# Improving Computer Supported Environmental Friendly Product Development by Analysis of Data

**Ileana Hamburg**
*Institut Arbeit und Technik, Wissenschaftszentrum Nordrhein-Westfalen*
*Gelsenkirchen, Germany, Tel: +49 2 09 17 07 265*
*E-mail: hamburg@iatge.de*

## Abstract

In this paper some proposals for a preventive environment protection, already in the product development process by analysing global environment aspects, market situation, strategy, philosophy and culture of the manufacturing companies as well as their customer behavior, are presented. Results of the research German group SFB 392 are included into the proposals.
An environmental friendly product development system is also presented in the paper including a module for supporting such analysis; it uses data mining and augmented fuzzy methods. The environment is an extension of a product development environment worked up by the Institut Arbeit und Technik, Gelsenkirchen, Germany, in cooperation with the University of Craiova and the Technical University "Gh. Asachi" Iasi, Romania.

Keywords: environmental friendly product development, data analysis, data mining, augmented fuzzy methods

## 1. Introduction

The increasing ecomomic growth carries a higher resource consumption and emission of materials which are damaging health and environment. This causality can be stopped only through a product-referring environment protection by considering products as indirect cause of environmental distructions. The products are developed by designers and engineers; they determine the functions, forms, effect principles and materials and decide about technical, economical and ecological properties of products. So a preventive environment protection has to be considered already in the product developement process. It can be done only by analysing global environment aspects, market situation, strategy, philosophy and culture of the manufacturing companies as well as their customer behavior.

In the second part of this paper we give some proposals in this direction as well as results of the research German group SFB 392 about the development of methods and work tools for an integral environmental friendly product development.

In part three we describe data mining and fuzzy methods as suitable tools to analyse global environmental policy, market, clients and companies specific aspects.

In part four we propose an environmental friendly product development system which includes a module for supporting such analysis and that uses data mining and fuzzy methods; it will be used in a pilot company of telecommunication components which is a partner of our research project. The environment is an extension of a product development environment worked up by the Institut Arbeit und Technik, Gelsenkirchen, Germany, in cooperation with the University of Craiova and the Technical University "Gh. Asachi" Iasi, Romania, within the project INCOT.

## 2. Computer supported environmental friendly product development

The first phase of product development, engineering design process, has a significant role for the productivity of manufacturing and for the determination of the total life cycle of a product.

The main tasks of the design engineer are to determine the properties of a product based on clients' requirements, on technical rules and on his creativity and to make them available to other experts in corresponding forms on different carriers. He should take into consideration requirements and wishes of company's management, marketing and controlling aspects. Because he can not be a specialist for design as well as for environment, ecological statistics or exploitation technology simultaneously, he needs to use software tools which analyze environmental policies, emissions, market data, clients behavior, previous similar products, products of competitors, etc. Many of the information used in this analysis (e.g. about long term environmental policy) are rather subjective, qualitative and have to be transformed in usable values and patterns to be used concretely by the design engineer for the development of a product.

The steps for such a methodology for an environmental friendly product development can be the following:

- the formulating of an environmental friendly strategy including specific aspects for determined group / parts of products, upgrading, refurbishing, etc.,
- analyse and estimation of market data, clients behavior, previous similar products, products of competitors,
- engineering design process taking into consideration results of analysis and estimations,
- analyse and evaluation of the new products, e.g. by using prototypes and product life modelling
- development of series production including optimization of the products and processes by using formal principles (e.g. further controlling, etc.).

The work within the research group SFB 392 has been supported by the German research community (DGB) since 1996. During the first work phase (1996-1998), a scientific basis for the development of environmental friendly products as well as the proof of a principle realisation of them were worked out. In the second work phase (1999-2001) the results were applied within prototypical implementations. A prototype of a product development environment has been developed where the product variants can be evaluated according to ecological, technical and economical criteria [1].

During the third phase (2002-2004) the methods and tools which have been developed in SFB 392 will be evaluated by users within a new special labor that will be set.

In our development environment presented in this paper we take into consideration some of the results of the SFB 392 [6].

In the next part of the paper we present data mining and fuzzy methods as suitable tools to analyse global, market, clients and companies specific aspects, according to the methodology described below.

## 3. Data mining and fuzzy methods

Data mining (DM) means the analysis of data, nontrivial extraction of implicit, previously unknown and potentially useful information from it and whose presentation is easily comprehensible and useful to users to achieve their objectives [7].

Such objectives for the management of data have emerged because more and more companies are moving a very large amount of data, e.g. for decision support to a centralized resource known as a data warehouse. Data are extracted from operational databases, which in a large organization can be dispersed among subsidiary units and centrally recorded ones. In theory "big data" can lead to much stronger conclusions for applications, but in practice many difficulties arise.

DM is an emerging field, but many of its basic principles are deducted from well-elaborated concepts of data bases, statistics, genetically algorithms, neural networks, fuzzy sets, etc.

In the following a summary of the main stages of a data mining process are given [4]:

- Definition of the goal for the data mining process which can be a business goal or an objective related to a company normal event, e.g. energy consumption in a process. In this phase, it is also planned how the discovery patterns should be used for finding the solution.
- Selection of data needed for the data mining project and the sources containing this data.
- Preparation of selected data involving e.g. cleansing the data, merging data sources, manipulation of existing data fields.

- Exploration of the prepared data, e.g. examining the minimum, maximum, average, etc., understanding the dependency between fields.
- Application of the pattern discovery interactive algorithm which allows also the transfer of user business knowledge to the discovery process. Rule or decision tree induction are the more established and effective data mining technologies being used. So rule induction is used to generate pattern that relate to the goal. The resulting patterns are generated as a tree with splits on data fields.
- Optimization of the previous stage in order to make a "good" prediction.
- Visual presentation of the developed patterns.

Many telecommunication companies, particularly in the USA (e.g. AT&Transnationals, GTE Telecommunications, AirTouch Communications, American Management Systems Mobile Communication Industry Group, Coral Systems of Longmont), have announced the use of data mining. With the hyper-competitive nature of this industry, a need to understand customers, to keep them and to model effective ways to market new products to these customers is driving a demand for data mining in telecommunications.

Referring to fuzzy methods, unlike a crisp (conventional) set which has members and non-members, fuzzy sets allow degrees of membership, represented by membership functions which take values between 0 and 1. Thus a fuzzy set can be defined by the fuzzy membership function from a universe of objects, X into the unit interval I = [0,1].
In order to exprime the extent to which, for instance, a particular property holds, fuzzy sets can be used to represent imprecise data.
The real power of fuzzy logic systems lies in the ability to represent a concept using a small number of fuzzy values which can be considered as values of a linguistic variable [5].

There are two main types of uncertainty in fuzzy set values:
a) Uncertainty due to measurement problems,
b) Uncertainty due to problems in assigning fuzzy set values to measurements.

The first type of uncertainty arises because it may be difficult or impossible to measure variables and/or there may be large errors in the measurements. In some cases only estimates may be available, possibly derived from measurements of another variable, and there may be limited information about how accurate these estimates are.

The second type of uncertainty arises due to a lack of information about the fuzzy set and the correspondence between measured values, whether qualitative (e.g. high or low) or numerical (e.g. 3.5 cm) and fuzzy set values.

A number of different techniques have been developed for representing information which is uncertain, inaccurate and/or inconsistent as well as imprecise, including augmented, intuitionistic and type 2 fuzzy sets. These approaches therefore give additional information, but at the price of slightly greater complexity.

Augmented fuzzy sets resolve the problem of representing information which is both imprecise and uncertain. by using two parameters rather than one i.e. an ordered pair of numbers between 0 and 1. The second parameter represents the degree of certainty or reliability of the information represented by the first value.

An augmented fuzzy set can be defined as a mapping from the space X of all possible points to an ordered pair of numbers on the unit interval I = [0, 1] i.e.

There are two different cases. The first element $a$ (x) of the ordered pair is always an ordinary fuzzy set membership function. The second value $b$ (x) represents the degree of certainty in the assessment of the first value. It can be either:
- a fuzzy set membership function,
- a probability.

In the membership function case $b(x)$ is the membership function of the fuzzy set: 'the value of the membership function is known with certainty'. In the probabilistic case $b$ (x) represents the probability of the set membership function taking the value $a(x)$.

Different expressions are obtained for the set operations of union and intersection according to whether $b(x)$ is a membership function or a probability. The union of a number of crisp sets is defined to include all the elements of any of the component sets. For fuzzy sets membership is replaced by membership functions, and the most commonly used generalisation of the union operator to fuzzy sets is through the maximisation operator, with each element in the union of fuzzy sets having the maximum membership function taken over all the fuzzy sets in the union.

The principle is similar for augmented fuzzy set, except there are now two values, the data itself and the degree of certainty with which it is known, which have to be taken into account, so that a simple maximum can no longer be used. Thus, an appropriate definition of the union operator for the augmented fuzzy set case will be an appropriate generalisation of maximisation of an ordered pair of functions. This can be treated as maximisation of a function of two variables or bicriteria optimisation.

There are many different ways of doing this, some of which give fairly complicated algorithms. The particular choice taken here is both computationally simple and relatively general. It is based on comparing the performance of the following augmented fuzzy sets to the default option companies).
a)   Maximisation of a simple function of a and $b$, possibly followed by maximising over $a$,
b)   Maximisation over $b$ and then over $a$,
c)   Maximisation over $a$ and then over $b$.
Option a) is the preferred option, but is only accepted if performance is satisfactory relative to the default option in terms of constants companies and d which can be chosen by the user to give the relative importance of maximising over $a, b$ or both together. If option a) is not found to satisfy appropriate conditons, option b) is then tested in an analogous way. If neither option is found to be satisfactory, then option c) is accepted. Sometimes it is useful to put additional conditions on the maximising values to ensure that they are the best overall values. In this case the product $ab$ is preferred to the sum $a+b$ of $a$ and $bs$ for the following reasons:
a)   To avoid the case where the highest overall sum is obtained with either of $a$ or $b$ zero.
b)   To give greater discrimination between different augmented fuzzy sets, since there are fewer pairs of numbers on the unit interval with a given product than a given sum.

Inference from a set of fuzzy rules involves fuzzification of the conditions of the rules, then propagation of the confidence values (membership values) of the conditions to the conclusions (outcomes) of the rules. The outcomes can be finally translated into conventional, crisp values (defuzzification).

There are a number of various methods of fuzzy inference and defuzzification [8].


## 4. An example

The first version of the integrated product development environment, presented in this section, has been developed within the EU-Tempus project INCOT for the design of machine tools, but, potentially, it can be applied anywhere in manufacturing industry. It is based on the Common Object Request Broker Architecture - CORBA and provides a communications infrastructure between the different applications which correspond to different tasks of the design engineer. CORBA has been chosen as a description language which allows incapsulation of the objects and supports communication between all the applications, rather than, as in many existing systems, just communication of the application with one or more data bases.

As shown in figure 1, the actual version of the design environment within our project will consist of a module for the analyse and estimation of market data, customers behavior, previous similar products, products of competitors (based on data mining and using augmented fuzzy methods [3]), a CAD module of calculation procedures (e.g. FABER) linked with the AUTOCAD system [2], a part for analyse and evaluation of the developed products (including a life modelling module) and many data bases of generic life cycle data, market and customers data, etc. The modular structure of the environment allows additional modules to be added or modules to be removed as desired, giving considerable flexibility.
For every module an agent is responsible for the presentation of data, receiving user events and controlling the user interface: Different types of software agents are illustrated in Figure 1 and include analysis agents (AA), CAD agents (CADA), evaluation and product life modelling agents (ELMA) and data-base agents (DBA). They are implemented by using different facilities of Java. For example the agents DBA use JDBC (Java Data Base Conectivity) and provide uniform access to a wide range of relational data bases. This agents also serve queries

from other users which can come in different formats and from different computers. The agents ELMA are supported by the Java API, which is a package of software giving features as the set of conventions used by Java applets, networking facilities (URLs, TCP and UDP sockets and IP addresses), security and JDBC.

In the following we present the module based on data mining. At the moment the module for predicting customers behavior is in development.

Our pilot company (which is a medium-sized, family-owned one) is interested in answering a wide variety of questions about customers with the help of data mining. For example:

- How does one retain customers and keep them loyal when competitors offer special offers and reduced costs?
- What characteristics make a customer likely to be profitable or unprofitable?
- What are the factors that influence customers to call more at certain times?
- What set of characteristics indicate companies or customers who will increase their line usage?
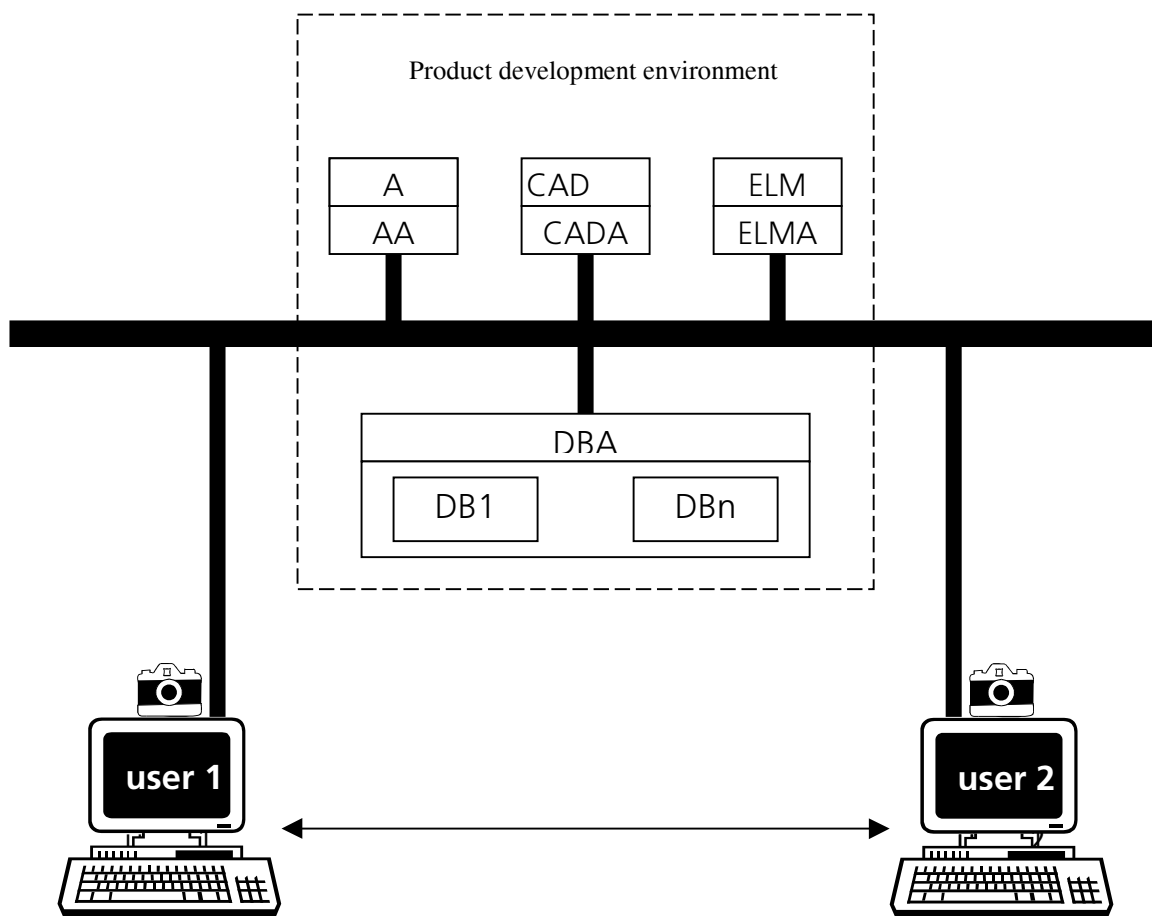
Figure 1: Product development environment

The project team wanted a controlled, scientific comparison of advanced data mining techniques such as neural networks and decision trees; this requires an investment in new software and training.

The company we work with was new to data mining and so we did a survey of software and decided to use XpertRule@Miner (Attar Software Limited: www.attar.com). One reason was that the decision tree and rule induction model was chosen. It did the best, apparently because this model was able to make good use of a wider range of variables to find many small pockets of profitable customers. Another reason was that a fuzzy logic implementation was developed in XpertRule in order to produce comprehensive features, to maintain ease of use and to integrate the fuzzy side with the non-fuzzy one of this software package. When inserting tasks, a Fuzzy

Rule mode of inference can be defined. This makes these tasks fuzzy. If the outcomes of a fuzzy task are not fuzzy (for example a list), then the task is defined to be of type List with fuzzy rules inference.

XpertRule allows also advanced facilities e.g. to implement the optimization of fuzzy rules and the fuzzification of discrete attributes.

In our project we use fuzzy reasoning to describe non numeric terms about the customers, e.g. "Loyalty of the customer is good". Loyalty may have many categories and we would need to define which of these are good.

We use client based data mining, e.g. the data to be mined is downloaded (extracted) and stored on the client machine (Windows 95 or NT). All the data preparation and mining is carried out on the client machine.

We are now at the stage of accumulation of data (creating data warehouses) and preparation of it and this are the biggest hurdles to begin the process of data mining.

## 5. Conclusion

In our project we would like to demonstrate how analysis of global environment aspects, of market situation, of behavior of companies customers can be integrated into software environments for product development. This fact supports the design engineer (if some of the data are imprecise, fuzzy) and make his work easier because he knows more precise requirements for the development of the product.

Our example shows that not only huge, multinational corporations could benefit from data mining, but also small- and medium-sized ones which have a good marketing team.

## *References*

[1] R. Anderl, H. John: Product evaluation integrated with product data models. Proceedings of the Product Data Technology Days- 9th Symposium, 2000, Noordwijk, Niederlande, pp.12-16.

[2] I. Hamburg, L. Padeanu: A flexible, user-friendly communication structure for engineering applications. Automated systems based on human skills, (Editors D. Brandt, J. Cernetic), VDI/VDE, 2000, Aachen, Germany, pp. 47-50.

[3] I. Hamburg, M. Hersh: The use of soft methods and data mining to support engineering design. Lucrarile stiintifice ale simpozionului international "Universitaria Ropet 2000", 19 - 20 octombrie, 2000, Petrosani, Romania, pp. 267-270.

[4] I. Hamburg, S. Pentiuc, S. Balanica: Improving cooperation in e-commerce by integrated mining methods in web-environments. 8th IFAC / IFIP / IFORS / IEA symposium on Analysis Design, and Evaluation of Human-Machine Systems, (Editor G. Johannsen), VDI/VDE, 2001, Kassel, Germany, pp. 165-169.

[5] M. Hersh, I. Hamburg: Fuzzy methods for efficient knowledge management: application to ´green´design processes. IASTED International Conference, (Editor M. Hamza), Goshen, 1999, Tretino, Italy, pp. 839-846.

[6] U. Lindeman, M. Mörtl: Ganzheitliche Methodik zur umweltgerechten Produktentwicklung. Konstruktion, November/Dezember 11/12, 2001, pp. 64-68.

[7] S. Weiss, N. Indurkhya: Predictive Data Mining. A Practical Guide. Morgan Kaufmann, 1998, San Francisco, California.

[8] L. Zadeh: Fuzzy Logic, Neural Networks, and Soft Computing. Communications of the ACM, Vol. 37, 1997, pp. 67-80.