Symbolic Data As Measures For Logistical Performances

Daniella De Vos Koen Vanhoof Hendrik Van Landeghem

Hasselt University
Institute of Mobility and Transportation/Data Analysis and Modelling
Wetenschapspark 5 Bus 6
B-3590 Diepenbeek
BELGIUM
e-mail: *daniella.devos@uhasselt.be*

Ghent University
Department of Industrial Management
Technologiepark 903
B-9052 Zwijnaarde
BELGIUM

Abstract: In this paper we are analysing logistical surveys. The surveys assess the level of best practices application within manufacturing companies and are processed in our prototype of a machine-learning system combining categorical data and numerical financial data building value factors. We extended the meta-data phase measuring dissimilarities between different modal variables. The main objective is to find patterns in the distribution of the results that can be translated in general rules.

Keywords: logistical performance, symbolic data

1. Relationship to Previous Work

We are working on surveys on concurrent engineering to explore the potential value of knowledge discovery methods in databases for information access. The data for this research is based on surveys on concurrent engineering (CE), from 1994 to 1999.

Table 1: *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

These surveys are based on a CE compliance checklist measuring in total 302 CE best practices. The questions (best practices) have been grouped into subjects. Table 1 describes all 30 subjects.
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 [Trends Top 5000].
We are working on an automated knowledge creation system [De Vos 2005]. This paper is a contribution to the Value factors phase and at the same time to the Interpretation & Classification phase which is extended with new possibilities.
Part of the input space data is coming from the concurrent engineering surveys; part is the financial figures looked up for each company. The financial data were used to classify the companies in three classes: class3 was accorded to those companies performing financially very well, class1 to the companies performing moderately and class2 is a class in between [De Vos 2006]. Both data, categorical and numerical data, are then 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. Evolution of Recent Work

2.1 Measuring dissimilarity

In [De Vos 2006] we continued to work with vectors on subject level. In this section each best practice is defined as a variable *y* for each training case i.e. a variable *y* is defined for all elements k of a set E (E = set of training cases per sector). This variable is termed set-valued with the domain γ if it takes its value in $P(\gamma) = \{U | U \subseteq \gamma\}$ that is the power set of γ. y_k is finite for each k, so *y* is called multi-valued i.e. all possible discrete values resulting from the mapping process per sector. The aggregated data per subject describe a group of individual best practices by set-valued or modal variables. So we continue to work with modal variables for each subject i.e. set-valued variables (for a formal definition see Bock 2000) with a measure distribution associated to y_k and their linear combinations Θ(k).
We are comparing two objects at the same level with each one another to obtain information about how they relate, how similar or dissimilar they are. As dissimilarity metric M(k) we are using the difference between the aggregated set-valued variables on subject level (Θ(k)) and the sum of the number of best practices applied Θ'(k) - both objects are quantifiable and expressed as percentages in order to have an equivalent scale to calculate the dissimilarity M(k) mathematically.

$$M(k) = \Theta(k) - \Theta'(k)$$

The equation above is calculated on subject level and describes the distance between two measured objects. The difference is quantified by the size of M(k) i.e. the distance between Θ(k) and Θ'(k), and the sign i.e. a positive or negative sign. So the distance describes precisely the difference between the actual measurements while "dissimilarity" might be an estimation of a distance we are not able to measure physically. Dissimilarity in this context expresses the dissimilarity between practicing a certain amount of best practices (Θ'(k)) and the output performance gained practicing a specific combination of best practices (Θ(k)) The dissimilarity yields a high number when the two objects differ in a high number of percentages and a low number otherwise.

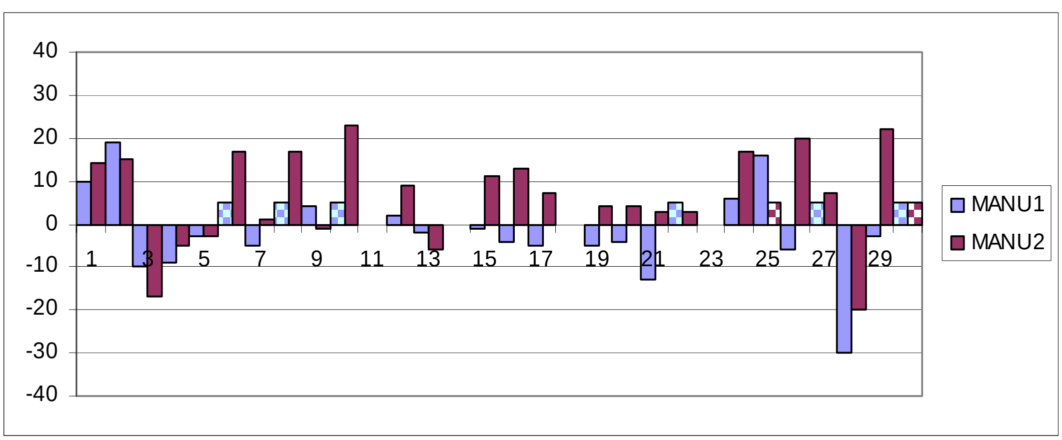
3.2 Bar Charts

A bar chart is used to graphically summarize and display the differences between groups of data. The range of data has been segmented into subject groups or subject bins. The vertical axis of the bar chart measures the dissimilarity M(k). The horizontal axis of the bar is labelled with the subject figure for each bin (see Table 1).

3.3 Deployment

The dissimilarity M(k) expressed in section 3.1 has been calculated for the training cases of the Belgian machinery sector as well as for the Belgian automotive sector (see Graph 1 for class3). The main objective is to find patterns in the distribution of the results that can be translated in rules. Following features were investigated:

1) The sign of M(k) : a positive sign indicates a surplus in logistical performance compared to a merely practising of practices, and vice versa.
2) The height of the bar equals the height of M(k) for that subject hence is a measure for the surplus gained in logistical performance.
3) The white gaps on the x axis represent gaps in the utilization of best practices for that particular subject.
4) A dissimilarity M(k) that equals zero indicates that there is no gain neither loss in the application of the specific best practice combination compared to practising a certain number of best practices for the subject under investigation (crossed texture in bar chart).



Graph 1: *Dissimilarity M(k) distribution for class3*

For class3 (cases MANU1 and MANU2) in graph 1, we noted:

1) For the most part the sign is positive, surely for MANU2, less for Manu1.
2) The height of the bar frequently reaches the level 10 up to 20 and higher (bar in full color).
3) Only 3 white gaps on 29 subjects measured (subject 11 is not measured).
4) The dissimilarity $M(k)$ equals zero for 6 (MANU1) and 2 (MANU2) subjects (crossed texture in bar).

Like wise we graphed the requested features for class2 and class1 and expressed them in rules. The rules differ much between class3 and class1, less for class2. Class3 has been more selective in choosing which efforts it decides to put into cash compared with class1. Hence competing the market successfully does not need to be related with the number of best practices applied or the size of the company (as big companies have more resources for applying best practices).

3. Conclusions

In this study we are measuring the difference between two modal variables. On the one hand we used the number of best practices applied. Maximizing this number is still recommended in the business application field. On the other hand we used the value factors resulting from our model. Latest modal variable stresses the application of best practices in patterns. Hence the hypothesis that applying as much best practices as possible is not the most efficient and effective strategy for a company. A careful selection in practicing which best practices to apply might be a less expensive strategy resulting in synergy between the selected best practices themselves. When choosing carefully one's skills one can gain efforts and still do very well financially.

References

Bock, H.H., Diday, E. (2000) Analysis of Symbolic Objects, Exploratory Methods for extracting Statistical Information from Complex Data, in : *Studies in Classification, Data Analysis, and Knowledge Organisation*, Vol. 15, Springer-Verlag, Berlin, 55-77.

De Vos D., Van Landeghem, H. (2005) 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.

De Vos D.., Van Landeghem, H., & Vanhoof, K. 2006 A Prototype to Predict the economical Sustainability of a Company Evaluating its Logistical Performance Using Knowledge Discovery Methods in Databases, in: *International Journal of Concurrent Engineering: Research and Applications CERA*, Vol. 14, Nb 4, SagePublications, 293-303.

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