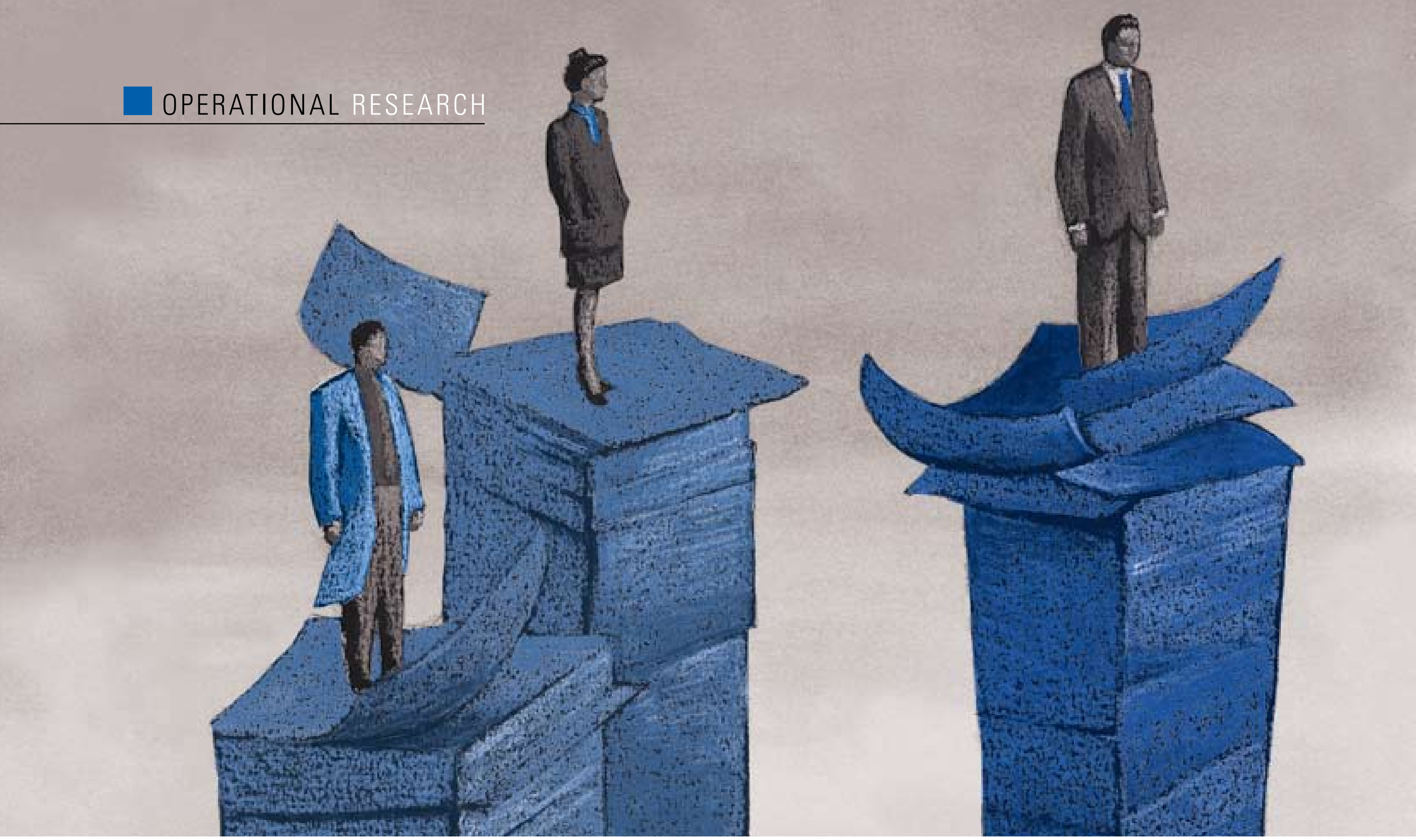


Image : Fotosearch

The subconscious *mined*

Beneath the surface of a myriad of business data, meaningful patterns can be found by the sophisticated data-mining techniques of neural networks.

by Riaan Olivier and Wim Gevers

The concept of self-organising feature maps (SOFMs) is a powerful tool that can be used in a variety of applications for data-mining and exploration, as well as decision support. In its basic form, an SOFM reduces the parameters that describe a specific application, making it easier for users to understand the underlying problem. By enabling the visualisation of the reduced parameters, it adds the strength of human visual interpretation to complex decision-making scenarios.

Applications of these tools may be found in bankruptcy prediction, control of industrial processes and in a variety of other instances with time-series data, as was the case in a study conducted at the University of Stellenbosch Business School (USB). This research examined how SOFMs could help solve the logistical planning of a complex product.

The product in question, an electrical distribution board, uses multiple components which are assembled in a customised configuration for each client. A final design may be made up from between 20 and 30 different items selected from a total of 554 components. Depending on the end-user's requirements, the different components can be used in various quantities. Over time, there is an apparently random pattern of component utilisation. Because of varying end-user requirements, probabilities that certain components will be used together with others are hard to establish through normal statistical methods.

How can one best decide which items to keep in stock and which to buy in as needed? The USB study set out to establish a practical methodology to support stockholding decisions. When an order is placed, there should be a high level of certainty that all components required to assemble the customer's configuration are in stock. Items not in stock may cause delays of up to four weeks, as these components are imported from Europe.

> **...it adds the strength of human visual interpretation into complex decision-making scenarios.**

Logistics in respect of such complex products often involve delicate trade-off decisions between easy administration, coping with delivery lead times, containing working-capital levels, and optimising revenues. In the example examined in the study, the unpredictable nature of component utilisation complicated these decisions. The problem was analysed by applying Kohonen's SOFM method.

New insights through visually displayed information

The study showed that the initial classification of stock items was not sufficiently accurate. It had to be changed as the data-exploration process revealed patterns that provided new insights. Once the reference data had been mapped, an accurate feature map evolved that could be used in further decision support. Multiple zones for conceptualising the stock-management problem became available.

After the model had been refined, new data that became available as the study progressed were tested. It became possible to see which components were changing their status, in which case stockholding had to be reconsidered. SOFM techniques were found to be an effective method for simplifying and supporting the stock-management decisions of the components of a complex product.

The real strength of SOFMs became apparent in the way they enable the user to analyse the data visually. The visual-interpretation capability of people can more easily make sense of information when it is displayed on two axes, which is precisely what SOFMs accomplish.

Neural networks and SOFM: excavating patterns from multidimensional data

SOFM, a data-mining technique based on the principles of neural networks, was first developed in its current form by Teuvo Kohonen in 1989. SOFM uses the neural transformation principle to create a two-dimensional map that makes the visualisation of information possible. It therefore transforms the original, multidimensional input space by reducing it to a two-dimensional grid.

Neural network techniques are modelled on the human brain and emulate the functioning of brain cells or neurons. Because of their ability to categorise complex information, these methods have become excellent tools for studying problems with multiple parameters and for reducing such problems to simpler levels.

Multidimensional data are available everywhere, mostly in the form of time-series data such as the marketing and financial information of companies. Various fields, such as finance, the medical environment, logistics and marketing, have already utilised neural networks to analyse problems.

Marketing: Marketing data can be notoriously multidimensional. SOFM has been used to study how the multiple parameters and data levels available from product movements and surveys can be simplified into a picture of market-growth potential.

Finance: The most popular application of SOFM to date appears to be bankruptcy prediction. Since much of the information relevant to a company is embedded in its financial ratios, time-series data of these ratios provide a perfect parameter base for SOFM applications.

Inventory management: Although such applications are more rare, neural networks have been used in inventory management models. In a particular case, consultants applied data-mining combined with neural networks to examine the stock system of a medical company in the United States. After analysis and re-modelling, they managed to cut the company's $1 billion stockholding by half without affecting statistical delivery times and probability of order fulfilment.

Probing the stockholding problem with SOFM

Component usage for a period of 18 months constituted the data used in this study. Each data point indicated the quantity of a particular item used in a particular month. The 554 components were therefore each represented by a time-series of 18 data points.

The initial understanding of the problem is shown in the table below. By interpreting utilisation data as best as possible, the researcher classified components into different categories according to how they moved.

Number of components	Classification	Description
40	Stable	Follow a stable demand pattern and can be described and estimated with usual statistical methods
44	Growing	Have shown increased demand in the last 3 to 6 months
81	Random	Have shown fair utilisation, but with a highly random pattern (this is a main area of concern)
52	Low	Have shown a highly random pattern, but with low utilisation
337	Non-stock	Have shown an even more random pattern; their low utilisation has not justified keeping them in stock

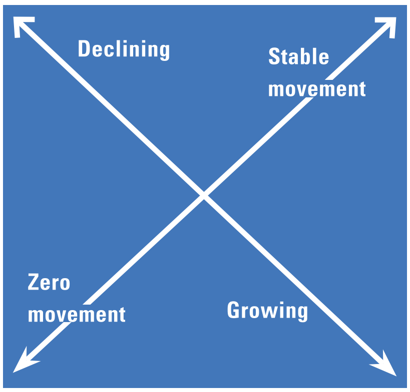
The Kohonen SOFM algorithm of SAS Enterprise Miner was applied to the problem in question.

There is no fixed rule for selecting the size of the network – the number of nodes on the output grid. A rule of thumb states that there should be at least as many output nodes as there are input vectors. The 554 input data points implied at least a 25x25 grid on the output map. For better spacing and discrimination between close-lying vectors, a 35x35 output map was selected. Although a 30x30 grid might have worked, the 35x35 grid indicated the clusters more clearly and showed up small differences between closely related input vectors.

By running the SOFM, visual interpretation of the data found scenarios where vectors were not classified in the same manner as other similar ones. Thus the initial visual method did not produce a consistent classification. Colour-coding of the initial classifications made it possible to show the positions on the map that could be associated with specific classifications.

It thus became a process of finding the area on the map where a specific colour dominates. If close inspection showed that validation vectors were incorrect, they could be reclassified. Several cases were found where the input vector was matched to a cluster of vectors that had another initial classification.

An example was where an input vector was classified as a random mover, but appeared on an output map node with a number of vectors classified as stable movers. It therefore became obvious that the initial estimation of several inventory items had been incorrect.

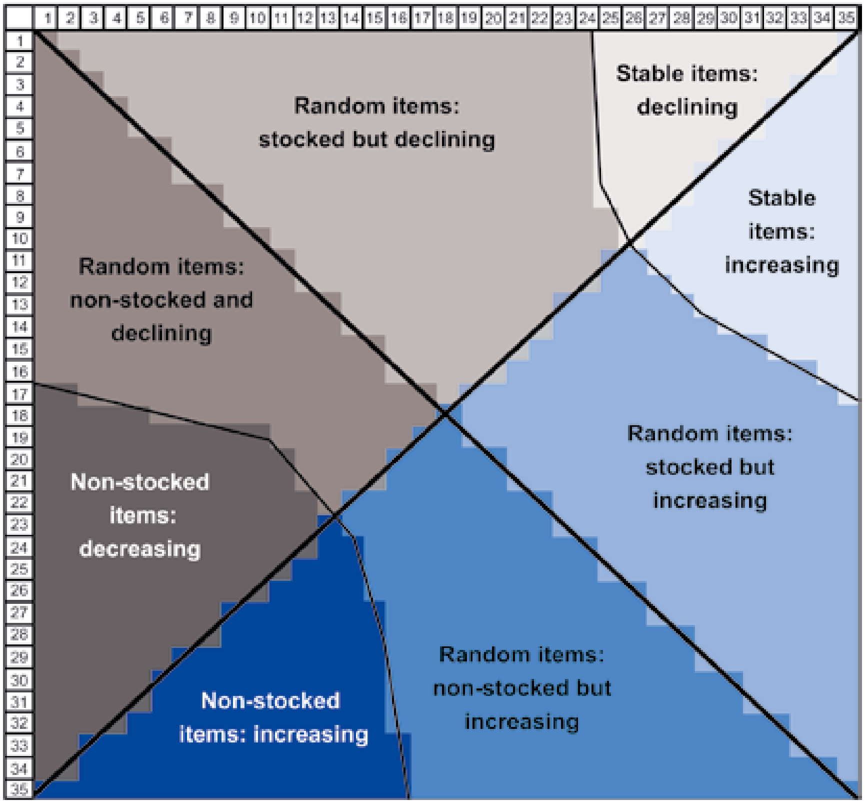


Underlying axis system

With these new insights vectors could be reclassified, and then re-analysed to explore and deal with new inconsistencies. Through this learning process, it soon became clear that the map had an underlying axis system that seemed to describe the data. The system appeared to run diagonally across the map, as shown above.

This finding made an improved classification system possible, which could now be further refined. Only after this process had gone through a few iterations was a result found that could be deemed sufficient for doing a final round of analyses and drawing conclusions. The picture that evolved from this process can be seen in the diagram opposite. It shows the 35x35 grid and the zones of vector clustering according to the final classification scheme.

A meaningful classification scheme, involving the eight zones on the map, could now be implemented instead of the original five classes.



A self-organising feature maps (SOFM) for stockholding

For each of the classes it is possible to define a separate and different stockholding regime.

A trained map to support decisions

Once such a trained map has been completed and all the prescribed steps fulfilled, it can be used for day-to-day management of the particular stockholding system by defining order quantities and re-order levels for each class of components. Every new order for products yields a resultant demand for components. As the demand for the components changes, it can simply cause components to appear within a specific area of the SOFM; based on their new positions and associated zones, the order quantities and re-order levels can be changed.

The trained map in the form shown above was subsequently used with three months of newer data, as mentioned earlier. This exercise clearly showed a few items that had moved from one zone to another, which necessitated the updating of stockholding decisions. These examples demonstrated how the trained map can be implemented to analyse and review the status of individual items on a continuous basis.

The final good news about a trained map? It does not require sophisticated software like SAS to deliver results. The underlying weights of the map, as produced by SAS, can be transferred to Microsoft Excel and month-to-month analyses can be carried out in Excel by informed administrative staff.

This study was conducted by Riaan Olivier (pictured) (riaan.olivier@schneider.co.za) as his MBA research project at the USB, with Prof Wim Gevers as supervisor. The research report, *A logistical application of the Kahonen self-organising feature map*, was presented in December 2005.