Contribution to Incomplete and Noisy Information Problem Solving by Artificial Intelligence Principles Applying

Stefania GALLOVA, *Member, IAENG*

Abstract— **Engineering or manufacturing domain knowledge is rarely in the form required by these systems. Furthermore, we have to solve a lack of information, presence of incomplete and noisy knowledge about solved complex diagnostic system from reliability, optimal and predictive manufacturing point of view. Decision support purposes require for the knowledge provider to know about primary „cause-effect" relationships but not be in a position to assert that other relationships are nonexistent. The use of maximum entropy inference in reasoning with uncertain information is commonly justified by an information-theoretic argument. This contribution deals with a possible objection to this information-theoretic justification, presents a probabilistic reasoning methodology and a maximum entropy application, which can estimate missing information by making some sort of global assumption and provide advice based on the knowledge available. Some diagnostic problems could be operated in a chaotic fashion. Achieved results showing how the fractal dimension and entropy describing the chaotic motion depend on the operating characteristics of the device. Diagnostic frequency waves could also lead to chaotic fluctuations in the time evolution of the transmitted intensities.**

Index Terms— **chaos theory, incomplete information, knowledge base, knowledge reconstruction, maximum entropy, noisy signal, uncertainty.**

I. Introduction

The solved knowledge base is very large and complex and represents a clue how to realize effective methodologies for real diagnostic system problem solving. The combination between object-oriented and knowledge-based approaches to application software system design is the most suitable for intelligent automation in the industrial engineering domain. In addition to the "*formal knowledge*" bounded to data, there exists a knowledge which is a result of a gained experience or the so called heuristic knowledge. This fact is particularly true for industrial engineering processes and activities.

From above mentioned point of view we have to solve a lack of information which is incomplete, uncertainty and vague. We often obtain relevant formally inexpressible information. Partially, it is possible to solve this problem also by fuzzy theory principles applying. There are many approaches to deal with lack of information. We introduce the procedure which is able to estimate missing values in incomplete fuzzy relations.

The methodology is also shown to be capable of encapsulating whatever knowledge is available. This paper also deals with the reconstruction of a knowledge base from observational data. The practical applicability of these methods is usually limited as it is impossible to estimate multidimensional probability distributions from observational data reliably.

There are static and dynamic theories of degrees of belief. It is argued that maximum entropy is a dynamic theory and that a complete theory of uncertain reasoning can be gotten by combining maximum entropy inference with probability theory. A probability theory is a static theory. This "*total*" theory – it seems – is much better grounded than are other theories of uncertain reasoning.

A great deal of theoretical effort has been devoted to understanding the complex behaviour observed in the experiments. Active diagnostic systems such as the machinery relevant diagnostic signals monitoring have attracted more attention than passive systems. Its study has shown instabilities, higher-order dynamical states and non classical statistics of the monitoring solved signals. Here is created a software product which is usable in real practice of machine equipments environment. A static theory concerns the consistency conditions for degrees of belief at a given time. A dynamic theory concerns how one's degrees of belief should change in the light of new information The standard information-theoretic justification of maximum entropy inference goes as follows – when presented with information in the form of a set of probability constraints, we want to infer a probability distribution that satisfies those constraints but that, of all such distributions, is the least biased [1],.[2].To be least biased is to minimize the amount of information contained in the distribution. The distribution must have at least as much information content as the constraints since it satisfies them. Among the distributions satisfying the constraints there is a unique one, the maximum entropy distribution that has minimum information. Hence maximum entropy inference selects the right distribution [3], [4], [5], [6].

II. Some Problems of Knowledge Reconstruction

One of the solved tasks here is to determine the knowledge

Manuscript received March 22, 2008. This work was supported by the Slovak Republic's Ministry of Education under Scientific Grant Project: Intelligent Approach to Automated Diagnostic Problem Solving of Machinery Equipment (Number: VEGA 1/4133/07).

Stefania GALLOVA. Author is with the Pavol Jozef Safarik University in Kosice, Srobarova 2, SK-041 80 Kosice, Slovak Republic;
e-mail: stefania.gallova@zoznam.sk; stefania.gallova@upjs.sk; phone: +421-904-584-426;

structure from observational data. The data available may consist of truth values for all nodes for some set of cases. It is assumed that it is known which nodes are goal hypotheses. The knowledge of joint distribution for all pairs of nodes enables one to find not only the structure of the corresponding graph, but also to order nodes and to determine their structure.

Here is a solved knowledge base system. Events are represented by nodes and relationships between events are represented by edges. The node type is identified according to the channel matrix. If nonzero means mutual information between variables X_i and X_k , then the X_i is input and X_k is output of the communication channel or vice versa. Properties of the communication channel are determined by matrices, either $[P(x_{ij} / x_{kl})]$ or $[P(x_{kl} / x_{ij})]$.

The x_{ij} and x_{kl}, $i, k = 1, 2, ..., n,$
$j, l = 1, 2, ..., m$
are values of variables X_i and X_k . It is assumed that all variables are metrical or ordinal and that the inputs for node X_k are nodes $X_1, X_2, ... ,X_r$. For the „AND“ node and m-valued logic, the following condition will be satisfied:

$$x_{kl} = min\ (x_{1p}, x_{2q}, ... , x_{rs}) \tag{1}$$

and for the „OR“ node, it will be determined:

$$x_{kl} = max\ (x_{1p}, x_{2q}, ... , x_{rs}) \tag{2}$$

where $p, q, l, ..., s = 1, 2, ..., m.$ Relations between output X_k and one of the inputs, e.g. X_r were investigated. They are output and input of communication channel [7],[8] [9] [10].

There are channel matrices $[P(x_{kl} / x_{rs})]$ for the „AND“ and „OR“ nodes. They are as follows:

$$\left[P(x_{k,l}/x_{r,s})\right]_{AND} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ p_{21} & p_{22} & 0 & \cdots & 0 \\ p_{31} & p_{32} & p_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & 0 \\ p_{m1} & p_{m2} & p_{m3} & \cdots & p_{mm} \end{bmatrix} \tag{3}$$

$$\left[P(x_{k,l}/x_{r,s})\right]_{OR} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & \cdots & p_{1m} \\ 0 & p_{22} & p_{23} & \cdots & p_{2m} \\ 0 & 0 & p_{33} & \cdots & p_{3m} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \tag{4}$$

If the orientation of the channel is available, i.e. the assignment of variables both to the input and output of the channel is known, then the pertinent matrix determines the node type. If the node type is known, one can determine the orientation of the pertinent channel. The channel matrix can be neither upper triangular nor lower triangular. Then, there are three possibilities of a channel matrix form: unit matrix, any permutation matrix, or genuine stochastic matrix.

One can show that there are the following relation holds:

$$R\ (X_i : X_k) = min\ [H(X_i), H(X_k)] \tag{5}$$

Where $H(X_i)$, $H(X_k)$ are entropies. In a special case, the channel matrix does not contain any 1.

The condition $0 \leq P(x_{kl} / x_{ij}) < 1$ is thus satisfied and the channel matrix is than a genuine stochastic matrix. In this case, the node is a Bayesian one, and the output is not deterministic but a random function of relevant inputs. The relation may be other stochastic by its nature or deterministic – logical, and only be seems to be stochastic because some inputs are not available.

There is a task - which of the two nodes associated with the edge is an input and which is an output of the channel. There is the possibility of finding the edge orientation in a special case of the channel matrix. Assume the channel matrix to be genuinely stochastic. Let it be a symmetrical one in addition, i.e. $p_{jl} = p_{lj}$ for any $j, l = 1, 2,..., m.$ It is a „doubly stochastic“ matrix. There is an expression for a communication channel with input X_i and output X_k and a doubly stochastic channel matrix:

$$H\ (X_k) \geq H\ (X_i) \tag{6}$$

This means that if both entropies are not equal, the higher entropy is the entropy of the output variable. „Shannon's entropy“ says nothing about channel orientation if the channel matrix is not symmetrical. In this case, there is an additional uncertainty resulting from the asymmetry of the channel matrix. The „uncertainty measure“ called relative entropy makes it possible to consider this additional uncertainty. The relative entropy is defined as:

$$H_G(X_i) = \sum_{j=1}^{m} P(x_{ij}) \log_2 \left[P(x_{ij})/w_j\right] \tag{7}$$

where w_j are the prior weights, $H_G(X_i)$ – the relative entropy can be interpreted as a „Gibbsian entropy“ (a concept known from statistical dynamics). The weights w_j as average values of elements of the jth column of the channel matrix are introduced here. The weights then correspond to probabilities of values of output if input values are uniformly distributed [10].

Relative entropies can be defined for both variables connected with the channel as:

$$H_G(X_i) = \sum_{i=1}^{m} P(x_{ij}) \log_2 \left[P(x_{ij}) \Big/ \sum_{l=1}^{m} P(x_{ij}/x_{kl}) \right] \tag{8}$$

$$H_G(X_k) = \sum_{k=1}^{m} P(x_{kl}) \log_2 \left[P(x_{kl}) \Big/ \sum_{j=1}^{m} P(x_{kl}/x_{ij}) \right] \tag{9}$$

If X_i is an input and X_k an output of the channel, the relative entropies satisfy the following inequality:

$$H_G(X_i) \leq H_G(X_k) \tag{10}$$

If all inputs are known, the node type can be determined. This approach has been tested using data achieved from simulated knowledge base for solved diagnostic expert system.

ISBN: 978-988-17012-9-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

Reconstruction technique shows how the introduced facts are used for knowledge base reconstruction. Start from the matrix $[T(X_i:X_k)]$. By the *chi-square test*, one eliminates the pairs X_i,, X_k for which the zero hypotheses, which are the hypothesis on statistical independence, cannot be refuted. These pairs correspond to elements 1 of the matrix $[T(X_i:X_k)]$ *chi-square*. Other elements have value 0. This matrix does not guarantee that all dependencies in the knowledge base have been detected. If the data set is not a representative ensemble and there is only a little information on some dependency, the dependency cannot be detected. On the other hand, an inappropriate choice of level of significance may result in some spurious dependencies.

A fuzzy number is a fuzzy set defined on the real line. The membership function of a fuzzy set defined on a truth space with interval [0,1] could be interpreted as the „meaning" of a label describing the degree of certainty in a linguistic manner. During the aggregation process, these fuzzy numbers will be modified according to given combination rules and will generate another membership distribution that could be mapped back into a linguistic term for the convenience of user or to maintain closure [2], [5], [6], [10].

III. A Maximum Entropy and Probabilistic Reasoning

Probabilistic reasoning is inherently an exponentially large problem. There are two fundamental problems associated with probabilistic reasoning – the amount of knowledge required and the size of the computation - which can be reduced by using knowledge which is in a very specific form, i.e. causal dependencies, and by propagating probabilities through a directed acyclic graph. Here are also following situations to solve. The knowledge provider has no information relating two or more events but would not like to assert that they independent or conditionally independent of each other. A possible way to proceed is to assume independence whenever the knowledge provided does not assert dependence. Other situation is if knowledge is available which cannot be propagated, e.g. expected values. If the knowledge is not in the form required for propagation, we are left with the problem that both the information required and the computations are potentially exponentially large, as in above mentioned problems.

It is a big problem to provide an exponentially large quantity of knowledge, but the maximum entropy methodology will accept the information that can be provided and will estimate the rest. The maximum entropy methodology will accept knowledge – pertaining to the probability of given situations – in whatever form it may be given, thereby also handling the second above mentioned situation. The maximum entropy methodology can only estimate this missing information by making some sort of global assumption. This is usually done by invoking the principle of insufficient reason and assuming that, in the absence of any information, the joint probabilities are uniformly distributed. Consequently, the distribution derived using the maximum entropy methodology is referred to as being minimally prejudiced.

The nature of the network is such that events are represented by nodes and relationships between events are represented by edges. The absence of an edge between two nodes is an assertion that there is some form of independence between the events. When an edge is present, specific information must be provided to fully specify the nature of the dependencies. Although propagation methods differ in detail, all require information which will enable the current probability of a node to be calculated from knowledge of the current probability of its antecedent nodes. The current (conditional) probabilities can be also calculated from the respective marginals which can each be found by summing the appropriate joint probabilities.

A following example illustrates a maximum entropy theory approach which is compare with traditional theories approach [7], [8], [10].

A three node graph is assumed. Causal information in example has these relations ($n_1 ® n_3$, $n_1 \rightarrow n_2$):

$$\begin{array}{ll} \text{node } n_1: & P(E_1) = k_1 \\ \text{node } n_2: & P(E_2 \mid E_1) = k_2 \\ & P(E_2 \mid \overline{E}_1) = k_3 \\ \text{node } n_3: & P(E_3 \mid E_1) = k_4 \\ & P(E_3 \mid \overline{E}_1) = k_5 \end{array} \tag{11}$$

The nodes n_1, n_2, n_3 represent events which can take these values: E_1, E_2, E_3 and their inversions. The knowledge given in above mentioned three node causal tree is not in the form required by the maximum entropy method and so it has been converted into the form below using the convention for numbering the states.

$$\begin{aligned} &\sum_0^7 P(S_i) = 1 \\ &(1-k_1) \sum_{i=0,1,2,3} P(S_i) - k_1 \sum_{i=4,5,6,7} P(S_i) = 0 \\ &(1-k_2) \sum_{i=0,1} P(S_i) - k_2 \sum_{i=2,3} P(S_i) = 0 \\ &(1-k_3) \sum_{i=4,5} P(S_i) - k_3 \sum_{i=6,7} P(S_i) = 0 \\ &(1-k_4) \sum_{i=0,2} P(S_i) - k_4 \sum_{i=1,3} P(S_i) = 0 \\ &(1-k_5) \sum_{i=4,6} P(S_i) - k_5 \sum_{i=5,7} P(S_i) = 0 \end{aligned} \tag{12}$$

The constraints given by equations (12) give rise to the following constraints/state matrix:

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ (1-k_1) & (1-k_1) & (1-k_1) & (1-k_1) & -k_1 & -k_1 & -k_1 & -k_1 \\ (1-k_2) & (1-k_2) & -k_2 & -k_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & (1-k_3) & (1-k_3) & -k_3 & k_3 \\ (1-k_4) & -k_4 & (1-k_4) & -k_4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & (1-k_5) & -k_5 & (1-k_5) & -k_5 \end{bmatrix} \tag{13}$$

which in turn gives rise to the following general solution for the probability distribution for the states:

ISBN: 978-988-17012-9-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

$$P(S_0)=e^{-l_0}\,e^{-l_1(1-k_1)}\,e^{-l_2(1-k_2)}\,e^{-l_4(1-k_4)}$$
$$\dots$$
$$P(S_7)=e^{-l_0}\,e^{-l_1(-k_1)}\,e^{-l_3(-k_3)}\,e^{-l_5(-k_5)} \tag{14}$$

Substituting the general solution given by equation (14) into the constraints in equations (12) produces the expressions:

$$\begin{aligned}
k_1 &= e^{-l_0}\,e^{-l_1(1-k_1)}\,e^{-l_2(-k_2)}\,e^{-l_4(-k_4)}\,(e^{-l_2}+1)\,(e^{-l_4}+1)\\
1-k_1 &= e^{-l_0}\,e^{-l_1(-k_1)}\,e^{-l_3(-k_3)}\,e^{-l_5(-k_5)}\,(e^{-l_3}+1)\,(e^{-l_5}+1)\\
k_2(1-k_2) &= e^{-l_2}\\
k_3(1-k_3) &= e^{-l_3}\\
k_4(1-k_4) &= e^{-l_4}\\
k_5(1-k_5) &= e^{-l_5}
\end{aligned} \tag{15}$$

By substituting the first, third and fifth equations from above mentioned equations (15) into each of the first four equations of equations (14) we have:

$$\begin{aligned}
P(S_0) &= k_1\,k_2\,k_4\\
P(S_1) &= k_1\,k_2\,(1-k_4)\\
P(S_2) &= k_1\,(1-k_2)\,k_4\\
P(S_3) &= k_1\,(1-k_2)(1-k_4)
\end{aligned} \tag{16}$$

By substituting the second, fourth and sixth equations from equations (16) into each of the last four equations of equations (15), we have:

$$\begin{aligned}
P(S_4) &= (1-k_1)\,k_3\,k_5\\
P(S_5) &= (1-k_1)\,k_3\,(1-k_5)\\
P(S_6) &= (1-k_1)\,(1-k_3)\,k_5\\
P(S_7) &= (1-k_1)\,(1-k_3)\,(1-k_5)
\end{aligned} \tag{17}$$

from where:

$$P(E_2)=\sum_{i=0,1,4,5} P(S_i)=k_1\,k_2+(1-k_1)\,k_3 \tag{18}$$

$$P(E_3)=\sum_{i=0,2,4,6} P(S_i)=k_1\,k_4+(1-k_1)\,k_5 \tag{19}$$

Following important results, which can be made about above mentioned example, are here. It must be noted that the result agrees with those derived from first principles. Sufficient information was available to fully describe the system. In fact, had the answers disagreed, the soundness of the maximum entropy approach would have been brought into question. If the information is incomplete, i.e. one or more of k_1, k_2, k_3, k_4, k_5 are missing, a solution from first principles is no longer possible. However, the maximum entropy method is not embedded and will still produce a result.

The result will be a minimally prejudiced distribution for the probabilities $P(S_i)$ for $i=0,1,2,3,4,5,6,7$, i.e. it will be the result which, whilst complying with the constraints, keeps the probability distribution *„as near as possible“* to the uniform distribution. This is almost certainly not the same as the distribution which would be achieved when everything was known for certain but, given the information available, it is arguably the best distribution to use.

There are grounds for hoping that methods can be found for containing the computation of a maximum entropy solution for practical situations. Above mentioned evidences are based on the fact that computing the maximum entropy solution is feasible for a reasonably large number of events. There is the speculation that practical domains can be partioned, or decomposed, into clusters whose individual sizes do not exceed this number. There is also a possibility that it is the form of the knowledge and not the method used to compute the solution which determines the size of the computation. Computing the advice is conceptually easy, although it is also potentially large, but it too should respond to the same methods for containing the computation.

In general, maximum entropy methods must be applied globally, but for the purposes of deciding whether or not it is wise to assume independence they could be used locally. Consequently, given the more usual directed acyclic graph, one could examine nodes which have more than one parent using maximum entropy, and determine the disparity between the minimally prejudiced solution and the one assuming independence. If the maximum entropy method could be used locally to determine the relationship between two events which would otherwise have to be assumed to be independence, this result could be added to the edge of the directed acyclic graph. By doing this whenever there was a significant discrepancy between the assumption of independence method and the maximum entropy method, one might produce a graph which have good agreement with global maximum entropy methods when processed by propagative methods. One would then have a computationally feasible method of producing a good approximation to a minimally prejudiced solution.

We have some results that can be made about observations. The information given is incomplete, i.e. one or more of k_i (i = 1,...,5) are missing, a solution from origin principles is no longer possible. The maximum entropy methodology is not impeded and will still produce a result. The result will be a minimally prejudiced distribution for the probabilities $P(S_i)$ for i = 0,...,7, i.e it will be the result which, whilst complying with the constraints, keeps the probability distribution as „near“ as possible to the uniform distribution. This is almost certainly not the same as the distribution which would be achieved when everything was known for certain but, given the information available, it is arguably the best distribution to use.

IV. Chaos Theory Principles Approach

In practice, data, information and knowledge about real machinery equipment and technological processes are incomplete and insufficient for implementation of relevant diagnostic system. Chaos theory seems to be very important tool to gain control of these problems effectively. Chaos theory provides very valuable complementary information,

ISBN: 978-988-17012-9-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

which is not possible to obtain by another means of analysis. We have higher quality of analysis at lower determinateness of inputs (achieved estimating and approximating trajectories). Missing useful information is possible to obtain by this way. Achieved experimental results are very promising.

We solve, within diagnostic signal spectrum analysis, the lack of information or noise. The effect of additive and multiplicative stochastic noise on dynamical processes was studied. We designed the techniques for effective dealing with noise and dynamics. We realize simulation and experimental models for investigation of links between multiwave mixing and its well establish techniques, terminologies and dynamical system approaches. A nonlinear dynamical analysis is applied to predictive diagnostic system. A primary task in this area is how to determine whether a solved dynamical system is chaotic and if it is chaotic how to characterize its chaotic evolution. Our approach to solve it is based on estimating dimensions, Lyapunov exponents and metric entropies. By this approach we verify the presence of chaos as underlying noisy spectra and in characterizing of chaos types in various systems.

Measurements of the metric entropy based on time-delay embedding vectors techniques seem to give two important values, i.e. both the maximum divergence rate and the maximum convergence rate of trajectories. The next task is how to interpret the measured characteristics of experimental data. Due to the high dynamical sensitivity of a system near a critical point, careful measurements of the frequency of the nearly of the periodic behaviour near a subcritical Hopf bifurcation point can be used to obtain information regarding system parameters. Solved diagnostic system can be described by semiclassical fields and additive, multiplicative stochastic noise. Semiclassical approximations provide the predictions of residual observable effects. The noise added to the solved system can disrupt the coherences that prevent chaotic evolution, thus permitting localization in real space and exponential divergence of trajectories as in the classically chaotic cases [11], [12], [13].

Surprising results on the very large influence of noisy microwaves (external and internal noise) with noise bandwidth roughly equal one half their frequency, were also included in partial solutions that cause „hesitations “in transitions between different instability regimes. Noise influence understanding from noise sources precise characterization point of view completes solved methodology approach. We experimentally observe the additive or multiplicative noise sources progress. Statistical fluctuation in observed frequency microwave dependences on others observed frequency microwaves. This fact influences lack of relevant information problem solving by higher quality of analysis.

Some diagnostic problems could be operated in a chaotic fashion. Achieved results showing how the fractal dimension and entropy describing the chaotic motion depend on the operating characteristics of the device. Experiments show how stimulated scattering induced by counterpropagating (diagnostic) frequency waves could also lead to chaotic fluctuations in the time evolution of the transmitted intensities.

The basic ideas of solved methodology are based on the following calculations. We investigate the effect of noise and the accuracy of the prediction using genetic algorithm. The more noise the system has the more difficult it is to discern the structure with more judgements. We introduce some noise levels with the testing purpose whether the genetic algorithm tool is still able to recognize the system and realize useful prediction. Experiments show that the genetic algorithm approach is able to learn the judged structure.

We realize a genetic algorithm scheme by this way:

Genetic_Algorithm ()
(<Initialize population>
while <Not (Stop condition)> do(
<Fitness evaluation>
<Selection>
<Reproduction and Mutation>
)
<Choose final solution>
)

A genetic algorithm is a stochastic computational model that seeks the optimal solution to an objective function. Fitness function, as well as the parameters of the fitness function, can affect the result of learning process. The fitness of an individual structure is a measure indicating how fitted the structure is. The search is performed through an iterating procedure applied to a „population of individuals“, i.e. a set of feasible solutions. Searching strategy is similar to biological evolution, i.e. better solutions are reproduced, whereas worse solutions are discarded. Thus, the search strategy is based on the possibility to discriminate between elements in order to resolve which is the good solution of the fitness function and therefore has a good chance of reproducing and generating new elements with its genetic inheritance.

We have a parameter ***Fmn*** **,** which is the best fitness of a string in the population for test problem example ***m*** after generation ***n***.A parameter ***pt*** is the number of test problems [10], [11]. The ***fitness*** of the genetic algorithm ***GAFs*** after generation ***n*** is given by:

$$Fitness[GA_{FS}] = \frac{1}{p_t}\sum_{m=1}^{p_t}\frac{\max(0, F_n^m - F_0^m)}{1 - F_0^m} \tag{20}$$

A general used formula for calculation of multidimensional vector data code is given by [14], [15]:

$$\frac{(k^2 - k + 1)}{R} = V_1 \times V_2 \times \cdots \times V_i \qquad with \quad (V_1, V_{2,} \quad \cdots \quad V_i) = 1 \tag{21}$$

We enumerate a set of nodes of V_1 ´ V_2 ´, … , ´ V_i (grid). Each node of the grid meet exactly ***R***-times. We consider fuzzy rule base of Takagi-Sugeno system form. The Takagi-Sugeno fuzzy model is described by a set of fuzzy implications, which characterize local relations of a system in the state space [16] to [21].

The controlled Takagi-Sugeno fuzzy system is described as follows:

ISBN: 978-988-17012-9-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

$$x(t+1)=\sum_{i=1}^{n} m_i(t)\, H_i\; x(t)\, v(t)$$
$$=\sum_{i=1}^{n} m_i(t)\, H_i\; x(t)+d\;\sin\left(\frac{p}{d}\;b\;x(t)\right) \tag{22}$$

The controller is taken as a sinusoidal function. In the simulation, the magnitude of the control input is experimentally chosen to be δ*=0.09*. Thus, $\| v(t) \|_\infty < \delta$, and can also be regarded as a control parameter. Without control, the Takagi-Sugeno fuzzy model is stable.

Let the solved dynamical system is generally described by the differential equation [13], [19]:

$$x(t)\equiv\frac{dx}{dt}=F(x,m) \tag{23}$$

where x is a vector whose components are the dynamical variables of the system, $F(x,\mu)$ is a nonlinear function of x and μ stands for the control parameter. Let $x_1(\mu)$ and $x_2(\mu)$ be two different solutions of equation (23). A critical point is therefore defined by the equation:

$$x_1(\mu_c) = x_2(\mu_c) \tag{24}$$

which is an implicit equation for μ_c. The Jacobian matrix $G(x, \mu)$ associated with $F(x,\mu)$ is defined through:

$$F(x+\delta x, \mu) = F(x,\mu) + G(x, \mu)\, \delta x + 0(\delta x)^2 \tag{25}$$

The fundamental property that can be proved by using the implicit function theorem is:

$$\det G(x, \mu_c) = O \tag{26}$$

This result implies critical slowing down. This is easily shown by a linear stability analysis of the solution $x_1(\mu)$.

We have the following assumption:

$$(x,\mu) = x_1(\mu) + \varepsilon x'(\mu) \exp(\lambda t) + 0\,\varepsilon \tag{27}$$

$$\det\{\lambda I - G(x_1, \mu)\} = O \tag{28}$$

where I is the unit matrix, the ε-exponent is interpreted as a dimension. Hence at $\mu = \mu_c$ one root (at least) of (28) will vanish, implying an infinite relaxation time. This result remains true if $x_1(\mu)$ and $x_2(\mu)$ are time-periodic solutions of relation (23). We realize the direct estimate of fractal dimension from experimental data (embedding theorem and related topics) with a particular attention to the effect of filtering on a chaotic signal.

All relevant quantities involved in the description of strange attractors require taking some limit (either in space or in time) [12] [20], [21]. The fractal dimension, for example, is related to the scaling behaviour of the (natural) invariant measure μ, when the observational resolution is increased. To be more specific, we first define a *partition* (covering) of the phase space, as a collection of disjoint open sets with variable size ε_i. In this way, we can associate a mass p_i to each element E_i of the partition as:

$$m_i=\int_{E_i} dm \tag{29}$$

The mass m_i can be evaluated from the fraction of points belonging to E_i, when a sufficiently large number of the set points, is generated according to the measure μ. From the obvious consideration that $m \sim \varepsilon^1$ in the case of a plane, it follows that the ε-exponent is to be interpreted as a dimension. Accordingly, we can define a *local* dimension α_i in terms of size and mass of the ***i-th*** element of the covering:

$$m_i \sim \varepsilon_i^{\alpha i} \tag{30}$$

where ε_i is assumed to be sufficiently small, and we let α_i explicitly depend on the index i. This is a crucial point that makes also a statistical approach most appropriate. Metric entropy using improves this approach. It is also worth noting that such a definition of dimension is not affected by the existence of multiplicative factors, since they only yield corrections to the leading scaling behaviour. Hence, they can, in principle, be neglected, although their contribution is often relevant in numerical simulations. The distinctive features of a numerical algorithm for the estimation of fractal dimensions are essentially related to the partition used to cover the set. Therefore, the task starts from the rules to generate a suitable covering (according to the probability of each element). This approach can be extended to metric entropies and Lyapunov exponents for problem solving improving.

The action of the map F in phase-space is translated into a shift of symbols in the associated space:

$$x_n \rightarrow x_{n+1}\;\text{Þ}\;\{\ldots, s_{n-1}, s_n, s_{n+1}, \ldots\} \rightarrow \{\ldots, s_n, s_{n+1}, s_{n+2}, \ldots\} \tag{31}$$

The time origin in the (doubly-infinite) symbol sequence is moved one place to the right. When a generating partition is not known, as in the case of experimental data or in most of computer simulations, it is possible to divide the searching space of size $\varepsilon_i < \varepsilon$ (they are usually taken all equal for simplicity) and evaluate the *metric entropy* as:

$$K(q)=\lim_{e\to 0}\;\lim_{N\to\infty}\frac{1}{N}\frac{1}{(1-q)}\ln\sum_{S_N} p^q\,(S_N) \tag{32}$$

where the limit ε→0 guarantees that a generating partition is finally obtained. The performed analysis assumes implicitly the knowledge of all coordinates of each attractor's point in phase space. This is certainly the case of all numerical simulations, but it is not always possible in experiments. Sometimes, just one variable can be measured at different times.

In this way, we construct the values and parameters of genes within solved domain chromosome. We solve metric entropy for each element or for solved chromosome within genetic algorithm running. This chromosome implementation by genetic algorithm approach is near to real diagnostic situation representation. Nowadays, some experiments were

ISBN: 978-988-17012-9-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

realized. Achieved results are very promising. By this way, we will avoid distorted fitness values arising from possibly different degrees of complexity of test problem examples [11],[12],. The fitness of an individual structure is a measure indicating how fitted the structure is.

Distributed noise was added to the genetic algorithm running from 1 to 100 % to see the noise effect on the prediction accuracy. Experiments show that the structure of the system unrecognisable was less than 3,85 % and not influence the requirement prediction function of solved diagnostic system.

We have the following time series:

$$x=x_1,x_2,, \ , \ x_n, \tag{33}$$

where dimension equals to *d.* and added coordinate *y* is expressed:

$$y = f[x(t-q)] \tag{34}$$

where q is time delay.

The distance of all nearest neighbour in dimension *d* is computed and then compared to the corresponding nearest neighbour in dimension *d+1.* If we have the case that the distance is almost the same, it means it is not a false nearest neighbour. By increasing the dimension value until there is no further decrease of the false nearest neighbour count, then the dimension *d* found is the minimum embedding dimension.

$$GA_{prediction} = \sqrt{\frac{\sum_{m=1}^{n}\left(x_m - x_m^{'}\right)^2}{n^2}} \tag{35}$$

where x_m represents the actual value and x_m’ represents the predictive value.

We provide also a contribution into inductive inference problem solving, that is mathematically well-based theory of algorithmic learning from incomplete information. „Learning“ belongs to the central issues of artificial intelligence likewise learnability is a fundamental characteristics of natural intelligence [11], [21].

V. Conclusion

The other property of created methodology based on maximum entropy which has not yet been explored is their ability to include noncausal information. This property is inherent to maximum entropy methods because they comply with the constraints imposed on them whatever the nature of these constraints – as long as they are mutually compatible.

A research within the solved problem demonstrates that it is relatively easy to estimate missing information using the maximum entropy methodology. Methodology need to rely on domain specific properties to permit simplification. There are difficult computational requirements of the maximum entropy methodology for large knowledge domains. Our future work will also focus on two main tasks. At first, we want to examine practical diagnostic applications within solved expert system for which the global maximum entropy solutions can be computed for example in polynomial time. Secondly, we want to find efficient approximate methodologies which give good agreement with global maximum entropy solutions for practical diagnostic applications not covered by the above mentioned.

In practice, we solve possible types and forms of incompleteness, uncertainty and lack of information within analysis process of solved diagnostic system. In real manufacturing situation we often classify data under uncertainty conditions. Commonly encountered problem is the aggregation of the clustering results with the expert judgements. Uncertainty contains uncertainty of classification objectives, basic data uncertainty or incompleteness, positional and form relationship uncertainty, quantity of clusters uncertainty and partitioning fuzziness. This paper presents also a genetic algorithm approach using to predict the chaotic system. Some noise levels were introduced with the purpose to test whether the genetic algorithm methodology is still able to recognize the system and make a useful prediction. The chaotic system structure can be learn by genetic algorithm as long as the noise in the chaotic system does not exceed a certain value. The more difficult is for the genetic algorithm to learn the underlying structure of the chaotic system (the higher or the lower level of the noise). Generally used methods developed in both fractal geometry and chaotic dynamics explain a multitude of diverse physical phenomena i.e. from trees to turbulence, cities to cracks, heart beat to fluid flow and much more [11], [13].

Forecast of future statements within diagnostic problems experimental modelling has been important and has brought relevant results to secure and predict optimal operation of technological processes in manufacturing context.

The research reported on this paper is supported by following scientific project of Slovak Grant Agency VEGA (Ministry of Education): *Intelligent Approach to Automated Diagnostic Problem Solving of Machinery Equipment* (Number: VEGA 1/4133/07).

References

[1] Anderson, S. K, Olesen, K. G., Jensen F. V: Hugin – a Shell for Building Bayesian Belief Universes for Expert Systems. In: Shaffer, W & Pearl, J. (eds.) Readings in Uncertain Reasoning, Morgan Kaufma, USA, 1990.

[2] Chilausky, R. L.: An Experimental Comparison of Several Many-valued Logic Inference Techniques in the Context of Computer Diagnosis of Soybean Disease. International Journal of of Man-Machine Studies. Department of Computer Science, pages 178-162, University of Illinois, Illinois, 1980.

[3] Cohen, P. A.: A Theory of Heuristic Reasoning about Uncerainty. The AI Magazine, Summer 1983, Report 84.

[4] Goldman, S. A., Rivest, R. L.: A Non-iterative Maximum Entropy Algorithm. In: Koval, L. N. & Lemmu, F. J. (eds.) Ucertainty in Artficial Intelligence, Vol.2, pages 133-148, North-Holland 1988.

[5] Hamming, R. W.: Coding and Information Theory. Prentice Hall Inc., Englewood Cliffs, N.J., 1980.

[6] Maung, I., Paris, J. B.: A Note on the Ifeasibility of some Inference Pocesses. Internatinal Journal of Intelligent Systems, pages 593-603, Vol 5, No 5, 1990.

[7] Rhodes, P. C., Karakoulas, G.: A Probabilistic Model-based-method for Diagnosis. Artificial Intelligence in Engineering, Vol. 6, No 2, 86-99, 1991.

[8] Rhodes, P. C., Garside, G. R.: Computing Marginal Probabilities in Causal Inverted Binary Trees Given Incomplete Information. Knowledge-Based-Systems, Vol. 10(4), 213-224, 1998.
[9] Shore, J. A., Johnson, R. W.: Axiomatic Derivaion of the Principle of Maximum Entropy and the Principle of Minimum Cross-entropy. IEEE Transactions on Information Theory, Vol. IT, 26-37, 1980.
[10] Gallová, Š.: A maximum Entropy Inference within Uncertain Information Reasoning. In: Information Processing and Management of Uncertainty in Knowledge-based Systems: Proceedings: Paris, Les Cordeliers, July 2-7, 2006. Paris: E.D.K., 2006. p. 1803-1810. ISBN 2-84254-112-X.
[11] Jantke,P.: Algorithmic learning from incomplete information. 5th International meeting of computer scientists. Smolenice, Czechoslovakia, 1988
[12] Boyd,R.W.,Raymex,M.G., Narducci,L.M.> Optical instabilities. U.Press Cambridge, 1986
[13] Lugiato,L.A., Narducci,L.M.,Abraham,N.B.> For a recent suggestion of a way
[14] Goldberg D. E.: Genetic Algorithms in Search, Optimization and Machine Learning. Addison-Wesley, reading, Massachusets, 1989
[15] Abraham, N. B.:, Lugiato, L. A; Narducci, L. M.: Instabilities in Active Media. J.O.Society Am.B.;1999
[16] Mandelbrot, B. B.: The Fractal Geometry of Nature. Freeman, San Francisco, 1982
[17] Li, Z; Park, J. B.; Joo, I. H.; Chen, G.; Choi, I. H.: Anticontrol of Chaos for Discrete TS Fuzzy Systems. IEEE Trans., Circ.-Syst.1, Vol.49, 2/2002
[18] Mandel, P.; Erneux, T.: Dynmic versus Static stability. Frontiers in Quantum Optics. A.Hilger, Bristol, Boston, 1986
[19] Badii, R.; Politi, A.: Strange Attractors. Phys.Letters, 104A, 303, 1984
[20] Gallová, Š.: Fault Diagnosis of Manufacturing Processes via Genetic Algorithm Approach. In: IAENG Engineering Letters, Volume 15 Issue 2, 2007. Pages 349-355. ISSN: 1816-093X.
[21] Gallová, Š.: Fuzzy Ontology and Information Access on the Web. In: IAENG International Journal of Computer Science, Volume 34 Issue 2, 2007. Pages 234-238. ISSN: 1819-656X.

ISBN: 978-988-17012-9-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)