Comprehensive Validation of the Regeneration Workload Forecast for Complex Capital Goods Using Data Mining

Steffen C. Eickemeyer
Institute of Production Systems and Logistics (IFA)
Leibniz University of Hanover
Hanover, Germany
eickemeyer@ifa.uni-hannover.de

Simon Steinkamp
Institute of Production Systems and Logistics (IFA)
Leibniz University of Hanover
Hanover, Germany

Cameron Tucker
George Washington University
Washington, D.C., USA

Sebastian Schäfer
MTU Maintenance GmbH
Hanover, Germany

***Abstract*— Capacity planning in the regeneration of complex capital goods faces major challenges because it is affected by a high level of uncertain workload information. A methodology is developed here to predict the regeneration workload on the basis of the CRISP-DM model using Bayesian networks. The forecasts are validated for the different capacity planning levels. The results support the conclusion that capacity planning can gain permanent benefits from the methodology developed.**

***Index Terms*— Bayesian networks, capacity planning, complex capital goods, damage library, data mining, forecast, regeneration**

I. Introduction

The aim of this article is to create the conditions for raising the logistics performance of regeneration processes. The processes combined under the heading of regeneration place maximum demands on planning and control as well as the availability of resources [1]. This is caused by many things in addition to the great complexity and large number of individual components, one of those being the fixed times for the start and end of the regeneration process, whose workload is not known in advance [2].

In order to do justice to the high logistical demands of capacity planning and control for regeneration processes, the information available must be used as early and as effectively as possible [2]. Because the regeneration workload required is unknown at the time of planning, adequate forecasting methods have to be developed [3]. Potential approaches for considering information affected by uncertainties can be found in the field of reaction methodology [4]. Such approaches are characterized by their ability to integrate current information from their surroundings on an ongoing basis and process that data to help the procedure [4]. Data mining methods can be used to supplement these approaches. These learning systems represent a promising means of deriving probabilities of occurrence in order to produce forecasts from existing data and in the end deduce reaction strategies [5].

II. Capacity Planning for the Regeneration of Complex Capital Goods

In order to increase the logistical performance of regeneration processes, a method for efficient capacity planning and coordination in the case of imprecise workload information is being developed within the scope of subproject D1 of the Collaborative Research Centre 871 (SFB 871) [3]. In contrast to production planning and control for conventional areas of production, capacity planning in regeneration is characterized by greater uncertainty with respect to the regeneration workload [6]. This uncertainty increases with the complexity of the goods requiring regeneration. To compensate for this, the response in practice is commonly to procure overcapacities in terms of personnel and plant and to keep large stocks because the regeneration providers face potential conventional penalties if they miss deadlines [7]. In order to counteract the effects of these cost drivers, the uncertainties in capacity planning must be reduced by providing reliable forecasts of the regeneration workload. When drawing up forecasts, it should be remembered that as the regeneration process progresses, the information on the actual damage situation and the repair steps necessary increases with time [6]. The options for action in capacity planning also change over time, so it becomes possible to match capacity and workload [8]. Whereas on the one hand the information relating to the actual status of the components is constantly increasing, on the other hand we see a decrease in the capacity coordination options. Accordingly, capacity planning is carried out for different time periods; we distinguish between long-, medium-, and short-term planning (Fig. 1) [9].

IJBAMR

		Capacity Coordination Options				
		Capacity Adjustment		Workload Adjustment	Workload Alignment	
		Adjusting the Workforce	Adjusting the Resources	Time Compensation		Technology Compensation
Planning Horizon	Short	More/less overtime Internal secondment of personnel	Changing the processing rate Postponing maintenance work	Outsourcing jobs/ sub-assemblies Accepting outsourced jobs / subas-semblies	Splitting batches Regeneration pool Prioritizing/ post-poning jobs or individual needs	Switching to other resources
	Medium	Additional shift/ short-time/ weekend working/ temporary staff	Reviving/ decommis-sioning plant	Outsourcing jobs Accepting outsourced jobs	Prioritizing/ postponing jobs	
	Long	Training/ recruitment/ dismissal of personnel	Procuring/ offloading plant			

Fig. 1. Options for coordinating capacity in a regeneration company (according to [3])

III. Generic Process Model and Damage Library

Subproject D1 of project SFB 871 initially involved developing a generic process model based on the structure of the SCOR model [10] [11]. Locating information is particularly important here because in logistical and technical terms the planning and control of regeneration processes depends on the provision, exchange, and processing of information. The consistent, comparable, and assessable process model serves as a starting point for product regeneration [11]. Based on this, a damage library in the form of a relational database was developed with the aim of specifying the information for capacity planning [12]. The content of the database is legacy job and process data that indicate the wear behavior, findings, and physical regeneration process [12]. This includes data on the lifetime (to date) of the goods requiring regeneration, the customer, and the regeneration procedure so far, which is made up of workstations, setup times, and processing times.

IV. Forecast of Regeneration Workload Using Data Mining

Using the CRISP-DM model for structured data mining projects as a basis, the defined aim is to draw up forecasts of the regeneration workload that are as precise as possible (Fig. 2) [13]. The purpose of these forecasts is to reduce the uncertainties in capacity planning and control and to stabilize the processes as a whole.

Within the data mining project, the damage library is used as a database. The data records required for the following steps are extracted and prepared. Furthermore, Bayesian networks have proved to be a suitable modeling method for this particular problem [8]. This is a graph-based method that permits the use of imprecise data and is able to produce forecasts even in the case of a considerable lack of information [8]. Bayesian networks consist of nodes and edges used in graph theory which are connected with each other and represent the conditional probabilities of different states (Fig 3). Forecasts can be drawn up (child nodes) depending on the input parameters (parent nodes). The predicted value (output parameter) changes depending on the attributes of the various input parameters [8] [14].

V. Constructing the Bayesian Networks

A Bayesian network is developed for each planning horizon taking into account the subdivision into long-, medium-, and short-term capacity planning (Fig. 3). Different amounts of information are available for the various planning stages. The demands placed on the predicted value also differ. This affects the input and output parameters chosen for the respective networks. Once the nodes have been established and the edges determined, the various discrete states of the nodes are defined and, if necessary, continuous states combined in subsets. Based on the data fed in, the conditional probabilities of the node states are eventually calculated, which concludes the network construction process [14]. If the state of one or more nodes for an individual type of capital good is known in one network, then this information can be fed into the network and processed such that the predicted value is modified. The more we know about one type of capital good, the more individual and more accurate is the workload forecast [8].

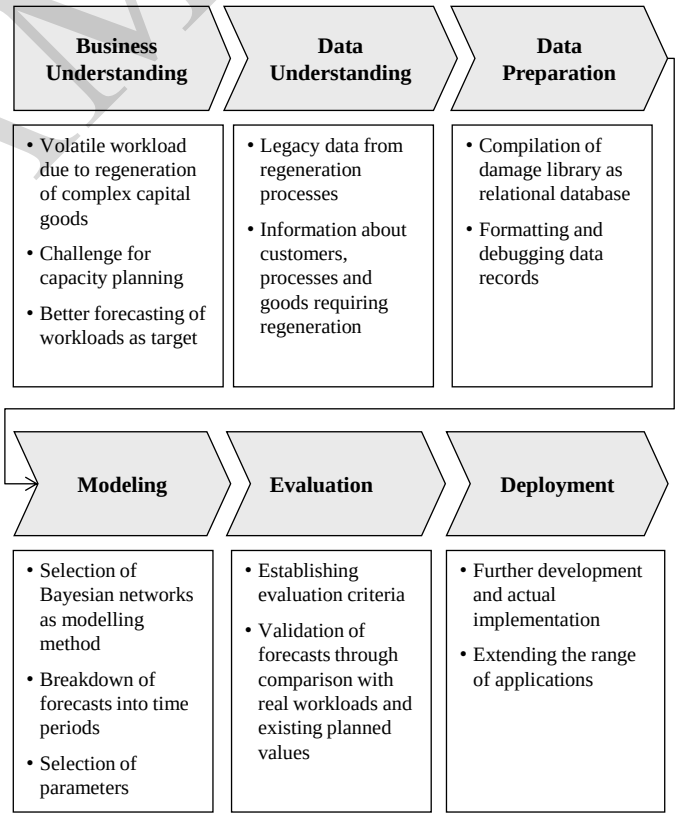


Fig. 2. Steps in the CRISP-DM model (according to [13])

IJBAMR

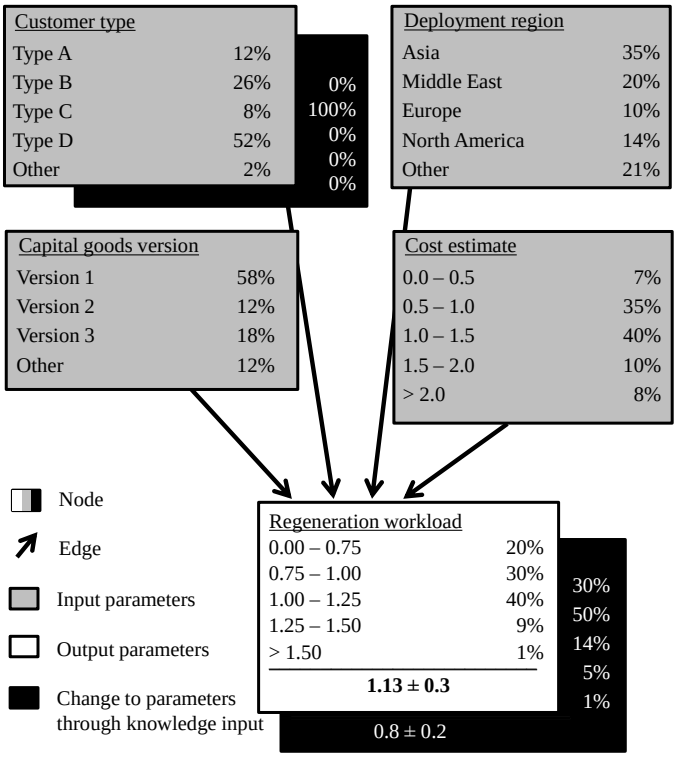


Fig. 3. Example of a Bayesian network (according to [8])

When it comes to long-term capacity planning, which is carried out annually, little is known about future regeneration assignments. Despite this limitation, a forecast that is as accurate as possible is critical at this stage if, for example, sales targets are to be adjusted or investment decisions are made. The input parameters include, for example, customer type, version (design/type), and the deployment region of the capital goods.

Medium-term planning, which is carried out several weeks before commencing the regeneration of the capital goods, encompasses a larger and more specific amount of information that is incorporated in the workload forecast. Additional input parameters implemented include information on age and the reason for the regeneration assignment.

The most accurate information concerning the status of the capital goods, and hence also the ensuing regeneration workload, only becomes available during the regeneration process itself and is beneficial for short-term capacity planning. This information includes both results from diagnoses, carried out prior to the regeneration process, and findings. On this planning level, the predicted regeneration workload relates to individual subassemblies and cost centers and no longer to the capital goods as a whole.

VI. Validating the Forecast

The Bayesian networks for the capacity planning were validated in collaboration with an industrial partner from the engine maintenance sector. To this end, suitable input and forecast parameters were identified for the Bayesian networks in addition to compiling a damage library. After that, regeneration workloads were fed into the networks, which are not included in the damage library, in order to generate workload forecasts. The forecasts were compared with the internal planned values of the industrial partner for the respective planning level. The industrial partner used fixed mean values, which are based on the workload figures of the preceding years, as the planning basis for the long- and short-term levels. An expert report was available as a comparative value for the medium-term planning; in the report the total workload is estimated from the assignment shortly before the start of regeneration.

The assessment of the forecasts was carried out through the comparison with the workload that actually occurred due to the regeneration assignments. After calculating the mean percentage and absolute deviations of the forecast values from the actual workloads plus the standard deviation, the results were divided into four different categories specified by the industrial partner:

- 0–10% deviation: very good forecast result
- 10–25% deviation: good forecast result
- 25–50% deviation: satisfactory forecast result
- > 50% deviation: unsatisfactory forecast result

Finally, which of the forecasting methods produces more accurate results more frequently was investigated.

The forecasts of the Bayesian networks have a distinct advantage over the rigid comparative values of the industrial partner, especially on the long- and short-term planning levels, because current environmental influences are taken into account more quickly. There are more forecasts in the very good and good categories, fewer in the unsatisfactory category. With respect to the medium-term capacity planning, it can be seen that the Bayesian network is able to compete with the forecasts of the industrial partner, which are drawn up for individual orders by a team prior to commencing regeneration. Better forecasts are achieved in almost 60% of cases and these frequently lie within the very good category.

The results on the short-term level are conspicuous in that their quality is lower than that of the other levels. Here, the predicted workload relates to a subassembly of the capital goods, which is allocated to a certain cost center, whereas the input parameters continue to use information related to the capital goods as a whole. By relating to general input parameters, only the validation of one subassembly is considered in this case.

The use of suboptimal input data on the short-term level is due to the fact that the industrial partner currently employs more control and less planning when it comes to individual capacities. Therefore, the data required for the planning are currently not yet available with sufficient quality. Both the results of the Bayesian networks and those of the industrial partner's planned values must therefore be treated differently and should at this point chiefly indicate potential that can be increased in the future.

IJBAMR

TABLE I. ASSESSMENT AND VALIDATION RESULTS OF BAYESIAN NETWORKS (BN) AND PLANNED VALUES OF INDUSTRIAL PARTNER (IP)

Planning	Long-term planning		Medium-term planning		Short-term planning	
Forecast	BN	IP	BN	IP	BN	IP
Percentage deviation	25.8%	62.2%	24.4%	19.6%	53.9%	68.5%
Standard deviation	36.9%	140.1%	32.1%	13.8%	66.8%	15.8%
Absolute deviation [h]	540	840	570	600	174	295
Standard deviation [h]	420	720	390	480	124	173
Proportion of forecasts with a percentage deviation in the range...						
0–10%	27.0%	26.4%	33.8%	26.7%	2.8%	0.0%
10–25%	38.1%	25.0%	29.2%	45.0%	25.0%	2.8%
25–50%	27.0%	29.2%	26.2%	25.0%	38.9%	2.8%
>50%	7.9%	19.4%	10.8%	3.3%	33.3%	94.4%
More accurate forecast	63.8%	31.0%	58.6%	39.7%	91.7%	8.3%

Despite this limitation, the forecasts produced show considerable progress over the previous planned values.

On the whole, it can be said that in most cases the forecasts of the Bayesian networks are superior to the planned values already available to the industrial partner and are generally at least on a par with those values. They provide a sound foundation for reliable capacity planning and can help to improve planning processes permanently (Fig. 4).

VII. CONCLUSION

The theme of this article has been the use of data mining to compile forecasts of the regeneration workload for complex capital goods in the case of imprecise workload information. The aim is to provide reliable planned values for capacity planning for different time periods. The application of Bayesian networks to real regeneration assignments has shown that the forecasts can be assessed as positive in principle. In addition, it could be established that the quality of the forecasting increases with the quantity of information fed in. The statements regarding the regeneration workload are convincing and represent an advantage over the planning data of the industrial partner. The forecasts are able to provide regeneration workload estimates that are more precise than those of the past so capacity planning can be placed on a firm footing and improved through this method. The quality of the forecasts can be improved by extending the damage library, i.e. the database. Looking beyond capacity planning, data mining using Bayesian networks has the potential to be able to create clear added value from existing information in many areas of business (e.g. calculating quotations).

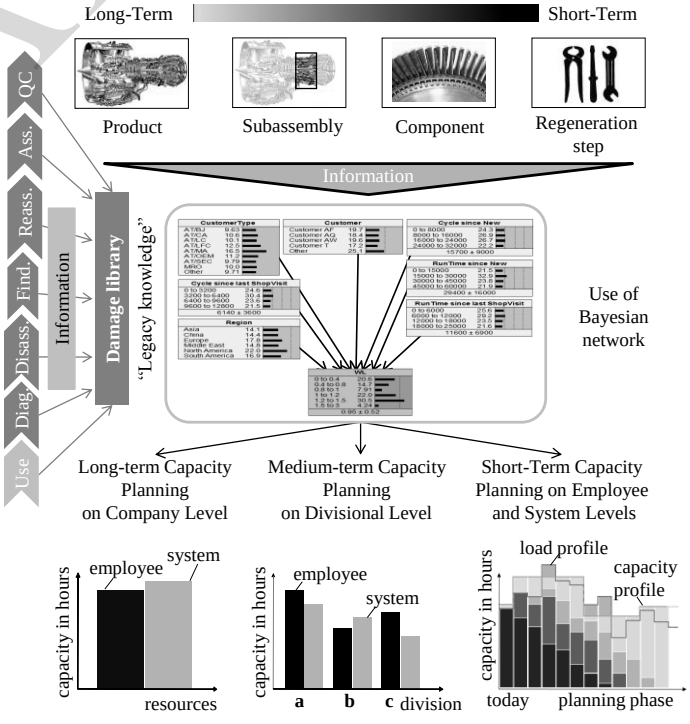


Fig. 1. Summary of product structure and target (according to [8])

IJBAMR

ACKNOWLEDGMENT

The authors would like to thank the German Research Foundation (DFG) for providing funding for this research project within the scope of the CRC 871 program. The authors would also like to thank the Graduate Academy of Leibniz University Hannover for providing funding for the validation of the theoretical results in the global industry.

REFERENCES

[1] Reményi, C., Staudacher, S., Becker, H., Dinc, S., Manejev, R., Fichtelmann, R.: Decentralized Job-shop Control for the Maintenance of Aircraft Engines. In: Proceedings of the 21st ICPR International Conference on Production Research, CD-ROM. Fraunhofer-Verlag, Stuttgart (2011)

[2] Reményi, C., Staudacher, S., Holzheimer, N., Schulz, S.: Simulation of the Maintenance Process in an Aircraft Engine Maintenance Company. In: Proceedings of the 44th CIRP Conference on Manufacturing Systems. Omnipress, Madison, Wisconsin (2011)

[3] Eickemeyer S.C., Nyhuis P (2010) Capacity planning and coordination with fuzzy load information. The Bus Rev, Camb16,1: 259-264

[4] Scholz-Reiter, B., Scharke, H.: Reaktive Planung. Ind. Manag. 16 (2), 21-26 (2000)

[5] Hippner, H., Grieser, L., Wilde, K.D.: Data Mining – Grundlagen und Einsatzpotenziale in analytischen CRM-Prozessen. In: Hippner, H., Hubrich, B., Wilde, K.D. (eds.) Grundlagen des CRM. pp. 783-810. Gabler, Wiesbaden (2011)

[6] Reményi, C., Staudacher, S., Holzheimer, N., Schulz, S.: Simulation von Reihenfolgeregeln in einem Triebwerkswartungsunternehmen. Product. Manag. 16 (1), 16-19 (2011)

[7] Matyas, K.: Taschenbuch Instandhaltungslogistik. Qualität und Produktivität steigern. pp. 195-196. Carl Hanser Verlag, Munich, Vienna (2005)

[8] Eickemeyer, S.C., Borcherding, T., Schäfer, S., Nyhuis, P.: Validation of data fusion as a method for forecasting the regeneration workload for complex capital goods. Prod. Eng. 7 (2-3), 131-139 (2013)

[9] Gehr, F., Hellingrath, P.: Logistik in der Automobilindustrie. Springer Verlag, Berlin, Heidelberg (2007)

[10] Supply Chain Council: SCOR – Supply Chain Operations Reference Model Version 9.0. Supply Chain Council, Cypress (2008)

[11] Eickemeyer, S.C., Doroudian, S., Schäfer, S., Nyhuis, P.: Ein generisches Prozessmodell für die Regeneration komplexer Investitionsgüter. ZWF 106 (11), 861-866 (2011)

[12] Eickemeyer, S.C., Goßmann, D., Wesebaum, S., Nyhuis, P.: Entwicklung einer Schadensbibliothek für die Regeneration komplexer Investitionsgüter. Ind. Manag. 28 (2), 59-62 (2012)

[13] Shearer, C.: The CRISP-DM model: The new blueprint for data mining. J. Data Warehous. 5 (4), 13-22 (2000)

[14] Heckerman, D.: Bayesian Networks for Data Mining. Data Min. Knowl. Discov. 1 (1), 79-119 (1997)

.

IJBAMR