasting transition points with square neural network topology with 8 neighbours applied while using different network load q. The definition of q - the network load parameter is given in [7]. The more precise the forecasted transition point will be, the more precise will be the result, returned by the Decision Support Module.

The network load q was changing incrementally from one to five. To calculate the system errors - Mean Absolute Error (*MAE*) and Logical Error (*LE*), a 10-fold cross validation method was applied, totally giving 50 system runs.

Table 1 contains the size of the network, the total number of neurons, as also supplies the number of iterations for initial organization and convergence process.

Table 1: Size of networks and number of iterations

Size	Neurons	Initial Organisation	Convergence process
5 x 5	25	1000 iter.	12500 iter.

The learning parameters, used for network organisation in each run, are given in Table 2. For each learning parameter the starting value and the minimal (last) value are supplied, as also the type of a function used for managing the parameter decline process.

Table 2: Learning parameters

Parameter	Start with	End with	Type of the function
Learning coefficient - η	0.9	0.01	Exponential
σ for Gaussian neighbourhood	0.5	0.01	Exponential

While testing the system in On-line mode, for each of five defined values of q a Mean Absolute Error and a Logical Error were obtained. The gathered results are accumulated in Table 3 and Table 4.

Table 3: On-line Mean Absolute Error – *MAE*

Topology	$q = 1$	$q = 2$	$q = 3$	$q = 4$	$q = 5$
SQR-8	2.050	2.167	2.408	2.216	2.416

Table 4: On-line Logical Error – *LE*

Topology	$q = 1$	$q = 2$	$q = 3$	$q = 4$	$q = 5$
SQR-8	13.5%	11.7%	14.5%	12.3%	17.0%

The obtained results show that created system with certain precision is able to predict transition points for new products, using a model, built on a historical demand data. In On-line mode System was able to make a logically correct decisions in at least 83.0% and at most in 87.7% of times. Thus the Mean Absolute Error lies close to 2 periods. Together with the specificity of the dataset obtained and the size of the network, it is possible to conclude that the created system can be used as a data mining tool to gain an additional knowledge for solving a planning policy management task. As well as for solving other tasks, connected with forecasting a value of a target parameter for a time dependent variable, followed by a Decision Support process.

5. Conclusions

For the practitioners of management of the product life cycle the knowledge, which describes in which phase the product currently is and when the transition between phases will occur, is topical. Such knowledge, in particular, helps to select between the cyclic and non-cyclic policy of planning supply chain operation.

In this paper, the task of forecasting the transition points between different phases of product life cycle is stated, and the structure of Data Mining driven Decision Support system, which helps to solve this task, is shown. The Decision Support Module functioning algorithm is presented. On the basis of the analysis of historical demand data for products it is possible to train the modular neural network based system, which will be able to forecast the transition points in life cycle of new products. Experimentally gathered results show that the created Data Mining driven Decision Support system has its potential and can process real demand data, forecasting possible transition points and reporting an analysis of expenses for cyclic and non-cyclic planning policies.

One aspect is that in the future it is necessary to examine the developed system on the data from different production fields, and, which is also important, to have a response from practitioners of supply chain management who will use these systems.

Another aspect, modest data volume that was used for practical experiments, is related to the fact, that it is necessary to have transition marks in historical data from experts and practitioners. The more products, the more complicated for human to make all these marks - in practice the amount of marked data will always be restricted. As a result, possible direction of future research is treatment of forecasting the transition points in the context of a semi-supervised learning [8]. In this case, there is a small set with marked transitions and also a large dataset in which transitions are not marked. In such a situation it is necessary to create a model, which will be able to apply the knowledge, gathered on the small set of marked data, to the new (test) data.

6. References

1. MERKURYEV Y., MERKURYEVA G., DESMET B. and JACQUET-LAGREZE E. Integrating Analytical and Simulation Techniques in Multi-Echelon Cyclic Planning.

Proceedings of the First Asia International Conference on Modelling and Simulation. IEEE Computer Society, pp. 460-464, **2007**, ISBN: 0-7695-2845-7.

2. CAMPBELL G. M. and MABERT V. A. Cyclical Schedules for Capacitated Lot Sizing with Dynamic Demands. Management Science **37**(4), pp. 409-427, **1991**.
3. KOTLER P. and ARMSTRONG G. Principles of Marketing. Prentice Hall, 11th edition, **2006**, ISBN: 978-0131469181.
4. DUNHAM M.H. Data Mining Introductory and Advanced Topics. Prentice Hall, **2003**, ISBN: 0-13-088892-3.
5. HAN J. and KAMBER M. Data Mining: Concepts and Techniques. San Francisco: Morgan Kaufman, 2nd edition, **2006**, ISBN: 1-55860-901-6.
6. HAND D. J., MANNILA H. and SMYTH P. Principles of Data Mining. The MIT Press, **2001**, ISBN: 0-262-04194-4.
7. PARSHUTIN S., ALEKSEJEVA L. and BORISOV A. Time Series Analysis with Modular Neural Networks. Scientific Proceedings of Riga Technical University // Computer Science **36**(5), pp. 162-169, **2008**. ISSN: 1407-7493.
8. ZHU X. Semi-supervised learning literature survey. Technical report 1530, Department of Computer Sciences, University of Wisconsin, **2005**.