Interpreting Logistical Performances using Trellis displays

Peer-reviewed author version

DE VOS, Daniella; VANHOOF, Koen & Van Landeghem, H (2006) Interpreting Logistical Performances using Trellis displays. In: International Conference on Multidisciplinary Information Sciences and Technologies. p. 433-437..

Handle: http://hdl.handle.net/1942/1360

Interpreting Logistical Performances using Trellis displays

Daniella De Vos[a,], Koen Van Hoof[a] and Hendrik Van Landeghem[b]
[a] *Hasselt University,, Institute of Mobility and Transportation/Data Analysis and Modelling , Wetenschapspark 5 Bus 6, B-3590 Diepenbeek, BELGIUM*
[b] *Ghent University, Department of Industrial Management, Technologiepark 903, B-9052 Zwijnaarde, BELGIUM*

Your abstract should be clear, descriptive and no longer than 250 words. It should usually contain the objectives, methods, results, and conclusions of your work. Please submit your full paper following the styles of this template. Full papers are limited to a maximum of 5 pages. Copy and paste from your original contents to this template and then use the InSciT2006 styles. The pull-down style menu is at the left of the Formatting Toolbar at the top of your Word window (for example, the style at this point in the document is "InSciT2006 PARAGRAPH"). Highlight a section that you want to designate with a certain style, then select the appropriate name on the style menu. The style will adjust your fonts and line spacing. Use italics for emphasis; do not underline. Do not change the font sizes or line spacing to squeeze more text into a limited number of pages. Please be certain to follow all submission guidelines when formatting an article or it will be returned for reformatting.

Keywords: Trellis displays, logistical performance, decision tool, knowledge discovery in databases.

1 BACKGROUND AND MOTIVATION

For the past two years we are working on benchmarking performance data to explore the potential value of knowledge discovery methods in databases for information access. Many of these data are still very under analysed. In particlular we are interested in exploiting machine-learning systems using an inductive supervised approach to predict a company's logistical performance related to its past financial state.

Benchmarking assessments produce a broad heterogeneous collection of data difficult in analysing and which goes beyond the classical framework in providing measurable information. So there is a need to describe more complex units or concepts. There is a need for carriers of aggregated information and new information at the same time. And above all there is a need to interpret these new carriers of information.

In this paper we are extending our previous work in this field which was concentrated on building new carriers of information, to recent work that deals with visual displays reputed in discovering structure in complex data. And we summarize this research with some conclusions and future perspectives.

2 RELATIONSHIP TO PREVIOUS WORK

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. Our approach is inductive and we are generating a knowledge model from examples. Different types of data are combined and transformes in meta-data or symbolic data. We developed the model in vector spaces.

2.1 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 [1]. 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. By classic engineering all processes follow a downstream departments information supplying and 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. 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.

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 [2]. 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.

Table1. Structure of CE best practices grouped into 30 subjects.

Subject	*Subject*
1. General scope of knowledge-Base	16. Design aids
2. Management's role	17. Design for manufacture and assembly
3. Continuous improvement	18. Rules-based engineering
4. Cultural change	19. Variety reduction
5. Pilot project	20. Design to cost
6. Departmental interface management	21. Visualization tools
7. Cross-functional teams	22. Computer-aided engineering
8. Organizational structure	23. Value analysis
9. Supplier's involvement	24. Monitoring and controlling progress
10. Purchasing's role	25. Computer-aided manufacturing
11. Customer's involvement	26. Statistical and quality methods
12. Employee involvement	27. Logistics support
13. Training	28. Electronic Data Interchange
14. Economical analysis	29. Product data management
15. Computerized tools	30. Group technology

The manufacturing companies were divided into 8 industry sectors. This paper represents research and results mainly on the Belgian machinery sector (9 training cases) and the Belgian automotive sector (7 training cases). We have augmented the data for both sectors comprehensively with financial figures from the years 1994, 1996, 1998 and 2002 [3].

2.2 Creation of Knowledge : a classification engine

We are working on an automated knowledge creation system [4] (Fig. 1.). This paper is a contribution to the Interpretation & Classification phase.

2.3 Processing Kernel : a combination of mixed data types

Part of the input space data is coming from the benchmarking surveys, part are the financial figures looked up for each company. Both data, categorical and numerical data, are processed in the feature selection phase, transformed and then combined in a mapping process resulting in a value factor for each best practice. 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. So we are generating 302 value factors, one for each best practice.

2.4 Prototype : a model in vector spaces

The knowledge creation system (Fig. 1.) has been implemented in vector spaces [5]. The learning process starts by selecting all answers for one subject for each training case and each training case represents a 'training vector' i.e. all answers for a subject belonging to the training case. In the transformation process each vector is divested form those answers with no meaning for this research. 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 and the GINI-Index function. Both functions are looking for information embedded within the data. The resulting value factors can then be used in different linear combinations or ranking procedures resulting in a score for each subject for each training case.This processing has been repeated for each subject and for all training cases.

After the learning phase and the generation of the model a test vector can be processed in the system in order to predict the concurrent engineering performance and hence to proclaime the economical performance of the company. 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.

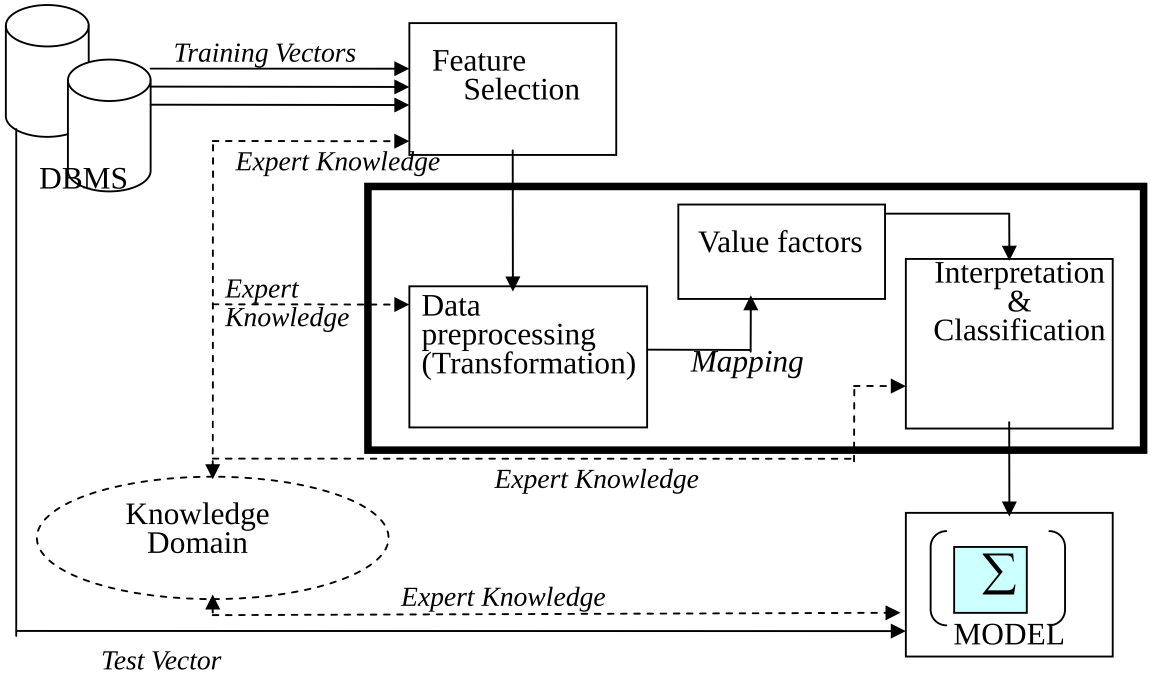

Fig. 1.Knowledge creation system architecture with classification engine in a supervised learning environment (Acomen 2005).

3 EVOLUTION OF RECENT WORK

Analysing and interpreting the value factors and their linear combinations still can be a daunting task seen the complex dependencies and the complex interactions among independent and dependent variables. So we looked for a suitable display technology promising visual pattern detection. The Trellis[1] display was used to reveal structure in our data.

3.1 Trellis Displays

The Trellis display is a framework for visualization of a multivariable database. The visual design reminisces a garden trellis work in which panels are laid out into rows, columns and pages. On each panel of the trellis, a subset of the data is graphed by a display method which in our case are scatterplots. Each panel shows per subject the relationship of the linear combinations of the value factors conditional on the class accorded to each company. So we extended the scatterplots to the explanatory variable 'class' trying out the mechanism which Trellis displays provide in understanding the interactions in our study of how a response depends on explanatory variables. Trellis was developed initially in the context of large data sets, but literature reviews [6] show that it is also useful for modelling small experiments. The concurrent engineering benchmarking surveys under investigation concern a large number of variables with a limited number of runs.

3.2 New Metrics and Usage

The value factors calculated in [5] are a new type of metric. This metric is quantifiable. To describe relationships it is often helpful to express them in quantifiable terms. And the quantification in this new metric lies in the information contribution revealed with formula's borrowed from the Information Theory.

This new type of metric is now assessed using graph analysis. This means that the hard assessment of qualifiable measures has been bypassed transforming them in quantifiable measures. Quantifiable measures have a much more interpretative value. The quantifiable measures are now tested on their spatial distribution and looked for patterns that link the economical well doing on their logistical performance.

Fig.2 up to Fig.10 represents a 'logistical dashboard' of trellis displays to highlight aspects of performance for companies with a poor, medium and high economical sustainability. So it is a multipanel

[1] Barley experiment from 1930 and Morris field revealed by Trellis displays in 1990s.

trellis dashboard display for logistical process decisions. It combines a variety of indicators in a coherently organized display. The data are gleaned from a selected number of subjects i.e. aggregations on best practice performance. One can easily see that while ….

4 CONCLUSIONS AND FUTURE PERSPECTIVES

In this research we leveraged 'vision' into a key component in our knowledge creation system creating stimuli to see structure in a graphed data set. Our emphasis has been on presenting the user with an overview of the structure of the result set rather than concentration on finding an individual relationship. We experimented with only one conditional variable, but the extension to many explanatory variables is under investigation and seems to be promising. It is credible to claim that visualization improves this type of research but still should be on a fair comparison with other wise equivalent alternatives.

There are many different information sources associated with logistical performance and there are many different kinds of information a user might like to know about logistical performance. This paper only claims to be a contribution to both aspects and was so far looking for mechanismes to help us to understand the contents of the data collection. We did not evaluate our system and the results of this research on a formal i.e. statistically significance. Instead we made a more informal approach and assessed in terms of more general properties.

REFERENCES

[1] Portioli-Staudacher, A., Van Landeghem, H., Mappelli, M., Redaelli, C. 1992. Implementation of concurrent engineering: a survey in Italy and Belgium. Robotics and Computer Integrated Manufacturing, 19(3), pp. 225-238.

[2] 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).

[3] Trends Top 5000, de 5000 grootste ondernemingen in België. Brussels Media Centre, www.trendstop.be.

[4] De Vos, D., Van Landeghem, H. "Transformation of Logistical Classical Data into Symbolic Data using Vector Spaces," 3 rd International Conference on Advanced Computational Methods in Engineering, Acomen 2005. Ghent,Belgium CD Rom only.

[5] De Vos, D.., Van Landeghem, H., & Van Hoof, K. A Prototype to Predict the Economical Sustainability of a Company Evaluating its Logistical Performance Using Knowledge Discovery Methods in Databases. International Journal of Concurrent Engineering: Research and Applications CERA, submitted.

[6] Trellis Display: What, Why, Who, When. Http://www.stat.ell-labs.com/project/trellis/wwww.html

[7] Goodchild, M. F. 1987. A spatial analytical perspective on geographical information systems. International Journal of Geographical Information Systems, 1(4), pp. 327-334.

[8] Goodchild, M. F., Haining, R., & Wise, S. 1992. Integrating GIS and spatial data analysis: problems and possibilities. International Journal of Geographical Information Systems, 6(5), pp. 407-423.

[9] The local government guide to geographic information systems: Planning and implementation. 1991. Washington, D.C.: PTI (Public Technology, Inc.): International City Management Association, 126 p.

[10] National Center for Geographic Information and Analysis. 1989. The research plan of the National Center for Geographic Information and Analysis. International Journal of Geographical Information Systems, 3(2), pp. 117-136.