Chapter 15

Intelligent data engineering

Olli Simula, Jaakko Hollmén, Kimmo Raivio, Miki Sirola, Timo Similä, Mika Sulkava, Pasi Lehtimäki, Jarkko Tikka, Jukka Parviainen, Jaakko Talonen, Golan Lampi, Mikko Multanen, Tuomas Alhonnoro, Risto Hakala

15.1 Failure management with data analysis

Miki Sirola, Jukka Parviainen, Jaakko Talonen, Golan Lampi, Tuomas Alhonnoro, Risto Hakala, Timo Similä

Early fault detection with data-analysis tools in nuclear power plants is one of the main goals in NoTeS-project (test case 4) in TEKES technology program MASI. The industrial partner in this project is Teollisuuden Voima Oy, Olkiluoto nuclear power plant. Data analysis is carried out with real failure data, training simulator data and design based data, such as data from isolation valve experiments. A control room tool, visualization tools and various visualizations are under development.

A toolbox for data management using PCA (Principal Component Analysis) and WRLS (Weighted Recursive Least Squares) methods has been developed [1]. Visualizations for e.g. trends, transients, and variation index to detect leakages are used. Statistically significant variables of the system are detected and statistical properties and important visualizations are reported. Data mining methods and time series modelling are combined to detect abnormal events.

X-detector tool based on feature subset selection has been developed. The idea is to do real-time monitoring and abnormality detection with efficient subsets. Measuring dependencies and cluster separation methods are used in variable selection in this visualization tool.

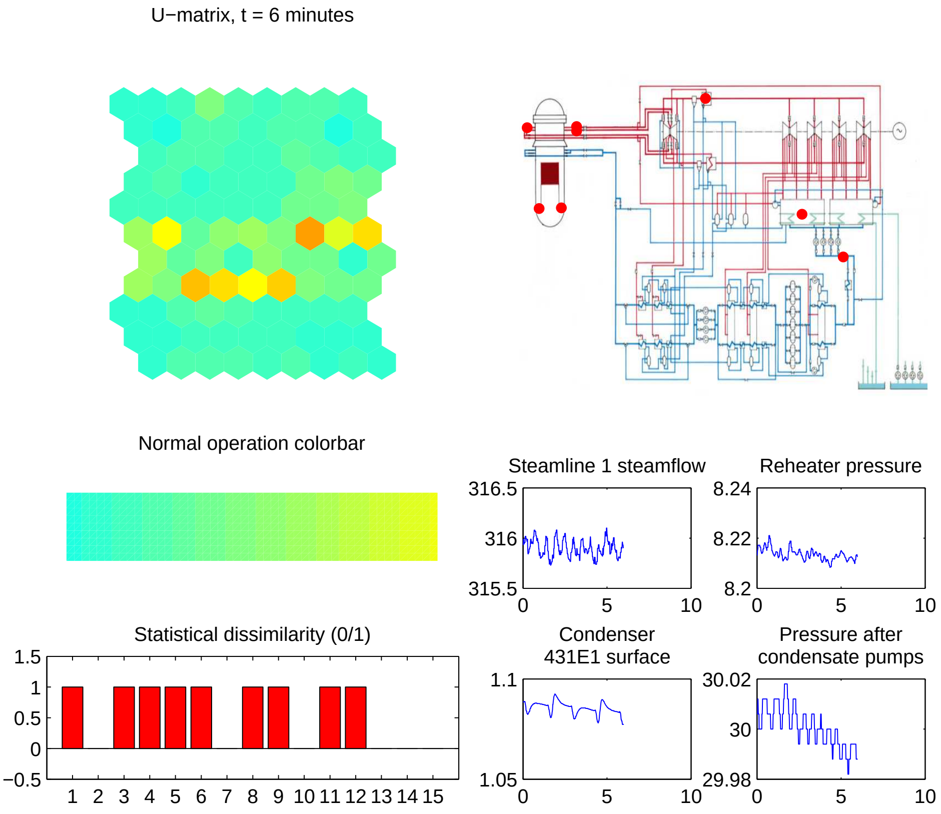

Figure 15.1: X-detector tool user interface: leakage in the main circulation pump. SOM visualization combined with statistical Kolmogorov-Smirnov test, process flow diagram and selected process variable graphs.

Decision support prototype DERSI for failure management in nuclear power plants is under development. It is a control room tool for operator or analysis tool for expert user. It combines neural methods and knowledge-based methods. DERSI utilizes Self-Organizing Map (SOM) method and gives advice by rule-based reasoning. The operator is provided by various informative decision support visualizations, such as SOM maps for normal data and failure data, state U-matrix, quantization error for both component level and state U-matrix, time-series curves and progress visualizations. DERSI tool has been tested in fault detection and separation of simulated data [2].

A separate study of process state and progress visualizations using Self-Organizing Map was also done [3]. All visualizations developed in the project will be collected to make a first proposal for wide monitoring screens.

References

[1] J. Talonen. Fault Detection by Adaptive Process Modeling for Nuclear Power Plant. Master's thesis, Helsinki University of Technology, 2007.

[2] M. Sirola, G. Lampi, and J. Parviainen. Failure detection and separation in SOM based decision support. In *Workshop on Self-Organizing Maps*, Bielefeld, Germany, 2007. WSOM.

[3] R. Hakala, T. Similä, M. Sirola, and J. Parviainen. Process state and progress visualization using self-organizing map. In *International Conference on Intelligent Data Engineering and Automated Learning*, Burgos, Spain, September 2006. IDEAL.

15.2 Cellular network performance analysis

Kimmo Raivio, Mikko Multanen, Pasi Lehtimäki

Structure of mobile networks gets more and more complicated when new network technologies are added to the current ones. Thus, advanced analysis methods are needed to find performance bottlenecks in the network. Adaptive methods can be utilized, for example, to perform hierarchical analysis of the networks, detecting anomalous behavior of network elements and to analyse handover performance in groups of mobile cells.

Combination of the Self-Organizing Map and hierarchical clustering methods can be utilized to split the analysis task into smaller subproblems in which detection and visualization of performance degradations is easier. The method consists of successive selection of a set of cellular network performance indicators and hierarchical clustering of them. Initially only a couple of key performance indicators are utilized and later some more specific counters are used. Thus, the root cause of degradation is easier to find [1]. The method can be utilized both in general network perfromance analysis and in more specific subareas like soft handover success rate [3].

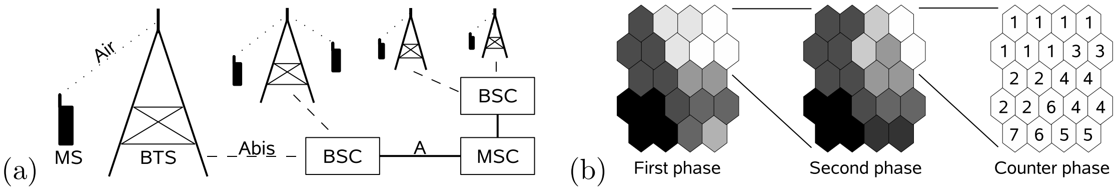

Figure 15.2: Architecture of a cellular network (a) and simple view of the hierarchical analysis algorithm (b).

In outlier detection as well neural as statistical methods can be used to find out network elements with decreased performance or otherwise anomalous traffic profile. Statistical approaches may include both parametric and non-parametric methods. An example of parametric method is Gaussian mixture model. Correspondingly, nearest-neighbor and Parzen windows are non-parametric methods. A neural method called Neural gas is very similar to the statistical approaches and it can be used also in this task [2].

It can be said, that neural and other learning methods can be utilized in the analysis of complicated performance degradation problems in cellular networks. The analysis tools can be built in a way to require only a minimal amount of knowledge of the network itself.

References

[1] M. Multanen, P. Lehtimäki, and K. Raivio. Hierarchical analysis of GSM network performance data. In *Proceedings of the European Symposium on Artificial Neural Networks (ESANN)*, pages 449–454, Bruges, Belgium, April 26 - 28 2006.

[2] M. Multanen, K. Raivio, and P. Lehtimäki. Outlier detection in cellular network data exploration. In *Proceedings of the 3rd International Workshop on Performance Analysis and Enhancement of Wireless Networks (PAEWN)*, Okinawa, Japan, March 25 - 28 2008.

[3] K. Raivio. Analysis of soft handover measurements in 3G network. In *Proceedings of the 9th ACM Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, pages 330–337, Torremolinos, Malaga, Spain, October 2 - 6 2006.

15.3 Predictive GSM network optimization

Pasi Lehtimäki, Kimmo Raivio

In this study, the focus is on the final step of the mobile network monitoring procedure, that is, on making adjustments to configuration parameters so that the amount of predictable, regularly occurring performance degradations or faults is minimized. In order to automate the configuration parameter optimization, a computational method to evaluate the performance of alternative configurations must be available. In data-rich environments like cellular networks, such predictive models are most efficiently obtained with the use of past data records.

In blocking prediction, the interest is to compute the number of blocked requests at different conditions. This can be based on the use of well known Erlang-B formula. The expected value for the number of blocked requests is obtained by multiplying the number of arriving requests with the blocking probability, leading to $B = \lambda p(N_c|\lambda,\mu,N_c)$. The expected value for the congestion time is $C = p(N_c|\lambda,\mu,N_c)$ and the expected value for the number of channels in use is $M = \sum_{n=0}^{N_c} np(n|\ \lambda,\mu,N_c)$.

In [2], it was shown that the Erlang-B formula does not provide accurate predictions for blocking in GSM networks if low sampling rate measurements of arrival process are used in the model. More traditional regression methods can be used for the same purpose with the assist of knowledge engineering approach in which Erlang-B formula and regression methods are combined. With the use of Erlang-B formula, the dependencies between B, C and M that remain the same in each base station system need not be estimated from data alone. The data can be used to estimate other relevant and additional parameters that are required in prediction. In [2] and [1], a method to use Erlang-B formula and measurement data to predict blocking is presented. The regression techniques are used to estimate the arrival rate distribution describing the arrival process during short time periods. The Erlang-B formula is used to compute the amount of blocking during the short time periods.

Suppose that the time period is divided into N_s segments of equal length. Also, assume that we have a vector $\boldsymbol{\lambda} = [0\ \ 1\Delta_\lambda\ \ 2\Delta_\lambda\ \ \ldots\ \ (N_\lambda - 1)\Delta_\lambda]$ of N_λ possible arrival rates per segment with discretization step Δ_λ. Let us denote the number of blocked requests during a segment with arrival rate λ_i with $B_i = \lambda_i p(N_c|\lambda_i,\mu,N_c)$, where $p(N_c|\lambda_i,\mu,N_c)$ is the blocking probability given by the Erlang distribution. Also, the congestion time and the average number of busy channels during a segment with arrival rate λ_i are denoted with $C_i = p(N_c|\lambda_i,\mu,N_c)$ and $M_i = \sum_{n=0}^{N_c} np(n|\lambda_i,\mu,N_c)$. In other words, the segment-wise values for blocked requests, congestion time and average number of busy channels are based on the Erlang-B formula.

Now, assume that the number of segments with arrival rate λ_i is θ_i and $\sum_i \theta_i = N_s$. Then, the cumulative values over one hour for the number of requests T, blocked requests B, congestion time C and average number of busy channels M can be computed with

$$\begin{bmatrix} \lambda_1 & \lambda_2 & \ldots & \lambda_{N_\lambda} \\ B_1 & B_2 & \ldots & B_{N_\lambda} \\ \frac{C_1}{N_s} & \frac{C_2}{N_s} & \ldots & \frac{C_{N_\lambda}}{N_s} \\ \frac{M_1}{N_s} & \frac{M_2}{N_s} & \ldots & \frac{M_{N_\lambda}}{N_s} \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_{N_\lambda} \end{bmatrix} = \begin{bmatrix} T \\ B \\ C \\ M \end{bmatrix} \tag{15.1}$$

or in matrix notation $\mathbf{X}\boldsymbol{\theta} = \mathbf{Y}$.

Now, the problem is that the vector $\boldsymbol{\theta}$ is unknown and it must be estimated from the data using the observations of $\mathbf{Y}$ and matrix $\mathbf{X}$ which are known a priori. Since the output vector $\mathbf{Y}$ includes variables that are measured in different scales, it is necessary to include weighting of variables into the cost function. By selecting variable weights according to their variances estimated from the data, the quadratic programming problem

$$\min_{\boldsymbol{\theta}} \quad \left\{ \frac{1}{2}\boldsymbol{\theta}^T \mathbf{H} \boldsymbol{\theta} + \mathbf{f}^T \boldsymbol{\theta} \right\} \tag{15.2}$$

$$w.r.t \quad 0 \leq \theta_i \leq N_s, \; i = 1, 2, ..., N_\lambda, \tag{15.3}$$

$$\sum_{i=1}^{N_\lambda} \theta_i = N_s \tag{15.4}$$

is obtained where $\mathbf{f} = -\mathbf{X}^T \mathbf{W}^T \mathbf{W} \mathbf{Y}$ and $\mathbf{H} = \mathbf{X}^T \mathbf{W}^T \mathbf{W} \mathbf{X}$ include the weighting matrix $\mathbf{W}$. In other words, the goal is to find the vector $\boldsymbol{\theta}$ that provides the smallest prediction errors for variables T, B, C and M.

The optimization problem could be solved for each of the N_d observation vectors separately, leading to N_d solution vectors $\boldsymbol{\theta}$ for hour h. Since we are interested in long-term prediction of blocking, we should somehow combine the solution vectors so that behavior common to all solution vectors are retained and non-regular properties of the demand are given less attention.

Let us denote the ith solution vector for hour h with $\boldsymbol{\theta}_h^{(i)}$ and the jth element of the corresponding solution vector with $\theta_{jh}^{(i)}$. Since $\theta_{jh}^{(i)}$ described the number of segments with arrival rate $\lambda = \lambda_j$ during ith observation vector at hour h, the probability for a random segment during ith observation period to have an arrival rate $\lambda = \lambda_j$ can be computed from $\theta_{jh}^{(i)}$ with $p_{jh}^{(i)} = \theta_{jh}^{(i)} / N_s$, where N_s is the number of segments in a period.

The probability for observing a segment with arrival rate $\lambda = \lambda_j$ at hour h would become

$$p_{jh} = \frac{1}{N_d N_s} \sum_{i=1}^{N_d} \theta_{jh}^{(i)}. \tag{15.5}$$

Now, the arrival rates λ_j and their probabilities p_{jh} for hour h form a probabilistic model. Let us define a column vector

$$\underset{seg \mapsto hour}{\boldsymbol{\theta}_h} = \mathbf{p}_h N_s \tag{15.6}$$

that maps the segment-wise candidate arrival rates λ_j to the total number of arrived requests T in a single one hour time period with

$$T = \boldsymbol{\lambda} \underset{seg \mapsto hour}{\boldsymbol{\theta}_h} . \tag{15.7}$$

Note that the parameter vector $\boldsymbol{\theta}_{h, seg \mapsto hour}$ can also be used to map the vector $\mathbf{B} = [B_1 \; B_2 \; \ldots \; B_{N_\lambda}]$ of segment-wise blocking candidates to the total number of occurrences of blocked requests during one period. Similarly, the cumulative values for the average number of busy channels and the congestion time can be computed.

References

[1] P. Lehtimäki. A model for optimisation of signal level thresholds in GSM networks. *International Journal of Mobile Network Design and Innovation*, 2008. (accepted).

[2] P. Lehtimäki and K. Raivio. Combining measurement data and Erlang-B formula for blocking prediction in GSM networks. In *Proceedings of The 10th Scandinavian Conference on Artificial Intelligence (SCAI)*, Stockholm, Sweden, May 26 - 28 2008.

15.4 Learning from environmental data

Mika Sulkava, Jaakko Hollmén

Data analysis methods play an important role in increasing our knowledge of the environment as the amount of data measured from the environment increases. Gaining an insight into the condition of the environment and the assessment of the its future development under the present and predicted environmental scenarios requires large data sets from long-term monitoring programs. In this project the development of forests in Finland has been studied using data from various forest monitoring programs. In addition, the global changes and drivers of the CO_2 exchange of forests have been studied based on eddy covariance data from a high number of sites around the world.

The work in this project includes collaboration with a high number of parties. During 2006–2007, there has been cooperation with two research units of the Finnish Forest Research Institute, University of Antwerp, and numerous researchers in the carbon cycling community all around the world. The latest journal contributions are joint work of a team of more than a dozen researchers from nine countries in three continents.

Plant nutrients play an integral role in the physiological and biochemical processes of forest ecosystems. The effects of nitrogen and sulfur depositions on coniferous forests have been studied using the Self-Organizing Map. It was concluded that evidence for deposition-induced changes in needles has clearly decreased during the nineties. The results of the effects of the depositions have been presented in conferences [1, 2].

Various environmental factors and past development affect the growth and nutritional composition of tree needles as they are aging. Different regression models have been compared to find out how these effects could be modeled effectively and accurately during the second year of the needles [3]. We found that sparse regression models are well suited for this kind of analysis. They are better for the task than ordinary least squares single and multiple regression models, because they are both easy to interpret and accurate at the same time.

Good quality of analytical measurements techniques is important to ensure the reliability of analyses in environmental sciences. We have combined foliar nutrition data from Finland and results of multiple measurement quality tests from different sources in order to study the effect of measurement quality on conclusions based on foliar nutrient analysis [4, 5]; see Figure 15.3. In particular, we studied the use of weighted linear regression

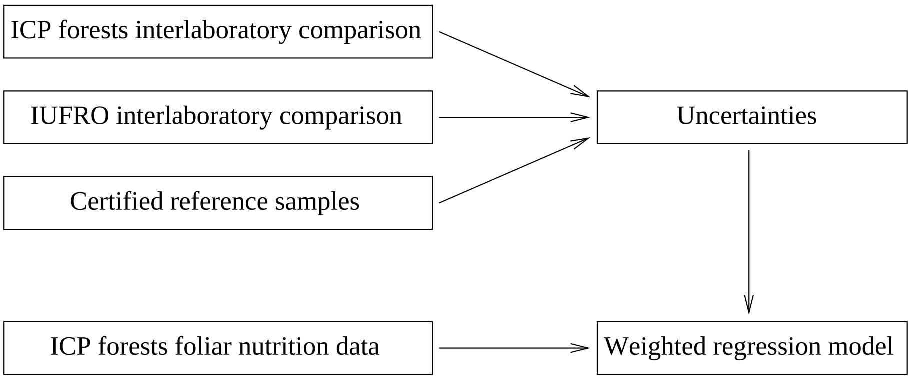
ICP forests interlaboratory comparison
IUFRO interlaboratory comparison
Certified reference samples
Uncertainties
ICP forests foliar nutrition data
Weighted regression model

Figure 15.3: Fusion of measurement quality metadata from three different sources and forest nutrition data made it possible to use weighted regression models for trend detection.

models in detecting trends in foliar time series data and showed that good precision of the measurement techniques may decrease the time needed to detect statistically significant trends in environmental time series by several years.

The dependencies between the atmospheric CO_2 exchange of the world's forests and different environmental factors and between the annual radial growth of coniferous trees and environment and properties of the trees have been studied since 2006. First results concerning the significance of photosynthesis in differences between yearly CO_2 exchange have been published lately [6, 7]. Also, the effects of nitrogen deposition on CO_2 exchange in forests have been studied [8].

Finally, the effects of environmental conditions on radial growth of trees has been studied. Methods for automatic detection of the onset and cessation of radial growth [9] and for model selection and estimation based on expert knowledge [10] have been developed.

References

[1] S. Luyssaert, M. Sulkava, H. Raitio, J. Hollmén, and P. Merilä. Is N and S deposition altering the mineral nutrient composition of Norway spruce and Scots pine needles in Finland? In Johannes Eichhorn, editor, *Proceedings of Symposium: Forests in a Changing Environment – Results of 20 years ICP Forests Monitoring*, pages 80–81, Göttingen, Germany, October 2006.

[2] Päivi Merilä, John Derome, Sebastiaan Luyssaert, Mika Sulkava, Jaakko Hollmén, Kaisa Mustajärvi, and Pekka Nöjd. How are N and S in deposition, in percolation water and in upper soil layers reflected in chemical composition of needles in Finland? In *Book of abstracts of Seminar on forest condition monitoring and related studies in northern Europe under the Forest Focus and ICP Forests programmes*, Vantaa, Finland, November 2007.

[3] Mika Sulkava, Jarkko Tikka, and Jaakko Hollmén. Sparse regression for analyzing the development of foliar nutrient concentrations in coniferous trees. *Ecological Modelling*, 191(1):118–130, January 2006.

[4] Mika Sulkava. Modeling how varying data quality affects the ability to detect trends in environmental time series. In Veli Mäkinen, Greger Lindén, and Hannu Toivonen, editors, *Summer School on Algorithmic Data Analysis (SADA 2007) and Annual Hecse Poster Session, Abstract proceedings*, volume B-2007-4 of *Series of Publications B*, page 104, Helsinki, Finland, May/June 2007. University of Helsinki, Department of Computer Science, Helsinki University Printing House.

[5] Mika Sulkava, Sebastiaan Luyssaert, Pasi Rautio, Ivan A. Janssens, and Jaakko Hollmén. Modeling the effects of varying data quality on trend detection in environmental monitoring. *Ecological Informatics*, 2(2):167–176, June 2007.

[6] S. Luyssaert, I. A. Janssens, M. Sulkava, D. Papale, A. J. Dolman, M. Reichstein, T. Suni, J. Hollmén, T. Vesala, D. Lousteau, B. Law, and E. J. Moors. Photosynthesis drives interannual variability in net carbon-exchange of pine forests at different latitudes. In *Proceedings of the Open Science Conference on the GHG Cycle in the Northern Hemisphere*, pages 86–87, Sissi-Lassithi, Greece, November 2006. CarboEurope, NitroEurope, CarboOcean, and Global Carbon Project, Max-Planck-Institute for Biogeochemistry.

[7] S. Luyssaert, I. A. Janssens, M. Sulkava, D. Papale, A. J. Dolman, M. Reichstein, J. Hollmén, J. G. Martin, T. Suni, T. Vesala, D. Lousteau, B. E. Law, and E. J. Moors. Photosynthesis drives anomalies in net carbon-exchange of pine forests at different latitudes. *Global Change Biology*, 13(10):2110–2127, October 2007.

[8] S. Luyssaert, I. Inglima, R. Ceulemans, P. Ciais, A. J. Dolman, J. Grace, J. Hollmén, B. E. Law, G. Matteucci, D. Papale, S. L. Piao, M. Reichstein, E.-D. Schulze, M. Sulkava, J. Tang, and I. A. Janssens. Unravelling nitrogen deposition effects on carbon cycling in forests. *Eos, Transactions, American Geophysical Union*, 88(52), December 2007. Fall Meeting Supplement, Abstract B32B-02.

[9] Mika Sulkava, Harri Mäkinen, Pekka Nöjd, and Jaakko Hollmén. CUSUM charts for detecting onset and cessation of xylem formation based on automated dendrometer data. In Ivana Horová and Jiří Hřebíček, editors, *TIES 2007 – 18th annual meeting of the International Environmetrics Society, Book of Abstracts*, page 111, Mikulov, Czech Republic, August 2007. The International Environmetrics Society, Masaryk University.

[10] Jaakko Hollmén. Model selection and estimation via subjective user preferences. In Vincent Corruble, Masayuki Takeda, and Einoshin Suzuki, editors, *Discovery Science: 10th International Conference, DS 2007, Proceedings*, volume 4755 of *Lecture Notes in Artificial Intelligence*, pages 259–263, Sendai, Japan, October 2007. Springer-Verlag.

15.5 Parsimonious signal representations in data analysis

Jarkko Tikka, Jaakko Hollmén, Timo Similä

The objective in data analysis is to find unsuspected and practical information from large observational data sets and to represent it in a comprehensible way. While utility is a natural starting point for any analysis, understandability often remains a secondary goal. A lot of input variables are available for a model construction in many cases. For instance, in the analysis of microarray data the number of input variables may be tens of thousands. It is impossible to evaluate all the possible combinations of input variables in a reasonable time. In this research, improved understandability of data-analytic models is sought by investigating sparse signal representations that are learned automatically from data. Naturally, the domain expertise is useful in many cases in validation of results, but it may also be biased by established habits and, thus, prevent making novel discoveries.

In a time series context, parsimonious modeling techniques can be used in estimating a sparse set of autoregressive variables for time series prediction [7]. We presented a filter approach to the prediction: first we selected a sparse set of inputs using computationally efficient linear models and then the selected inputs were used in the nonlinear prediction model. Furthermore, we quantified the importance of the individual input variables in the prediction. Based on experiments, our two-phase modeling strategy yielded accurate and parsimonious prediction models giving insight to the original problem.

The problem of estimating sparse regression models in a case of multi-dimensional input and output variables has been investigated in [4]. We proposed a forward-selection algorithm called multiresponse sparse regression (MRSR) that extends the Least Angle Regression algorithm (LARS) [1]. The algorithm was also applied to the task of selecting relevant pixels from images in multidimensional scaling of handwritten digits. The MRSR algorithm was presented in a more general framework in [5]. In addition, experimental comparisons showed the strengths of MRSR against some other input selection methods. The input selection problem for multiple response linear regression was formulated as a

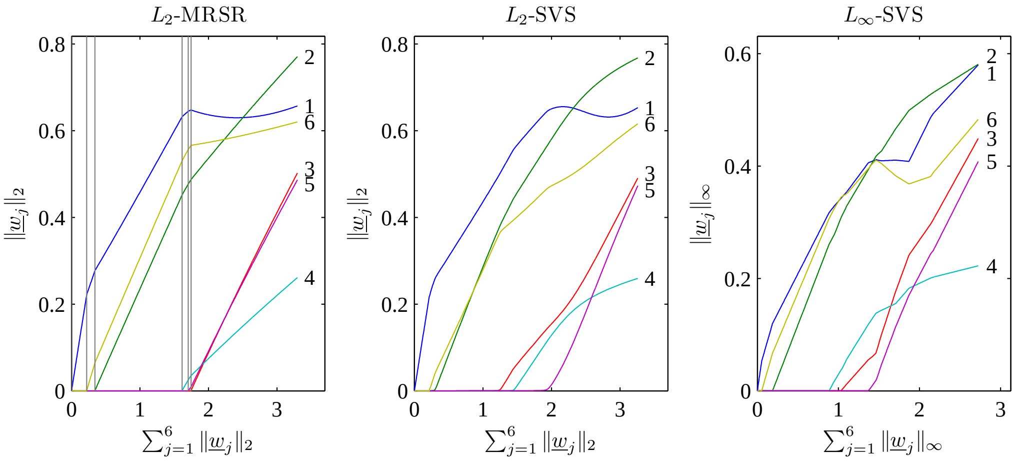

Figure 15.4: Solution paths of the importance factors of input variables. In the subfigure on the left panel, vertical lines indicate the breakpoints of the MRSR algorithm, i.e the points where a new input variable is added to the subset of selected input variables. All the solution paths end to the ordinary least square solution.

convex optimization problem to minimize the error sum of squares subject to a sparsity constraint in [6]. The proposed simultaneous variable selection (L_2-SVS) method is related to L_∞-SVS method [10]. We also reported an efficient algorithm to follow the solution path as a function of the constraint parameter. In Figure 15.4, the solution paths of MRSR, L_2-SVS, and L_∞-SVS are illustrated using a data set, which includes six input variables. The most important inputs are x_2, x_1, and x_6 according to all the three methods. The multiresponse sparse regression is studied further in [2, 3].

The artificial neural networks are an appropriate choice to model dependencies in nonlinear regression problems, since they are capable to approximate a wide class of functions very well. A disadvantage of neural networks is their black-box characteristics. We have developed input selection algorithms for radial basis function (RBF) networks in order to improve their interpretability [8, 9]. A backward selection algorithm (SISAL-RBF), which removes input variables sequentially from the network based on the significance of the individual regressors, was suggested in [9]. The calculation of ranking of inputs is based on partial derivatives of the network. Only 15% of the available inputs were selected by the SISAL-RBF without sacrificing prediction accuracy at all in the case of real world data set [9]. In [8], each input dimension was weighted and a sparsity constraint was imposed on the sum of the weights. The resulting constrained cost function was optimized with respect to the weights and other parameters using alternating optimization approach. The optimum weights describe the relative importance of the input variables. Applications to both simulated and benchmark data produced competitive results.

References

[1] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani. Least angle regression. *Annals of Statistics*, 32(2): 407–499, 2004.

[2] T. Similä. Majorize-minimize algorithm for multiresponse sparse regression. *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2007)*, Vol. II, pp. 553–556, Honolulu, HI, USA, April 2007.

[3] T. Similä. *Advances in variable selection and visualization methods for analysis of multivariate data.* PhD Thesis, Helsinki University of Technology, 2007.

[4] T. Similä and J. Tikka. Multiresponse sparse regression with application to multidimensional scaling. *Proceedings of the 15th International Conference on Artificial Neural Networks (ICANN 2005)*, Vol. 3967 (part II) of Lecture Notes in Computer Science, Springer, pp. 97–102, Warsaw, Poland, September, 2005.

[5] T. Similä and J. Tikka. Common subset selection of inputs in multiresponse regression. *Proceedings of the 19th International Joint Conference on Neural Networks (IJCNN 2006)*, pp. 1908–1915, Vancouver, Canada, July, 2006.

[6] T. Similä and J. Tikka. Input selection and shrinkage in multiresponse linear regression. *Computational Statistics & Data Analysis*, 52(1): 406–422, 2007.

[7] J. Tikka and J. Hollmén. Sequential input selection algorithm for long-term prediction of time series. *Neurocomputing.* Accepted for publication.

[8] J. Tikka. Input selection for radial basis function networks by constrained optimization. *Proceedings of the 17th International Conference on Artificial Neural Networks (ICANN 2007)*, Vol. 4668 of Lecture Notes in Computer Science, Springer, pp. 239–248, Porto, Portugal, September, 2007.

[9] J. Tikka and J. Hollmén. Selection of important input variables for RBF network using partial derivatives. *Proceedings of the 16th European Symposium on Artificial Neural Networks (ESANN 2008)*. In press.

[10] B.A. Turlach, W.N. Venables, and S.J. Wright. Simultaneous variable selection Technometrics, 47(3): 349–363, 2005.