The Self-Organizing Map in Industry Analysis

Olli Simula[1], Petri Vasara[2], Juha Vesanto[1], Riina-Riitta Helminen[2]
[1] Helsinki University of Technology
Laboratory of Computer and Information Science
P.O.Box 2200, FIN-02015 HUT, Finland
[2] Jaakko Pöyry Consulting
Jaakonkatu 3, FIN-01620 VANTAA, Finland
e-mail Olli.Simula@hut.fi, Petri.Vasara@poyry.fi,
Juha.Vesanto@hut.fi, Riina-Riitta.Helminen@poyry.fi

3rd November 1997

Abstract

The Self-Organizing Map (SOM) is a powerful neural network method for the analysis and visualization of high-dimensional data. It maps nonlinear statistical relationships between high-dimensional measurement data into simple geometric relationships, usually on a two-dimensional grid. The mapping roughly preserves the most important topological and metric relationships of the original data elements and, thus, inherently clusters the data. The need for visualization and clustering occurs, for instance, in the data analysis of complex processes or systems. In various engineering applications, entire fields of industry can be investigated using SOM based methods. The data exploration tool presented in this chapter allows visualization and analysis of large data bases of industrial systems. Forest industry is the first chosen application for the tool. To illustrate the global nature of forest indsutry, the example case is used to cluster the pulp and paper mills of the world.

1 Introduction

The Self-Organizing Map (SOM), developed by professor Kohonen [9], is one of the most popular neural network models. The SOM implements a nonlinear projection from the high-dimensional space of input signals onto a low-dimensional array of neurons. The array forms an elastic net that during learning folds onto the "cloud" formed by the input data. The net approximates the probability density function of the input data. This means that the neurons tend to drift to where the data is dense, while there are only a few neurons where data is sparsely located. The mapping tends to preserve the topological relationships of the input signal domains. Due to this topology preserving property, the SOM is able to cluster input information and their relationships on the map. The SOM also has a capability to generalize, i.e. the network can interpolate between previously encountered inputs.

The most important applications of the SOM are in the visualization of complex processes and systems and discovery of dependencies and abstractions from raw data [8, 23, 26]. Especially the latter operation, data exploration or data mining, has recently become important because of the increasing amounts of measured information and data.

In industrial and engineering applications, the most straightforward applications of the SOM are in analysis and monitoring of complex process or machine states, otherwise difficult or even impossible to detect and interpret. The SOM algorithm is based on the unsupervised learning principle, i.e. the training is entirely data-driven and little or no *a priori* information about the input data is required. The SOM can be used for pattern recognition and clustering of data without knowing the class memberships of the input data. The SOM can thus be used to automatically detect features inherent to the problem. This is a clear advantage when compared with the artificial neural networks (ANN) methods based on supervised learning which require that the target values corresponding to the data vectors are known.

The SOM has been successfully applied in various engineering applications [10] covering areas such as pattern recognition, full-text and image analysis, financial data analysis, process monitoring and control, and fault diagnosis [17, 18, 21]. The ordered signal mapping property of the SOM has also proven to be useful in certain telecommunications tasks, e.g. in signal detection [15] and adaptive resource allocation problems [20].

Knowledge discovery in databases (KDD) is an emerging area of research in artificial intelligence and information management. The purpose of KDD is to find new knowledge from databases in which the dimension, complexity or the amount of data has so far been prohibitively large for human observation alone. Some typical tasks of KDD are classification, regression, clustering, summarization and dependency modeling. The algorithms that are employed in these tasks include decision trees and rules, nonlinear regression and classification methods such as feed-forward networks and adaptive splines, example based methods such as k-Nearest Neighbors (kNN), graphical dependency models and relational learning [3]. Analysis of major global industries is an example of an application for these techniques.

In this chapter, the applicability of the SOM based methods in knowledge discovery is discussed. Special emphasis is placed on industrial applications in which a lot of different types of information are available from databases and automation systems. Thorough analysis of the industry field requires, for instance, the integration of knowledge originating from different sources. A data mining tool ENTIRE utilizing the SOM algorithm is described. As a case study, the analysis of the world pulp and paper industry is considered.

2 The Self-Organizing Map in Knowledge Discovery

2.1 The Self-Organizing Map

The SOM algorithm resembles vector quantization (VQ) algorithms, such as k-means [1], and is closely related to principal curves [5]. The important distinction from VQ techniques is that the neurons are organized on a regular grid and along with the selected neuron also its neighbors are updated, so that the SOM performs an ordering of the neurons. In this respect the SOM is a multidimensional scaling method projecting data from input space to a lower, typically 2-dimensional output space [27].

A SOM consists of neurons organized on an array. The number of neurons may vary from a few dozen up to several thousand. Each neuron is represented by an n-dimensional weight vector, $\mathbf{m} = [m_1, \ldots, m_n]$, where n is equal to the dimension of the input vectors. The neurons are connected to adjacent neurons by a neighborhood relation, which dictates the topology, or structure, of the map. Typically a rectangular (figure 1) or hexagonal (figure 2) neighborhood is used.

The SOM is trained iteratively. In each training step, one sample vector $\mathbf{x}$ from the input data set is chosen randomly and the distance between it and all the weight vectors of the SOM is calculated using some distance measure, e.g. Euclidian distance. The neuron whose weight vector is closest to the input vector $\mathbf{x}$ is called the Best-Matching Unit (BMU):

$$||\mathbf{x} - \mathbf{m}_c|| = \min_i \{||\mathbf{x} - \mathbf{m}_i||\}, \tag{1}$$

where $\|\cdot\|$ is the distance measure.

After finding the BMU, the weight vectors of the SOM are updated so that the BMU is moved closer to the input vector in the input space. The topological neighbors of the BMU are also treated in a similar way. This adaptation procedure stretches the BMU and its topological neighbors towards the sample vector as shown in figure 1. The SOM update rule for the weight vector of the unit i is:

$$\mathbf{m}_i(t+1) = \mathbf{m}_i(t) + h_{ci}(t)[\mathbf{x}(t) - \mathbf{m}_i(t)], \tag{2}$$

where t denotes time. The $\mathbf{x}(t)$ is the input vector randomly drawn from the input data set at time t and $h_{ci}(t)$ the neighborhood kernel around the winner unit c at time t. The neighborhood kernel is a non-increasing function of time and of the distance of unit i from the winner unit c. It defines the region of influence that the input sample has on the SOM.

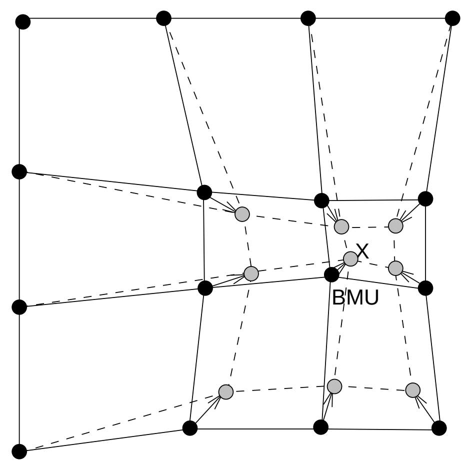


Figure 1: *Updating the best matching unit (BMU) and its neighbors towards the input sample marked with* x. *The solid and dashed lines correspond to situation before and after updating, respectively.*

2.2 Knowledge Discovery Using the SOM

The SOM has several beneficial features which make it a useful methodology in Knowledge Discovery in Databases (KDD). It follows the probability density function of the underlying data, it is readily explainable, simple and - perhaps most importantly - highly visual. The SOM functions as an effective clustering and data reduction algorithm and can thus be used for data cleaning and preprocessing. Integrated with other methods it can be used for rule extraction [22] and regression [16].

An important property of the SOM is that it is very robust. Naturally, the SOM suffers from the presence of any kind of flaws in the data, but the degradation of performance is graceful. An outlier in the data only affects one map unit and its neighbors. The outlier is also easy to detect from the map, since its distance in the input space from other units is large. The SOM can even be used with partial data, or data with missing data component values.

Three application areas important in KDD are highlighted in greater detail below.

2.2.1 Visualization

Because of the important role that humans have in KDD, visualization is a data mining method in itself in addition to being essential in reporting the results, or creating the knowledge [2]. The different SOM visualizations offer information of correlations between data components and of the cluster structure of the data. The illustrations can be used to summarize data sets and to compare them.

The SOM can be efficiently used in data visualization due to its ability to represent the input data in two dimensions. In the following (figure 2), several ways to visualize the network are introduced using a simple application example, where a computer system in a network environment was measured in terms of utilization rates of the central processing unit (CPU) and traffic volumes in the network. The SOM was used to form a representation of the characteristic states of the system.

- *The unified distance matrix (u-matrix)* method by Ultsch [23] visualizes the cluster structure of the SOM. Firstly, a matrix of distances (u-matrix) between the weight vectors of adjacent units of a two-dimensional map is formed. Secondly, some representation for the matrix is selected, for example a grey-level image [7]. The u-matrix of the example system is shown in figure 2a. The lighter the color between two map units is, the smaller is the relative distance between them. On the left side, there is a large uniform area, which corresponds to idle state of the computer system. The top right corner forms a clearly separated area, which corresponds to high CPU load in the system.

- *Component plane representation* visualizes relative component values in the weight vectors of the SOM. The illustration can be considered as a "sliced" version of the SOM, where each plane shows the distribution of one weight vector component. Using the distributions, dependencies between different process parameters can be studied. For example, Goser *et al.* [4] have used this kind of visualization to investigate parameter variations in VLSI circuit design.

  The component planes of the example system are presented in figure 2c. The colors of map units have been selected so that the lighter the color is, the smaller is the relative component value of the corresponding weight vector. It can be seen, for instance, that the components #1, #2 and #6 (read blocks per second, written blocks per second and write I/O percentage of CPU usage) are highly correlated.

- *Sammon's mapping* is an iterative algorithm [13] to project high-dimensional vectors in two dimensions. The nonlinear mapping tries to preserve the relative distances between input vectors. The algorithm can be used to visualize the

SOM by mapping the values of the weight vectors onto a plane. To enhance the net-like look, the neighboring map units may be connected to each other with lines to show the topological relations. Since the SOM tends to approximate the probability density of the input data, the Sammon's mapping of the SOM can be used as a very rough approximation of the form of the input data. The Sammon's mapping of the example system is illustrated in figure 2d. According to the mapping, the SOM seems to be well-ordered in the input space.

Sammon's mapping can also be applied directly to data sets, but because it is computationally very intensive, it is too slow for large data sets. However, the SOM quantizes the input data to a small number of weight vectors, which lightens the burden of computation to an acceptable level.

- *Data histogram* shows how input data are clustered by the SOM. In other words, it shows how many input vectors belong to clusters defined by each map unit. The histogram is formed using a trained SOM and a data set: for each data set vector, BMU is determined, and "hit counter" of that unit is increased by one. The histograms may be visualized in many ways. In our example, we have used squares of different sizes: the larger the square is, the larger is the counter value. The data histogram of the example application is shown in figure 2e.

- *Operating point and trajectory* can be used to study the behavior of a process in time. The operating point of the process is the BMU of the current measurement vector. The location of the point on the topologically ordered SOM can be easily visualized and used to determine the current process state. If also the history of the process is of interest, a sequence of operating points in time forming a trajectory can be studied. The trajectory shows the movement of the operating point, which in some cases may be a very useful piece of information. A piece of trajectory of the example system is illustrated in figure 2b. Also, the key areas of the SOM have been identified and labelled on the map. The trajectory starts from the normal operation area and moves through a disk intensive phase to high load area.

2.2.2 Clustering

Clustering is one of the main application areas of the SOM. The neurons of the SOM are themselves cluster centers, but to accomodate interpretation the map units can to combined to form bigger clusters. A significant advantage with this approach is that while the Voronoi regions of the map units are convex, the combination of several map units allows the construction of non-convex clusters.

A common strategy in clustering the units of the SOM is to calculate a distance matrix between the reference vectors and use a high value of the matrix as an indication of a cluster border [24, 11, 23]. In 3D visualization of such a matrix,

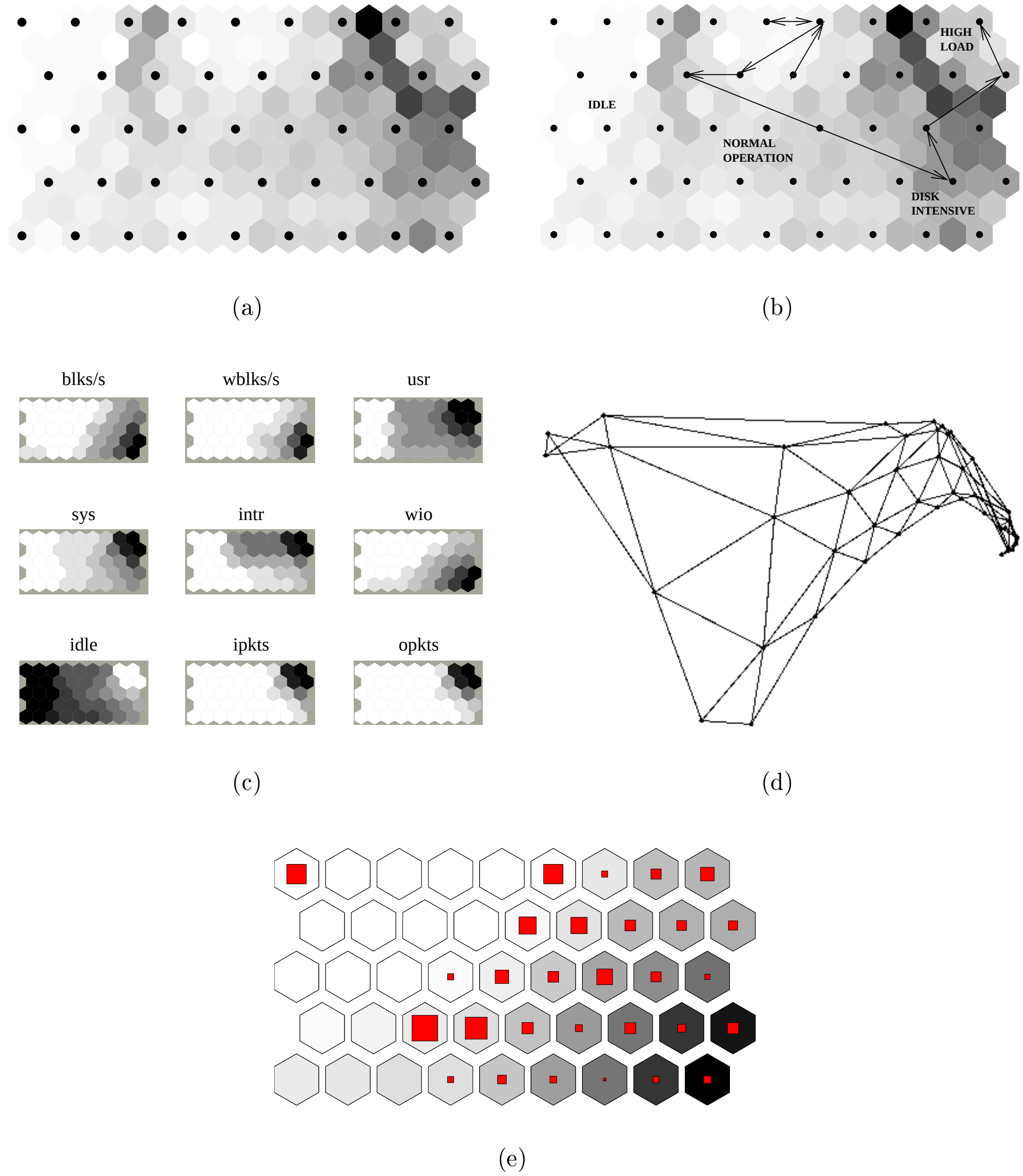


Figure 2: *Different visualizations of the SOM. U-matrix presentation (a), trajectory on top of labeled u-matrix (b), planes representation (c), Sammon's mapping (d) and data histogram on top of a component plane (e). In figures (a) and (b) the black spots denote map units.*

e.g. the u-matrix, the clusters will appear as "valleys". The problem then is how to determine which map units belong to a given cluster. For this, agglomerative and divisive algorithms are typically used, e.g. in [14, 25]. In addition to distance, some other joining criteria can be used, for example that the joined clusters are required to be adjacent [14].

Another quite interesting option is to use another SOM to cluster the map units. This kind of structure is often referred to as a hierarchical SOM. Usually a "hierarchical SOM" refers to a tree of maps the lower levels of which act as a preprocessing stage to the higher ones. As the hierarchy is traversed upwards, the information becomes more and more abstract. Hierarchical self-organizing networks were first proposed by Luttrell [12]. He pointed out that although adding extra layers to a vector quantizer yields a higher distortion in reconstruction, it also effectively reduces the complexity of the task. Another advantage is that different kinds of representations are available from different levels of the hierarchy.

The SOM can be used for classification purposes by assigning a class for each reference vector and deciding the class of a sample vector based on the class of its BMU. However, it should be noted that if the class memberships of the training data are known, using the SOM for classification purposes is not sound, since the SOM does not take into account the known class memberships and cannot therefore optimize the class boundaries appropriately. In such cases the Learning Vector Quantizer (LVQ), a close relative of the SOM, or another method of supervised classification should be used [9].

2.2.3 Modeling

The problem of system modeling is one of high practical importance. A traditional way to approach modeling is to estimate the underlying function globally. In the last decade, however, local models have been a source of much interest because in many cases they give better results than global models [19]. This is especially true if the function characteristics vary throughout the feature space.

The elastic net formed by the SOM in the input space can be interpreted as an implicit lookup model of the phenomena that produced the training data. The lookup model can be used for sensitivity analysis [6]. An expansion is to fit local models for each map unit. The local models can be constructed in various ways ranging from using the best example vector to splines and small MLPs. Usually, local models are kept simple, such as weighted averages of the example vectors or linear regression models [26].

3 Requirements for a Forest Industry Analysis Tool

3.1 From the General to the Specific: the Neo-Generalist Between Sectoral Expert, Amateur and Polymath

In an age other than ours, it was possible to be a polymath. Now, the sheer amount of information in each field makes this description hard to apply. Instead, we have a progression from the British "cult of the amateur" towards the generalist with domain expertise on on a broad but shallow basis. A global consulting company such as the Jaakko Pöyry Group[1] is a good laboratory for detecting the appearance of a new type of knowledge worker. The company's 35-year history is illustrative: a tradition of domain expertise was over the years slowly fused with a dose of management consulting. Then, the management consulting wing split off in a separate company (Jaakko Pöyry Consulting) under the group aegis. Finally, in the latter company the need for "neo-generalist" consultants with a combination of broad but shallow domain expertise and general, meta-level analyst skills is becoming apparent.

3.2 Future Direction of Knowledge Discovery in Industry Analysis

Much of the work done at Jaakko Pöyry Consulting can be put under the vague heading of "industry analysis", involving all relevant aspects of the forest industry, from markets and financial analysis to environmental issues and forestry. The arrival of the new type of consultant/knowledge worker is heralded by new types of problems gaining in relevance.

Characteristics of the problems inherent in this type of analysis include:

- Many dimensions of data.

  The number of relevant fields in different Jaakko Pöyry pulp and paper mill databanks, of which there are many, linked types, rarely drops below 30.

- Many categories of dimensions of data which have to be dealt with both separately and in combination.

  When the above mentioned databanks are used, the output is mostly a combination of at least two categories of information (technical data on equipment, time dependent production numbers, resource usages, production costs). These are of interest by themselves, but the combination is where the true value lies.

[1] A market leader in management and engineering consultancy for the forest industry worldwide.

- Non-existent or poor theoretical foundation.

  Rules of thumb abound, and even such a basic component as the pulping process is far from completely understood in the sense of physical modeling.

- Incomplete data.

  Even with an extensive global data collection network, it is nearly impossible to find out many highly relevant pieces of information. Some are protected by commercial interest, others are not measured in the same way or at all across the board, etc. The market shares of individual companies for different products in different countries are among these items, and environmental data is perhaps shrouded in the greatest veil of secrecy of all.

- Correlations between variables are not obvious.

  The prices for different products or resources (e.g. pulp and recovered paper; pulp and fine paper) have a logical connection but often seem to be the output of a chaotic system. The modern pulp mill is a very complex piece of equipment, extending over a large area and volume, rich in feedback loops, constantly being rebuilt and modified. The effects of changes in one area on another may be great but hard to predict.

3.3 Characteristics of the Forest Industry

The characteristics noted above are common to the data in many industries. Attempting to put the key features of the forest industry at its current stage of development in a nut shell is a hard task. Figure 3 gives not a nut shell but an onion shell view of the forest industry - peeling away layer after layer of complexities.

The chosen order of these layers is not self-evident, and the onion shell idea is correct in describing the unfolding levels of complexity but incorrect in eliminating the interplay of factors. However, more and more, the questions asked by the industry are described of the type described in the example below:

An executive wants to know the competitiveness of his packaging business units in Europe and North America, competing with local players in France and contemplating the construction of a new corrugated board mill in the same country because of the supply of recovered paper. The energy consumption of the planned mill is rising in importance with the possibility of EU-wide harmonised CO_2-taxes. Elsewhere, the reduction of the discharges of nitrogen oxides to meet emission limits is proving difficult because of technical problems not easily solvable with current knowledge. The environmental element in the decision-making is complemented by the need to take a stand on the producer responsibility issue: how should the responsibility for paper recovery be shared by different players such as municipalities, collectors, merchants, retail stores and the paper industry?

If there is a less subtly hidden message in the figure, it is perhaps the location of environmental issues at the core.

3.4 Requirements for a Computerised Tool for a Forest Industry General Analyst

Questions asked by companies with the need to have a global overview, to take into account the long lifetime (up to 30 years) of an investment and to try to assure a long-term supply of fibre, energy and water put a heavy strain on the resources of the analyst. Computer tools are the basis of most analysis assignments, but the bag of tricks of the "neo-generalist" is of special importance. Presenting a set of possible elements from the field of mathematics and artificial intelligence, we have:

- Extracting production rules

  If "company concentrates on newsprint" AND "company is situated in Canada" then with probability X% "company does not use recovered paper"

- Forecasting with exact numbers

  What is the price of market pulp for the next quarter in Asia?

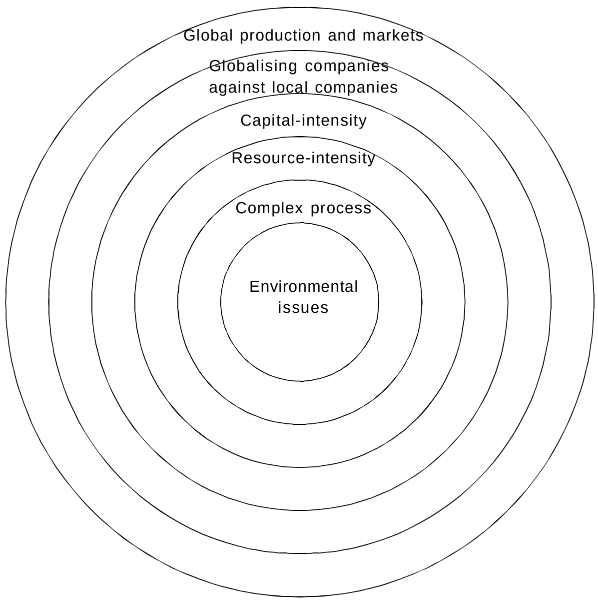


Figure 3: *The forest industry onion shell.*

- Visualisation

  When Sweden took the path of making internal changes in the process and Finland chose external treatment of effluents, how did the Scandinavian effluent discharge emission landscape change?

- Classification

  The Finnish major company Enso decided to acquire the Holtzmann paper company in Germany, thus reinforcing its publication paper sector. Find an acquisition analogous in effects on the company-internal geographical and production balance.

- Time dimension/trajectories

  How has the state of the Finnish company UPM-Kymmene changed when it has emerged (as the result of a series of fusions and acquisitions) to become the biggest forest sector company in Europe?

- Cutting across and combining problem dimensions

  When the Swedish company SCA acquired the German company PWA, how did the acquisition change the company, when we combine a) financial and environmental b) environmental and technological c) all three aspects?

- Correlation hunting

  Taking the top-150 pulp and paper companies in the world, is there a link between return on investment and product sortiment?

All of these features are useful. However, exact forecasts are perhaps the most researched area in forest industry analysis - and the predictions are no more correct than elsewhere. Experiments conducted with Jaakko Pöyry Consulting databases resulted in highly stochastic rules with a generous amount of preconditions. Single experiments are of course not conclusive, and rule extraction is a task worth pursuing. Rules could even be extracted from backpropagation networks, thus combining a limited degree of numeric forecasting and rule extraction. The last five features above, however, are intriguing questions rarely asked because of their difficulty. They may be tackled using the Self-Organizing Map (SOM).

3.5 A SOM of ENTIRE tasks: Mapping the Consultant's Work Load Trajectory

We can make the characteristics above into a series of questions about one company. The example chosen for the set is the Finnish large company Enso.

- How should Enso's investments in the future be directed to best match the group's product sortiment to market demands?

- How is the result of the merger of two state-owned forest industry companies, Enso and Veitsiluoto, different from the two starting points, when the enviro-techno-economic profile is examined?
- Does the path of formation of today's Enso (acquitions and fusions) represent a clear strategy or has the company changed strategic direction as a result of circumstances?
- How does the Enso acquisition of the Holtzmann company change its recovered paper consumption profile?
- How well do the emission changes for Enso's pulp mills match the overall trend of changes in, on one hand, Finland and Sweden, and, on the other hand, the Northern hemisphere?

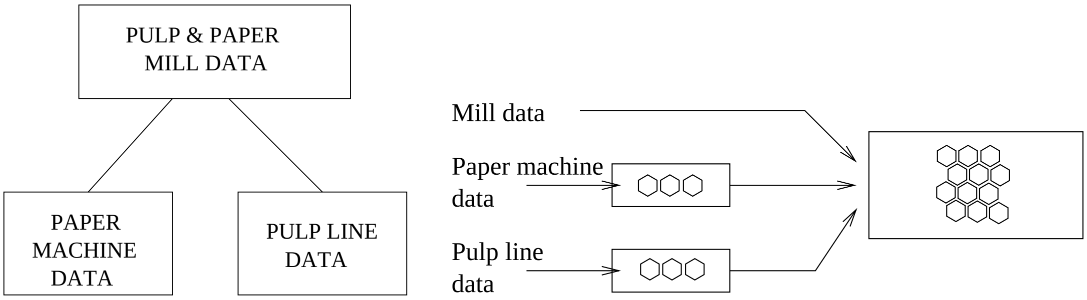
In answering questions, an array of databases is needed, financial, technological and environmental data are combined in different ways and both static and dynamic data must be considered. The consultants can use SOMs in each case.

Taking a set of projects completed by consultant teams at Jaakko Pöyry Consulting and classifying them using a suitably chosen descriptive vector (including e.g. client type, country, project financial overhead) would lead to a meta-SOM of tasks performed using SOMs. The number of projects in the 1990's is around 2000 so far; not all of them are obvious candidates for SOMs, but a suitable set could be found and a Jaakko Pöyry Consulting trajectory SOMmed.

4 Study of Pulp and Paper Technology

In this case study, the SOM is used to analyze data from the world pulp and paper industry. Three data sets were used including information on over 4000 pulp and paper mills, and the over 11000 paper machines and pulp lines in them. The first one included information on the production capacities of the mills, the second included information on the technology of the paper machines and the third on the technology of the pulp lines.

Each mill could contain several paper machines and pulp lines and, therefore, a hierarchical structure of maps was used (see figure 4). At first two low level maps were constructed from the paper machine and pulp line data sets. These maps provided a clustering of the different machine types. The technology map was trained using the mill-specific info in the mill data set and the data histograms from the two low level maps.



(a) Data sets (b) Hierarchical map structure

Figure 4: *(a) There were three technological data sets: one of mill production capacities, one of paper machines and one of pulp lines. Each mill could contain several paper machines and pulp lines. (b) The hierarchical map structure. Data histrograms from the two smaller maps were utilized in the training of the third map. The arrows show which data sets were used in training the maps.*

4.1 Paper machines and pulp lines

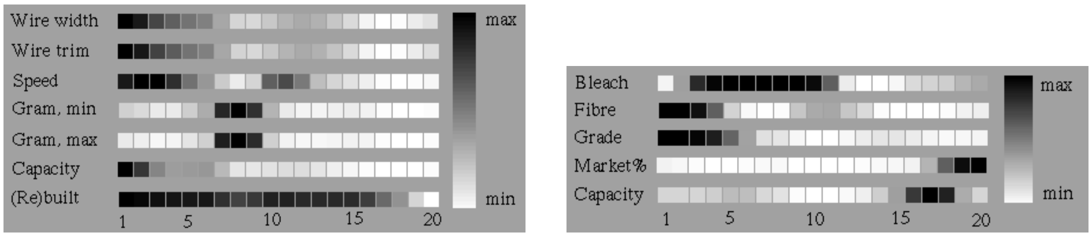
In the construction of the paper machine map seven features were used: wire width, wire trim width, speed, minimum and maximum grammages, present technical production capacity and the year the machine was (re)built. For the pulp lines five features were used: bleaching type, fibre type, main pulp grade type, market capacity as a percentage of the total pulp production and the total pulp production capacity itself. The units and ranges are given in table 1. Prior the training each component was linearly scaled so that its variance was equal to one.

The vector component values in the map units can be seen from the component plane representations in figure 5. By visual inspection of the paper machine map the three first components (wire width, wire trim width and speed) have a strong correlation, as do the next two components (minimum and maximum grammage). In the pulp line map no such global correlations can be seen.

The clustering of the maps was based on visual inspection of the u-matrixes shown in figure 6 supplemented by the knowledge of the distribution of component values. From the paper machine map six clusters were extracted corresponding to different types of machines. The clusters and their descriptions are shown in table 2. The first cluster correspondes to new, high capacity machines with high speed and high value for wire width. The second cluster includes slightly older big or average sized machines. The fourth cluster resembled the second with the difference that the machines are a bit smaller and about five years older. The third cluster consists of machines with big paper weight. The two last clusters got the largest number of hits. They correspond to smallest and oldest paper machines. Based on the properties of the clusters, the paper machines can be divided into three different types:

Table 1: *Paper machine feature vector components (top) and pulp line feature vector components (bottom). The second column gives the range of values of the component.*

Component	value range
Paper machine data	
Wire width (mm)	0 – 10160
Wire trim width (mm)	0 – 9450
Speed (m/min)	0 – 2800
Grammage, min (g/m^2)	0 – 999
Grammage, max (g/m^2)	0 – 9999
Capacity (1000 t/a)	0.1 – 525
(Re)built (year)	1863 – 1996
Pulp line data	
Bleaching code	1,2,3
Fibre code	1,2,3,4
Main pulp grade code	1,2,3,4,5
Market pulp capacity (%)	0 – 100
Total capacity (1000 t/a)	1 – 1075



(a) Paper machine data components

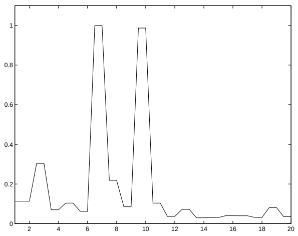
(b) Pulp line data components

Figure 5: *The component planes of the paper machine (a) and pulp line (b) maps. The name of the component on the right and the corresponding values in the map units from 1 to 20 in the middle. All components are presented with a grayscale with black representing the maximum value and white the minimum.*

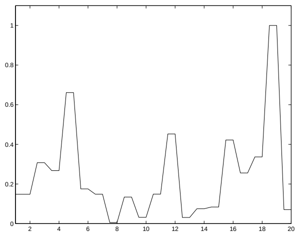
1. The new machines with highest capacity and speed and widest wire (cluster 1).
2. The paper machines with big paper weight (cluster 3).
3. The majority of the paper machines belong to the third type. Their capacity and size decrease steadily with increasing age (clusters 2, 4–6).

The pulp line map can be divided into five clusters listed in table 2. The pulp lines in the first cluster make bleached pulp and use waste paper for fibre. The second cluster corresponds to pulp lines making unbleached pulp with varying fibre and main grade types. The third cluster has pulp lines which produce bleached pulp mostly from wood or other virgin fibres with chemical main grade. The fourth cluster consists of pulp lines with high capacity and an average market ratio. The pulp is bleached, fibre mostly wood and grade mostly chemical. The last cluster consists of pulp lines with high market ratio. All in all, there are three major pulp line types:

1. Those using waste paper for pulp.
2. Those making unbleached pulp, further divided into three subtypes:
   (a) wood fibre and semichemical pulp grade
   (b) wood fibre and either chemical or mechanical pulp grade
   (c) other virgin fibers with chemical pulp grade
3. Those using chemical pulp grade to produce mainly bleached pulp from wood fibre. This type could further be divided into three subtypes: those with small capacity, those with high capacity and those with a high market ratio of production.



(a) Paper machines u-matrix



(b) Pulp line u-matrix

Figure 6: *The u-matrix presentations as of the paper machine (a) and pulp line (b) maps, with the map unit on the x-axis and the value of u-matrix on y-axis. Clusters on map are separated by peaks in the u-matrix. Clusters are listed in table 2.*

4.2 Mill techonology

The mill data set consisted of the total pulp and paper production capacities, the number of paper machines, the number of coaters and the percentage of total production of different pulp and paper types for each mill. They were augmented by the data histograms from the paper machine and pulp line maps. For each mill the two data histograms were constructed using the paper machines and pulp lines in

Table 2: *Clusters of the paper machine map (top) and pulp line map (bottom). The second column gives the units that belong to a certain cluster. The third column gives the number of samples that were projected to the cluster. See also figures 5 and 6.*

Cluster	Units	Hits	Description
Paper machine map			
1	1,2	454	Wide wire and high speed, small paper weight, high capacity, new machines.
2	3–6	1024	Average sized machines, over ten years old on the average.
3	7–9	467	Narrow wire, slow speed, big paper weight, small capacity.
4	10–12	1033	Average wire width and speed, over 15 years old on the average.
5	13–16	2648	Small machines, over 15 years old on the average.
6	17–20	3139	Small and old machines.
Pulp line map			
1	1–4	557	Pulp made from waste paper.
2	5–11	993	Unbleached pulp, both fibre and grade vary.
3	12–15	762	Bleached pulp, wood or other virgin fibres, chemical main grade.
4	16–18	303	Bleached pulp, mostly wood fibre and chemical main grade, average market ratio, high capacity.
5	19,20	364	Big market ratio, mostly wood fibre and chemical main grade.

that particular mill. To emphasize the importance of high capacity the histograms were weighted by the production capacities of the machines/lines. After this the histograms were smoothed by convolving them with a symmetric vector $[0.3, 1, 0.3]$[2]. Finally the histograms were normalized.

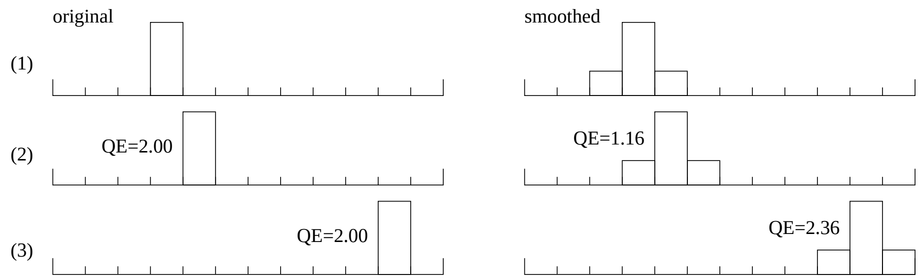


Figure 7: *On the left side three different histograms 1,2 and 3. Since in the SOM similar vectors are projected to close lying map units, the histograms 1 and 2 should be more similar than histograms 1 and 3. However, using euclidian distance between histograms this is not so. Instead the quantization errors (QE) are equal. This is corrected by convolving the histograms with a symmetric matrix as seen on the right side.*

All in all the technology data set consisted of 4205 75-dimensional vectors each corresponding to one mill. A 40x25-sized SOM was trained with this data.

The u-matrix of the mill technology map is shown in figure 8 and part of the component planes in figure 9. Several correlations between components can be seen from the component plane representation, such as that the mills with high paper production capacity usually also have a high pulp production and that newsprint production corresponded to high thermomechanical pulp production. Of special interest are correlations between different pulp and paper types. From figure 9 it can also be seen that for some components the high values were divided between clearly different types of mills. For example there are two types of mills that produce chemimechanical pulp: those that also produce wood containing paper and those that concentrate purely on producing pulp.

An analysis of mill types was performed by first dividing the map into 31 initial clusters. After this all components were separated to three ranges and it was investigated which clusters had component values in the two highest thirds. Finally each cluster was examined to see what kind of mill type it corresponded to. During this procedure some of the initial clusters were combined. For example the SmallInd cluster was combined from several initial clusters corresponding to mills producing various industrial papers. The only truely separating factor between the initial clusters was that they received data from different units of the paper machine map: units 13 – 20 corresponding to old and small paper machines. The cluster analysis resulted in the 20 different mill types described in table 3.

[2]Since the neighboring units of the SOM have similar data items projected to them, the histograms (1) and (2) in figure 7 should be considered more similar than histograms (1) and (3).

A couple of important notices are in place here. Firstly, the set of 20 mill types is by no means definitive. Depending on the desired precision, several other clusters could be extracted from the map. For example the PULP cluster can be divided into three subtypes based on the pulp type and the production capacity. Secondly, the clustering is heavily influenced by the choice and scaling of the components. The mill types in this study reflect only the information present in the data components used. By changing part of the components, or even scaling them differently, a different set of mill types could be obtained. Thirdly, the analysis made so far details the different clusters only in very sketchy terms. To get a better view of the different types, a more detailed in-cluster or in-type analysis would be beneficial.

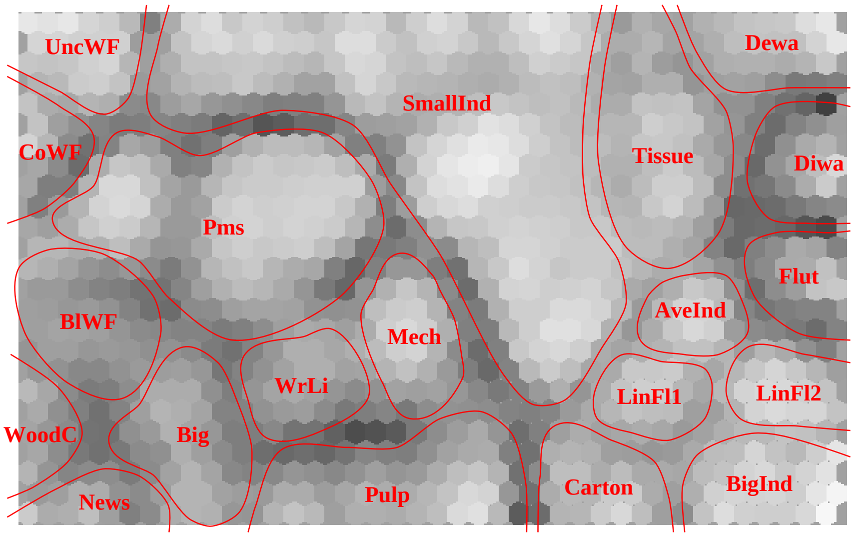


Figure 8: *The u-matrix of the mill technology map. Black corresponds to high value of the u-matrix, white to small value. The mill types have been marked on the map with lines and labeled as in table 3.*

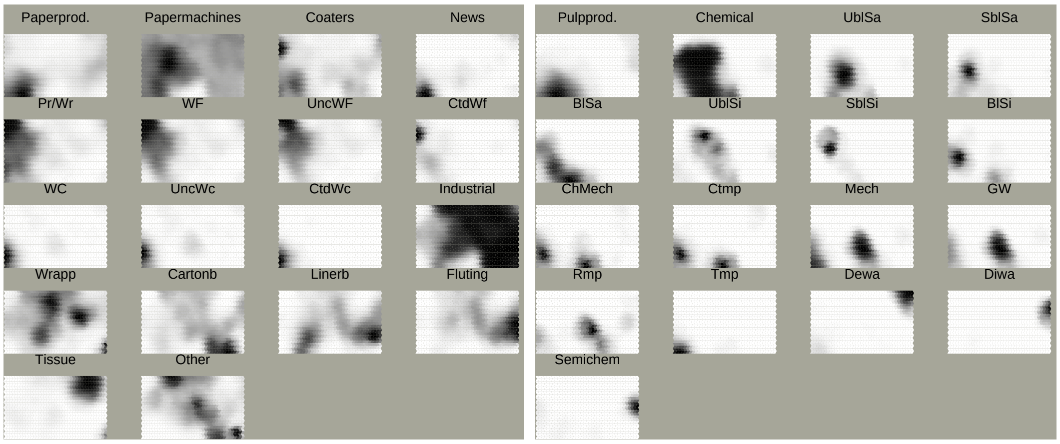
4.3 Geographical areas

It is interesting to note that some mill types are typical for certain geographical regions. For the analysis of different geographical areas the data was separated to 11 sets each consisting of pulp and paper mills in a certain area. The data sets were projected on the map and, based on the resulting histograms, some conclusions can be drawn for each region, as listed in table 4. The same approach can be directly used for comparing and analysing different companies.

Four of the histograms are shown in figure 10: Scandinavia, North America, Far Asia and China. Scandinavia and North America represent technologically advanced regions. Scandinavia in particular has mostly new, high capacity mills the majority of which produce printing/writing papers and pulp. North America has in addition a large number of old and small industrial paper mills. Far Asia on the other hand

Table 3: *Pulp and paper mill types. The clusters refer to the areas marked on the u-matrix of the mill technology map in figure 8.*

Type	Description
UncWF	Uncoated woodfree paper, either new big machines or older and smaller.
SmallInd	Various industrial papers, machines old and small.
Tissue	Tissue paper, average wirelength and speed.
Dewa	Some tissue paper, but especially high deinked waste paper usage.
CoWF	Many coaters, coated woodfree paper, machines of average capacity and speed.
Pms	Many paper machines including some high capacity paper machines, uncoated woodfree but also various industrial papers, unbleached and semibleached sulphite pulp.
Diwa	Cartonboard, linerboard and fluting papers, disperged waste paper for pulp.
BlWF	Uncoated woodfree paper, bleached chemical pulp from wood fibre.
Big	High capacity, many machines and coaters, woodfree paper or linerboard, pulp is chemical (sulphate), machines are big.
WrLi	Wrapping paper and linerboard, unbleached sulphate pulp, big paper machines.
Mech	Various industrial papers, pulp is unbleached and mechanical: groundwood or rmp.
AveInd	Cartonboard, linerboard and fluting paper, average capacity machines.
Flut	Fluting paper, semichemical unbleached pulp, average capacity paper machines.
WoodC	Wood containing paper, chemimechanical or mechanical pulp from wood fibre, average capacity paper machines.
News	High paper capacity: newsprint, thermomechanical pulp, high capacity paper machines and pulp lines.
Pulp	No paper production, but large pulp production, high capacity pulp lines, big market percentage.
Carton	Cartonboard and other papers, big weight in paper machines.
LinFl1	Linerboard and fluting paper, small to average sized machines.
LinFl2	Linerboard and fluting paper, high capacity machines.
BigInd	Cartonboard, wrapping, tissue and other papers, high capacity machines.



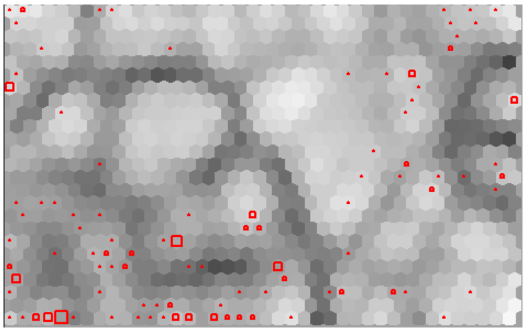
(a) Paper production (b) Pulp production

Figure 9: *The paper production (a) and pulp production (b) related component planes of the mill technology map. For each component the black corresponds to the highest value and white to the smallest.*

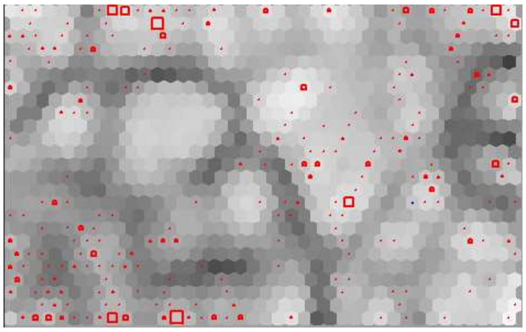
is a growing region with mostly average or small capacity mills, though the paper machines themselves are big. China is a special case; the mills have many machines and they produce both industrial and printing/writing papers. Both Far Asian and Chinese printing/writing paper is almost exclusively wood containing.

Table 4: *Different geographical areas and the main mill types they have.*

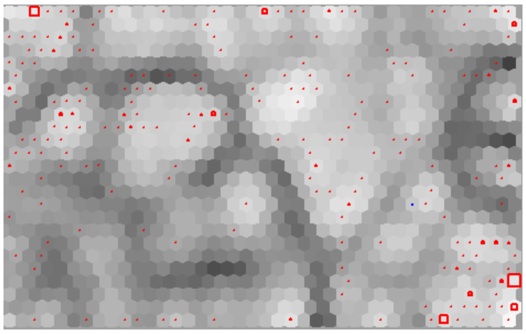
Region	Mills	Description
Scandinavia	149	Big capacity mills, newsprint and pulp-only mills but relatively little industrial paper.
Western Europe	1004	Even spread of all mill types, special notice on the many mills using disperged waste paper.
North America	759	Printing/writing paper production resembles that of Scandinavia, but in addition quite a lot of old SmallInd mills.
Eastern Europe	302	Industrial papers; old SmallInd mills and mills making mechanical pulp. Also some mills in Diwa cluster.
Latin America	533	Even spread of all mill types, special notice of mechanical pulp.
Near and Middle East	65	Industrial papers, mills in Diwa and BigInd clusters.
Africa	106	Mainly industrial papers.
China	370	Many paper machines per mill, woodfree paper, some high-capacity industrial paper mills and several small pulp-only mills.
Japan	221	Even spread of all mill types, many mills using deinked waste paper.
Far Asia	665	Woodfree and various industrial papers, many of them with high capacity machines.
Oceania	31	Mostly new machines, otherwise an even spread of all mill types, many mills in Diwa cluster.



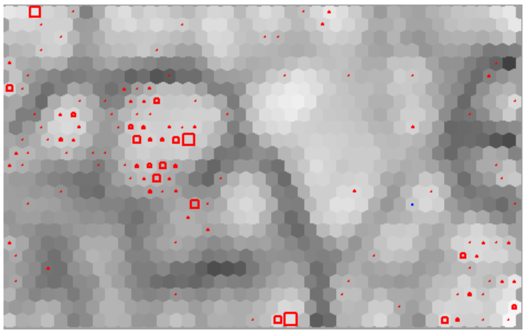
(a) Scandinavian pulp and paper mills



(b) North American pulp and paper mills



(c) Far Asian pulp and paper mills



(d) Chinese pulp and paper mills

Figure 10: *The data set histograms of four different geographical regions on the u-matrix of the pulp and paper mills map. The bigger the square the more mills were projected to that unit on the map.*

5 Future Directions

The Self-Organizing Map is a versatile tool for exploring data sets. It is an effective clustering method and it has excellent visualization capabilities including techniques which use the weight vectors of the SOM to give an informative picture of the data space, and techniques which use data projections to compare data vectors or whole data sets with each other. The visualization capabilities of the SOM make it a valuable tool in data summarization and in consolidating the discovered knowledge. The SOM can also be used for regression and modeling or as a preprocessing stage for other methods. The many abilities of the SOM together with its robustness and flexibility are a combination which makes the SOM an excellent tool in knowledge discovery and data mining.

At present, our implementation of the SOM, the ENTIRE, is clearly an expert tool. An understanding of the SOM fundamentals and a modicum of domain expertise are of essence for efficient utilisation of ENTIRE's potential. Thus, it is a tool for a "neo-generalist" with at least a neural network veneer. In developing ENTIRE further, several directions are possible. They can be pursued simultaneously, but the amount of resources available makes a focus necessary. Goals include

- Further improvements in visualisation.

  Using 3D-graphics in virtual reality environments is a particular interest.

- Injecting a degree of expertise into the tool.

  ENTIRE could be developed in an application-specific direction, resulting in a help desk in the form of small expert systems. This expertise can be used both as a reference during SOM processing and, perhaps more importantly, in the interpretation phase.

- Improving clustering, autolabeling, correlation hunting.

  In the more general realm of research, an aid in the visual hunt for correlation between variables, in the form of a primitive "reporter" summarising links between SOM layers, would speed up analysis and ensure a smaller rate of missed connections. Combining this with an improved autolabeling function and, more fundamentally, autoclustering, would yield benefits to the user. It would simultaneously provide a pivot point for research: an application and an application area to test new concepts on. This type of cross fertilization between use in industry and research at university is only possible given a consistent vision and enough time for the cooperation.

In the case study, the world pulp and paper technology was investigated. A hierarchical structure of SOMs was used to combine data from the different data sets. Such use of multiple interpretation layers introduces some additional error due to necessary generalizations but, on the other hand, provides a structured solution to data fusion.

A study combining economical, environmental and technological data has been made to produce a comprehensive view of the whole pulp and paper industry [26]. The case study has added value, as it transforms a long-talked about idea in the forest industry (combining economy, technology and environment in decision-making) into a concrete example.

The results achieved so far have been encouraging. However, much work is still needed in the postprocessing stage and the interpretation of results. The development and automated usage of algorithms that cluster the units of the SOM will be an essential part of future work. This may be accomplished by the use of the hierarchical maps or with fuzzy interpretation rules.

The use of ANNs in industrial applications is often simplified by straightforward due to the availability of large amounts of measurement data. The unsupervised learning principle of the SOM is desirable property and noise and distortion in data can partly be compensated by the robustness of the algorithm. In this chapter the forest industry has been considered as a case study. However, the methods used are applicable to other fields of industry as well. The ENTIRE tool could easily be modified to a SOM of ENTIRE tasks in e.g. steel or telecommunications industry. In addition, the behavior of industry fields could be simulated, for instance, in changing environments.

References

[1] C. M. Bishop. *Neural Networks for Pattern recognition.* Oxford University Press, 1995.

[2] R. J. Brachman and T. Anand. The process of knowledge discovery in databases. In U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, editors, *Advances in Knowledge Discovery and Data Mining.* AAAI Press / MIT Press, 1996.

[3] U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, editors. *Advances in Knowledge Discovery and Data Mining.* AAAI Press / The MIT Press, California, 1996.

[4] K. Goser, S. Metzen, and V. Tryba. Designing of basic integrated circuits by self-organizing feature maps. *Neuro-Nimes*, 1989.

[5] T. Hastie and W. Stuetzle. Principal curves. *Journal of the American Statistical Association*, 1989.

[6] J. Hollmén and O. Simula. Prediction models and sensitivity analysis of industrial production process parameters by using the self-organizing map. In *1996 IEEE Nordic Signal Processing Symposium Proceedings (NORSIG'96)*, pages 79–82, 1996.

[7] J. Iivarinen, T. Kohonen, J. Kangas, and S. Kaski. Visualizing the clusters on the self-organizing map. In Christer Carlsson, Timo Järvi, and Tapio Reponen, editors, *Proc. Conf. on Artificial Intelligence Res. in Finland*, number 12 in Conf. Proc. of Finnish Artificial Intelligence Society, pages 122–126, Helsinki, Finland, 1994. Finnish Artificial Intelligence Society.

[8] S. Kaski. *Data Exploration Using Self-Organizing Maps.* PhD thesis, Helsinki University of Technology, 1997.

[9] T. Kohonen. *Self-Organizing Maps.* Springer, Berlin, Heidelberg, 1995.

[10] T. Kohonen, E. Oja, O. Simula, A. Visa, and J. Kangas. Engineering applications of the self-organizing map. *Proceedings of the IEEE*, 84(10):27 pages, October 1996.

[11] M. A. Kraaijveld, J. Mao, and A. K. Jain. A nonlinear projection method based on kohonen's topology preserving maps. *IEEE Transactions on Neural Networks*, 6(3):548–559, May 1995.

[12] S. P. Luttrell. Hierarchical self-organizing networks. In *Proceedings of the ICANN'89*, 1989.

[13] J. W. Sammon, Jr. A nonlinear mapping for data structure analysis. *IEEE Transactions on Computers*, C-18(5):401–409, May 1969.

[14] F. Murtagh. Interpreting the kohonen self-organizing feature map using contiguity-constrained clustering. *Pattern Recognition Letters*, 16:399–408, April 1995.

[15] K. Raivio, O. Simula, and J. Henriksson. Improving decision feedback equalizer performance using neural networks. *Electronics Letters*, 27(23):2151 – 2153, November 1991.

[16] H. Ritter, T. Martinetz, and K. Schulten. *Neural Computation and Self-Organizing Maps.* Addison-Wesley Publishing Company, 1992.

[17] O. Simula, E. Alhoniemi, J. Hollmén, and J. Vesanto. Monitoring and modeling of complex processes using hierarchical self-organizing maps. In *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS'96)*, volume Supplement, pages 73–76, May 1996.

[18] O. Simula and J. Kangas. *Neural Networks for Chemical Engineers*, volume 6 of *Computer-Aided Chemical Engineering*, chapter 14, Process monitoring and visualization using self-organizing maps. Elsevier, Amsterdam, 1995.

[19] A. C. Singer, G. W. Wornell, and A. V. Oppenheim. A nonlinear signal modeling paradigm. In *Proc. of ICASSP*, 1992.

[20] Haitao Tang and Olli Simula. The optimal utilization of multi-service scp. In *Intelligent Networks and New Technologies*, pages 175–188. Chapman & Hall, 1996.

[21] Viktor Tryba and Karl Goser. Self-Organizing Feature Maps for process control in chemistry. In T. Kohonen, K. Mäkisara, O. Simula, and J. Kangas, editors, *Artificial Neural Networks*, pages 847–852, Amsterdam, Netherlands, 1991. North-Holland.

[22] A. Ultsch. Self organized feature maps for monitoring and knowledge acquisition of a chemical process. In S. Gielen and B. Kappen, editors, *Proc. ICANN'93, Int. Conf. on Artificial Neural Networks*, pages 864–867, London, UK, 1993. Springer.

[23] A. Ultsch and H.P. Siemon. Kohonen's self organizing feature maps for exploratory data analysis. In *Proc. INNC'90, Int. Neural Network Conf.*, pages 305–308, Dordrecht, Netherlands, 1990. Kluwer.

[24] M. J. van Gils. *Peak Identification in Auditory Evoked Potentials using Artificial Neural Networks.* PhD thesis, Eindhoven University of Technology, 1995.

[25] A. Varfis and C. Versino. Clustering of socio-economic data with kohonen maps. *Neural Network World*, 1992.

[26] J. Vesanto. Data mining techniques based on the self-organizing map. Master's thesis, Helsinki University of Technology, 1997.

[27] F. W. Young. Multidimensional scaling. In S. Kotz, N. L. Johnson, and B. C. Read, editors, *Encyclopedia of Statistical Sciences*, volume 5, pages 649 – 659. Wiley, New York, NY, 1985.