

Knowledge Based Modelling of Complex Interconnected Systems

Boriana Vatchova
Institute of Information and Communication
Technologies
Bulgarian Academy of Sciences
Sofia, Bulgaria
boriana.vatchova@gmail.com

Mo Adda
School of Computing
University of Portsmouth
Portsmouth, UK
alexander.gegov@port.ac.uk

David Sanders
School of Engineering
University of Portsmouth
Portsmouth, UK
david.sanders@port.ac.uk

Alexander Gegov
School of Computing
University of Portsmouth
Portsmouth, UK
mo.adda@port.ac.uk

Abstract - This paper considers the use of intelligent control for modelling complex interconnected processes. The latter usually have many inputs and outputs and can be found in different application areas. Some parts of the inputs are measurable and others are not because of the presence of stochastic environmental factors. This is the reason why processes of this kind operate under uncertainty. The latter is addressed in this paper by intelligent systems that use probabilistic and fuzzy network structures.

Keywords- logical probability functions, logical probabilistic networks, fuzzy networks

I. ATTRIBUTES OF COMPLEX PROCESSES

Data mining can be considered as coherent merging of information from multiple sources. In this sense, data mining is part of a knowledge discovery process [1, 2, 3].

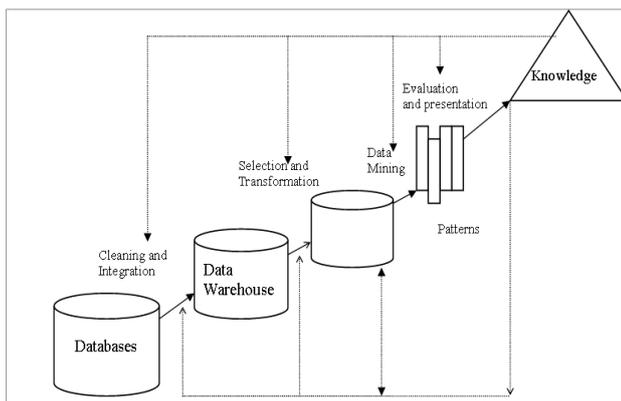


Fig.1 Data mining as part of knowledge discovery

The existing methods for knowledge presentation in intelligent systems can be classified as follows:

- **Logical formulas**

Logical models for knowledge presenting are based on the concept of a formal system $M = (T, P, A, F)$ where T is set of base elements, P is set of syntaxes rules, A is set of true axioms and F is set of rules for inferences.

- **Production rules**

Production models are combinations of production rules which are related to each other in form of the following type: If {antecedent} Then {consequent}.

- **Semantic Nets**

Semantic nets are another way of representing knowledge. They are familiar as their graphical visualization in form of nodes and vertices.

- **Frames**

Frames or list of facts are the next form for knowledge presenting in intelligent systems. They are used when investigated data are characters or words.

The existing methods for knowledge extraction which use expert knowledge or data mining are usually subjective and intuitive, not suitable for real time and not suitable for complex processes. In this paper, the existing data mining methods [4, 5, 6] are supplemented by methods of random functions theory and multi-valued logic [7, 8]. The models of the processes are knowledge bases of production rules or multi-layer network structures which include probability of occurrence.

Most existing methods for modelling complex systems are based on deterministic mathematics [9] or probability theory whereby the assumption is that all input variables are entirely known or completely unknown. Similarly, most of the existing methods for simulating complex systems are based on fully centralised or completely decentralised concepts whereby the assumption is that all input variables are considered as a single entity or multiple entities. Both assumptions presume the use of computational algorithms that do not adequately reflect the complexity attributes of the system to be modelled such as non-linearity, uncertainty, dimensionality and structure. To address the shortcomings of most existing methods, this paper uses a hybrid approach that builds on other recent works by the authors. This approach can potentially improve model performance indicators for complex systems such as feasibility, accuracy, efficiency and transparency. The one-to-one

relationships between the above system complexity attributes and the associated model performance indicators are given in Table 1,[10].

TABLE 1. COMPLEXITY ATTRIBUTES AND PERFORMANCE INDICATORS

System complexity attribute	Model performance indicator
non-linearity	feasibility
uncertainty	accuracy
dimensionality	efficiency
structure	transparency

There are usually two types of complexity:

Quantitative

- Large number of inputs
- Large number of state variables
- Large number of outputs
- Large number of sub-processes

Qualitative

- Non-linearity
- Non-stationary
- Uncertainty
- Disturbances

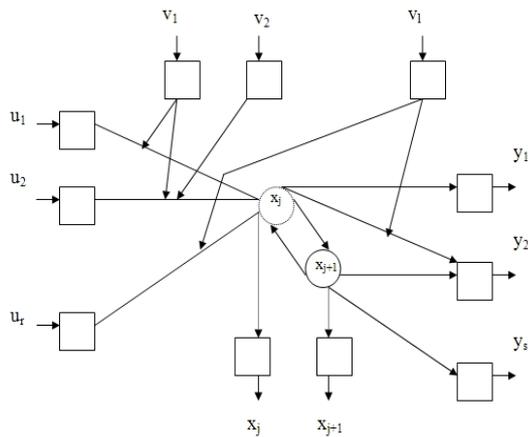


Fig.2 Model of complex process

Figure 2 shows a complex process with many inputs, state parameters and outputs, where $U(u_1, u_2, \dots, u_r)$ are controllable inputs; $V(v_1, v_2, \dots, v_l)$ are parameters of the environment (disturbances); $X(x_1, x_2, \dots, x_d)$ are state variables; $Y(y_1, y_2, \dots, y_s)$ are outputs. Figures 3, 4, and 5 show different kinds of multi-stage complex processes whose stages are interconnected in a sequence, in parallel or through feedback loops

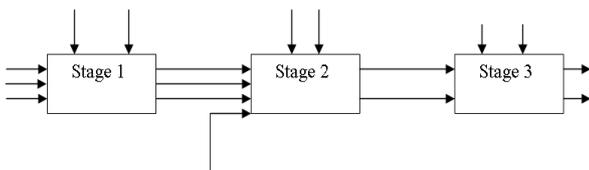


Fig.3 Complex process with sub-processes in a sequence

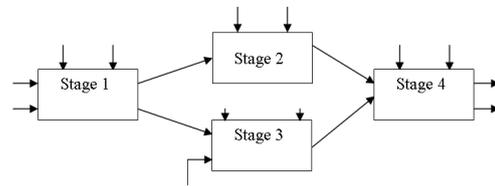


Fig.4 Complex process with sub-processes in parallel

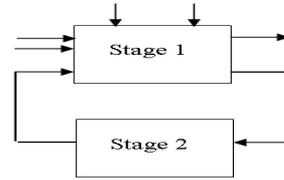


Fig.5 Complex process with a feedback loop

The main problems for many of the existing models of complex processes have been low feasibility, low accuracy, low efficiency and/or low transparency.

II. MODELS WITH LOGICAL PROBABILITY FUNCTIONS

The main model of this type is knowledge bases of production rules with probability of occurrence [7].

TABLE 2 LOGICAL PROBABILISTIC FUNCTIONS IN TABLE FORMAT

Grouping sequence set	GLN ₁	GLN ₂	GLN ₃	GLN ₄	
Frequency of occurrence	p{GLS ₁ }	p{GLS ₂ }	p{GLS ₃ }	p{GLS ₄ }	
LX ₁	LX ₁₁	LX ₂₁	LX ₃₁	LX ₄₁	
LX ₂	LX ₁₂	LX ₂₂	LX ₃₂	LX ₄₂	
LX ₃	LX ₁₃	LX ₂₃	LX ₃₃	LX ₄₃	
Ly ₁	a ₁	p{Ly ₁₁₁ }	p{Ly ₁₂₁ }	p{Ly ₁₃₁ }	p{Ly ₁₄₁ }
	a ₂	p{Ly ₁₁₂ }	p{Ly ₁₂₂ }	p{Ly ₁₃₂ }	p{Ly ₁₄₂ }
	a ₃	p{Ly ₁₁₃ }	p{Ly ₁₂₃ }	p{Ly ₁₃₃ }	p{Ly ₁₄₃ }
	a ₄	p{Ly ₁₁₄ }	p{Ly ₁₂₄ }	p{Ly ₁₃₄ }	p{Ly ₁₄₄ }
	a ₅	p{Ly ₁₁₅ }	p{Ly ₁₂₅ }	p{Ly ₁₃₅ }	p{Ly ₁₄₅ }

LPF in analytical form, see equation 1, [7, 8]

$$Ly_1, p\{Ly_1\} = F\{LX_1, LX_2, LX_3\} \quad (1)$$

where LX_1, LX_2, LX_3 are inputs and Ly_1 and Ly_2 are the outputs; a_1, a_2, a_3, a_4 and a_5 are logical values with a meaning for very small, small, medium, big and very big.

P_{abcd} is probability (frequency) of occurrences for the three inputs and one output, see Table 2.

LPF is a knowledge base, whose knowledge is a system of production rules ordered by the columns. Each column from Table 2 is a production rule or a knowledge base [7]. For example, from column $j=2$, for grouping sequence set GLN_2 with probability of occurrence $p\{GLS_2\}$, the associated rule is

$$\text{If } \{LX_1 = LX_{21}, LX_2 = LX_{22}, LX_3 = LX_{23}\} \quad (2)$$

Then

$\{Ly_1=a_1$ with probability $p(Ly_{121})$

$Ly_1=a_2$ with probability $p\{Ly_{122})$

$Ly_1=a_3$ with probability $p(Ly_{123})$

$Ly_1=a_4$ with probability $p(Ly_{124})$

$Ly_1=a_5$ with probability $p(Ly_{125})\}$.

III. MODELS WITH PROBABILISTIC NETWORK STRUCTURE

The inputs to the models with dominant data sets are logical values of the inputs which form the first layer of the network. The set of dominant data sequence sets is the intermediate layer of the network. The set of outputs defines the output layer of the network. Connections between the elements of these layers correspond to the relations between the elements of the presented sets and the frequency of occurrence [8].

In Figure 6, the inputs to the models with network structure are logical values of the inputs Lx_{ij} , which form the first layer of the network.

The set of dominant data sequence sets GLN_r is the intermediate layer of the network. The set of the outputs Ly_{eq} defines the output layer of the network

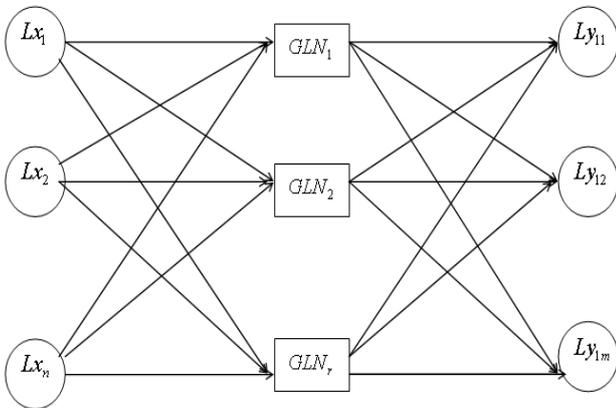


Fig.6. Data occurrence frequency model with network structure

The following relations are introduced here:

R_{LxGLNr} - is the relation between sets of the inputs and dominant grouping sequence sets;

R^*_{LxGLNr} - is the relation between the frequency of occurrence of elements of the inputs and intermediate layers;

$R_{GLNrLy_{eq}}$ - is the relation between the logical values of the elements of intermediate layer and output layer

$R^*_{GLNrLy_{eq}}$ is the relation between the frequency of occurrence of the elements between the intermediate and the output layer.

Using the network model, logical values and probability of occurrences of the outputs are calculated for each combination of logical values of measurable inputs

$$Ly = R_{GLNr} Ly \times R_{LxGLNr} \times Lx \quad (3)$$

$$p\{Ly\} = R^*_{GLNr} Ly \times R^*_{LxGLNr} \times p\{Lx\} \quad (4)$$

where symbol 'X' is multiplication.

IV. MODELS WITH FUZZY NETWORK STRUCTURE

A fuzzy system with r rules, m inputs $x_1 \dots x_m$ taking linguistic terms from the input sets $\{A_{11}, \dots, A_{1r}\}, \dots, \{A_{m1}, \dots, A_{mr}\}$ and n outputs $y_1 \dots y_n$ taking linguistic terms from the output sets $\{B_{11}, \dots, B_{1r}\}, \dots, \{B_{n1}, \dots, B_{nr}\}$ can be described by the following if-then rules [10,11].

Rule 1: If x_1 is A_{11} and ... and x_m is A_{m1} , then y_1 is B_{11} and ... and y_n is B_{n1} (5)

.....
Rule r: If x_1 is A_{1r} and ... and x_m is A_{mr} , then y_1 is B_{1r} and ... and y_n is B_{nr}

For compactness, the linguistic terms of the inputs and the outputs for the four nodes above can also be represented by positive integers. In this case, 'small', 'low' and 'light' are represented by '1', the linguistic terms 'medium', 'average' and 'moderate' are represented by '2' whereas the linguistic terms 'big', 'high' and 'heavy' are represented by '3'. Boolean matrices are novel formal models for fuzzy systems that can represent nodes in a fuzzy network.

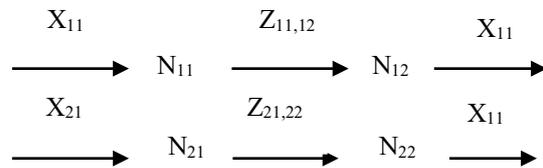


Figure 7: Feedforward rule based model with network structure

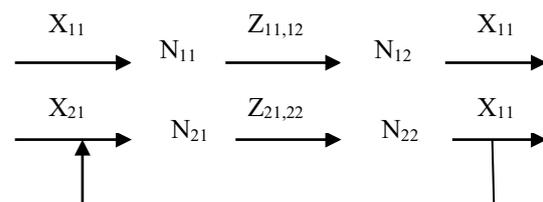


Figure 8: Feedback rule based model with network structure

The structures in Figures 7 and 8 are feedforward and feedback formal models for a fuzzy network with a node set $\{N_{11}, N_{12}, N_{21}, N_{22}\}$, an input set $\{x_{11}, x_{12}, x_{21}, x_{22}\}$ and an output set $\{y_{11}, y_{12}, y_{21}, y_{22}\}$. These structures specify the location of nodes as well as their inputs and outputs.

Rule 1 for N_{11} : If x_{11} is small, then y_{11} is low (6)

Rule 2 for N_{11} : If x_{11} is medium, then y_{11} is high (7)

Rule 3 for N_{11} : If x_{11} is big, then y_{11} is average (8)

Rule 1 for N_{12} : If x_{12} is low, then y_{12} is moderate (9)

Rule 2 for N_{12} : If x_{12} is average, then y_{12} is heavy (10)

Rule 3 for N_{12} : If x_{12} is high, then y_{12} is light (11)

Rule 1 for N_{21} : If x_{21} is small, then y_{21} is average (12)

Rule 2 for N_{21} : If x_{21} is medium, then y_{21} is low (13)

Rule 3 for N_{21} : If x_{21} is big, then y_{21} is high (14)

Rule 1 for N_{22} : If x_{22} is low, then y_{22} is heavy (15)

Rule 2 for N_{22} : If x_{22} is average, then y_{22} is light (16)

Rule 3 for N_{22} : If x_{22} is high, then y_{22} is moderate (17)

The connections between the nodes $N_{11}, N_{12}, N_{21}, N_{22}$ from Equations (6)-(17) can be described by the connection set $\{z_{11,12}, z_{21,22}\}$. In this case, the first connection is identical with the output from N_{11} and the input to N_{12} whereas the second connection is identical with the output from N_{21} and the input to N_{22} . The if-then rules for the fuzzy network nodes $N_{11}, N_{12}, N_{21}, N_{22}$ from Equations (6)-(17) can be described by the Boolean matrices in Equations (18)-(21).

$$N_{11}: \begin{matrix} y_{11} & & & \\ x_{11} & 1 & 2 & 3 \end{matrix} \quad (18)$$

$$\begin{matrix} 1 & 1 & 0 & 0 \\ 2 & 0 & 0 & 1 \\ 3 & 0 & 1 & 0 \end{matrix}$$

$$N_{12}: \begin{matrix} y_{12} & & & \\ x_{12} & 1 & 2 & 3 \end{matrix} \quad (19)$$

$$\begin{matrix} 1 & 0 & 1 & 0 \\ 2 & 0 & 0 & 1 \\ 3 & 1 & 0 & 0 \end{matrix}$$

$$N_{21}: \begin{matrix} y_{21} & & & \\ x_{21} & 1 & 2 & 3 \end{matrix} \quad (20)$$

$$\begin{matrix} 1 & 0 & 1 & 0 \\ 2 & 1 & 0 & 0 \\ 3 & 0 & 0 & 1 \end{matrix}$$

$$N_{22}: \begin{matrix} y_{22} & & & \\ x_{22} & 1 & 2 & 3 \end{matrix} \quad (21)$$

$$\begin{matrix} 1 & 0 & 0 & 1 \\ 2 & 1 & 0 & 0 \\ 3 & 0 & 1 & 0 \end{matrix}$$

Boolean matrices as the ones presented above are very suitable for formal modelling of fuzzy systems [11]. Besides this, Boolean matrices work well with other

formal models which can take into account connections among nodes in fuzzy networks.

For example, the four nodes $N_{11}, N_{12}, N_{21}, N_{22}$ from Equations (6)-(17) can be described by the grid structure with two levels and two layers in Equation (22).

$$\begin{matrix} & \text{Layer1} & & \text{Layer2} & \\ \text{Level 1} & N_{11}(x_{11}, y_{11}) & & N_{12}(x_{12}, y_{12}) & \\ \text{Level 2} & N_{21}(x_{21}, y_{21}) & & N_{22}(x_{22}, y_{22}) & \end{matrix} \quad (22)$$

$$\begin{matrix} & \text{Layer 1} & & \\ \text{Level 1} & z_{11,12} = y_{11} = x_{12} & & \\ \text{Level 2} & z_{21,22} = y_{21} = x_{22} & & \end{matrix} \quad (23)$$

V. CONCLUSION

Several types of intelligent systems for modelling complex interconnected processes models have been presented in this paper. In particular, these models make use of logical probability functions, probabilistic network structure and fuzzy network structure.

The main potential of the presented models is the more adequate presentation of the considered interconnected processes. This can help with the model feasibility, accuracy, efficiency and transparency that has been a problem for many of the existing methods.

REFERENCES

- [1] Hopgood A., "Knowledge -Based Systems for Engineers and Scientists", Second Edition, CRC-Press, p.57-85, 1993.
- [2] D. Ruan, G. Chen, E.Kerre, G.West (Edts.), „Intelligent Data Mining: Techniques and Applications (Studies in Computational Intelligence)", Springer-Verlag, Berlin,Hidelberg, 2010.
- [3] Han J., M.Kamber, "Data Mining Techniques", Morgan Kaufmann Publisher, 2005.
- [4] Guida G., C.Tasso. "Design and Development of Knowledge Based Systems. From Life Cycle to Methodology". John Wiley & Sons, p.3-30, 1994.
- [5] Lee J. (Ed.), "Software Engineering with Computational Intelligence, Studies in Fuzziness and Soft Computing", Springer, 2003.
- [6] Tim J.M., "Artifitial Intelligence A systems Approach" Infinity Science Press LLC, Hingham Massachusetts, New Delhi,2007.
- [7] Vatchova B., "Derivation and Assessment of Reliability of Knowledge for Multifactor Industrial Processes", PhD Thesis, 167 pages, Bulgarian Academy of Sciences, Sofia, 2009, (in Bulgarian).
- [8] Vatchova B.,A.Gegov," Production rule and network structure models for knowledge extraction from complex processes under uncertainty", Recent Contributions in Intelligent Systems, Springer, 2017, ISBN 978-3319414379, p.379-390.
- [9] Terzieva, Y. Pavlov, N. Dobrinkova, E. Paunova-Hubenova: "A Study on Integration of ICT Resources in STEM Education - Utility Assessment", 11th annual International Conference of Education, Research and Innovation (ICERI'18), Seville, Spain, 12 - 14 November, 2018. (in print).
- [10] Gegov A., Fuzzy Networks for Complex Systems, Springer, 2010.
- [11] Gegov A., Petrov N., Sanders D., Vatchova B. Boolean matrix equations for node identification in fuzzy rule based networks. 4, 21, International Journal of Knowledge-based and Intelligent Engineering Systems - Volume 21, issue 4, 2017, ISSN: 1875-8827, p.69-83.