No. 788

Sensitivity Analysis in Data Mining

by

Masato Koda

July 1998

SENSITIVITY ANALYSIS IN DATA MINING

Masato Koda
Institute of Policy and Planning Sciences
University of Tsukuba
1-1-1 Tennoudai, Tsukuba Science City, 305-8573 Japan
Tel: +81-298-53-5222
Fax: +81-298-55-3849
koda@shako.sk.tsukuba.ac.jp

Abstract

Implications of recent development of sensitivity and uncertainty analyses, which may be relevant to applications in data mining are explored. Variance-based approach to complex models that involve both parametric (aleatory) and subjective (epistemic) uncertainty are discussed, and also a new stochastic (i.e., aleatory) sensitivity analysis methodology is outlined for neural computations.

1. INTRODUCTION

Sensitivity and uncertainty analyses are now widely recognized as an essential tool of analysis for model outputs of complex systems (e.g., the readers are referred to the proceedings of SAMO'95 and SAMO'98, and references therein). The sensitivity and uncertainty analyses are usually used to understand the behavior of a model, and the coherence between a model and an underlying system. In this study, we would like to concentrate on the sensitivity analysis that is relevant to data mining, i.e., a science of hypothetico-deductive analysis of models built on and/or derived from data. Hence, the sensitivity analysis here refers to a general assessment of how much a given model (numerical or otherwise, which is obtained from data mining tools and techniques) depends on each or some of its input parameters.

In general, the sensitivity analysis of complex models for decision analysis is computationally intensive because the analysis will require a large number of model evaluations, which involve two types of model uncertainty due to parametric (or aleatory) and subjective (or epistemic) uncertainties. Model output involves the subjective uncertainty attributable to our imperfect knowledge on the underlying system and the parametric uncertainty that is stochastic in nature and may not be reducible. In this sensitivity analysis study, we will mostly focus on the parametric uncertainty, behind which lies a variety of different problem settings encountered in practical situations.

To illustrate the sensitivity analysis for data mining applications is, however, a difficult task since its effective use has been hindered by a number of causes. A major difficulty may be an epistemological reason in that many different things are meant by sensitivity analysis in many different user communities. For a chemical engineer, for example, sensitivity analysis could be the process of moving and replacing components in the design of a chemical plant to investigate how a fault tree analysis for that plant would change. For a business analyst, it is most likely to be the derivative of the future stock price (i.e., model output) of a predictive model with respect to changing assumptions on economic scenarios (i.e., input parameters). For a knowledge engineer, the analysis may be related to the robustness and reliability of the data mining code with respect to different industry knowledge. There is no doubt that it will be extremely difficult to convince data mining practitioners of the benefit of new sensitivity analysis tools unless this epistemological barrier is overcome.

In most of the conventional data mining applications, the sensitivity analysis is attributed to exploring rather straightforward "what-if" questions. Sensitivity analysis is

complementary to a variety of techniques in data mining, and an ideal sensitivity analysis should be capable of global quantitative analysis. In fact, according to Saltelli (1998), it should

- apportion quantitatively the outcome uncertainty to the input factors (parameters)
- allow the factors to be uncertain over non negligible regions
- be effective independent of the model
- be computationally affordable.

Such tools and methods are currently available (e.g., SAMO'98). In the subsequent development, however, we will elaborate on the implications of the existence of such a methodology, rather than describing each method in detail.

Sensitivity analysis also plays a crucial role in model verification and validation. In data mining applications, the sensitivity analysis may be utilized to validate a model to be used for predictions of future behaviors, or behaviors under conditions not previously encountered. An epistemological approach to this may use the term corroboration (Oreskes *et al.* (1994)). A piece of evidence is said to corroborate a hypothesis (or model) when it does not contradict it. Then, the sensitivity analysis should be capable of corroborating not only a hypothesis obtained from traditional statistical analysis, but also new discoveries (i.e., models in the form of theories, rules, artificial neural networks, decision trees, etc.) derived from data mining techniques. In this paper, we would like to illustrate an epistemological foundation to the sensitivity analysis in data mining.

The organization of the paper is as follows: In section 2, we will briefly review the local and global sensitivity analysis methods and also related variance-based techniques. In Section 3, we will discuss two variance-based methods, i.e., Sobol' sensitivity indices (Sobol' (1990)) and Fourier Amplitude Sensitivity Test (FAST). In Section 4, we will outline the stochastic sensitivity analysis method for neural computation. The method belongs to the local sensitivity analysis approach and a numerical example is also included in Section 4. In the final section, conclusions will be summarized.

2. LOCAL AND GLOBAL SENSITIVITY ANALYSES AND VARIANCE-BASED METHODS

In general, there are two quite different approaches to the sensitivity analysis: i.e., the local sensitivity analysis and the global sensitivity analysis, respectively. In the local sensitivity analysis, the local response of model outputs, obtained by varying parameters

$$\frac{Var_X\{E(Y \mid X)\}}{Var(Y)} \tag{1}$$

where Y denotes the model output and X the input parameter, $E(Y|X)$ denotes the expectation of Y conditional on a fixed value of X, and Var_X denotes the variance taken over all possible values of X. The statistical quantity (1) is well-known in the Design Of Experiment (DOE) study (see Box, Hunter, and Hunter (1978)). The DOE analysis also implies the decomposition of the model response into terms of increasing order of correlated factor combinations (i.e., main effects, two-way and higher multi-way interactions). The similar decompositions are familiar in the DOE to derive information on the importance of the factors. In simulation this is known as "what-if" analysis, and DOE uses regression analysis, also called as ANalysis Of VAriance (ANOVA) including Generalized Least Squares (GLS) (Kleijnen (1987)(1995)). In ANOVA, the output variance is decomposed into partial variances of increasing order of dimensionality.

The Sobol' sensitivity indices, original extension of DOE by Sobol' (1990), are similar to FAST in the sense that the total variance of model output is assumed to be made up of correlation terms of increasing dimensionality. Hence, Sobol' indices may also provide a global quantitative sensitivity information. Both Sobol' and FAST methods are based on the (ANOVA-like) variance-based sensitivity measure (1), and useful to ascertain if a (small) subset of input variances may account for (most of) the output variance. FAST computes only the first-order terms (main effects) and, hence, is computationally appealing. Different from the FAST method, the Sobol' method evaluates each interaction term by computing a multi-dimensional integral via Monte Carlo techniques. It is further extended that the necessary computation involves only one Monte Carlo integration per each parameter (Homma and Saltelli (1996)).

Given the above discussion, in next section, we will briefly review the new development in variance-based method, i.e., the Sobol' indices and the FAST method. The variance-based method is superior to other methods in that it is essentially similar to the computation of main effects alone, whereas the input-output relationships (mappings) required in other methods, without any evident simplification, would tend to be NP-complete with computational complexity scaling as s^n where s denotes the number of sample values of the input in each of n-dimensional space. Note that substantial evidence exists to support the conjecture of the variance-based techniques that higher order interactions of the inputs are expected to have a negligible impact on the output variance. Although no claim is made that the variance-based method is optimal for general data mining applications, it has been successfully applied to several

one at a time with holding others fixed to nominal (reference) values, is investigated. There exists a well-defined sensitivity measure in the local sensitivity analysis, and the analysis involves an intensive computation of the sensitivity coefficients defined as first-order partial derivatives or functional derivatives, depending on whether the parameter is constant or function of some independent variables. The readers are referred to the work of Koda (1981)(1982)(1992) and also of Koda and co-workers (1979)(1982) for typical expositions. All the local analysis is run and corresponding sensitivity derivatives are evaluated only at the given central (reference) point, possibly normalized by the nominal values or by their standard deviations, in the input parameter space. Thus, the volume of the search region in the parameter space explored for the local analysis is nil.

The global sensitivity analysis, on the other hand, attempts to search a whole (finite or even infinite) parameter space, and the variation of the model output induced by an input parameter is evaluated globally, i.e., averaged over the variation over the sampled space of input parameters. A pragmatic, empirical approach has dominated the analysis due to the intrinsic difficulty in defining an effective sensitivity measure over a finite search region of the total space of parameter variations. In this approach, robust methods based on Monte Carlo regression and correlation analyses are often used (see Helton (1993) for an extensive review). Methods such as Standardized Regression Coefficients (SRC), Partial Correlation Coefficients (PCC), Standardized Rank Regression Coefficients (SRRC), Partial Rank Correlation Coefficients (PRCC), etc., are used in the approach. We may note that most of these methods are qualitative in nature than quantitative; for instance, SRC is quite model dependent and it gives information on the linear regression model that is used to describe the system, not on the system model itself.

A global, nonlinear sensitivity analysis method referred to as the Fourier Amplitude Sensitivity Test (FAST) is available. The original FAST method was developed in the 70's by Cukier and co-workers (1973)(1978), and further improved by Koda, McRae, and Seinfeld (1979). The FAST method explores the multi-dimensional space of input parameters by a single search curve that sweeps throughout the entire parameter space and thus avoids a multi-dimensional (Monte Carlo) integration over the same sample space. It gives a quantitative sensitivity measure defined by a fractional contribution to variance computed from the terms in the Fourier expansion of the model output. Thus, FAST is an effective global quantitative sensitivity analysis method.

The FAST method belongs to the so-called variance-based techniques, whose sensitivity measure can be expressed as (e.g., Iman and Hora (1990), McKay (1996))

nontrivial problems (e.g., SAMO'98).

3. SOBOL' SENSITIVITY INDICES AND FOURIER AMPLITUDE SENSITIVITY TEST (FAST)

In this section, we will briefly describe two variance-based sensitivity techniques, i.e., the Sobol' and FAST methods, which are both based on the ANOVA-like sensitivity measure (1). The Sobol' method will be introduced first since its development for computing higher order interaction terms is more natural than FAST.

3.1. Sobol' Sensitivity Indices

Let us consider, for the sake of simplicity, a model described by a transfer function $y=f(x)$, where y is a scalar output, and $x=(x_1,x_2,\Lambda\ ,x_n)$ is the vector of n-dimensional input parameters with each x_i normalized and varies from *0* to *1*. It should be noted that the explicit form of the scalar transfer function f may not be available and, more generally, output y be obtained by the prediction of a data mining code (i.e., neural network, decision or regression tree, etc.). Let us further assume that the input vector x is a stochastic vector characterized by a joint probability density function $p(x)$, which is assumed to be known. In practice, the parameters are affected by both stochastic (i.e., aleatory) and subjective (i.e., epistemic) uncertainty which reflects our imperfect knowledge of the underlying system. As a result, the output y becomes a random variable described by its own probability density function, which can be estimated independently.

In the present system model that involves both parametric and structural uncertainty, the primary analysis may be reduced to an expected value over stochastic uncertainty; for example, the r-th moment of y is computed from

$$< y^{(r)} >= \int_{K^n} f^r(x)p(x)dx \qquad (2)$$

where K^n denotes the n-dimensional unit cube of (normalized) input parameters. Note that the computation of (2) requires the evaluation of multi-dimensional integrals. Various pattern search procedures can be employed to allocate sampling points over the domain of the input parameters. In Monte Carlo analysis, for example, sample points are randomly selected via different sampling strategies, e.g., crude Monte Carlo, Latin Hypercube Sampling (LHS), etc., and low discrepancy sequence or quasi-random numbers may be utilized for efficient computation (Sobol' *et al.*(1992)).

The main idea in Sobol' approach is to decompose *f(x)* into summands of increasing order of dimensionality as follows:

$$f(x_1,x_2,\Lambda\ ,x_n)=f_0+\sum_{i=1}^{n}f_i(x_i)+\sum_{1\le i<j\le n}f_{ij}(x_i,x_j)+\Lambda+f_{1,2,\Lambda\ ,n}(x_1,x_2,\Lambda\ ,x_n) \quad (3)$$

where f_0 is a constant. Under assumptions described in Sobol' (1990), the sensitivity indices $S_{i_1,i_2,\Lambda\ ,i_s}$ $(1\le i_1<i_2<\Lambda<i_s\le n$ and $s=1,2,\Lambda\ ,n)$ can be introduced as

$$S_{i_1,i_2,\Lambda\ ,i_s}=D_{i_1,i_2,\Lambda\ ,i_s}/D \quad (4)$$

where D is the total variance of $f(x)$ defined by

$$D=\int_{K^n}f^2(x)dx-f_0^2\ . \quad (5)$$

Partial variances $D_{i_1,i_2,\Lambda\ ,i_s}$ are computed from each of the term in (3) as

$$D_{i_1,i_2,\Lambda\ ,i_s}=\int_0^1\int_0^1\Lambda\int_0^1 f^2_{i_1,i_2,\Lambda\ ,i_s}dx_{i_1}dx_{i_2}\Lambda\ dx_{i_s}\ . \quad (6)$$

Then, using the orthogonality of all summands in (3), it is straightforward to derive

$$D=\sum_{i=1}^{n}D_i+\sum_{1\le i<j\le n}D_{ij}+\Lambda+D_{1,2,\Lambda\ ,n} \quad (7)$$

or equivalently, dividing (7) by D and using (4), we have

$$\sum_{i=1}^{n}S_i+\sum_{1\le i<j\le n}S_{ij}+\Lambda+S_{1,2,\Lambda\ ,n}=1 \quad (8)$$

where, for instance, S_1 denotes the sensitivity measure corresponding to the main effect of input x_1 (i.e., the fractional contribution of x_1 to the variance of $f(x)$), S_{12} is the interaction effect due to x_1 and x_2, which cannot be explained by the sum of the individual effects of x_1 and x_2, and so on.

The applicability of Sobol' sensitivity indices (4) to a large class of data mining problems is dependent on the practical feasibility of multi-dimensional integration via Monte Carlo method. In general, separate Monte Carlo integration is needed to compute any effect, be it first or higher order, and the method may not be computationally effective. Homma and Saltelli (1996) suggested a technique to partition input parameters into two subsets, one containing a given parameter x_i alone, and its complementary x_{-i} containing all other parameters. In this case, the total variance D is given by

$$D = D_i + D_{-i} + D_{i,(-i)} \tag{9}$$

where $D_{i,(-i)}$ denotes the interaction effect between x_i and x_{-i}, and D_{-i} denotes the sum of all the D_{i_1,i_2,Λ,i_s} terms which do not include the index i. Then, using the relationship in (9), the new sensitivity measure S_i^T called the Total Sensitivity Index (TSI) can be introduced as

$$S_i^T = (D_i + D_{i,(-i)})/D = 1 - D_{-i}/D = 1 - S_{-i}. \tag{10}$$

It is important to note that $S_{-i} = D_{-i}/D$ in (10) can be computed by just one Monte Carlo integration for each corresponding parameter. Thus, TSI provides an effective sensitivity measure of the total effect of a parameter, including all the possible synergetic terms arising from the coupling of the parameter with all others. For the efficient computation of TSI, the bootstrap method (Efron and Tibshirani (1993)) may be used as resampling techniques in Monte Carlo integration.

3.2. Fourier Amplitude Sensitivity Test (FAST)

Using the Fourier analysis techniques, it is also possible to decompose the total variance of model output into partial variances similar to those given by the Sobol' sensitivity indices (4). In view of the computational complexity of the *n*-dimensional Fourier transformation, Cukier *et al.* (1973) proposed the Fourier analysis along a single (space-filling) search curve that sweeps throughout the entire parameter space K^n. The search curve is a one-dimensional manifold generated by a set of incommensurate frequencies $\{\omega_i : i=1,2,\ldots,n\}$ through periodic transformations $x_i = x_i(s) = G_i(\sin\omega_i s)$, where s is a scalar variable varying over the range $-\infty < s < +\infty$, and G_i are appropriate functions. Each (integer) frequency is associated with one of input parameters and the frequencies determine the overall length of search curve and the necessary number of trial solutions. As s varies, all the parameters simultaneously change values accordingly, and systematically explore their range of uncertainty. We can thus write as $f(x)=f(x(s))=f(s)$.

An advantage of FAST is that the evaluation of (first-order) sensitivity measures can be computed independently for each parameter using just one simulation because all the terms in a Fourier expansion are mutually orthogonal. By the ergodic Weyl's

theorem (e.g., Cukier *et al.* (1978)), the average central moments defined in (2) can be effectively computed by evaluating the model output along the search curve defined as a line integral:

$$< y^{(r)} >= \frac{1}{2\pi} \int_{-\pi}^{\pi} f^r(s) ds . \tag{11}$$

Corresponding to (5), the output variance of any *n*-dimensional input model is given by

$$D = < y^{(2)} > - < y^{(1)} >^2 . \tag{12}$$

We may expand the output $y=f(s)$ in a Fourier series:

$$y = f(s) = \sum_{-\infty}^{+\infty} \{A_j \cos js + B_j \sin js\} \tag{13}$$

where A_j and B_j are Fourier coefficients over a mono-dimensional domain of integer frequencies $j \in Z$, where Z indicates the set of all integer numbers. Then the power spectrum of the Fourier series expansion of $f(s)$ is defined as $\Lambda_j = A_j^2 + B_j^2$. By evaluating the spectrum for the fundamental (integer) frequency ω_i and associated harmonics, i.e., $p\omega_i$ for $p \in Z^0$, we can estimate $\sigma_{\omega_i}^2$, the contribution of the output variance arising from the uncertainty of *i*-th parameter as

$$\sigma_{\omega_i}^2 = \sum_{p \in Z^0} \Lambda_{p\omega_i} = 2 \sum_{p=1}^{+\infty} \Lambda_{p\omega_i} \tag{14}$$

where Z^0 indicates the set of all relative integer numbers except zero. By summing $\Lambda_j, j \in Z^0$, we can obtain a Fourier estimate of the total variance

$$D = \sum_{j \in Z^0} \Lambda_j = 2 \sum_{j=1}^{+\infty} \Lambda_j . \tag{15}$$

Note that (15) and (12) provide the equivalent quantity, and the ratio $\sigma_{\omega_i}^2 / D = S_i'$, denoted by S_i', is the basic sensitivity measure in FAST; i.e., its magnitude reflects the influence of *i*-th parameter has on the output, and it does not, in principle, depend on the choice of the set of frequencies used in the computation. Thus, FAST enables the decomposition of the output variance made on a mono-dimensional domain, whatever the dimension of the input space may be. It can be further shown that S_i' is essentially

equivalent to the Sobol' first-order sensitivity index S_i^T.

Koda, McRae, and Seinfeld (1979) improved the FAST method to utilize symmetry properties of the Fourier expansion (13) for efficient computation. Recently, Saltelli, Tarantola, and Chan (1997) extended FAST further to be able to cope with nonlinear and non-monotonic models, where regression-based approach inevitably fails. It may be concluded that both Sobol' indices and FAST are model independent and applicable to generic models utilized in data mining applications.

4. NEURAL COMPUTATION BASED ON STOCHASTIC SENSITIVITY ANALYSIS

In this section, we will briefly describe the recent development of a stochastic sensitivity analysis method for neural computation introduced by the present author (Koda (1995)(1997a)(1997b)). The method belongs to the local sensitivity analysis approach described in Section 2, in which an intensive computation of (first-order) sensitivity derivatives is crucial.

The noise-based, stochastic learning algorithm is called as the Subconscious Noise Reaction (SNR) and, in the discrete-time formulation, SNR is based on the following stochastic network model:

$$x_i(t+1) = \sum_j w_{ij}(t) S_j(t) + \xi_i(t+1), \qquad i,j=1,2,\cdots,n \tag{16}$$

where $x_i(t)$ denotes the internal state of the i-th unit, $w_{ij}(t)$ is the time-dependent connection weight between the i-th and j-th units, $S(t)$ denotes the signal, and $\xi(t+1)$ is the normalized Gaussian white noise. We specify the nonlinear signal (logistic) function as $S_i(t) = 1/[1+\exp\{-x_i(t)\}]$.

We consider the minimization of the scalar performance functional $J[x(T)]$ where T is the time of interest, which is usually some fixed (final) time $(T>t)$. Instead of the standard partial derivative formulation of gradient descent algorithm, a new functional derivative formulation is used. The main reason for this is that the gradient (i.e., sensitivity) computation involves multiple time variables, since time-dependent connection weights, $w_{ij}(t)$, can be correlated against the performance functional which is evaluated at a different point in time, T. Thus, the formulation of gradient descent algorithm in terms of the stochastic sensitivity analysis may lead to a novel theoretical

framework for designing stochastic learning algorithms.

Based on a local functional derivative sensitivity analysis, a relevant gradient descent method is formulated as follows:

$$w_{ij}^{\tau+1}(t) = w_{ij}^{\tau}(t) - \mu \frac{\delta J[x(T)]}{\delta w_{ij}^{\tau}(t)} \tag{17}$$

where the superscript τ denotes the index for the learning time and μ is the learning rate, which is usually a small positive number ($\mu > 0$). In (17), $\frac{\delta J[x(T)]}{\delta w_{ij}^{\tau}(t)}$ denotes the functional derivative (see, for example, Koda and Seinfeld (1982)).

4.1. SNR Learning Algorithm

Stochastic sensitivity problems for SNR can be solved by using the variational approach developed and typically demonstrated by the present author (e.g., Koda (1995)(1997a)(1997b)). By using the variational approach, the following sensitivity lemmas are straightforward results:

Lemma 1 (State Sensitivity): For the stochastic neural network defined by (16), the state functional derivative sensitivity coefficients are expressed as follows:

$$\frac{\delta x_i(T)}{\delta w_{jk}^{\tau}(t)} = \frac{\delta x_i(T)}{\delta \xi_j(t+1)} S_k(t) \tag{18}$$

for $t=0,1,2,\ldots,T-1$.

In the present study, as formulated in (2), we are interested in the expected value of $J[x(T)]$, i.e., $\langle J[x(T)] \rangle$. In order to evaluate $\langle J[x(T)] \rangle$, we have

Lemma 2 (Performance Sensitivity): For the stochastic neural network defined by (16), the performance functional derivative sensitivity coefficients are expressed as follows:

$$\frac{\delta J[x(T)]}{\delta w_{ij}^{\tau}(t)} = J[x(T)] \xi_i(t+1) S_j(t) \tag{19}$$

for $t=0,1,2,\ldots,T-1$.

The detailed proof of these lemmas is omitted here, but it is important to note that the Lemma 2 is valid in a statistical sense, i.e., as a mean or average. The functional derivative sensitivities (18) and (19) are obtained as a result of the local sensitivity analysis described in Section 2 (e.g., Koda (1997b)).

Using these two lemmas in the functional derivative formulation of the gradient descent method, i.e., inserting (19) into (17), we can obtain the SNR learning algorithm.

Theorem (Subconscious Noise Reaction): For the stochastic neural network defined by (16), the learning algorithm

$$w_{ij}^{\tau+1}(t) = w_{ij}^{\tau}(t) - \mu J[x(T)]\xi_i(t+1)S_j(t) \tag{20}$$

guarantees a monotonic decrease in the average value of the performance functional, $\langle J[x(T)] \rangle$.

It should be noticed that the present SNR algorithm, i.e., (20), leads to a non-symmetric learning regime. Furthermore, the introduction of ubiquitous noise term $\xi(t+1)$ into the learning process may help avoid absorption into local minima.

A central role in the SNR learning algorithm described above is played by ubiquitous noise. In general, such noise in each individual unit is small and the resulting variation of the performance functional may not be noticeably large. However, (20) indicates that the iterative modification of connection weights can be performed via computation of a stochastic correlation of ubiquitous noise $\xi(t+1)$ and the performance functional $J[x(T)]$. This computation can be executed locally without any synchronous transmission of information backward along network connections, e.g. back-propagation.

4.2. Numerical Example

The convergence behavior of the SNR algorithm is studied and compared with that of the standard back-propagation-through-time (BTT) algorithm. In order to adapt the SNR algorithm to layered networks, we interpret $x_i(t)$ as denoting the internal state at the i-th unit in the t-th network layer: This is the spatially unfolded network model. Hence, hereafter, t represents an index for the t-th network layer; for example, $t=T$ denotes the output layer. Similarly, $w_{ij}(t)$ denotes the connection weights between the j-th unit in the t-th layer and the i-th unit in the $(t+1)$-th layer. A three layered network $N(7;n;1)$ is used for the numerical study (Fig. 1). The network has six input units in the zeroth layer, $n-1$ hidden units in the first layer, and one output unit in the second layer. In addition, auxiliary (threshold) units in the zeroth and first layers, which always hold the constant value ($=1.0$), give threshold values to the units in the first and second layers, respectively. The value of n ranges from 2 to 12 in this study. We specified that the network performance function is a summation of the square of the difference (error) between the network and desired outputs.

For learning patterns (i.e., sets of input and desired output vectors), we selected a

mirror symmetry (SYM) pattern. The SYM pattern requires at least three hidden units for learning in order to detect mirror symmetry in an input vector of six bits; when an input vector is symmetric—for example, *(1,0,0,0,0,1), (0,1,0,0,1,0), (0,0,1,1,0,0),* etc.—the output is *1*; otherwise, the output is *0*.

We recall that no noise was injected in the case of BTT simulation, while, in that of SNR, Gaussian white noise was injected. Simulation results (Figs. 2 and 3) show that the SNR's convergence behavior is slower than that of the BTT. The BTT always converges regardless of the complexity of the network architecture (i.e., the number of hidden units), when the number of hidden units n varies over values between 4 and 12. On the other hand, the SNR's convergence is sensitive to the number of hidden units and optimal performance may be achieved by using a network architecture with $n=6$ or 7 (Fig. 2.). This result seems to be natural, since the SNR is running in a stochastic environment where even a steady state should be interpreted as a stochastic equilibrium; i.e., each network state vibrates randomly around the equilibrium state.

If a correct network architecture is selected, two learning algorithms may have basically the same capabilities (in a stochastic sense). Thus, we conclude that the (local) stochastic sensitivity analysis may have a potential to be applied to the neural computation. To overcome the slower convergence of SNR and exploit its algorithmic simplicity will be the subject of future study.

5. CONCLUSIONS

A model for complex decision analysis (i.e., data mining) involves two types of uncertainty. It involves parametric or stochastic uncertainty, which arises because the model is built on the data obtained from the underlying system that behaves in many different ways. Subjective uncertainty, on the other hand, arises from a lack of knowledge on the part of the analysts carrying out data mining with respect to the underlying system itself or appropriate theories to be used in assessing the model behavior. Thus, stochastic uncertainty is a property of the system model under study while subjective uncertainty is a property of the analysts carrying out data mining.

In this paper, the stochastic sensitivity and uncertainty analyses, which are complementary to a variety of data mining modeling techniques, are reviewed. Variance-based sensitivity analysis methods are introduced and it is concluded that the Sobol' indices and FAST may provide a (truly) global quantitative measure for data mining applications. Variance-based method is superior in that it is model independent and applicable to generic models utilized in data mining. It is also numerically superior

to other methods. In addition, a new comprehensive theoretical framework is illustrated for the (local) stochastic sensitivity analysis, and its applications to noise-based learning outlined for potential neural computation.

Little, however, is touched on the subjective (i.e., epistemic) uncertainty. This is due to the lack of availability of effective analysis methods to treat such uncertainty. Because the uncertainty characterization of any decision analysis models has the prescribed dual nature, data mining activities must be planned, designed, and implemented very carefully. Variance-based sensitivity analysis methods in conjunction with techniques in design of experiment (DOE) may be used to design, implement, and validate models in order to ascertain informed accuracy and draw intelligent decisions from data mining.

Acknowledgement

The author would like to thank H. Okano of IBM Tokyo Research Laboratory for assistance in numerical computation. He is also grateful to A. Saltelli of Environment Institute, EC Joint Research Center, for inviting him to SAMO'95 and SAMO'98 and updating him of recent development in sensitivity analysis.

References

Box, G.E.P., Hunter, W.G., and Hunter, J.S. (1978) Statistics for Experimenters. An Introduction to Design, Data Analysis and Model Building. Wiley & Sons.

Cukier, R.I., Fortine, C.M., Shuler, K.E., Petshek, A.G., and Schaibly, J.H. (1973) "Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I Theory." *J. Chem Physics*, **59**, 3873-3878.

Cukier, R.I., Levine, H.B., and Shuler, K.E. (1978) "Nonlinear sensitivity analysis of multi-parameter model systems." *J. Comp. Physics*, **26**, 1-42.

Efron, B., and Tibshirani, R. (1993) An Introduction to the Bootstrap. Chapman & Hall.

Helton, J.C. (1993) "Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal." *Reliability Eng. and Syst. Safety*, **42**, 327-367.

Homma, T., and Saltelli, A. (1996) "Importance measures in global sensitivity analysis of nonlinear models." *Reliability Eng. and Syst. Safety*, **52**, 1-17.

Iman, R.L., and Hora, S.C. (1990) "A robust measure of uncertainty importance for use in fault tree system analysis." *Risk Analysis*, **10**, 401-406.

Kleijnen, J.P.C. (1987) Statistical Tools for Simulation Practitioners. Marcel Dekker.

Kleijnen, J.P.C. (1995) "Verification and validation of simulation models." *European J. Oper. Research*, **82**, 145-162.

Koda, M. (1981) "Sensitivity analysis of time-delay systems." *Int. J. Syst. Sci.*, **12**, 1389-1997.

Koda, M. (1982) "Sensitivity analysis of atmospheric diffusion euation." *Atmospheric Environment*, **16**, 2595-2601.

Koda, M. (1992) "Sensitivity analysis of stochastic dynamical systems." *Int. J. Syst. Sci.*, **23**, 2187-2195.

Koda, M. (1995) "Stochastic sensitivity analysis method for neural network learning." *Int. J. Syst. Sci.*, **26**, 703-711.

Koda, M. (1997a) "Neural network learning based on stochastic sensitivity analysis." *IEEE Trans. Syst., Man, and Cybern. Part B: Cybernetics*, **27**, 132-135.

Koda, M. (1997b) "Stochastic sensitivity analysis and Langevin simulation for neural network learning." *Reliability Eng. and Syst. Safety*, **57**, 71-78.

Koda, M., and Seinfeld, J.H. (1982) "Sensitivity analysis of distributed parameter systems." *IEEE Trans. Aut. Control*, **AC-27**, 951-955.

Koda, M., Dogru, A., and Seinfeld, J.H. (1979) "Sensitivity analysis of partial differential equations with application to reaction and diffusion processes." *J. Comp. Physics*, **30**, 259-282.

Koda, M., McRae, G.J., and Seinfeld, J.H. (1979) "Automatic sensitivity analysis of kinetic mechanism." *Int. J. Chemical Kinetics*, **11**, 427-444.

McKay, M.D. (1996) Variance-Based Methods for Assessing Uncertainty Importance in NUREG-1150 Analyses. Los Alamos National Laboratory.

Oreskes, N., Shrader-Frechette, K., and Belitz, K, (1994) "Verification, validation, and confirmation of numerical models in the earth sciences." *Science*, **263**, 641-646.

Saltelli, A. (1998) "Are variance-based sensitivity analysis methods healthy for your models?" Proc. SAMO'98, 259-263.

Saltelli, A., Tarantola, S., and Chan, K.P.-S. (1997) Sensitivity Analysis of Model Output: an Imrovement of the FAST Method. Report EUR 17338, Environment Institute, EC Joint Research Center.

SAMO'95 (Sensitivity Analysis of Model Output) (1995) Proc. Int. Symposium on Sensitivity Analysis of Model Output, European Commission, Joint Research Center, Beligirate, Italy.

SAMO'98 (Sensitivity Analysis of Model Output) (1998) Proc. Second Int. Symp. on Sensitivity Analysis of Model Output, European Commission, Joint Research Center and University of Venice, Venice, Italy.

Sobol', I.M. (1990) "Sensitivity estimates for nonlinear mathematical models." *Matematicheskoe Modelironanie*, **2**, 113-118 (in Russian), translated in *Mathematical Modeling and Computational Experiments*, **1**, 407-414. 1993.

Sobol', I.M., Turchaninov, V.I., Levitan, Y.L., and Shukhman, B.V. (1992) Quasirandom Sequence Generators. Keldysh Institute of Applied Mathematics, Russian Academy of Sciences.

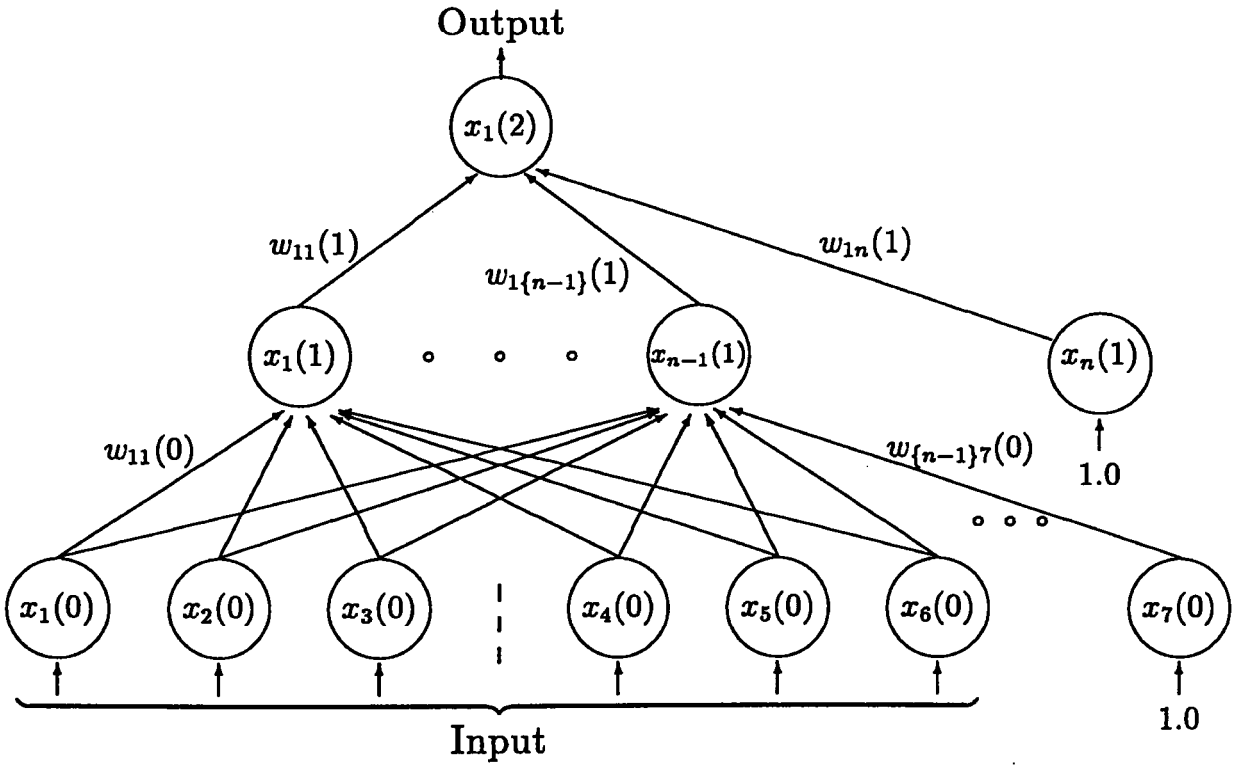

Figure 1. N(7; n; 1) layered network

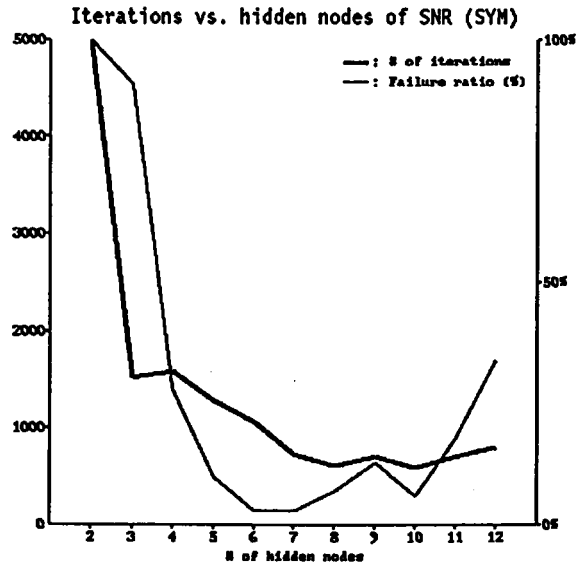

Figure 2. Average convergence time and failure ratio of the SNR algorithm for a mirror symmetry pattern, plotted against the numbers of hidden units

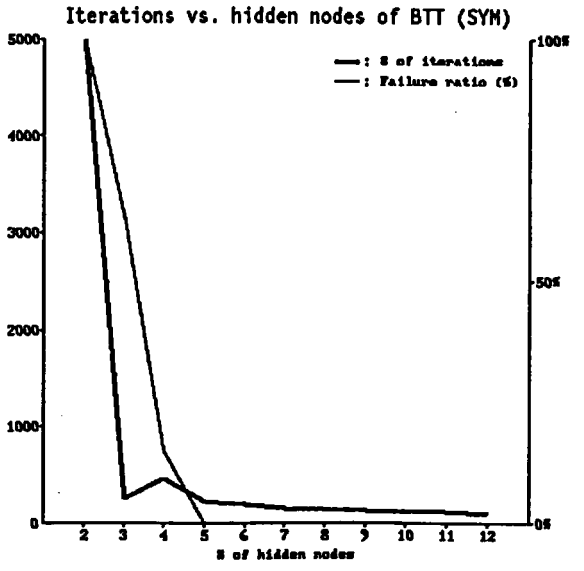

Figure 3. Average convergence time and failure ratio of the BTT algorithm for a mirror symmetry pattern, plotted against the numbers of hidden units