EnviroInfo 2005 (Brno)
Informatics for Environmental Protection - Networking Environmental Information
Copyright © Masaryk University Brno, Brno 2005. ISBN: 80-210-3780-6

A New Approach to Solve Data Defects in Material Flow Networks

Leticia Arco[1], Isis Bonet[1], Lourdes García[1], Jorge Marx Gómez[2] and
Claus Rautenstrauch[2]

Abstract

Eco-balances are the basis for assessing production impacts on environment and creation of environmental reports. Creation of an eco-balance is supported by the software tool Umberto®. In Umberto material and energy flows, required for the production of inventory analysis, will be designed by means of material flow networks. This tool allows modelling the material flow networks using Petri Nets. During the creation and evaluation of material flow networks there appear some defects, which inhibit or make it more difficult to establish realistic statements towards environmental impacts. Therefore, in the following paper the defects in material flow networks will be discussed in detail. The classification of defects is followed by the analysis of defect causes and possible solution attempts to their removal will be presented. We propose a new solution attempt using Petri Nets, Neural Networks and Case Based Reasoning in order to solve missing data process and incorrect data in material flow networks.

1. Eco-balancing and Material Flow Networks

Eco-balances are the basis for assessing production impacts on environment and creation of environmental reports. According to DIN EN ISO 14040 eco-balancing is divided into four phases: goal and scope definition, inventory analysis, impact assessment and interpretation (Marx et. al. 2004).

Creation of an eco-balance is supported by the software tool Umberto®. In Umberto material and energy flows, required for the production of inventory analysis, will be designed by means of material flow networks. Material flow networks base on Petri nets, which consist of knots (transitions, places) and edges (arrows). Transitions symbolize transformation processes of material and energy. They are specified by input-output-equations. Places represent stores or serve the distribution of material flows. Raw materials and supplies, preliminary products, semi-finished products and by-products, waste and emissions are considered as materials. Arrows between places and transitions illustrate material flows. Arrows between two knots of the same type are prohibited. In order to create material flow network information about the process structure (including sub-processes) is necessary. Besides, subnets serve as the refinement of the material flow network (Marx et. al. 2004).

In order to make reliable statements about environmental impacts with the help of eco-balances it is, in principle, necessary to consider all material and energy flows. However, during the creation and evaluation of material flow networks there appear some defects, which inhibit or make it more difficult to establish realistic statements towards environmental impacts. Therefore, in the following paper the defects in material flow networks will be discussed in detail. The classification of defects is followed by the analysis of defect causes. Subsequently possible solution attempts to their removal will be presented. A

1 Central University of Las Villas. Faculty of Mathematics, Physics y Computer Science. Center of Studies of Informatics. Carretera a Camajuaní km 5 ½, Santa Clara, Villa Clara, Cuba. e-mails: leticiaa@cei.uclv.edu.cu, isisb@cei.uclv.edu.cu, lourdes@fce.uclv.edu.cu

2 Otto-von-Guericke-Universität Magdeburg. Faculty of Computer Science. Institute for Technical and Business Information Systems, P.O. Box 4120, 39016 Magdeburg, Germany, e-mail: gomez@iti.cs.uni-magdeburg.de, rauten@iti.cs.uni-magdeburg.de

possible solution attempt using the combination of the advantages of Petri Nets, Neural Networks and Case Based Reasoning will be presented. Missing process data and incorrect data in the material flow networks will be solved by our solution.

2. Defect Types

Data defects in material flow networks can be roughly divided into missing data and incorrect data.

2.1 Missing Data

Modelling a material flow network requires information about the structure of the modelled (sub-) processes and corresponding material including their quantities resp. calculation rules. If this information is not available or only partly available, it is said that data are missing.

Missing data can be classified more precisely in:

- missing process data,
- missing process steps and
- missing pre-chains resp. post-chains.

With missing process data the structure of the (sub-) process, which should be modelled, is well known (e.g. technical flow diagram), but there is a lack of some (or all) relevant materials or their quantities resp. calculation rules. The structure of the material flow network can be modelled, although, the specification of places, transition or arrows, which is necessary for the material flow network calculations, is not available (see Figure 1).

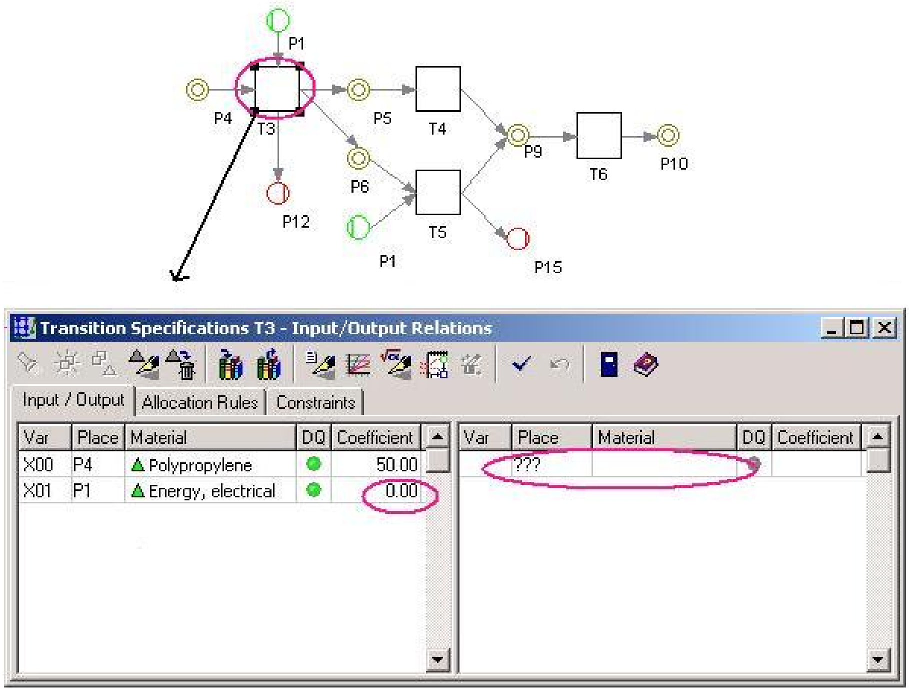

Fig. 1: Missing process data

Copyright © Masaryk University Brno, Brno 2005. ISBN: 80-210-3780-6

Missing process steps means that information about process structure of a (sub-) process is missing. A detailed modelling of the process as a subnet is, therefore, not possible, only a representation as a "black box" in form of a transition can be modelled.

If a material flow network contains incomplete information about the production of a used material (production of raw materials resp. preliminary products and also by-products, wastes and emissions), it is said that pre-chain is missing. Similarly missing post-chain, which means that there is no information regarding future usage of a material. In a material flow network knots and edges that are connected directly or indirectly with the help of transition in the pre-area (post-area) are classified as pre-chain (post-chain). The case when in the middle of the material flow net-work partial chains are missing will be classified as missing process steps.

On the one hand the reason for missing data is company's secrets. Publication of such data endangers the existence of an enterprise. On the other hand the measurability of data often is problematic. While calculating material flow network and inventory analysis, missing data cannot be taken into consideration. This results in the fact that only a limited assessment of the inventory analysis data concerning the environmental burdens is possible. The influence of the missing data can be only guessed. In this case a precision of the eco-balance will be pretended such that it does not actually exist. Mistakes in inventory analysis will be carried over to the impact assessment and interpretation (Ciroth 2001).

2.2 Incorrect Data

Mistakes during acquisition, calculation or aggregation of data lead to incorrect data. It is valid for both: in-house acquired and supplied from outside data.

Measurement errors can occur during data acquisition. They are an outcome of negative outside influences and can be subdivided in gross, systematic and random errors (Papula 1997). Gross errors appear because of, for example, reading errors, transfer errors or improper working measuring instruments. These errors are avoidable. Inaccurate measuring methods or incorrect measuring instruments lead to systematic errors. Thereby the measured values are falsified in the same way. Random errors are based on non-controllable negative outside influences like slight variation in temperature or air pressure.

During the calculation of data, wrong or incorrect calculation formulas can cause incorrect data. Furthermore, incorrect data can be a result of over-simplification. After the calculation of the material flow network incorrect data entail a faulty inventory analysis. Based on this, the impact assessment does not correspond to reality but looks correct or is "plausibly" reasonable. As a result the problem is to detect incorrect data.

2.3 Possible Defect Combinations

Various defect types and defects of one type can occur in a material flow network simultaneously. The occurrence of one defect type does not exclude the occurrence of other defect types and also has no influence on their characteristics.

3. Solution Attempts

3.1 Missing Data

There are various possible solution attempts for missing data:

Copyright © Masaryk University Brno, Brno 2005. ISBN: 80-210-3780-6

- The environmental management standards include statutory basis that could force a publication of data, in order to know about external process steps, pre- or post-chains and an acquisition of data by the eco-balance originator. The input-output-balances are available for the company (=customer), but in practice it seemed to be unreal if economic constrains are taken into consideration.
- To integrate input-output-data into supply chain management in order to enable an improvement of data availability for eco-balance production (Seuring 2000).
- To remove data gaps due to missing data is the use of library data (e. g. Umberto-library) or (public) databases. The disadvantages are: the number of available data is limited, and library data are average values, which normally deviate from real ones.
- To estimate values using fuzzy logic, but is inapplicable because it would mean inaccurate calculations rather than data gaps removing.
- To use knowledge acquisition and information fusion from (Kunze/Rösner 2001). The expenditure of this is often disproportional to favourable results. The second problem is the information content of the documents because of company secrets required data is not accessible.
- To use machine learning in order to find out the values for missing data from known transition specifications. But, a great number of the known transition specifications of similar technical processes are necessary, which is often available only at competitors. Besides, the received data can again be defective.

3.2 Incorrect Data

There are various possible solution attempts for incorrect data:

- To increase density of measuring points in conjunction with sample calculations.
- To use preventive measures.
- To compare measured data and data from independent measurements.
- If aggregate data are used it is necessary to control aggregate criteria. The assumptions should be managed if statistical methods are used.

4. A Possible Solution using Artificial Intelligence Techniques

4.1 Neural Networks and Petri Nets

Nowadays, neural networks are used in different fields. Classification is one of problems where they are commonly used. The Back-propagation algorithm is one of the most widely-used for training feed-forward neural networks because of its simplicity and capability to extract useful information from the examples and implicitly store it in their weighing connections (Hilera 1995; Yu 1995).

This algorithm has some limitations in its practical use that are generally approached and accepted by researchers. Some of these limitations are that its convergence toward a state of minimum error can be extremely slow, mainly if the size of the network is not big enough regarding the size of the problem. Next, it can standby in local minima before finishing the learning of all the examples, and finally, it is almost impossible to select the design of the network before hand (Hilera 1995).

Due to its complexity and slow process, a lot of software is developed to help the designers of these networks in the design and implementation of Multilayer Perceptrons (MLP). New training algorithms are implemented to achieve results similar to the traditional ones, in a short time.

Petri Nets are alternative tools for the study of non-deterministic, concurrent, parallel, asynchronous, distributed or stochastic systems. They can model systems in an easy and natural way. Furthermore, the Petri Nets approach can be easily combined with other techniques and theories such as object-oriented programming, fuzzy theory, neural networks, etc. These modified Petri Nets are widely used in

Copyright © Masaryk University Brno, Brno 2005. ISBN: 80-210-3780-6

computing, manufacturing, robotic, knowledge based systems, process control, as well as in other kinds of engineering applications (Li/Yu 2000). Since Petri Nets offer advantages to model systems and can interact with other techniques easily, it would be advantageous to model neural networks starting from Petri Net models, which allow not only the design adjustment but also the initialization of the neural network weights. Following the algorithm proposed by Xiaoou Li and Wen Yu in (Li/Yu 2000), we can model a neural network starting from a Petri net with the application of weighty production rules in the algorithm. See the following example.

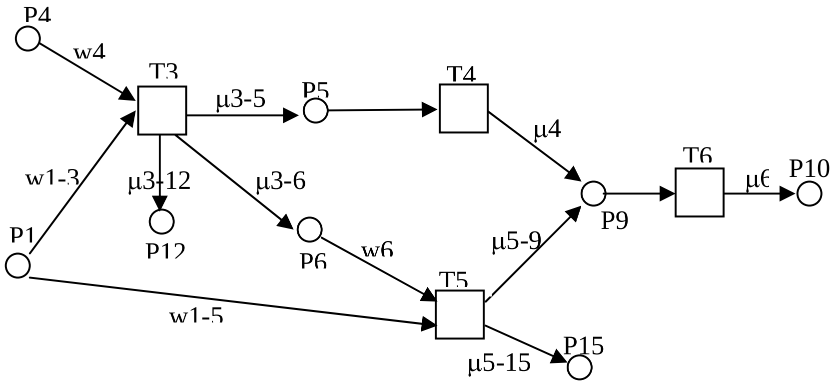

Fig. 2: Example of Petri Net

The following weighting production rules are obtained starting from the example of Figure 2 and following the algorithm described in (Li/Yu 2000):

1. If P1 and P4 then P5 (w1, w4)
2. If P1 and P4 then P6 (w1, w4)
3. If P1 and P4 then P12 (w1, w4)
4. If P5 or P6 then P9 (μ4, μ5-9)
5. If P6 and P1 then P15 (w6, w1)
6. If P9 then P10 (μ6)

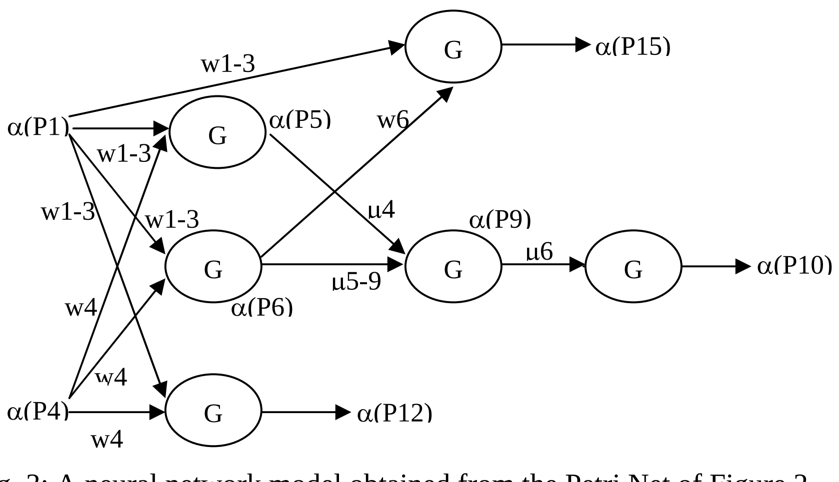

Fig. 3: A neural network model obtained from the Petri Net of Figure 2

Copyright © Masaryk University Brno, Brno 2005. ISBN: 80-210-3780-6

The learning algorithm of the neural networks obtained is the same as the backpropagation of multilayer neural networks. The main idea is that all layer weights can be updated through the backpropagation algorithm if certainty factors of all sink places are given (Yu 1995).

A complex neuronal network can be divided into several sub-networks starting from the modular design of an original Petri Net. The designed sub-networks will correspond to the real application sub-processes.

4.2 Case Based Reasoning

Case Based Reasoning (CBR) allows us to solve problems where reasoning is carried out starting from an associative memory with an algorithm to determine a measure of similarity among objects. Well-known solutions for a set of previously solved or not solved problems of the application domain can be the basis to solve a new problem. This method is different from others because it uses the stored information in the memory of the system in its original shape (López de Mántaras/Plaza 1997).

The bottom line of the CBR is to recover, adapt and validate the solutions found in previous experiences in their relation to a current problem.

CBR has different ways of using the cases:

- Antecedents may be used for reasoning.
- Old cases may be used to explain new situations.
- Old cases may be used to criticize new solutions.

This method may be advantageous if:

- rules are difficult to formulate and cases are available,
- rules are easily formulated but expensive to use,
- the formalization of problems is similar in cases with similar solutions,
- the usefulness of solutions is easy to verify or
- cases are available.

The CBR admits the use of the effort made in solving a problem for the further solution of any other new problem. It also lets us use previous experiences that have been successfully solved in the justification of other solutions. Non-successful ones can be used to anticipate problems. Another advantage would be that by recalling previous examples you can explain decisions already made. Besides, no interviews with experts are need, thus making the CBR a less complex procedure in the acquisition of knowledge. It allows for incremental learning. It allows for solutions in domains that are not entirely involved. Finally, cases may help determine which features in the problems are relevant (López de Mántaras/Plaza 1997).

It is necessary, however, to remember that the system does always explore the entire solution space. For that reason, optimal solutions cannot be found. Another disadvantage is that these systems call for a considerably large and well-selected case-base. Consistency among various cases is also difficult to maintain. It turns very difficult to find an appropriate similarity functions when solutions rely on their quality.

5. Hybrid System

Possible defects in material flow networks were approached in section 2. We propose a new hybrid system to solve two of these problems: incorrect data and missing process data. With this system we can design a neural network starting from the Petri net to model a specific material flow network application. This system combines this neuronal network with a case based reasoning module.

Copyright © Masaryk University Brno, Brno 2005. ISBN: 80-210-3780-6

The initial modelling Petri net will allow us to design a feed-forward neural network with the backpropagation learning algorithm, as described in section 4.1. It is even possible to create several neural networks starting from only one Petri net, taking into account its possible modules. To split neural networks in modules makes their training and performance easier.

This possible solution needs the initial construction of a case base to admit the training of designed neural networks. Therefore, it is necessary to store previous experiences of the problem to be solved.

It is necessary to combine neuronal networks with a case based reasoning module when there is missing process data. An alternative is to recover the cases similar to the problem and estimate the possible values of non-valued features. The recovery of similar cases can be done using near-neighbour algorithm, decision trees or associative memory methods.

How to correct the incorrect data? It is also necessary to use a case based reasoning module to solve this problem. Neural network outputs are compared with similar cases in the case base and corrections are made.

The combination of neural networks together with the CBR module offers multiple solutions to experts. The recovery of similar cases can be used in the readjustment of neural network outputs, which, in turn, can be used in the network self-learning. It is necessary to build appropriate similarity functions for each application in order to obtain an appropriate performance of the combination of both techniques. The quality of the case base is fundamental, for the training of the neural networks as well as for recovering similar cases and readjusting neural network outputs. Elements that can contribute to increase the quality of the case base are the application of relevant feature selection techniques, the automatic generation of cases using reasoning for analogy, the learning based on explanation, and the genetic algorithms to build hypothetical cases.

6. Conclusions

Eco-balances are the basis for assessing production impacts on the environment and for the creation of environmental reports. That's why we are proposing a solution for the detection and correction of defects in the material flow networks.

The suggested hybrid system combines Petri nets and neural networks with case based reasoning. We use the advantages of Petri nets in order to overcome the neural network deficiencies concerning their original design and definition of their initial weighs. Our solution solves incorrect data and missing process data defects using neural networks and case based systems together. Recovered similar cases have allowed the readjusting the network solutions, as well as the correction of cases with missing or incorrect data. Another advantage would be to propose several solutions to experts. A disadvantage of this proposal is the construction of an appropriate case base for the problem to solve.

Bibliography

Ciroth, A. (2001): Fehlerrechnung in Ökobilanzen. Dissertation, TU Berlin.

Hilera, J. (1995): Redes neuronales artificiales: Fundamentos, modelos y aplicaciones. Ed. Addison-Wesley Iberoamericana.

Kunze, M., Rösner, D. (2001): Eine xml-basierte Werkbank für das Document Mining. In: Lobin, H. (Hrsg.) Proceedings der GLDV-Frühjahrstagung 2001, Universität Gießen, pp. 131-140.

Li, X., Yu, W. (2000): Dynamic Knowledge Inference and Learning under Adaptive Fuzzy Petri Net Framework, IEEE Transactions on Systems, Man, and Cybernetics – Part C: Applications and reviews, Vol. 30, No. 4.

Copyright © Masaryk University Brno, Brno 2005. ISBN: 80-210-3780-6

López de Mántaras, R., Plaza, E. (1997): Case-Based Reasoning: An Overview, AI Communications Journal, Vol. 10, No 1, pp 21-29.

Marx Gómez, J., Prötzsch, S., Rautenstrauch, C. (2004): Data Defects in Material Flow Networks - Classification and Approaches, Cybernetics and Systems: An International Journal (CBS).

Papula, L. (1997): Mathematik für Ingenieure und Naturwissenschaftler, Band 3, 2nd ed. Vieweg, Braunschweig/Wiesbaden.

Seuring, S. (2000): MANAGEMENT WISSEN - Stoffstrommanagement und Supply Chain Management. In: Umwelt Vol. 30, 6, pp. 30-31.

Yu, X., Chen, G., Cheng, S. (1995): Dynamic Learning Rate Optimization of the Backpropagation Algorithm, IEEE Transaction on Neural Networks, Vol. 6, No. 3, pp 669-677.

Copyright © Masaryk University Brno, Brno 2005. ISBN: 80-210-3780-6