ROBUST SOFT SENSORS BASED ON INTEGRATION OF GENETIC PROGRAMMING, ANALYTICAL NEURAL NETWORKS, AND SUPPORT VECTOR MACHINES

Arthur Kordon, Guido Smits[1], Elsa Jordaan[1], and Ed Rightor

The Dow Chemical Company
Freeport, TX
USA

Dow Benelux V.V.[1]
Terneuzen
The Netherlands

Abstract – A novel approach for development of inferential sensors based on integration of three key computational intelligence approaches (genetic programming, analytical neural networks, and support vector machines) is proposed. The advantages of this type of soft sensors are their good generalization capabilities, increased robustness, explicit input/output relationships, self-assessment capabilities, and low implementation and maintenance cost.

1. INTRODUCTION

Soft (or inferential) sensors assume that there is an empirical relationship between some easily measured and continuously available process variables and some critical parameters related to process quality like molecular distribution. Due to the nature of empirical models, the development time and cost of soft sensors is significantly lower in comparison to fundamental model building or new hardware design. This is the driving force behind the growing interest in industry toward inferential sensing.

A first wave of soft sensors based on the "classical" back-propagation neural network approach is in use in different areas of manufacturing since the early 1990's [1]. The common methodology of building neural net soft sensors and the practical issues of their implementation have been discussed in detail [2]. However, along with the benefits that soft sensors have shown in numerous industrial applications, several performance and long-term operation issues have appeared. Most of the problems are related to some limitations that are typical for soft sensors based on neural nets. Due to their sometimes ineffective, non-parsimonious structure and poor generalization capability outside the range of training data, their performance is very sensitive to specific process conditions. As a result of this reduced robustness there is a necessity of frequent re-training. The final effect of all of these problems is an increased maintenance cost and gradually decreased performance and credibility.

Robustness toward process variability, industrial data quality, and inadequate modeling are key issues for reliable and mass-scale application of inferential sensors. Several machine learning approaches have the potential to contribute to the solution of this important problem. Stacked analytical neural networks (internally developed in The Dow Chemical Company) allow very fast model development of parsimonious black-box models with confidence limits. Genetic Programming (GP) can generate explicit functional solutions that are very convenient for direct on-line implementation in the existing process information and control systems [3]. Recently, Support Vector Machines (SVM) give tremendous opportunities for building empirical models with very good generalization capability [4]. At the same time, each approach has its own weaknesses that reduces the implementation space and makes the task for robust soft sensor design based on separate computational intelligence techniques difficult.

An alternative, more integrated approach for soft sensor development is proposed in this paper. It combines a nonlinear sensitivity and time-delay analysis based on Stacked Analytical Neural Nets with outlier detection and condensed data selection driven by the Support Vector Machines. The derived soft sensor is generated by GP as an analytical function. The integrated methodology amplifies the advantages of the individual techniques, significantly reduces the development time, and delivers robust soft sensors with low maintenance cost. The advantages of the proposed methodology have been demonstrated in several successful applications in The Dow Chemical Company.

2. INDUSTRIAL REQUIREMENTS FOR ROBUST SOFT SENSORS

The value and the large potential of soft sensors in more effective process monitoring and control are well understood in industry. Unfortunately, the current state of the art of soft sensors still requires specialized software (with the inevitable version upgrades), additional efforts for on-line model performance assessment, and above all, highly qualified (Ph.D. level) specialists for development and maintenance. As a result, very often the real development time and maintenance cost have significantly

0-7803-7282-4/02/$10.00 ©2002 IEEE

exceeded the initial expectations and led to reduced credibility. In order to address these issues and to define the design criteria for soft sensors with increased robustness, an analysis of industrial requirements and realistic expectations is needed.

There is a very simple and clear criterion for a mass scale acceptance of soft sensors in industry – they must have the same level of reliability, ease of use, and maintenance efforts as the hardware sensors. This is the natural way to integrate the new technology within the existing work processes and support infrastructures in manufacturing. The defined criterion can be separated into the following key requirements toward soft sensors:

- Robust, fast, and cost effective development process

The assumption is that soft sensor development has to be more effective than the alternative approaches (hardware sensor design or fundamental model building). Of special importance is the requirement to significantly reduce the development time while improving the consistency and performance of delivered empirical models. Another critical factor is to make the development process user-friendly with minimal tuning parameters and specialized knowledge.

- Low sensitivity to process changes

Process changes driven by different operating regimes, equipment upgrades, or product demand fluctuations are more of a rule than an exception. It is unrealistic to expect that all the variety of process conditions will be captured by the training data and reflected in the developed soft sensor. The potential solution is in modeling approaches with better extrapolation capabilities at least 20 % outside the training range.

- Performance self-assessment capability

Usually soft sensors infer the most critical parameters in industrial processes and as such require estimates with a very high level of reliability. It is necessary to include elements of self-assessment of prediction quality. A prospective approach is to use combined predictors [5] and their statistics as a confidence indicator of the soft sensor's performance.

- Low cost of ownership and maintenance

The experience from "classical" neural net-based soft sensors shows that the lion share of maintenance cost is in frequent-re-training and especially in model re-design. The expectation is that by using non-black-box models with increased robustness the need for re-training will be significantly reduced. Another factor that contributes to cost of ownership reduction is the ease of on-line implementation. Of special interest are the explicit functional models, generated by GP. They are well understood by process engineers, directly applicable in the control system, and do not require specialized knowledge for maintenance.

3. SELECTED COMPUTATIONAL INTELLIGENCE APPROACHES FOR DEVELOPMENT OF ROBUST SOFT SENSORS

It is very difficult to satisfy the defined requirements for industrial soft sensors by a specific empirical technique only. However, several machine learning approaches can effectively resolve some specific issues and become the building blocks of an integrated methodology for robust soft sensor development. Of special interest are the following three approaches – analytical neural nets, Support Vector Machines (SVM), and Genetic Programming (GP).

3.1 Analytical Neural Networks

Analytical neural networks are based on a collection of individual, feedforward, single layer neural networks where the weights of the input to hidden layer have been initialized according to a fixed distribution such that all hidden nodes are active. The weights of the hidden to output layer can then be calculated directly using least squares. Advantages of this method are: it is fast and each neural network has a well defined, single, global optimum. Each of these networks have a known Vapnik-Chernovenkis (VC) dimension, so collections with a given complexity can be developed and optimum use can be made of Statistical Learning Theory. Time delays between inputs are handled through convolution functions. In addition, the use of a collection of networks gives more robust models that include confidence limits based on the standard deviation of stacked neural nets.

Analytical neural networks contribute to the soft sensor development process by allowing an extensive nonlinear sensitivity analysis and input feature selection. They allow for a fast feasibility test of the model development process and they deliver models that have confidence limits associated with predicted outputs.

3.2 Support Vector Machines

Support Vector Machines have become an active field of research in recent years. This type of learning machine implements the Structural Risk Minimization principle, which has its foundation in Statistical Learning Theory and is particularly useful for learning with small sample sizes[4]. One of the key features is the use of kernel functions. This enables the method, not only to use non-linear mappings of the input data, but also overcomes the curse of dimensionality. Furthermore, through the

0-7803-7282-4/02/$10.00 ©2002 IEEE

introduction of a special loss function, the ε-insensitive loss, the model is defined in terms of a subset of the learning data, called the support vectors. Varying the size of ε influences the number of support vectors and therefore allows direct control over the complexity of the model.

The SVM method is a very robust method and has a unique contribution to the soft sensor development by means of automatic outlier and novelty detection. The fact that the SVM model is a sparse representation of the learning data allows the extraction of a condensed data set based on the support vectors. Finally, by using certain types of kernels, the extrapolation capabilities of the model can be increased dramatically, especially by incorporating prior information [6]. All these features combined pave the way to the development of robust soft sensors.

3.3 Genetic Programming

The third approach of interest to soft sensor development is GP with its capability for symbolic regression [3]. GP-generated symbolic regression is a result of simulation of the natural evolution of numerous potential mathematical expressions. The final results is a list of "the best and the brightest" analytical forms according to the selecting objective function. Of special importance to industry are the following unique features of GP[7]:

- no *a priori* modeling assumptions
- derivative-free optimization
- few design parameters
- natural selection of the most important process inputs
- parsimonious analytical functions as a final result.

The last feature has double benefit. On one hand, a simple soft sensor often has better generalization capability, increased robustness, and needs less frequent re-training. On the other hand, process engineers and developers prefer to use non-black box empirical models and are much more open to take the risk to implement inferential sensors based on functional relationships. An additional advantage is the low implementation cost of such type of soft sensors. It can be applied directly into the existing Distributed Control Systems (DCS) avoiding additional specialized software packages, typical for neural net-based inferential sensors.

At the same time there are still significant challenges in implementing industrial soft sensors generated by GP: function generation with noisy industrial data, dealing with time delays, sensitivity analysis of large data sets, to name a few. Of special importance is the main drawback of GP – the slow speed of model development due to the inherent high computational requirements of this method. For real industrial applications the calculation time is in order of days, even with the current high-end PCs.

4. INTEGRATED METHODOLOGY FOR ROBUST SOFT SENSOR DEVELOPMENT

The objectives of the proposed integrated methodology are to satisfy the defined criteria for successful industrial soft sensors, i.e., to reduce development time, to deliver a soft sensor with the best generalization capability, and to minimize the implementation and maintenance cost. The main blocks of the methodology and the related process of data reduction are shown in Figure 1.

The main purpose of the first main block is to reduce the number of inputs to those with the highest sensitivity toward the output. Another objective is to test via simulation the hypothesis whether some form of nonlinear relationship between the selected inputs and the output exists. This is a critical point in the whole methodology, because if a neural net model cannot be built, the soft sensor development process stops here. The conclusion in this case could be that if a universal approximator, like a neural net, cannot capture a nonlinear relationship, there would be no basis for variable dependence and no need to look for other methods. The sensitivity analysis is based on stacked Analytical neural nets. A big advantage of this type of neural nets is the reduced development time. Within a couple of hours, the most sensitive inputs are selected, the performance of the best neural net models is explored, and the data for the computationally intensive symbolic regression (GP-function generation) step is prepared. Typically, thirty stacked neural nets are used to improve generalization and estimate neural net model agreement error. This step begins with the most complex structure of all possible inputs. During the sensitivity analysis, decreasing the number of inputs, gradually reduces the initial complex structure. The sensitivity of each structure is the average of the calculated derivatives on every one of the stacked neural nets. The procedure performs automatic elimination of the least significant inputs and generates a matrix of input sensitivity vs. input elimination.

Another important task performed by the analytical neural networks is to deal with time delays. The classical approach to handle time series by neural nets is to add additional inputs for the previous time steps [8]. Unfortunately, this technique increases the dimensionality of the neural net significantly. This increase in the dimensionality of the input vectors has a large impact on the number of required data points for proper model identification. The problem is even bigger in the case of GP modeling. Therefore, it would be desirable to include

0-7803-7282-4/02/$10.00 ©2002 IEEE

information from previous time-steps without increasing the dimensionality of the input to the

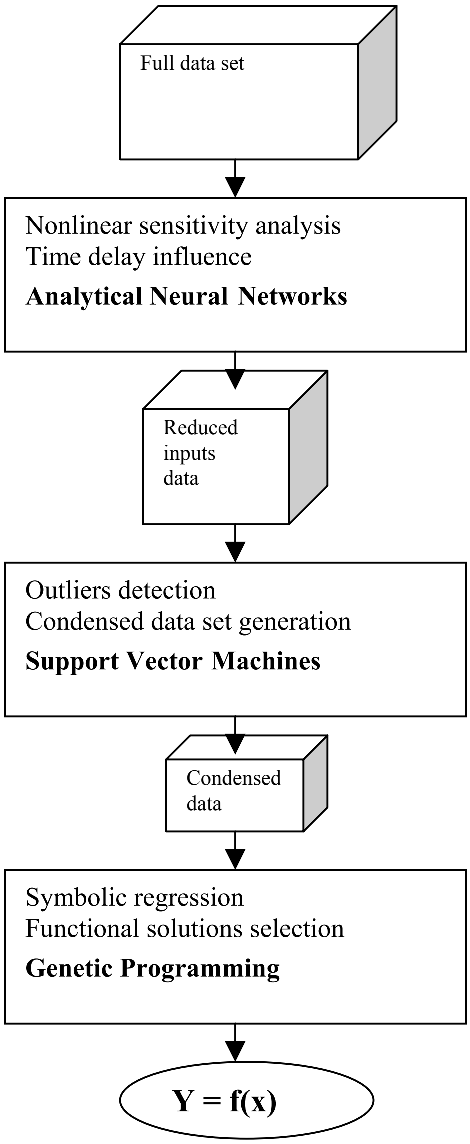

Figure 1. Main blocks of an integrated methodology for robust soft sensor development.

network. This can be achieved by performing a convolution on the input using an appropriately shaped function. As a result of the first block of the integrated methodology, the size of the full data set is reduced to the number of the most sensitive inputs.

The purpose of the next block, based on SVM, is to further reduce the size of the data set to only those data points that represent the substantial information about the nonlinear model. Outliers' detection is the first task in this process. For outlier detection, we make use of the fact that the data points containing important information are identified by the SVM method as support vectors. When the weight of a data point is non-zero, it is a support vector. The value of a support vector's weight factor indicates to what extent the corresponding constraint is violated. Non-zero weight factors hitting the upper and lower boundary indicate that their constraints are very difficult to satisfy at the optimal solution. Such data points are often so unusual with respect to the rest of the samples, that they might be considered as outliers. An outlier detection tool, using the SVM method, typically constructs several models of varying complexity. Data points with a high frequency of weight values on the boundaries are assumed to be outliers.

One of the main advantages of using SVM as a modeling method is that the user has direct control over the complexity of the model (i.e., the number of support vectors). The complexity can be controlled implicitly or explicitly. The implicit method controls the number of support vectors by controlling the acceptable noise level. To explicitly control the number of support vectors, one can either control the ratio of support vectors or the percentage of non-support vectors. In both cases, a condensed data set that reflects the appropriate level of complexity is extracted for effective symbolic regression.

An additional option in this main block is to deliver a soft sensor based on SVM. Some recent results show [6], that SVM models based on mixed global and local kernels have very good extrapolation features. If a soft sensor, generated by GP does not have acceptable performance outside the range of training data, the SVM-based inferential model is a viable on-line solution.

The final block of the integrated methodology for soft sensor development uses the GP approach to search for potential analytical relationships in a condensed data set of the most sensitive inputs. The search space is significantly reduced by the previous steps and the effectiveness of GP is considerably improved. The set of possible functions that can be generated in GP is the set of all possible functions that can be composed from the list of available terminals $T = \{X_1, X_2, ...X_n\}$ and the set of available functions $F = \{F_1, F_2, ...F_m\}$.

Various parameter settings control the type and complexity of equations that are generated. The most important parameters are the list of available functions as well as the list of available inputs. Another parameter that is quite important in controlling the average complexity of the equations being generated is the probability for function selection (default value equals 0.6). This parameter controls what the probability is to grow a specific branch of a tree by selecting a function or terminating the branch by selecting a terminal (a number or a variable) as the next node. The larger this probability

0-7803-7282-4/02/$10.00 ©2002 IEEE

value is, the higher the complexity of the functions being generated.

The final result of symbolic regression is a list of several analytical functions and subequations that satisfy the best solution according to a defined objective function. The analytical function selection for the final soft sensor on-line model is still more of an art than a well-defined procedure. Very often the most parsimonious solution is not acceptable due to specific manufacturing requirements. It is preferable to deliver several potential functions with different levels of complexity and let the final user make the decision. The generalization capabilities of each soft sensor are verified for all possible data sets. Of special importance is the performance outside the training range. It is also possible to design a model agreement-type confidence indicator based on stacked symbolic predictors.

Some of the advantages of the proposed methodology will be illustrated with an industrial application for an emission estimation soft sensor.

5. INDUSTRIAL APPLICATION

Soft sensors for emission estimation are one of the most popular application areas and a viable alternative to hardware analyzers. Usually an intensive data collection campaign is required for empirical model development. However, during on-line operation the output measurement is not available and some form of soft sensor performance self-assessment is highly desirable. Since it is unrealistic to expect that all possible process variations will be captured during the data collection campaign, a soft sensor with increased robustness is required.

Such type of soft sensors, based on the proposed integrated methodology, was developed and implemented in one of The Dow Chemical Company plants in Freeport, TX. The key results from implementation of the main blocks are as follows:

A representative data set from eight potential process input variables and the measured emission as output included 251 data points for training and 115 data points for testing. The test data is 140% outside the range of the training data which by itself is a severe challenge for the extrapolation capability of the model. As a result of the nonlinear sensitivity analysis based on the Analytical Neural Networks, the data set was reduced to five relevant inputs. The performance of such type of potential model with five inputs, 10 neurons in the hidden layer, and a model disagreement indicator based on the standard deviation of 30 stacked predictors is shown in Figure 2. The possibility for nonlinear model building and the potential of the model agreement indicator for performance self-assessment are clearly demonstrated.

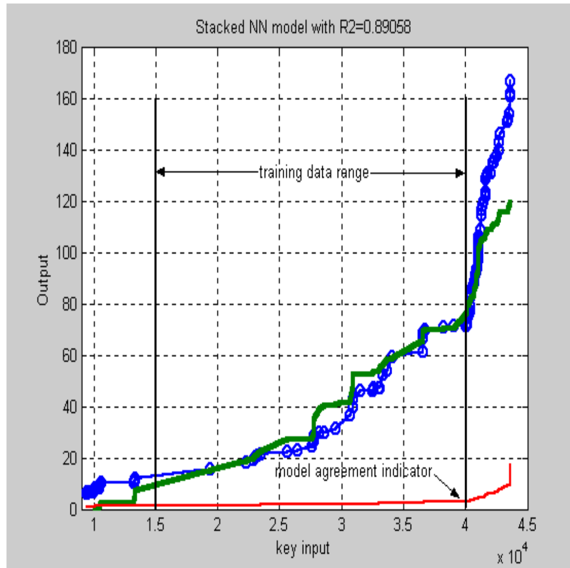

Figure 2. Performance of a Stacked Analytical Neural Net model with model agreement indicator.

The extraordinary extrapolation capability of a potential empirical model based on SVMs is shown in Figure 3. The model is based on a mixture of a second order polynomial global kernel and an RBF local kernel with width of 0.5 in a ratio of 0.95. An additional benefit from this phase of the integrated methodology is that the model is based on 34 support vectors only.

As a result, the representative data set for deriving the final symbolic regression model is drastically reduced to only 8.44% of the original training data set. As it is shown in Figure 4, the performance of the GP-generated model, based on the condensed data set, is comparable with the other two approaches.
The initial functional set for the GP includes: {addition, subtraction, multiplication, division, square, change sign, square root, natural logarithm, exponential, and power}. Function generation takes 20 runs with population size of 500, number of generations of 100, number of reproductions per generation of 4, probability for function as next node of 0.6, parsimony pressure of 0.05 and correlation coefficient as optimization criterion. Eight symbolic predictors with different number of inputs and nonlinear functions were selected in a stacked model. The average value is used as the soft sensor prediction and the standard deviation is used as a model disagreement indicator. The soft sensor for emission estimation is in operation in Freeport, TX since August 2001.

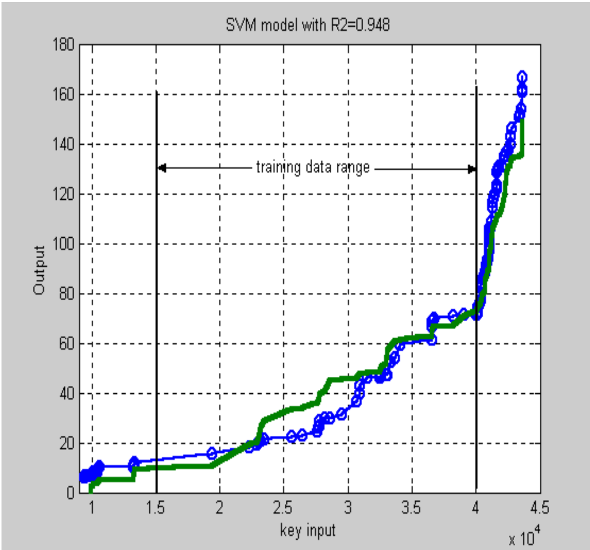

Figure 3. Performance of an SVM model using a mixture of polynomial and RBF kernels.

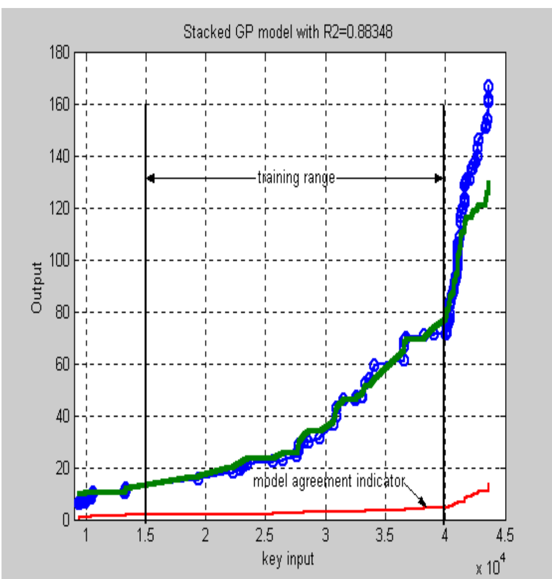

Figure 4. Performance of a Stacked Symbolic Regression model with model agreement indicator.

6. CONCLUSIONS

A novel integrated methodology for robust inferential sensing has been defined and successfully applied for fast and effective development of a soft sensor for emission estimation in The Dow Chemical Company. The proposed methodology is based on using different computational intelligence components (stacked analytical neural nets, genetic programming, and support vector machines). The driving force behind the need of integration is the requirement of industry for soft sensors with increased robustness. The illustrated application shows the main advantage of the proposed methodology – significant reduction of the training data set by nonlinear sensitivity analysis and Support Vector Machines. The final on-line solution, generated by GP, is based on a very compact and robust stacked empirical model with self-assessment capability that requires minimal re-training and maintenance cost. The success of this application in a complex industrial application demonstrates the great potential of the integrated approach as a very effective complement to neural net-based soft sensors.

REFERENCES

[1] Neelakantan R. and J. Guiver, *Applying Neural Networks*, Hydrocarbon Processing, 9, pp. 114-119, 1998.
[2] Qin, S., *Neural Networks for Intelligent Sensors and Control - Practical Issues and Some Solutions*. In *Neural Systems for Control*, Academic Press, New York, 1996.
[3] Koza, J. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*, MIT Press, Cambridge, MA, 1992.
[4] Vapnik, V. *Statistical Learning Theory*, Wiley, NY, 1998.
[5] Sharkey A.(Editor) *Combining Artificial Neural Nets*, Springer, London, 1999.
[6] Smits G. and E. Jordaan, *Using Mixtures of Polynomial and RBF Kernels for Support Vector Regression,*(submitted to Machine Learning), 2001.
[7] Kordon A. K. and G. F. Smits, *Soft Sensor Development Using Genetic Programming,* Proceedings of GECCO'2001, San Francisco, pp. 1346 – 1351, 2001.
[8] Medsker L. and L. Jain (Editors), *Recurrent Neural Networks: Design and Applications*, CRC Press, Boca Raton, FL, 2000.