Categorization of Information in Supervisory Control Systems*

Voula C. Georgopoulos and Peter P. Groumpos

Laboratory for Automation and Robotics,
Department of Electrical Engineering
University of Patras, Patras
GREECE

Abstract: Most of today's industrial processes being of complex and hybrid nature (consisting of both continuous and discrete event parts), require by necessity a supervisor for control. Since the supervisor is at the upper level of a two-level hybrid system, it is required to process and make decisions on a large amount of nonhomogeneous information. In this paper we discuss the categorization of information through neural network interfaces and the issues surrounding the format of the input data to the network, along with a measure of the information content of the data eventually reaching the supervisor.

Peter P. Groumpos received his Ph.D in Electrical Engineering from the State University of New York at Buffalo, in 1978. He is Professor at the Division of Systems & Control and Director of the Laboratory for Automation and Robotics, the University of Patras. He was formerly on the faculty at Cleveland State University, USA, 1979-1989. He was director of the Communication Research Laboratory (1981-1986) and a member of the Technical Committee of the Advanced Manufacturing Center (1985-1987). For the academic year 1987-88 he was a Fulbright visiting scholar at the University of Patras. He is an associate Editor for the International Journals: a) Computers and Electrical Engineering and b) Studies in Informatics and Control. He is a member of the Honorary Societies Eta Kappa Nu and Tau Beta Pi. He is the Co-ordinator of the ESPRIT Network of Excellence in Intelligent Controls and Integrated Manufacturing Systems. He has puublished over 70 journal, conference papers and technical reports. He has been the principal investigator on many R&D projects both in the USA and in Greece. His main research interests are intelligent manufacturing systems and CIM, process control, hierarchical large-scale control systems and adaptive control.

Voula C. Georgopoulos is a Research Fellow at the Laboratory for Automation and Robotics of the Department of Electrical and Computer Engineering, the University of Patras. She received a Diploma in Electrical Engineering from the Democritus University of Thrace in 1984, a M.Sc. degree in Electrical Engineering and Computer Science from M.I.T. in 1987, and a Ph.D degree from Tufts University in 1992. She was a member of technical staff of the MITRE Corporation in the USA from 1988 to 1994. She also taught courses at Northeastern University and Tufts University. She has published more than 30 papers in journals and conference proceedings. Her interests are in neural networks, signal processing and hybrid systems. She is a member of Tau Beta Pi and Sigma Xi.

1. Introduction

Industrial systems are by nature complex and they consist of a number of nonhomogeneous subsystems. They involve problems of different time-scales, they are continuous, or discrete-time and often mixed. Of particular interest in such an environment is a two-level hierarchical system where a supervisor at the upper level supervises and co-ordinates the operation of a number of processes/machines in the first layer, as shown in Figure 1. It should be noted here that two-level hierarchical structures have been used for extensively modelling and controlling large scale systems [1].

The supervisor is embedded in a dynamic environment where it continually interacts with and reacts to its environment and it must reason about events and actions.The information that the supervisor receives from the lower level in an industrial environment is in the form of discrete events, i.e. a sequence of events describing the system operation. Thus, the supervisor along with its interfaces can be considered an information representation and processing which in general must [2]:

- integrate in a coherent way the information and knowledge on the processes and machines
- synthesize the available data coming from the processes and machines

* This work was supported by the EC ESPRIT programme, Project SESDIP, P.8924

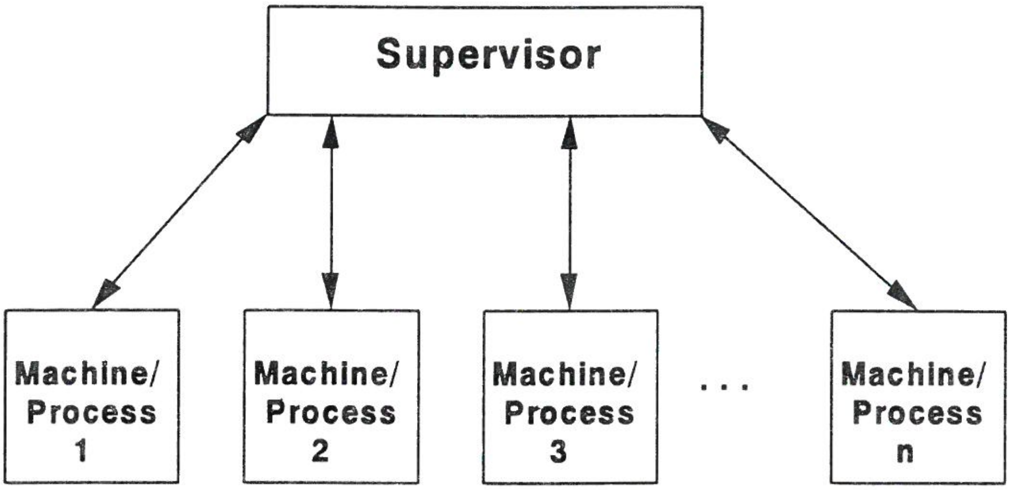

Figure 1. Two–level Hierarchical System

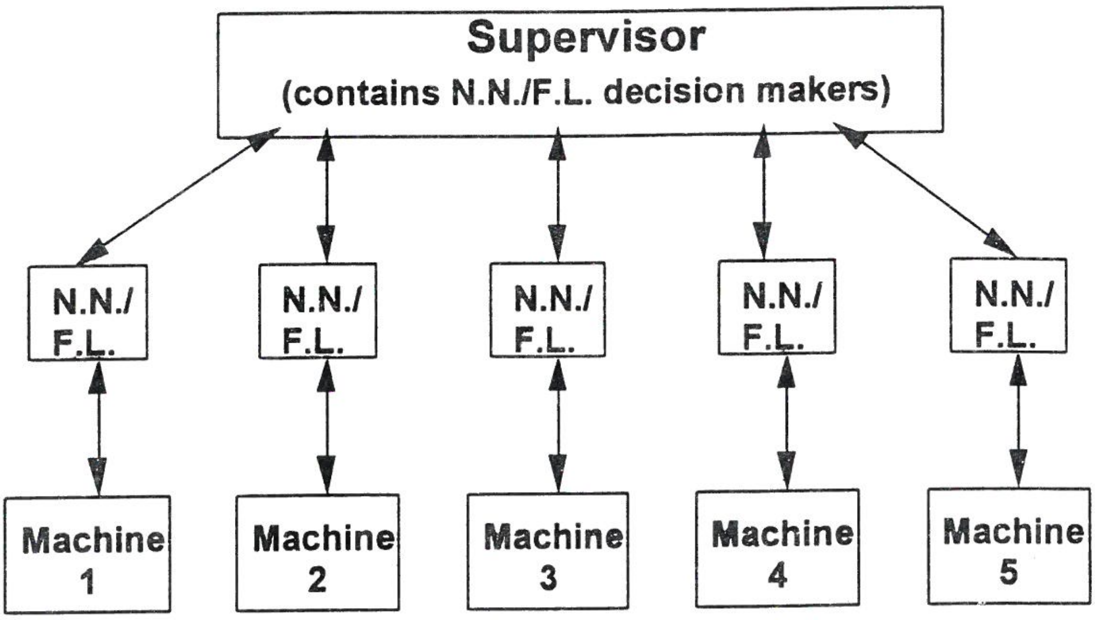

Figure 2. Supervisory Control Model

- explain any event in terms of cause-consequence trees of past related events,
- localize any anomalies and inform through their causes,
- mask and hierarchize dynamically the alarms in function of the operational context, and
- generate dynamically the required actions in normal or problematic conditions even in case of multiple failures.

There is a large gap between the generation of this information (data) at the machine/process level and its being properly processed and effectively comprehended at the supervisory level to be able to perform the above tasks. Some of the issues surrounding this problem are discussed in the following sections:

- What kinds of interface architectures can be used so that the necessary data at the machine/process level can be translated into appropriate events at the supervisor?
- What are the choices of representation of input information at each level, signal (in time, in frequency, in space, combinations of the three, etc.) so that the interface architectures can generate discrete events that can be easily processed and understood by the supervisor?
- What is the information content of the data (or variables/features) so that a minimum amount of information (something like a basis) that is critical to the operation of each subsystem is available to the supervisor?

This paper discusses these issues within a general model shown in Figure 2, where the processes/machines that the supervisor handles are dissimilar. In this case there is a need for breaking down the information into "equivalent" categories. There is a number of different machines attached to the supervisor, each one through a type of interface that translates the incoming data from each machine into discrete events for the supervisor. Each individual machine has its own set of states and dynamic rules and consequently its own amount of information "units", n_1, n_2, n_3, n_4, and n_5. The amount of information by which each machine is represented in the supervisor should be "equivalent" to the information from other machines. In other words, if one of the machines is very critical in the entire process, more units of information will be used to represent the machines in the supervisor than other less critical machines. We have identified the interface in this case to be a neural network along with fuzzy logic (N.N./F.L.) since we are particularly interested in neural networks due to their information transformation abilities, their adaptability to handling new and unusual situations, and to their modular-hierarchical structure.

2.The Role of Neural Networks in the Categorization Process

Since in the supervisor the information that is processed is symbolic (in terms of discrete events), neural networks (along with fuzzy logic) can also be used as an intermediate "interface" structure that will convert signal from the lower level to symbols to be processed at the upper level. This is because they have a nonlinearity built-in making them highly suitable for interfaces from continuous signals to discrete events.

Two categories of neural networks can be used for unified signal/symbol (in our case hybrid) representation and processing. The first category includes distributed networks, such as pattern recognition type neural networks using back-error propagation learning. The information is stored in a distributed way as it becomes freely organized during learning. The output of the network must be unique so that the network becomes stable after learning and the neurons in the hidden layers are not in general attached to any meaning or association, thus the information is not localized. In distributed models each concept is distributed over several nodes and each node can participate in representing more than one concept.

Pattern recognition neural networks can be considered "interfaces" from continuous-state systems to discrete event (symbolic) systems. The events belong to a set of attributes, each of which is represented by a symbol (or for the case of the fuzzy cognitive map, a concept). The symbol is connected to a continuous (sensory) signal x through a series of transformations:

$$z = \alpha((x))$$

where $\alpha(u)$ is a nonlinear transformation in general. It partitions the state space so that each region of the partition is associated with a symbol. A simple example of a neural network which performs such a function is a classification network as shown in Figure 3. This can be achieved using a multilayer standard backward-error propagation artificial neural network which, if sufficient number of neurons in the hidden layer ,can be used to approximate any mapping function [3]. The input to the system is a two-element vector x whose elements x_1 and x_2 take on continuous values. The output of the system takes on one of the three discrete values z={a,b,c} depending on which class the input signal belongs to. This example is presented here since it shows the ability of the neural networks to deal with hybrid control problems, such as autonomous switching and autonomous jumps [4] where abrupt changes are made in either vector fields or subsystem states when certain boundaries of the state space are "hit".

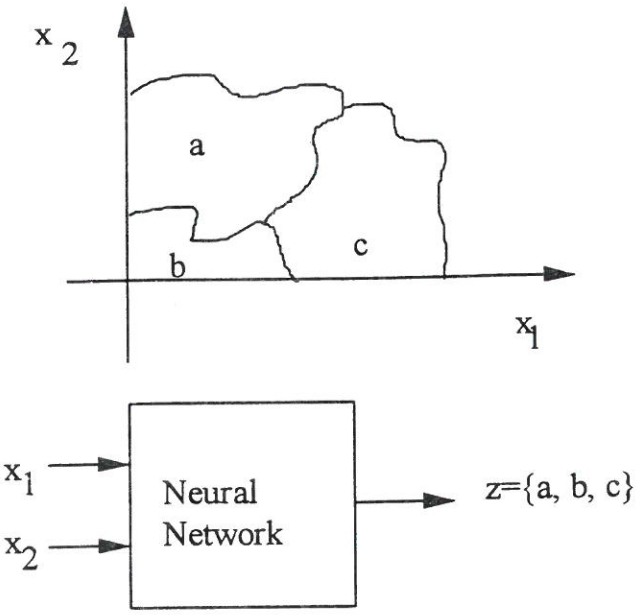

Figure 3. Classification by A Neural Network

The second category of neural networks that can be used for unified signal/symbol representation and processing includes localized neural networks, such as fuzzy cognitive maps [5] where each unit represents a concept and connections and weights are linked to causal relationships between concepts. Stability in these networks are not important, oscillations allow more than one potential answer. In localized neural networks each node of system corresponds to a single concept and vice versa.

Neural networks have been used extensively in the literature for decision- making, classification, recognition, identification, and optimization. Due to their highly parallel structure, they are able to perform these tasks in many cases even in real time and are able to perform global tasks with local interconnections. In addition to these advantages, neural networks have a number of desirable characteristics of information representation and processing schemes within this supervisory control environment [6]:

- adaptability
- modularity
- ability to weight information and
- efficiency

In a complex system, such as an industrial system, it may be too complex to be able to directly classify incoming raw data into discrete event so that they are processed by the supervisor. In such a case a "preprocessing" stage, where an appropriate number of relevant features is extracted from the raw data. Neural networks that perform feature extraction can be used prior to the neural network interfaces that perform a classification leading to "discrete events" to be processed by the supervisor. In these networks while it is essential that the information contained in the input vector is sufficient to determine the output class, the presence of too many input features can burden the training process and can produce a neural network with more connection weights than those required by the problem.

Introducing model abstractions through selective categorization means introducing tractability of decision making inference at the expense of decision quality [7]. Categorization of characteristics, creates generalizations that hide potentially relevant details in a decision model. If computational and representational resources were free or inexpensive, there would not be a need to remove the detail through categorization. Under conditions of limited resources, however, a supervisor may find that representing objects and events in the systems of the lower level in too much detail may require subsequent spending of intolerable computational time in computing optimal decisions. A classification that is too detailed may contain information that is irrelevant to the decision to be made, thus causing the supervisor waste cognitive effort without gain. On the other hand, a categorization model that is too abstract may overlook details thus biasing the decision- making process and perhaps ignoring a particular set of actions.

Thus, an important consideration should be how to represent input data represented so that the essential features can be extracted. There are two areas of focus that can be helpful in such a case. The first is the choice of domain in which to represent the knowledge and the second is to measure the information content of a given variable or feature of a system.

3. Domain of Representation of Input Data

There is a number of considerations in selection of the optimum input signal representation to a neural network. It is fundamental that the inputs to a neural network stage are comprehensible, concise and sufficient so that the neural network should be able to perform its function [8]. Of course, the input format is application driven. Input data representation can be in spatial form (precise location of an object in a manufacturing system), time, frequency or combinations of time and frequency, space and frequency, etc.

1. **Waveform Representations**: This is a data sequence representing a sampled signal in time. This can show effects such as periodicities, echoes, onset and turn-off of phenomena, etc.
2. **Frequency-Domain Representations**: The frequency content of a signal can be represented through the use of a set of fixed orthogonal functions as bases, such as DFT, FFT, Cosine/Sine, and Hadamard transforms. In such representation harmonics can be seen, noise sources, etc.
3. **Joint-Domain Representations**: In cases of signals whose characteristics as a whole vary with time or space, it is appropriate to represent themby using joint domain representations. With these a signal can be represented as a joint function of time and frequency, time and space, space and frequency, etc. Most popular are time-frequency distributions which were introduced to analyze signals with rapidly changing characteristics. The underlying idea of time-frequency distributions is to devise a joint function of time and frequency describing the energy density or intensity of a signal in time and frequency. Desired properties of a time-frequency distribution are: real distribution, integration over time and frequency of the distribution results in the energy of the signal, integration over time yields the spectral density, integration over frequency yields the instantaneous power, and the first-order moments of the distribution are the group delay and the instantaneous frequency of the signal. Examples of time-frequency distributions that obey all or some of the desired properties are the spectrogram, wavelets, and the Wigner Distribution.

Neural networks can be trained to process any one of these types of input signal representations.

4. Characterization of Information Content

It is important to be able to measure how useful a variable or feature is; in other words, what its information content is. In addition, due to the large number of variables available from each subsystem attached to the supervisor, it is desirable that the information reaching the supervisor is free from redundancies that will slow down and even prohibit in some cases the decision- making process.

A neural network performing classification trained to classify patterns from a set of different classes with the backpropagation algorithm, can be considered as a system that reduces the initial uncertainty of the information contained in the input vector. In the ideal case the final uncertainty will be zero (i.e. the class will be certain) [9]. However in a complex system, such as an industrial system, the final uncertainty can be higher due to insufficient input information or suboptimal operation (insufficient network size, training, etc.). It is important to be able to measure the information content of the various features of data with respect to the data. In the case that their information content is insufficient to reduce uncertainty, more features or more informative ones should be found. The uncertainty of a given class is measured by its *entropy* [10].

Since we are interested in categorizing the variables from each of the systems in the lower level so that a certain event is detected, it is important to associate a certain probability $p_1(\mathrm{i})$ of occurrence of each class of each variable. In a manufacturing environment, it is very difficult to have a priori knowledge of the probabilities of occurrence of a certain variable. Thus, each variable is observed over a period of time with N being the total number of observations, and the number of occurrences of each "event," where the variable belongs to the *ith* class of the set of classes $C_1 = \{1,2,..,i,...,m\}$ is $n_1(\mathrm{i})$. And the probability $p_1(\mathrm{i})$ can be defined as:

$$p_1(\mathrm{i}) = \lim_{N\to\infty} \frac{n_1(\mathrm{i})}{N}$$

The entropy $H(\mathrm{x}_1)$ of the variable x_1 estimated from N observations of the variable x_1:

$$H(\mathrm{x_i}) = -\sum_{i=1}^{m} \frac{n_1(\mathrm{i})}{N} \log\left(\frac{n_1(\mathrm{i})}{N}\right)$$

which leads to:

$$H(\mathrm{x}_1) = \log N - \frac{1}{N}\sum_{i=1}^{m} n_1(\mathrm{i}) \log\left(n_1(\mathrm{i})\right)$$

The joint entropy of the two variables x_1 and x_2 is similarly given by:

$$H(\mathrm{x}_1,\mathrm{x}_2) = \log N - \frac{1}{N}\sum_{i=1}^{m}\sum_{j=1}^{k} n_{1,2}(\mathrm{i,j}) \log\left(n_{1,2}(\mathrm{i,j})\right)$$

where $n_{1,2}(\mathrm{i,j})$ being the number of event happenings "x_1 belongs to the i th class of C_1" and "x_2 belongs to the j th class of C_2" and $C_2 = \{1,2,..,j,...,k\}$.

And the conditional entropy of variables x_1 and x_2 is:

$$H(\mathrm{x}_1|\mathrm{x}_1) = \log N - \frac{1}{N}\sum_{i=1}^{m} n_{1,2}(\mathrm{i,j}) \log\left(n_{1|2}(\mathrm{i|j})\right)$$

A non-negative quantity which shows the intensity relationship between the two variables x_1 and x_2 is the *mutual information* (MI) or *transinformation* of x_1 to x_2 given by:

$$\begin{aligned} I(\mathrm{x}_1:\mathrm{x}_2) &= H(\mathrm{x}_1) + H(\mathrm{x}_2) - H(\mathrm{x}_1,\mathrm{x}_2) \\ &= H(\mathrm{x}_1) - H(\mathrm{x}_1|\mathrm{x}_2) \\ &= H(\mathrm{x}_2) - H(\mathrm{x}_2|\mathrm{x}_1) \end{aligned}$$

Since the conditional entropy will be less that or equal to the initial entropy of a given variable, the mutual information shows to what extent the uncertainty is decreased by introducing a second variable.

A necessary and sufficient condition for x_1 and x_2 to be statistically independent is that their mutual is equal to zero. The value of $I(\mathrm{x}_1:\mathrm{x}_2)$ is maximum, when one variable depends *only* on the other.

From the above, it can be seen that the MI can measure the way the two are connected. Mutual information can be used as a method for measuring the information that a group of input variables provides about the outputs. This method can be used to eliminate less useful input variables, thus reducing the number of weights of a network and improving the networks generalization ability [11].

5. Conclusions and Future Work

We have presented in this paper some issues surrounding the categorization of information in a two-level supervisory control industrial system.

We have examined neural networks role playing in the categorization process in terms of feature extractors and classifiers , essentially performin a signal-to-symbol conversion (from data to discrete events) so that the supervisor handles a minimum and sufficient amount of information. Two areas of interest that might be helpful in this process we have also presented. The first is the choice of the domain in which to represent the knowledge and the second is to measure the information content of a given variable or feature of a system.

The issues and methods discussed here will be used in a system functionality model that has been proposed in [6] where the controls and information in complex hierarchical system are in parallel considered in a sortof integrated information- control concept [12]. This is a fundamental step towards establishing suitable techniques for hierarchical multi-level systems for use in Flexible Intelligent Manufacturing Systems.

REFERENCES

1. JAMSHIDI, M., **Large Scale Systems: Modeling and Control**, NORTH HOLLAND, Amsterdam, 1983.
2. TAYMANS, E., **S&F: A Method of Knowledge Representation and Expert System for Hybrid Process Control**, ADPM'94 Symposium, Brussels,23- 24 Nov. 1994.
3. POGGIO, T. and GIROSI, F. , **Networks for Approximation and Learning**, Proc. IEEE, Vol. 78, Sept. 1990, pp. 1481-1497.
4. BORKAR, V., BRANICKY, M. and MITTER, S.A., **Unified Framework for Hybrid Control**, LIDS Seminar MIT, February 22, 1994.
5. KOSKO, B., **Neural Networks and Fuzzy Systems**, Englewood Cliffs, New Jersey, PRENTICE-HALL, 1992.
6. GEORGOPOULOS,V.C. and GROUMPOS, P.P., **Information Representation and Processing in An Industrial Environment**, accepted for presentation at IFAC ACASP'95 Workshop, Budapest , June, 1995.
7. POH, K.L. et al, **Dynamic Construction and Refinement of Utility-Based Categorization Models**, IEEE TRANSACTIONS ON SYSTEMS, MAN AND CYBERNETICS, Vol. 24(11), 1994, pp. 1653 - 1663.
8. HUNG, S.Y., **Digital Neural Networks**, PRENTICE-HALL, Englewood Cliffs, NJ, 1993.
9. BATTITI, R.,**Using Mutual Information for Selecting Features in Supervised Neural Net Learning**, IEEE TRANSACTIONS ON NEURAL NETWORKS, Vol. 5, No.4, July 1994.
10. SHANNON, C.E., **A Mathematical Theory of Communication**, THE BELL SYSTEM TECH. JOURNAL, Vol. 17, 1948, pp. 623-656.
11. BONNLANDER, B.V. and WEIGEND, A.S. , **Selecting Input Variables Using Mutual Information and Nonparametric Density Estimation**, Proceedings of the '94 International Symposium on Artificial Neural Networks, Tainan, The People's Republic of China, December 15-17, 1994, pp. 42-50.
12. GROUMPOS, P.P., **Structural Dynamic Hierarchical Stabilization and Control of Large-Scale Systems**, in B.C. Leonde (Ed.) Control and Dynamic Systems 22, Decentralized Distributed Control and Dynamic Systems, ACADEMIC PRESS, 1985, pp. 117-163.