

A Knowledge Discovery Method to Predict the Economical Sustainability of a Company

Peer-reviewed author version

DE VOS, Daniella; Van Landeghem, H. & VANHOOF, Koen (2006) A Knowledge Discovery Method to Predict the Economical Sustainability of a Company. In: CONCURRENT ENGINEERING-RESEARCH AND APPLICATIONS, 14(4). p. 293-303.

DOI: 10.1177/1063293X06072616

Handle: <http://hdl.handle.net/1942/1505>

A Knowledge Discovery Method to Predict the Economical Sustainability of a Company

D. De Vos*, H. Van Landeghem and K. Van Hoof*

*Institute of Mobility and Transportation/Data Analysis and Modelling, University Hasselt,

Wetenschapspark 5 Bus 6, B-3500 Diepenbeek, Belgium

Department of Industrial Management, Ghent University, Technologiepark 903,

B-9052 Gent, Belgium

daniella.devos@uhasselt.be - mobile +32(0)474 231064

Abstract - In this paper we are building a prototype of a machine-learning system using an inductive supervised approach to predict the logistical performance of a company. Focus lies on the learning phase, the handling of different types of data, the creation of new concepts in order to provide better measurable information. In this system numeric financial data are combined with categorical data creating symbolic data, distinguishing the phase of model generation from examples, and the phase of model classification & interpretation. The system has been implemented in vector spaces. Our data are benchmarking surveys on concurrent engineering measuring the usage of in total 302 best practices in Belgian manufacturing companies. The general purpose for implementing a best practice is the statement that the company will improve his product processing and that this way the company will establish his economical existence on the market. Our model processes a limited number of predefined steps generating value factors for the 302 best practices. The best practices are grouped into 30 subjects, the value factors combined in linear combinations. These value factors and their linear combinations are then subject to pattern interpretation relating concurrent engineering performance to past financial state of the company but also to an economical well doing of the company on a longer term i.e. we also refer to the sustainability of the company on the market.

Keywords: machine learning, knowledge discovery in databases, symbolic data, logistical performance, decision tool

I. INTRODUCTION

Extracting knowledge from data with the aim of understanding and explaining phenomena and to support decision making is the common goal of statistical data analysis and machine-learning. The semiautomatic methods for locating interesting information from data collections are still looked for and are still very useful. This is for certain within the field of process management and logistical management where a lot of data is still under analysed. Many logistical managers are still looking for “the” decision tool focussing the keys to success. Our prototype does not yet claim this superior position but though we are looking for new measurement concepts predicting logistical performances related to financial performances.

There is a need to describe more complex units or concepts that go beyond the classical framework in order to provide better measurable information. This is what meta-data do, or so called second-order objects or symbolic data. They are carriers of aggregated information and new information at the same time. In the field of machine learning several methods have been proposed to extend inductive approaches from statistical data analysis to data represented as symbolic data.

As the empirical data often consists of different types of data (numerical, categorical...) the builder of a machine learning system has to pay attention to the handling of these different types of data inside the same system. In this paper we face the problem of both handling numerical and categorical data creating symbolic data, distinguishing the phase of model generation from examples, and the phase of model classification & interpretation.

II DOMAIN, CONSTRAINTS AND GOALS

According to [1] the general goal of knowledge discovery is to obtain useful knowledge from the data collections. The user has to have a solid understanding of the domain in order to select the right subsets of data, suitable classes of patterns and good criteria for interestingness of the patterns.

A. The domain: a solid understanding

The data for this research is based on surveys on concurrent engineering (CE), from 1994 to 1999. 48 manufacturing companies in Italy and 64 manufacturing companies in Belgium have participated in these surveys [2]. The objective of the surveys was to investigate the degree of implementation of CE in those countries and to compare a concurrent engineering situation with a classic engineering situation. Classic engineering, also known as sequential engineering, is characterized by downstream departments supplying information to design only after a product has already been designed, verified and prototyped, in order to change what design engineering did wrong, or what could have been improved. In classic engineering, the various functions such as design, manufacturing, and customer service are separated. On the contrary, in CE all functional areas are integrated within the design process. In this case information continuously flows back and forth among all functions. During the design process CE draws on various disciplines to trade off parameters such as manufacturability, testability and serviceability, along with customer performance, size, weight and cost. The integration of other functional areas within the design process helps to discover hard-to-solve problems at the design stage. Thus when the final design is verified, it is already manufacturable, testable, serviceable, and of high quality [2].

These surveys are based on a CE compliance checklist. The checklist (called SEGAPAN checklist) measures how much of in total 302 CE best practices are being used. The questions (best practices) have been grouped into subjects regarding different CE practices. For this research we used the grouping of [3]. Each question has to be answered by ‘Yes’ or ‘No’. Table 1 describes all 30 subjects. The questionnaire has been filled through personal interviews with design and manufacturing engineers of the companies participating in the survey, to be sure that all questions were precisely understood.

Table 1. Structure of CE best practices.

Subject	Description
1. General scope of knowledge-Base	Is the CE concept used?
2. Management’s role	Responsibilities of management
3. Continuous improvement	Is there a conscious drive to improvement?
4. Cultural change	Is there a coaching process to overcome resistance to change?
5. Pilot project	How a pilot project is used to initiate change
6. Departmental interface management	Are integration mechanisms applied to facilitate CE?

7. Cross-functional teams	How are teams assembled and guided
8. Organizational structure	
9. Supplier's involvement	To what extend are suppliers part of the design process, and are they up to it?
10. Purchasing's role	Its role in part standardization
11. Customer's involvement	Are they actively solicited for their input and are they kept informed?
12. Employee involvement	Gauging up to what level employees participate
13. Training	The extend of employee training programs
14. Economical analysis	Are financial yardsticks used?
15. Computerized tools	Are computerized tools used and resulted this in benefit?
16. Design aids	Use of CAD, computer-aided software engineering system?
17. Design for manufacture and assembly	Use of DFX methods and their results
18. Rules-based engineering	Is A.I. used to support computerized development
19. Variety reduction	Standardization, Pareto, classification, ...
20. Design to cost	Use of formal cost models to design
21. Visualization tools	Use of modeling tools and prototyping
22. Computer-aided engineering	Are finite element methods used?
23. Value analysis	Is V.A. used and up to what extend
24. Monitoring and controlling progress	Use of Project Network Techniques, Critical Path Analysis, scheduling
25. Computer-aided manufacturing	Use of Computer-Aided Process Planning technique, graphical-based numerical control, ...
26. Statistical and quality methods	QFD, Taguchi, Ishikawa, 7 tools of quality
27. Logistics support	Use of reliability methods
28. Electronic Data Interchange	Use of EDI
29. Product data management	Use of shared databases PDMS
30. Group technology	Use of GT

The manufacturing companies were divided into 8 industry sectors. This paper represents research and results on the Belgian automotive sector (7 training cases) and the Belgian machinery sector (9 training cases). We have augmented the data for both sectors comprehensively with financial figures from the years 1994, 1996, 1998 and 2002 [4].

B. Research constraints

As mentioned in the introduction heading, we are working in a supervised learning environment this way building a machine learning system (Figure 1) in order to discover knowledge stored within the data. So our approach is inductive and we are generating a model from examples (so called training cases).

All levels in the construction of an automated knowledge creation model have to respect the overall objective, i.e. which factors do we have to consider in order to learn successfully.

This high level objective can be translated into a limited number of processing steps:

1. Defining the finite uniform input space.
2. Transforming data elements into a predefined format taking into consideration the sequence of the questions (and answers).
3. Functional mapping of the transformed data to value factors.
4. Creating a reduced set of meaningful symbolic data concepts.

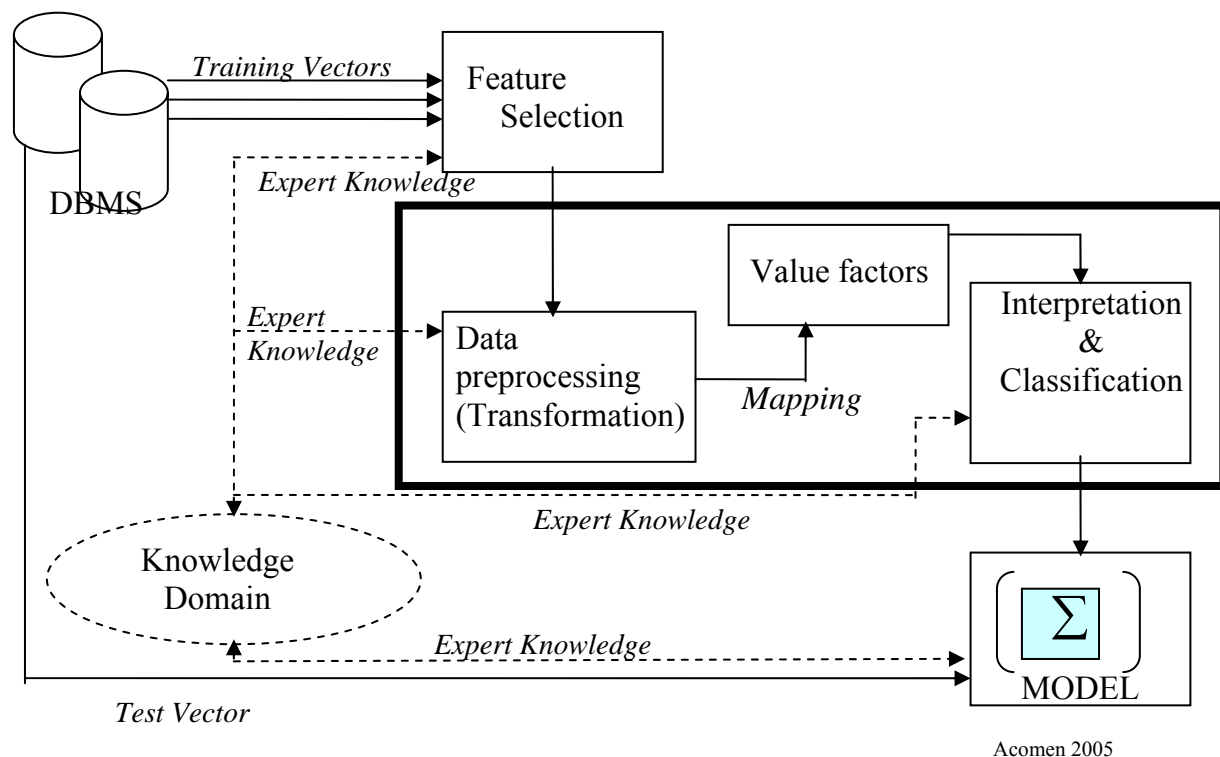


Figure 1. Knowledge creation system architecture with classification engine in a supervised learning environment.

Each step has a direct or indirect connection with the knowledge domain and each step receives input of knowledge from this knowledge domain and in return each step funds the knowledge domain with knowledge. This way the model can grow in knowledge performance. At this moment this research has been reduced to a one-cycle session i.e. no iterations in different sets of training cases have been investigated and compared, as this needs a larger input space.

In this research all questions for each subject have been included. As the number of questions for each subject varies, we need a ‘pre-processing’ phase in order to get all data in a predefined format known to the classification engine i.e. jumps to questions lower down the questionnaire, questions not representing a best practice but only with a guiding function, ... all these are taken into consideration and filtered out.

C. Research Goals

A best practice is a practice that has proven its quality, i.e. an equilibrium between what technically and economically can be achieved. So the general purpose for implementing a best practice is the statement that the company will improve his product processing and that this way the company will establish his economical existence on the market. This is what this paper, what this research intends to look for: a relationship between patterns of practising best practices and the financial state of the company

A first analysis of this goal description reveals also that

1. When referencing ‘patterns of practising best practices’, we do not estimate companies practising all best practices as ideal reference point.
2. When referencing the financial state of the company we refer to its past financial performance but also to an economical well or not well doing on a longer term i.e. we also refer to the sustainability of the company on the market.
3. When looking for a relationship we are focussing in the first place on a positive relationship, i.e. a good score on the best practices side should predict a fair well doing on the financial side. As one will see, this will not be the case. And indeed, the relationship can also be negative; i.e. the company makes a lot of efforts on the best practices side but still the financial results are poor or beneath expectations.

III KNOWLEDGE DISCOVERY IN DATABASES, MACHINE LEARNING AND STATISTICS

The concept of knowledge discovery in databases (KDD) or data mining is becoming increasingly popular as a business information management tool where it is expected to reveal knowledge structures that can guide decisions of limited certainty.

A. Close links between KDD, machine learning and statistics

Extracting knowledge from data with the aim of understanding and explaining phenomena and to support decision making is the common goal of statistical data analysis, machine learning and database management systems. Knowledge discovery in databases (KDD), often called data mining, combines methods of all three areas and aims at the discovery of useful information from collections of data. Both in industry and science there seems still to be a lack of methods for efficient analysis of large but also of small data sets. There is a suspicion that there might be nuggets of useful information hiding in under analyzed data. And therefore semiautomatic methods for locating interesting information from data collections are still looked for and are still very useful.

There is also an increasing need to extend standard statistical descriptive measures to more complex data that go beyond the classical framework. There is a need to describe more complex units or concepts in order to get more accurate information and to summarize data sets in databases. This is what meta-data do, or so called second-order objects or symbolic data as opposed to single individuals stored in tabular form in a database management system. In the field of machine learning several methods have been proposed to extend inductive approaches from statistical data analysis to data represented as symbolic data. In [5] we are working with decision-trees resulting in sequences of attribute-couples. Decision tree learning or rule induction is one of the main components in several data mining algorithms.

When building machine learning systems increasing attention is given to handling different types of data (numerical – discrete or continuous -, categorical data) within the same system. In [6] we are facing the problem of both handling numerical and categorical data creating symbolic data, distinguishing the phase of model generation from examples, and the phase of model classification & interpretation. This

partitioning of the full modelling generation into separate modelling phases creating new concepts, has been applied and extended within this research.

It is important to notice that data mining is relatively less concerned with identifying the specific relations between the involved variables. Instead, the focus of KDD is on producing a solution that can generate useful predictions. And the prediction lies in the predicted classifications. So specific logistical survey input data are analysed in order to reveal information on the financial well doing of a company (classification on finance).

B. Processing mixed classical input data

Part of the input space are “Yes” and “No” answers coming from the benchmarking questionnaires, i.e. categorical variables. The role of these “Yes” and “No” answers has been converted into Boolean expressions responding the machine learning environment. Categorical variables can be classified: a best practice is practised or it is not. Categorical data also called nominal variables therefore only allow qualitative classification; i.e. a best practice belongs to the distinct category “practised” or “not practised”, but it cannot be quantified and we neither can rank these categories.

The other part of the input space is the financial data looked up for each company under investigation. The financial figures selected are: turnover, added value, net profit after taxes, material investments¹, labour cost as a percentage of the added value. These figures are available in the annual statement of accounts published to the Belgian National Bank and are used in financial analyses. To this selection we added the “rating factor within the sector” calculated by Trends Top 5000 [4]. For each company under investigation we looked up the selection of financial figures for the fiscal years 1994, 1996, 1998, and 2002. This selection on fiscal years is based on the date of benchmarking survey, the availability of Trends Top 5000 financial figures and our research goals which are formerly described in §II.C. The financial figures are representing numerical variables i.e. quantitative expressions that can be mathematically transformed.

¹ In annual statement of accounts language : all material immovable assets purchased during the fiscal year.

So we end up investigating mixed measures. Both data are going through a pre-processing phase of selection and transformation and are combined in a mapping function resulting in a value factor for each best practice (Figure 2).

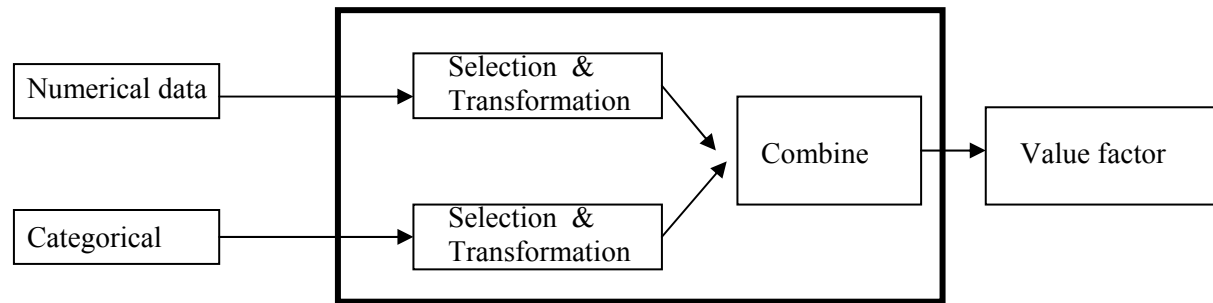


Figure 2. Mixed measures processing structure.

C. Creating new symbolic data

The purpose of this research is to transform the categorical and numerical data into meta-data giving more information i.e. providing more measurable information in fewer carriers.

Firstly we transformed the exact financial figures into meaningful relative ratios²: added value as a percentage of the turnover, turnover as a percentage of the net profit after taxes, labour cost as a percentage of the added value (available), turnover related to material investments, added value to material investments. Next ‘the rating in the sector’ as well as all relative figures (ratios calculated) are ranked from “best doing³” to “worst doing” for each company and this way each ratio is assigned to a ranking figure. So we continue to work with the ranking figures instead of the ratios. Then all ranking figures are summarized per company and scaled in a 20-80% rule, i.e. splitting up the range 0-100% into 3 parts: 100-80% financial best doing companies are assigned to class3, 80-20% financial well doing companies are assigned to class2, 20-0% financial poor doing companies are assigned to class1. The last

² Transforming exact figures into relative ratios is common in financial analyses as to make comparison between the companies under investigation conceivable e.g. divide ‘added value’ into ‘turnover’.

³ Depending on the relative ratio “best” doing can be the highest figure but can also be the lowest figure. E.g. “best” doing for the ratio “added value as a percentage of the turnover” is that company with highest result because this company achieves to obtain the highest result in added value per euro sold in goods. This company is assigned to the highest ranking figure.

aggregation puts the assigned classes for each company across all fiscal years in a table next to the year of benchmarking (Table 2). In this final table the financial figures of each company are levelled into the final meta concept class (class3, class2 and class1) taking into consideration the financial evolution of

	1994	1996	1998	2002	Year of survey	Class
Comp1	2	2	2	2	2002	2
Comp2	1	1	1	1	1999	1
Comp3	3	2	3	2	1998	3
Comp4	2	1	2	2	1998	2
Comp5	2	3	3	2	2002	3
Comp6	1	2	2	1	2002	2
Comp7	2	1	2	2	2002	2
Comp8	1	2	1	2	1999	2
Comp9	1	1	1	2	1999	1
Comp10	2	2	2	3	1998	2

Table 2 Classification according financial input space for the Belgian machinery sector (company names are not mentioned for privacy reasons).

the company compared to the year of benchmarking.

At this stage the financial numerical data transformed into a class carrier and the Boolean expressions for each best practice can be mapped using a mapping function. The mapping results in a value factor, i.e. a quantitative measure for each best practice to be used in linear combinations of the Boolean expression per best practice multiplied with the value factor of each best practice and this per subject and per company. Depending on their further usage value factors need to be normalized or can be used as is. The best practices are ranked according their value factor and allotted a ranking figure. This ranking figure is used in a linear combination per company across all best practices per subject. So different carriers can be calculated aggregating information.

The different carriers for each company per subject represent data of data or so called meta-data. Each carrier is a compression of classical data into a new concept that can be ranked. These meta-data represent a new type of data and are called 'symbolic data'. They extend the classical tabular model of data representation retrieved from the database management system (DBMS) and they also extend the classical statistical variables. Each concept (value factor, linear combination) is a weighted value for each related descriptive value (i.e. best practice, respectively each company related to the subject). The entire

classical data processing from feature selection, pre-processing, mapping up to classification represents a model to predict the CE performance of a company and this CE performance is closely related to the financial performance of the company.

Next to classical data analysis, these new concepts are ready for a symbolic data analysis [7]. The value factor per best practice, each linear combination is more than a single individual or data field like in classical data analysis, i.e. each value factor (respectively each linear combination) takes into account the ‘yes’ or ‘no’ performance of all training cases (companies) which in their turn has been classified this way taking into consideration the financial performance of each company as well.

IV. THE PROTOTYPE WORKED OUT IN VECTOR SPACES

The knowledge creation system in Figure 1 has been implemented in a vector space (Figure 3). The learning process starts by selecting all answers for one subject for each training case. Each training case represents a ‘training vector’

$$(x_{11}, x_{12}, x_{13}, \dots, x_{1n}) \in R^n \quad (\text{Eq. 4.1})$$

(Eq. 4.1) represents the vector $\overrightarrow{X_1}$ i.e. all answers for a subject for the first training case

$$(\overrightarrow{X_i}) \quad i \in \{1, 2, \dots, m\} \quad (\text{Eq. 4.2})$$

with i the addressed training case. The set of values of i covers all in survey manufacturing companies within the automotive sector in Belgium (respectively within the machinery sector in Belgium) being used in this training set up of the model.

Each vector $\overrightarrow{X_i}$ goes through a ‘pre-processing’ phase i.e. a transformation phase resulting in a vector $\overrightarrow{X'_i}$. This vector has been divested of those answers with no meaning for this investigation (jumps to questions lower down the questionnaire, questions not representing a best practice but only with a guiding function,...). Finally the vector $\overrightarrow{X'_i}$ with categorical data is ready to be mapped with the function θ . As

mapping function we used functions from the decision classifier's field i.e. we used the entropy information gain formula from the Information Theory (Eq. 4.3 and 4.4, [8] and the GINI-index function (Eq. 4.4, [9]). Both functions are looking for information embedded within the data.

The Information/entropy gain is formulated as Eq. (4.3 and 4.4).

$$GAIN = \sum_{j=1}^k \frac{n_j}{n} INFO(t) \quad (Eq. 4.3)$$

and $INFO(t) = - \sum p(j|t) \log p(j|t) \quad (Eq. 4.4)$

(Note: with $[p(j|t)]$ the relative frequency of class j at node t)

The GINI-index is formulated as Eq. (4.4)

$$GINI(t) = 1 - \lim_{x \rightarrow \infty} \sum_j [p(j|t)]^x \quad (Eq.4.5)$$

(Note: with $[p(j|t)]$ the relative frequency of class j at node t)

In this combining phase we stay at level 1 of the node levels. The class recording is based on the numeric financial data. This mapping of categorical and numerical data results in a value factor for each best practice (\vec{V}). The value factors can then be used in different linear combinations or ranking procedures resulting in a score $\vec{V} \cdot \vec{X}_i$ for each subject for each training case. This processing is repeated for each subject and for all training cases. Figure 3 gives an overall view of the processing in vector spaces.

After the learning phase and the generation of the model, a test vector \vec{X}_1 can be processed in the system in order to predict the concurrent engineering performance for each subject.

$$\vec{V} \cdot \vec{X}_i \quad (Eq. 4.6)$$

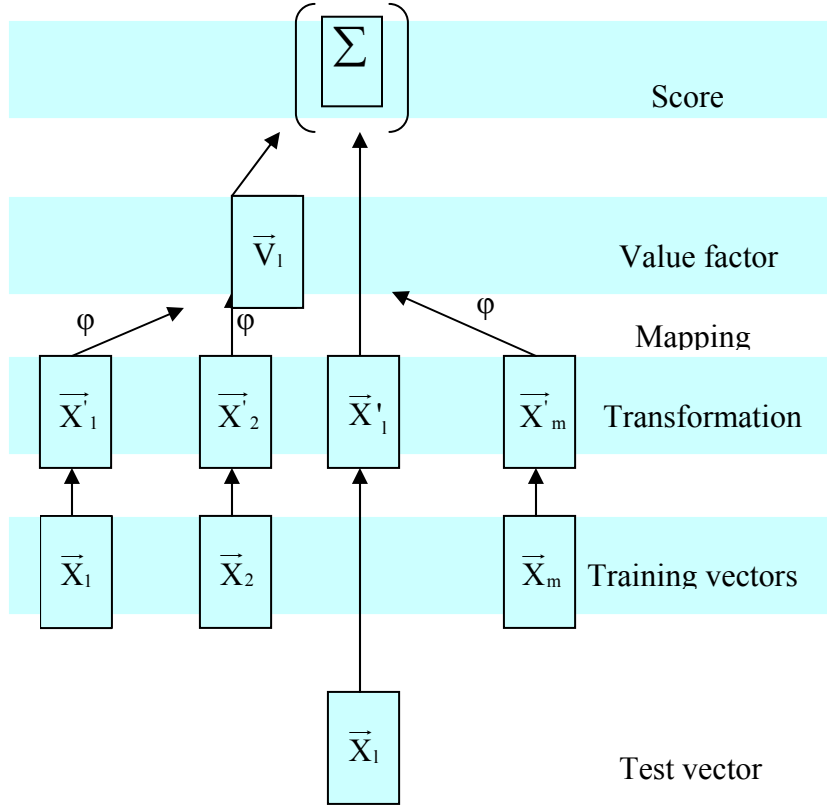


Figure 3. Architecture of the transformation of logistical classical data into symbolic data using vector spaces.

This categorical data processing no longer processes the combining of numerical and categorical data but instead utilizes the new concept i.e. value factors this way creating linear combinations for each subject of the test case.

V. DEPLOYMENTS AND CONCLUSION

This paragraph depicts the interpretation and classification phase (Fig. 1). We used tables and graphs to facilitate the interpreting of the new symbolic data.

Domain (4) Cultural Change					
Class		Ri	ri%	Total number of bp applied	Size
	(max number of bp)	32	100	7	
3	Comp3	22	69	5	120
3	Comp5	21	66	4	

2	Comp1	26	81	6	950
2	Comp4	32	100	7	2100
2	Comp6	12	38	3	
2	Comp7	13	41	2	
2	Comp8	17	53	3	125
1	Comp2	8	25	2	800
1	Comp9	25	78	6	12000
	Comp10	0	0	0	150
	Comp11	21	66	4	600
	Comp12	21	66	4	150
	Comp13	25	78	5	

Table 3 Results for subject (4) Cultural Change for the Belgian machinery sector.

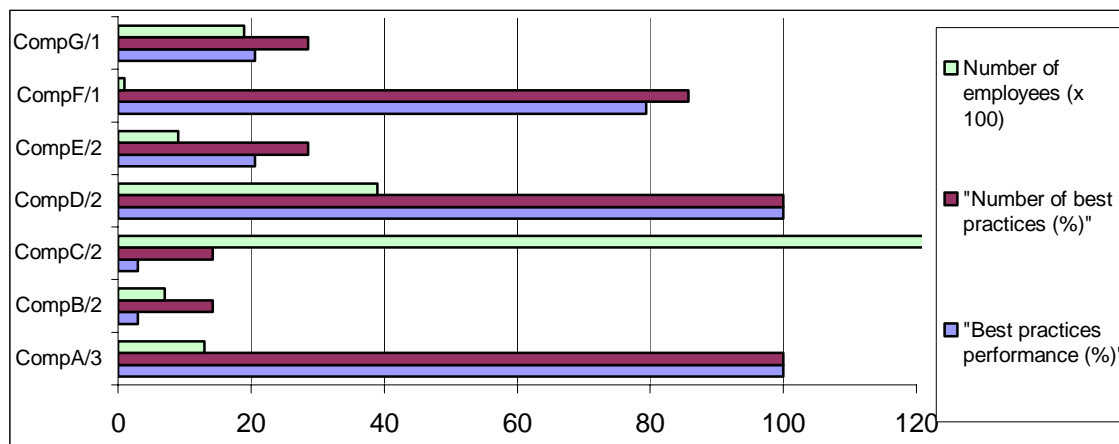
In tables 3 and 4 we listed the training results for the Belgian machinery sector respectively for the Belgian automotive sector for subject (4) Cultural change. The results in this paragraph are all based on the linear equations using the ranking figure (ri equation) allotted to each best practice and this for the entropy information gain function as mapping function. Using the GINI-index function resulted in the same ranking figure. We also listed the number of best practices applied, the maximum number of best practices that could be scored for the subject under investigation as well as the number of employees at the year of survey for each company (when available). For the graphs we used percentages for ri and the number of best practices applied in order to have a similar scale for both results.

domain (4) Cultural Change					
Clas		Ri	ri%	Total number of bp applied	Size
s	(max number of bp)				
3	CompA	24	73	5	1300
2	CompB	18	55	3	650
2	CompC	7	21	1	24000
2	CompD	29	88	5	3900
2	CompE	11	33	2	875
1	CompF	21	64	3	120
1	CompG	29	88	5	1900

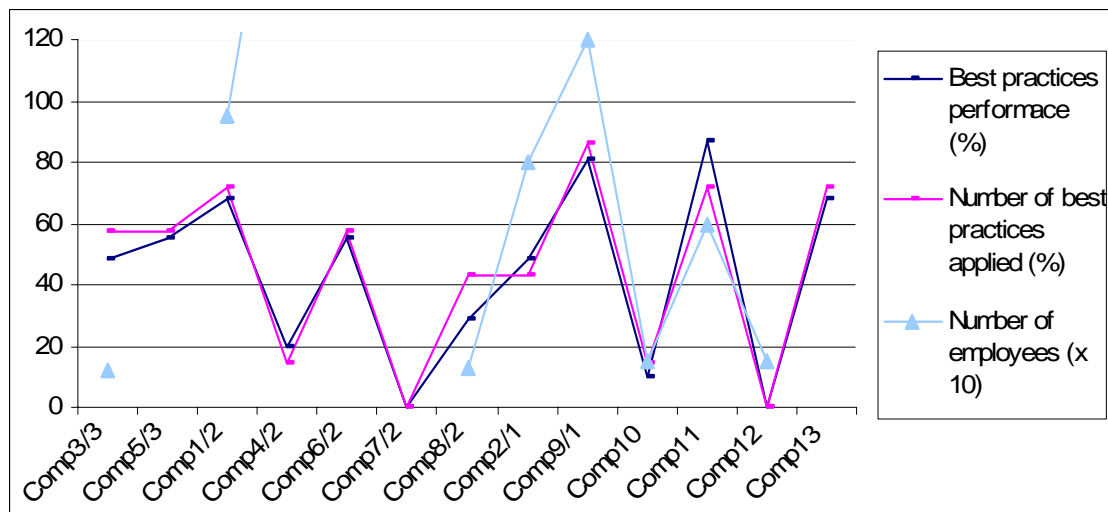
Table 4. Results for subject (4) Cultural Change for the Belgian automotive sector.

Analyzing the tables we notice that big companies are doing well on best practices performance but that this large effort is not at the same level as their financial performance (class2 and class1). There is no linear relation between the number of best practices practiced and the linear ri equations (best practices performance).

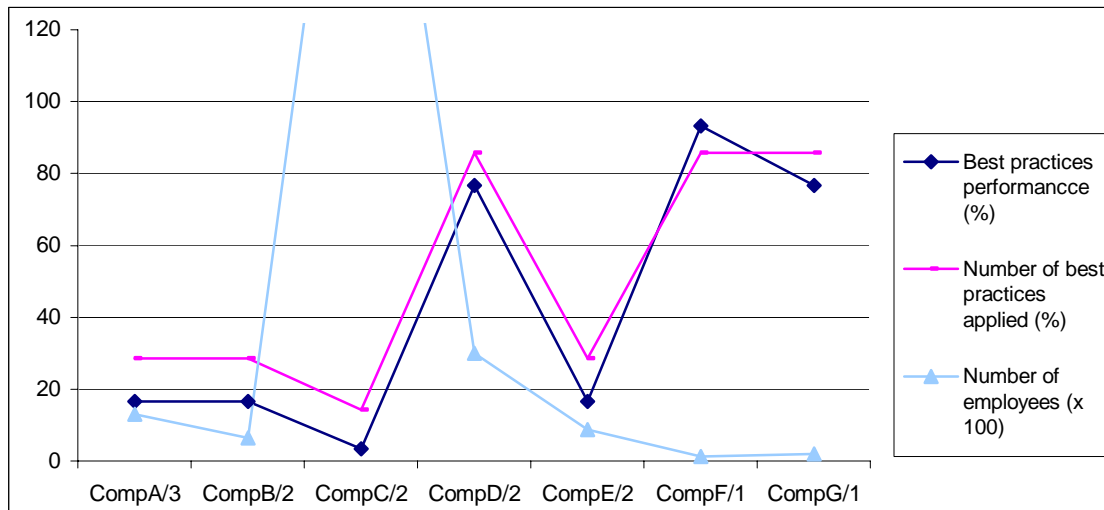
There are differences between the sectors but sometimes the differences are already very pronounced between the companies within the same sector (see graph 1 on the continuous improvement for the automotive sector).



Graph 1 Results on (3) Continuous improvement for the Belgian automotive sector (last character of each training case represents allotted class).

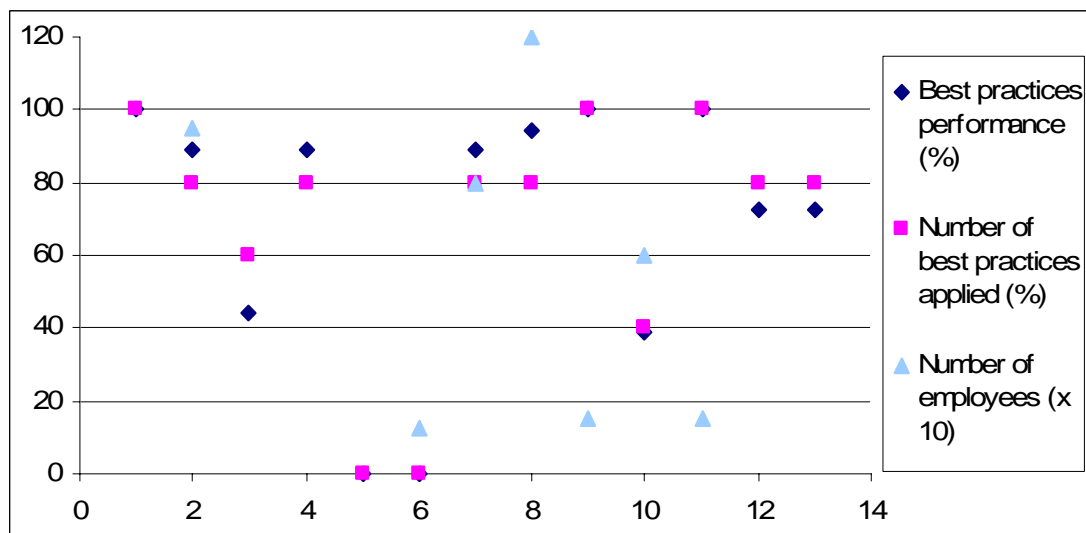


Graph 2 Results on (12) Employees involvement for the Belgian machinery sector (last character of each training case represents allotted class).



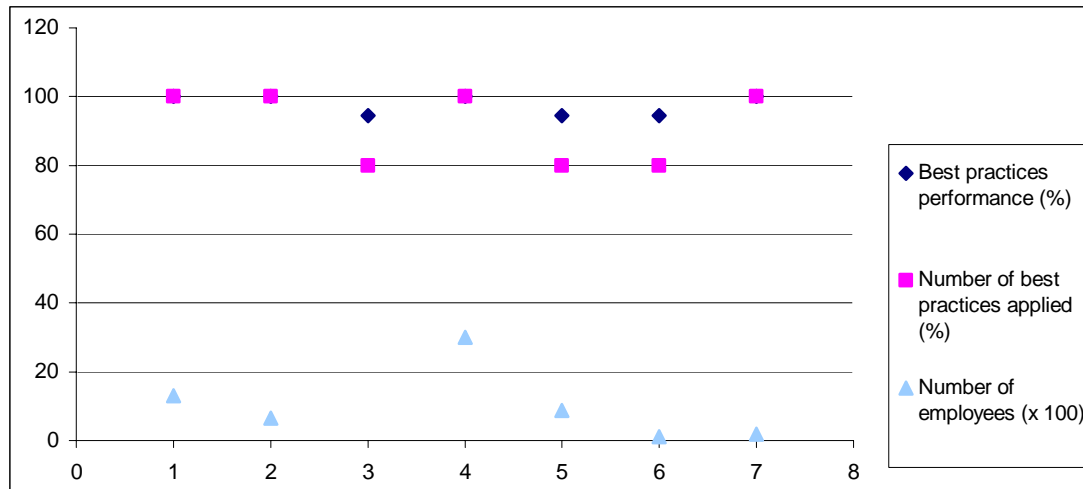
Graph 3 Results on (12) Employees involvement for the Belgian automotive sector (last character of each training case represents allotted class).

In graph 2 and 3 we notice that big companies do not realize to put specific best practices into practice (this could be related to the sector). Graph 4 and 5 reveals once more different patterns between sectors as well as the non-linear relation between performance of best practices and the number of best practices applied. In the automotive sector best practices performance is at a 100% or close to, even for those companies practicing 4/5 of all best practices related to this subject (computer-aided design).



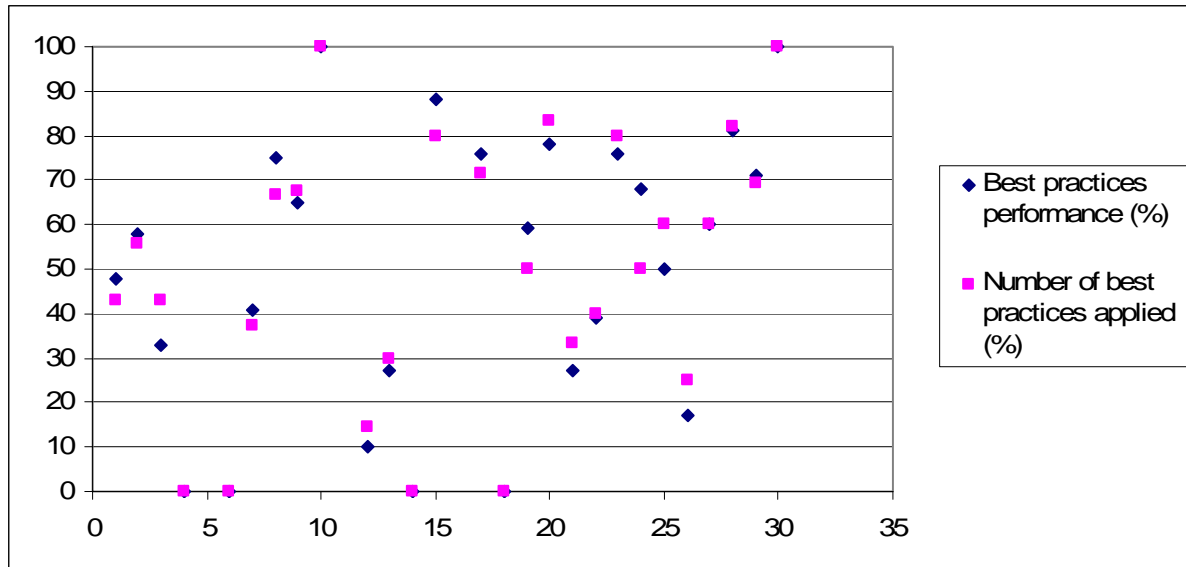
Graph 4 Results on (22) Computer-aided design for the Belgian machinery sector

For the machinery sector we have some test cases: Comp10 up to Comp13. For the financial performance Comp10 was assigned to class2 (with a lot of potential to go to class3) but Comp10 was not included in the selection as training case. As for the best practices performance, this test case was doing moderate in general (Graph 6), but analyzing more carefully one can notice some very good results for particular



Graph 5 Results on (22) Computer-aided design for the Belgian automotive sector

best practices subjects like (8) Organizational structure, (10) Purchasing role, (15) Computerized tools, (17) Design for manufacture and assembly, (20) Design to cost, (23) Value analysis, (28) EDI (Electronic Data Interchange), (29) Product data management and (30) Group technology. It seems that this small company with only 150 employees made a very careful selection in its skills: brilliant in a limited number, neglecting some and moderate on the rest. And it often succeeds in having a better score for best practices performance compared to number of best practices applied.



Graph 6 Results on best practices performance (in %) and number of best practices (in %) for Comp10 of the Belgian machinery sector as test case on all subjects (x values)

Some results and hypotheses:

- Big companies score very well in best practices CE performance but they do not succeed in putting all these efforts in cash.
- One does not have to be brilliant on the whole line of CE best practices to compete the market successfully hence the reluctance of SME's to surveys because benchmarking studies claim them to score on all.
- When choosing carefully one's skills one can gain doing efforts (higher performance with less number of best practices applied) and still do very well economically .
- There are different patterns of best practices CE performance between sectors but also within sectors between the companies themselves.

These results and hypotheses are preliminary and indicative. There is still a lot of noise in the model that has to be identified before we can start calculating levels of probability and predict with a statistical accepted level confidence limits. So the model is ready for further refinement: using other functions,

regrouping the questions into new subjects calculating the corresponding value factor, modelling a firm interpretation base ... From a theoretical point of view the model is based on some simple ideas and it also provides a clear intuition of what learning from examples is about. On the other hand the processing can be automated which makes it ready to be used as a prototype in benchmarking applications.

REFERENCES

- [1] Mannila, H. *Data mining: machine learning, statistics, and databases*, Department of Computer Science, University of Helsinki.
www.cs.helsinki.fi/~mannila/
- [2] Portioli-Staudacher, A., H. Van Landeghem, M. Mappelli, and C. Redaelli. 2003. "Implementation of concurrent engineering: a survey in Italy and Belgium" *Robotics and Computer Integrated Manufacturing* 2003; 19(3): 225-238.
- [3] De Wilde, H. 1993. *A Gap Analysis Knowledge-Base for Minimizing the Time-to-Market*, thesis, UGent Faculty of Engineering (promoter professor A.K. Kocchar).
- [4] Trends TOP 5000, de 5000 grootste ondernemingen in België, Brussels Media Centre. www.trendstop.be and published editions.
- [5] De Vos, D. and H. Van Landeghem. 2004. "Automated Knowledge Creation from Logistical Benchmarking Data", *International Conference on Cybernetics and Information Technologies, Systems and Applications CITSA* Orlando, Florida pp. 369-374.
- [6] De Vos, D. and H. Van Landeghem. 2005. "Transformation of Logistical Classical Data into Symbolic Data using Vector Spaces", 3rd *International Conference on Advanced Computational Methods in Engineering Acomen*, Ghent, Belgium, CD Rom only.
- [7] Diday, E. 1993. *An Introduction to Symbolic Data Analysis*. Institut National de Recherche en Informatique et en Automatique INRIA, N° 1936 Rapport de recherche.
- [8] Quinlan, JR. 1992. *Programs for Machine Learning*, New York: Morgan Kaufmann Publishers.
- [9] Vipin, K. and J. Mahesh . *Tutorial on High Performance Data Mining*, Department of Computer Science, University of Minnesota.
www.users.cs.umn.edu/~mjoshi/hpdmtdut/sld001.htm

AUTHOR BIOGRAPHIES

DANIELLA DE VOS holds a degree of Business Information Engineer (University Antwerp). She also studied philosophy (University Brussels) and biology (University Ghent). She has good experiences with industrial informatics and information banking finances . In 2002 she started her scientific career at the University Ghent. She was responsible for the functional design of an online benchmarking tool in the department of Industrial Management (Faculty of Engineering UGent). Later on she started her Ph. D. on benchmarking surveys, mainly Concurrent Engineering surveys. Since October 2005 she is related to the

Transportation Research Institute (IMOB) at Hasselt University in Belgium. She is teaching working colleges in Java and Visual Basic.Net and is doing research for her PhD. She published articles on her research of logistical benchmarking surveys and concurrent engineering surveys.

H. VAN LANDEGHEM holds a degree of Civil Engineer (Dipl-Ir) in Electro-Mechanics (UGent) and a M.Sc. in Industrial Management (UGent). He obtained his Ph.D. from Ghent University in 1984, based on work performed at the Georgia Institute of Technology in Atlanta (U.S.). He worked several years in industry as manufacturing consultant and in industrial informatics. He joined the BMT group in 1989, where he became Plant manager of the Sheet Metal division and finally the MIS Group Director. He currently is a Full Professor at Ghent University, as Head of the Department of Industrial Management (Faculty of Engineering Sciences). He is also director of the CIM_CIL Technology Transfer Center, a knowledge center for logistics and supply chain systems. He was elected Fellow of the World Academy of Productivity Science in 2006. He is Fellow of the European Academy of Industrial Management (AIM), and International Member of the Institute of Industrial Engineers (IIE). He is also an associate member of the Engineering Committee (CAWET) of the Academy of Arts and Sciences of Belgium. H. Van Landeghem is an expert in logistic systems and their application within companies, as well as in the application of Lean Manufacturing concepts. He published in International Journal of Production Research, European Journal of Operations Research, International Journal of Production Economics, International Journal of Operations and Production Management, International Journal of Operations Management, Robotics and Computer Integrated Manufacturing. He speaks frequently on conferences and training symposia for companies.

K. VAN HOOFF Koen Vanhoof attained a master in Physics in 1982 and a master in Computer Science in 1985 at the Katholieke Universiteit Leuven. His major research interests lie in the areas of data mining, statistics, knowledge engineering and modelling, computational intelligence methods, decision

support systems and soft computing applications to information management, marketing and finance. He has authored and/or co-authored over 20 peer-reviewed journal articles and about 6 book chapters and 60 conference papers on his research topics. He is co-editor of the International Journal of Information Theory and Applications. He has been appointed as a guest professor at Jagilionski University (Cracow, Poland), the University of Antwerp (Antwerp, Belgium), the University of Maastricht (Maastricht, the Netherlands), the University of Economics (Sofia, Bulgaria), the Technical University (St. Petersburg, Russian Federation) and the Academy of Economics (Wroclaw, Poland). He is Head of the Department of Applied Economics and project leader of the Data mining research group at Hasselt University in Belgium.