



HAL
open science

A Data Mining Approach to Support Capacity Planning for the Regeneration of Complex Capital Goods

Melissa Seitz, Maren Sobotta, Peter Nyhuis

► **To cite this version:**

Melissa Seitz, Maren Sobotta, Peter Nyhuis. A Data Mining Approach to Support Capacity Planning for the Regeneration of Complex Capital Goods. IFIP International Conference on Advances in Production Management Systems (APMS), Sep 2019, Austin, TX, United States. pp.583-590, 10.1007/978-3-030-29996-5_67. hal-02460468

HAL Id: hal-02460468

<https://inria.hal.science/hal-02460468v1>

Submitted on 30 Jan 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

A Data Mining Approach to Support Capacity Planning for the Regeneration of Complex Capital Goods

Melissa Seitz¹, Maren Sobotta¹ and Peter Nyhuis¹

¹Institute of Production Systems and Logistics (IFA)
Leibniz University Hannover
An der Universität 2
D-30823 Garbsen, Germany
www.ifa.uni-hannover.de

Abstract. With regard to the recommissioning of damage caused inoperable complex capital goods, a high logistics efficiency is a very important competitive factor for regeneration service providers. Consequently, fast processing as well as a high schedule reliability need to be realized. However, since the required regeneration effort for future damages may vary and is usually indefinite at the time of planning, capacity planning for the regeneration of complex capital goods has to deal with a high degree of uncertainty. Regarding this challenge, the evaluation of prior regeneration process data by means of data mining offers great potential for the determination of load forecasts. This paper depicts the development of a data mining approach to support capacity planning for the regeneration for complex capital goods focusing on rail vehicle transformers as a sample of application.

Keywords: Capacity Planning, Data Base, Data Mining, Complex Capital Goods, Logistics Efficiency.

1 Introduction

Complex capital goods consist of numerous expensive components [1]. Due to the high value of goods and components, regeneration processes (maintenance, repair and overhaul) are used to increase the product service life along with their value-adding potential [2]. Some manufacturers of complex capital goods, such as rail vehicle transformers, thus additionally offer those regeneration services. However, there is a challenging uncertainty in terms of planning and synchronizing the capacities required for the regeneration processes that needs to be managed in order to achieve high schedule adherence and reliable delivery times [1, 3]. This uncertainty is caused by the fact that the regeneration effort varies due to the good's condition and is usually not known when the planning is conducted [1, 4, 5].

The increasing availability of data due to progressive digitalization provides the opportunity to gain new, valid and relevant information from extensive databases, which may

be used to support business decisions [6, 7]. Research by Eickemeyer has shown that data mining methods may support capacity planning when there is uncertain load information. Data mining methods may be used as a tool in planning capacities by generating load forecasts regarding the anticipated expenditures for the regeneration tasks that are to be conducted. The underlying data need to be provided by a database of already completed regeneration orders which includes information about the damage incident, diagnosis as well as conducted regeneration measures and the related expenditures. [6, 8, 9]

With regard to the knowledge discovery in data bases (KDD) process by Fayyad et al. [8] current research is thus focused on developing a database and identify a suitable data mining method for generating reliable load forecasts to improve capacity planning in the regeneration of rail vehicle transformers [10].

Following the KDD process, the first step is to define the objective and identify available data [8]. During data selection, relevant data are chosen and integrated in a database [9]. These data are also referred to as target data and over the course of the KDD process, they form the core for the gained data based knowledge. In the next step, this data is prepared and pre-processed. [8] This step is critical because the data quality strongly influences the analysis results. During the subsequent transformation, the pre-processed database is transformed into a suitable database schema. [11]

Following this introduction section 1, section 2 focusses on the development of the database model regarding the regeneration of rail vehicle transformers in detail. In this database all relevant data collected during the regeneration processes of prior regeneration orders is stored. It contains information about the damage incident, diagnosis, measures used and scope of the regeneration. In order to gain new information from this data, section 3 describes the next step in KDD process that focuses on identifying an appropriate data mining approach to generate load forecasts.

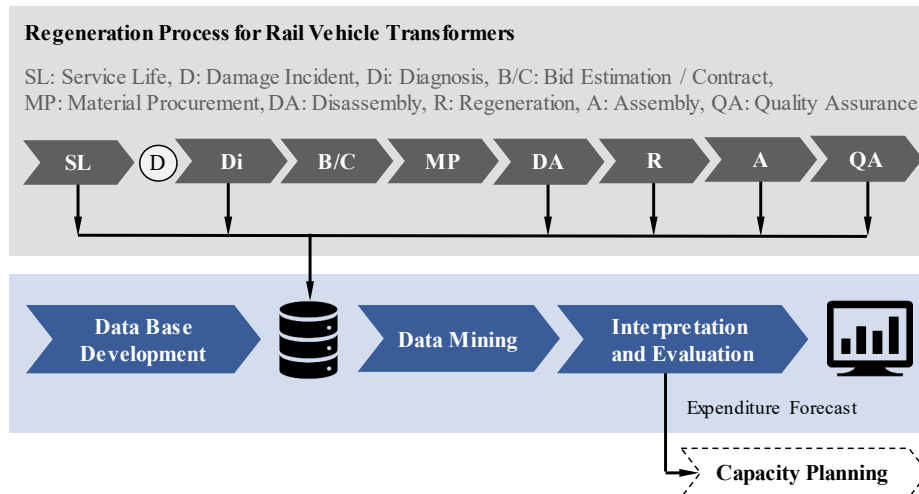


Fig. 1. Regeneration process for rail vehicle transformers (above) and approach to support capacity planning by means of data mining (below), based on [10]

For the sample of application described in this paper, data mining is used to generate reliable forecasts regarding the anticipated expenditures for the regeneration tasks to support capacity planning when regenerating rail vehicle transformers. With the aid of these forecasts, capacity requirements can be better planned and measures for aligning capacities can be introduced early on. Since new and different information becomes available as the regeneration process progresses, load forecasts can be re-generated over the course of planning [6]. The determination of a suitable data mining approach to support capacity planning when regenerating rail vehicle transformers is depicted in section 3. Figure 1 briefly depicts the regeneration process of rail vehicle transformers [10] and summarizes the approach of the research described in this paper.

2 Development of a Rail Vehicle Regeneration Database

In order to be able to analyze historical data and gain new information, there has to be a relevant data basis [8, 9]. Therefore, a database of damages and corresponding regeneration tasks was created for rail vehicle transformers as the addressed sample of application.

A relational database was used as a model. In contrast to a large hierarchically structured data sheet, the data in a relational database is acquired in groups for individual topics. In technical language, these topics are referred to as entities. Subsequently, relational databases are more flexible than hierarchical databases. This flexibility applies both with regards to expanding the database and to changing its structure. [12, 13]

The regeneration database was built in accordance with the database design process developed by Steiner [12]. The sequence of the individual steps for the selected design process are depicted in Figure 2 and are clarified below.

Define Function: The regeneration database for rail vehicle transformers is responsible for systematically acquiring information relevant to damages in order to generate load forecasts and regeneration measures that are to be executed.

Form Entity Sets: The regeneration database for rail vehicle transformers is comprised of the following entities: transformer, project information, damage incident, diagnosis results and regeneration expenditure.

Define Relationships: The entities of a relational database can be linked with one another in a number of ways. These links are defined by their relationship type. Types of relationships result from the combination of the following types of association: simple association '1' (exactly one linked data set), conditional association 'c' (no or exactly one linked data set), multiple association 'm' (at least one linked data set), or multiple conditional association 'mc' (any number of linked data sets) [12].

Every manufactured rail vehicle transformer is assigned to a project. A project is a customer's collective order and consequently can include multiple transformers (1-m). A damage incident recorded in the database belongs to exactly one transformer. Theoretically, a transformer can be damaged a number of times during its lifecycle, whereby a transformer might have to be assigned to a number of damage incidents (1-m). For the diagnosis of damage incidents in the observed rail vehicle transformer application, the individual analyses are only conducted when instructed by the customer. The results of

a test may in turn be identical for a number of transformers (1-c). This implies, that there can be either a diagnosis result or no diagnosis result for a saved damage incident and that a diagnosis result can be allocated to at least one damage incident. In contrast, exactly one expenditure is registered for the repair work entailed in regenerating a rail vehicle transformer, whereas the specific timespan required for the repair can theoretically be allocated to multiple damage incidents (1-m).

Define Identification Keys: A clear identification key has to be set in order to classify the entities and the data sets contained within them [12, 13]. In our sample of allocation, this is the transformer's article number and the project number. For damage incidents, case numbers are created and internal job numbers are used to unmistakably identify the diagnosis results and expenditures.

Global Normalization: In the global normalization step, the conceptual data model is converted into a physical data model with corresponding spreadsheets. To do so, auxiliary entities may be added [12]. To concretize the diagnosis, the results of the individual analyses are separately recorded in the database. The mechanical and electrical tests as well as the oil test are therefore assigned separate entities. On the relationship level, these are described as 1-c because the three named tests are conducted for each individual transformer only when instructed by the customer. Consequently, there can either be a result or there can be no result. The three test entities can also be assigned different individual results and are therefore defined by a 1-m relationship. Moreover, in the database, the regeneration expenditure is subdivided into four separate auxiliary entities (winding, prefabrication, final assembly). Figure 3, depicts the entity block diagram that results following these transformation steps. The individual entity sets are linked via so-called foreign key attributes [14].

Define Local Attributes: The local attributes describe the information that corresponds to the entities and thus define the content of the data sets, to which specific values can be assigned. Furthermore, as previously described, individual attributes serve as identification keys or foreign keys [12, 14]. In addition to the article number, the project number is also given for every transformer. A project is described by both its project number and the attributes customer and site of operation. The project number attribute

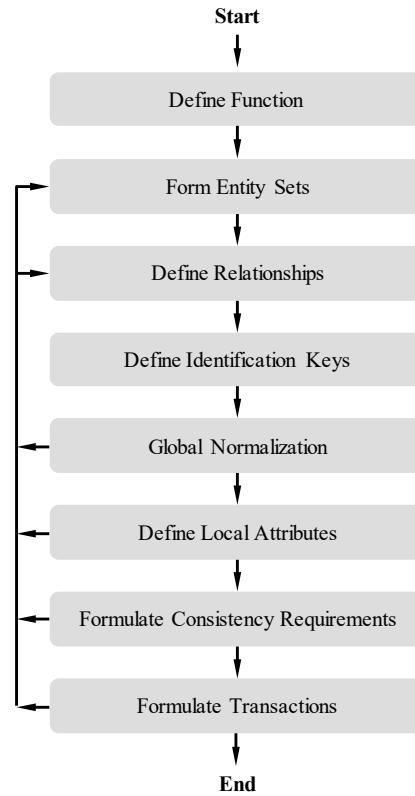


Fig. 2. Process for designing a database [12]

consequently is the foreign key between the transformer and project information entity sets. The damage incidents with the identification key damage case number are further described for example, by means of the service duration up to the time at which damages occurred. The damage categories include attributes such as module failure location, component failure location and cause of failure. The diagnosis results entity includes the possible analyses with their corresponding results. For electrical testing, possible attributes entail results of resistance or ratio testing. The damage case number is added as a foreign key attribute to link the entity sets damage incident, diagnosis results and regeneration expenditure.

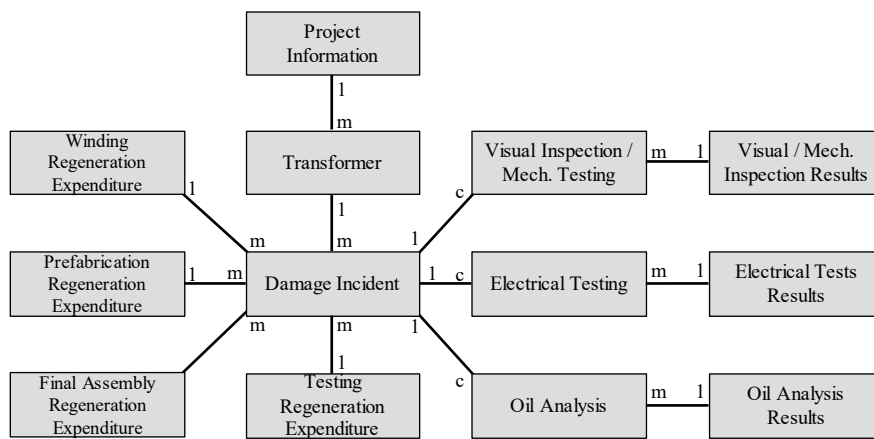


Fig. 3. Entity relationship diagram for the physical data model of the regeneration database

Formulate Consistency Requirements: During this step in the design process, we had to formulate requirements to ensure the consistency of data in the database. Among these, is the necessity for well-defined coding [12]. Thus, for example, in the transformer spreadsheet, a project number can only be entered when a corresponding data set is found in the project information spreadsheet. Moreover, the data quality can be ensured by setting specific value ranges for the individual attributes.

Formulate Transactions: When in operation, the database is continually expanded by all of the regeneration orders that arise. To ensure data consistency, operations need to be defined in the form of transactions that uniquely specify entering, editing or deleting data sets. Processually, access rights are also set for individual persons or groups [12].

3 Determination of a Suitable Data Mining Approach

According to the KDD process, once the database that is to be analyzed is created, a suitable data mining method has to be selected [8]. In view of the targeted implementation in the industrial practice, we can derive the following criteria for selecting an approach:

- *Forecast quality*: The method should be able to generate reliable load forecasts with a high predictive value for planning capacities.
- *Robustness*: The method should also be able to process incomplete input information. This criterion is particularly relevant because usually, not all of the information or only imprecise information about the damage incident is available at the start of a regeneration process.
- *Scalability*: The method should be applicable to a variant quantity of data in the damages database. This requirement ensures that the method will deliver valid forecasts even with a growing amount of data.
- *Transparency*: Users should be able to understand and interpret the application of the method and the targeted results. This requirement should ensure that users can identify discrepancies, for example resulting from input errors, and undertake appropriate correction measures.
- *Time Expenditure*: The time required for the user to prepare and operate the method should be reasonable and proportionate to the obtained benefit.
- *Flexibility*: The method should be able to be adjusted to changed conditions as quickly and easily as possible.

Within the framework of the research activities, different approaches from the fields of statistics and artificial intelligence were examined with regard to their suitability for determining load forecasts. Figure 4 briefly summarizes the results of our literature-based evaluation.

○ Requirements not met ◐ Requirements met slightly ◑ Requirements met partially ◒ Requirements met predominantly ◓ Requirements met completely	Statistics		Artificial Intelligence			
	KN	DA	BN	CBR	KNN	FL
Flexibility	◑	◐	◓	◑	◐	◑
Transparency	◑	◐	◑	◓	◐	◑
Time Expenditure	◑	◑	◓	◑	◐	◑
Forecast Quality	◐	◐	◑	◑	◑	◑
Scalability	◓	◓	◓	◓	◓	◓
Robustness	◐	◑	◓	◐	◑	○

KN: k-next Neighbor, DA: Discriminant Analysis, BN: Bayesian Networks, CBR: Case-based Reasoning, KNN: Artificial Neuronal Networks, FL: Fuzzy Logic

Fig. 4. Evaluation of possible data mining approaches

According to this assessment, the approaches from the field of artificial intelligence are particularly characterized by a higher potential in terms of forecasting accuracy. For a sustainable application in industry, especially transparency and time expenditure as well as robustness are of decisive importance. That is why Bayesian networks are chosen as the most promising approach when comparing approaches from the field of artificial intelligence.

Bayesian networks are among the visual models and result from a combination of graph and probability theories. The mathematical model underlying Bayesian networks is the Bayesian theorem, which links conditional and marginal probabilities of individual events. The symbolic or graphic representation of Bayesian networks is easily understood by users and transparent. [15, 16] Moreover, the forecasts of Bayesian networks are fairly accurate even with a minimal amount of data in the training set. Bayesian networks also fulfil the robustness requirement, since they can process incomplete data sets. Bayesian networks are therefore suitable for modelling and drawing conclusions when there is uncertainty resulting from unknown or incomplete information. [17] With regard to processing input information, Bayesian networks can work with continual or discrete variables [16]. Compared to simulation models, conducted queries are thus quickly answered by the Bayesian networks' analytical calculation [17]. To sum up, Bayesian networks fit the requirements described above comparatively best and will thus be implemented to determinate load forecasts to support capacity planning when regenerating rail vehicle transformers.

4 Summary and Outlook

Planning capacities for the regeneration of complex capital goods such as rail vehicle transformers is a tremendous challenge due to the imprecision of load information [1]. Methods from the field of data mining have proven to be useful tools to support planning processes for regeneration process [6]. Since logistics efficiency is a very important and competitive factor in regeneration business, more reliable planning benefits sales and profit [6, 18].

In order to support planners by means of data mining, it is necessary to have an underlying, suitable database with already conducted regeneration orders [6, 9, 10]. We thus detailed the process we took to develop a database model for regenerating rail vehicle transformers as well as the selection of a suitable data mining approach to generate load forecasts as a basis for capacity planning and throughput time estimation.

Following research activities will focus on determining load forecasts by Bayesian networks using real data from the regeneration industry to verify the general evaluation of this approach with regard the specific sample of application. Subsequently, the load forecasts may be used to enable automated decision support for capacity adjustments.

Furthermore, additional fundamental research activities may also address an empirical comparative analysis of different data mining approaches using data from different samples of application to review the so far literature-based evaluation.

5 Acknowledgments

The authors kindly thank the German Research Foundation (DFG) for the financial support to accomplish the research projects T3 "Capacity planning and quotation costing for transformer regeneration by means of data mining" within the Collaborative Research Centre (CRC) 871 – Regeneration of Complex Capital Goods.

References

1. Eickemeyer S.C., Nyhuis P. (2010): Capacity planning and coordination with fuzzy load information. *The Business Review* 16:259–264.
2. Uhlmann, E.; Bilz, M.; Baumgarten, J. (2013): MRO-Challenge and Chance for Sustainable Enterprises. *Procedia CIRP* 11:239–244.
3. Kuprat, T.; Nyhuis, P. (2016): Designing Capacity Synchronization within the Regeneration of Complex Capital Goods. *Universal Journal of Management* 4/10: 581-586.
4. Hoffmann, L.-S.; Kuprat, T.; Kellenbrink, C.; Schmidt, M.; Nyhuis, P. (2017): Priority rule-based planning approaches for regeneration processes. *Procedia CIRP* 59:89-94.
5. Gassner, S. (2013): Instandhaltungsdienst-leistungen in Produktionsnetzwerken. Mehrzielentscheidung zwischen Make, Buy, Concurrent Sourcing und Cooperate. Wiesbaden: Springer Gabler.
6. Eickemeyer, S.C. (2014): Kapazitätsplanung und -abstimmung für die Regeneration komplexer Investitionsgüter. Garbsen: PZH-Verlag TEWISS - Technik und Wissen GmbH.
7. Cabena, P.; Hadjinian, P.; Stadler, R.; Verhees, J.; Zanasi, A. (1998): *Discovering Data Mining – From concept to implementation*. New Jersey: Prentice Hall.
8. Fayyad, U.; Piatetsky-Shapiro, G.; Smyth, P. (1996): From Data Mining to Knowledge Discovery in Databases. In: *AI Magazine* 17/3: 37–54.
9. Maimon, O.; Rokach, L. (2010): Introduction to Knowledge Discovery and Data Mining. In: O. Maimon und L. Rokach (Hrsg.): *Data Mining and Knowledge Discovery Handbook*. 2. Auflage. New York: Springer Science+Business Media: 1–15.
10. Seitz, M.; Sobotta, M.; Nyhuis, P. (2018): Einsatz von Data Mining im Regenerationsprozess von Schienenfahrzeug-Transformatoren. Potenziale für die Kapazitätsplanung und Angebotskalkulation. In: *Zeitschrift für wirtschaftlichen Fabrikbetrieb*, 113/12.
11. Cleve, J.; Lämmel, U. (2016): *Data Mining*. 2. Auflage. Berlin, Boston: De Gruyter Oldenbourg.
12. Steiner, R. (2014): *Grundkurs Relationale Datenbanken. Einführung in die Praxis der Datenbankentwicklung für Ausbildung, Studium und IT-Beruf*. Wiesbaden: Springer Vieweg.
13. Elmasri R.; Navathe, S. (2005): *Grundlagen von Datenbanksystemen*. München: Pearson Studium Informatik.
14. Cordts, S.; Blakowski, G.; Brosius, G. (2011): *Datenbanken für Wirtschaftsinformatiker*. Wiesbaden: Vieweg+Teubner.
15. Schiaffino, S.; Amandi, A. (2009): Intelligent User Profiling. In: M. Bramer (Hrsg.): *Artificial Intelligence, An International Perspective*. Berlin: Springer: 193–216.
16. Sebastiani, P.; Abad, M.M.; Ramoni, M.F. (2010): Bayesian Networks. In: O. Maimon und L. Rokach (Hrsg.): *Data Mining and Knowledge Discovery Handbook*. 2. Auflage. New York: Springer Science+Business Media: 175–208.
17. Uusitalo, L. (2007): Advantages and challenges of Bayesian networks in environmental modelling. In: *Ecological Modelling* 203/3:312–318.
18. Lucht, T.; Kämpfer, T.; Nyhuis, P. (2019): Characterization of supply chains in the regeneration of complex capital goods. *International Conference on Competitive Manufacturing (COMA 19) Proceedings*: 444-449.