Advances in Smart Systems Research ISSN 2050-8662
Vol. 4. No. 1 : pp.7-16 : isrp15-003

A Novel Robust Algorithm for Information Security Risk Evaluation

Zne-Jung Lee[1*], Chou-Yuan Lee[2]

[1]Dept. of Information Management, Huafan University, Taiwan
johnlee@cc.hfu.edu.tw

[2]Dept. of Information Management, Lan Yang Institute of Technology, Taiwan
yuan@mail.fit.edu.tw

Abstract

As computer becomes popular and internet advances rapidly, information systems are used extensively in organizations. Various information systems such as attendance systems and accounting systems have already replaced manual operations. In such a drastic change, information security risk management issue encountered by organizations becomes increasingly significant. Risk evaluation is the core of information security risk management, and organizations use it to diminish the risks within information systems. Risk evaluation mainly focuses on the assessments of confidentiality, integrity and availability. Moreover, vulnerability of information systems and threats to the outside are also included in the scope of consideration. In this paper, a novel robust algorithm for information security risk evaluation is proposed. In the proposed algorithm, a modified TSK fuzzy model with the robust loss function is used to diminish the influence of noises or outlier first. Furthermore, the I-index is used to determine the best number of fuzzy rules in the proposed algorithm. From simulation results, the proposed algorithm outperforms other existing approaches.

Keywords: Robust Algorithm, Fuzzy Model, Risk Evaluation, Information Security Risk Management

1. Introduction

Along with the development of information systems, information security issues have attracted much attention from information systems researchers. The information security problems faced by organizations are becoming more and more serious. A good information security risk management plays an important role to protect information systems. Furthermore, the core of information security risk management is risk evaluation [1]. Risk evaluation ascertains the threat and vulnerability associated with assets. The threat means the likelihood of occurrence and the vulnerability represents the level of impact [2]. Organizations can use risk evaluation to gain an in-depth understanding the risks within an information system, and provide sufficient

controls for decision makers to reduce these risks [3-5]. Recently, information security risk evaluation is being developed by qualitative and quantitative methods. Qualitative risk evaluation methods include factor analysis, logical analysis, historical comparative method, Delphi method, and Analytic Hierarchy Process (AHP). The major disadvantages of qualitative risk evaluation methods are rely heavily on subjective, because they are based on judgment, intuition, and experiences [2,6,7]. Quantitative risk evaluation methods use statistics data to build models. Typical methods of quantitative risk evaluation include cluster analysis, time series model, regression model, and decision tree [2,6]. Above mentioned methods have some drawbacks, because they are too cumbersome to implement [1,7]. Recently, some researchers have done on fuzzy methods to ameliorate the subjective nature of quantitative risk evaluation methods [1,6-9]. Because information security risk evaluation has the characteristic of nonlinearity and subjectivity, it still needs to diminish the influence of noises or outlier and automatically decide the number of fuzzy rules for fuzzy methods to conduct information security risk evaluation [7]. In this paper, a novel robust algorithm for information security risk evaluation is proposed. In the proposed algorithm, a modified TSK fuzzy model with the robust loss function is used to diminish the influence of noises or outlier first. Moreover, the *I*-index is used to determine the best number of fuzzy rules in the proposed algorithm.

This paper is outlined as follows. Section 2 first describes the risk evaluation of information security, and then introduces the concept of Takagi, Sugeno and Kang (TSK) fuzzy model. The proposed algorithm is introduced in Section 3. Section 4 highlights the simulation results of the proposed algorithm. From simulation results, it is shown that the proposed algorithm outperforms other existing approaches. Concluding remarks are presented in section 5.

2. The Concept of Information Security Risk Evaluation and TSK Fuzzy Model

The proposed robust algorithm is based on TSK fuzzy model for evaluating the information security risk evaluation. In this section, the basic concepts of information security risk evaluation and TSK fuzzy model are introduced.

2.1. The Introduction of Information Security Risk Evaluation

Information is a type of asset which is the same as other important business assets. Because it is essential to the operation of an organization, it needs to be properly protected. This is especially important as operation environments become increasingly interconnected. Information security is protecting information from a variety of threats to ensure continuity of operations, minimize operations risk, earn a substantial return on investment and win greater business opportunities [1,2]. Information security risk evaluation is a basis for the development of safety measures to ensure information security [2]. Information security risk evaluation is related to the asset, threat, and vulnerability [1-5]. The

asset is anything that has value to the organization. The threat means the damages to the organizations underlying harmful event, and it is related to the likelihood of occurrence (LOO). The vulnerability is the weakness that the asset could be used, and it means the level of impact (LOI). The valuation of asset (VOA) is the individual evaluation and grading of each information asset based on confidentiality (C), integrity (I) and availability (A) attributes in ISO/IEC 27001. The attribute of confidentiality and its value is shown in Table 1. The attribute of integrity and its value is shown in Table 2. The attribute of availability and its value is shown in Table 3. Traditionally, the valuation of asset is calculated as Eq.(1).

$$\mathbf{VOA} - Max\ (\mathbf{C}, \mathbf{I}, \mathbf{A}) \tag{1}$$

After the valuation of asset is identified, the information asset risk can be calculated. Risk evaluation (RE) is done based on the valuation of asset, likelihood of occurrence and level of impact [1]. The information security risk evaluation is formulated as Eq.(2).

$$\mathbf{RE} - f(\mathbf{VOA}, \mathbf{LOO}, \mathbf{LOI}) \tag{2}$$

The risk evaluation consisted mainly of information value of asset, level of threat and level of vulnerability grading. These input variables are interrelated with VOA, LOO and LOI, and the output variable is risk evaluation. In this study, the VOA is obtained from Eq. (1). The attribute of LOO is considered as Table 4. The attribute of LOI is considered as Table 5. There are 5 attributes for LOO, and 16 attributes for LOI. Each attribute of LOO and LOI could be with noise or outliers, and its value is a real variable between 1 and 4.

2.2. The Concept of TSK Model

The well-known TSK fuzzy model is one of the most efficient fuzzy models [10]. It is to decompose the input space into fuzzy regions and to approximate the input-output relations as the expansion of piecewise linear partition [11]. Typically, a TSK fuzzy model consists of IF-THEN rules. The form of the i^{th} rule is represented as the following form:

$$\mathbf{R^i}: \text{If } x_1 \text{ is } A_1^i(\theta_1^i) \text{ and } x_2 \text{ is } A_2^i(\theta_2^i), \cdots, x_n \text{ is } A_n^i(\theta_n^i) \text{ then } y^i = \mathbf{1} + a_1^i x_1 + \cdots + a_n^i x_n \tag{3}$$

For $i = 1,2,\cdots,\mathbf{C}$ where C is the number of rules, $x_j (j = 1,2,\cdots,n)$ is the input, $A_j^i(\theta_j^i)$ is the fuzzy set of the ith rule for x_j with the adjustable parameter set θ_j^i, $\mathbf{y^i}$ is the output of the fuzzy rule $\mathbf{R^i}$, and $\overline{\alpha^i} - \left(1, a_1^i, \cdots, a_n^i\right)$ is the parameter set in the consequent parts. The predicted output of the fuzzy model is inferred as

$$y = \frac{\sum_{i=1}^{c} y^i w^i}{\sum_{i=1}^{c} w^i} \tag{4}$$

where $w^i = \min_{j=1,2,\cdots,n} A_j^i\left(\theta_j^i; x_j\right)$ is the ith rule's firing strength. In Eq.(3), it is necessary to identify the parameter set of the premise parts (i.e. θ_j^i) and consequent parts (i.e. a^i) of TSK fuzzy model. Furthermore, the number of rules must be also determined.

3. The Proposed Algorithm

The proposed algorithm is to obtain a function $\mathbf{f}$ from a set of observations where $\{(x^{\rightarrow}(1), y_1 1), (x^{\rightarrow}(2), y_1 2), \cdots, (x^{\rightarrow}(N), y_1 N)\}$ with $\vec{x}(i) \in R^m$ and $\vec{x}(i) = [x_1(i), x_2(i), \cdots, x_3(i)]$ is the i^{th} input vector. Let e_{ij} be the error between the j^{th} desired output and the output of the i^{th} rule with the j^{th} input data. It is defined as follow.

$$e_{ij} = y_j - f_i\left(\vec{x}(j), \vec{a}^i\right), \; i=1,2,\cdots,C \text{ and } j=1,2,\cdots,N \tag{5}$$

Where y_j is the j^{th} desired output, and C and N are the numbers of fuzzy rules and of the training data, respectively. The objective function is defined as

$$J = \sum_{i-1}^{c} \sum_{j-1}^{N} u_{ij}^z \rho_i\left(e_{ij}^z\right) - \sum_{i-1}^{c} \left[\sum_{j-1}^{N} w_{ij} u_{ij}\right] \tag{6}$$

subject to:

$$\sum_{i=1}^{c} u_{ij} = 1, \quad \text{for} \quad 1 \le j \le N \tag{7}$$

Where u_{ij} is the firing strength of the i^{th} rule for the j^{th} training pattern, $\rho_i(\cdot)$ is a robust loss function associated with cluster i, and w_{ij} is the weight function and obtained as $w_{ij} = \partial\rho\left(e_{ij}^z\right) \Big/ \partial e_{ij}^z$. In the proposed algorithm, the tanh-estimator is used as the robust loss function of $\rho_i(\mathbf{Y})$ and is defined as [12,13].

$$\rho_1 i.(Y) - \blacksquare(\quad 1/z\, \gamma^1 z \quad if\ u < |\gamma| < a, \omega 1/z\, r^1 z + c_1 1/c_1 z \ \text{In}\ [\cosh [(c_1 2\,(b-a))] /c\text{Osh}\ [(c$$

(8)

Where $\mathbf{a}$ and b are time-dependent cutoff points, and two constants $\mathbf{c_1}$ and $\mathbf{c_2}$ are set as $\mathbf{c_1 = 1.73}$ and $\mathbf{c_2 = 0.93}$ [12,13]. Let the residual of the N number of training data be sorted, i.e.,

$$|r|_1 \le |r|_2 \le \cdots \le |r|_{(1-\sigma)N} \le \cdots \le |r|_N \tag{9}$$

Where $|r|_i$ is the absolute value of i^{th} error, and σ is an upper bound of the percentage of outliers to be tolerated. Let training data include at least (1-σ)N good data, and $\mathbf{a}$ and b are set as $\mathbf{a} = |\gamma|_{(1-\sigma)N}$ and $\mathbf{b = 2a}$. To deal with outliers, the derivative of the robust loss function is defined as follow [13]:

$$\varphi_i(Y) - \begin{cases} 1 & if \ \left|\frac{r}{\hat{s}}\right| \le \tilde{c_1}, \\ \dfrac{\tilde{c_2} - \left|r/_{\hat{s}}\right|}{\tilde{c_2} \quad \tilde{c_1}} & if \ \tilde{c_1} \le \left|\frac{r}{\hat{s}}\right| < \tilde{c_2}, \\ 1e^{-4} & otherwise. \end{cases} \tag{10}$$

Where $\hat{\mathbf{s}} = \dfrac{\mathbf{IQR}}{\mathbf{1}}.349$ is a robust estimate that is derived from the Gaussian distribution, $\tilde{\mathbf{c_1}}$ is set as 2.5, and $\tilde{\mathbf{c_2}}$ is set as 3. The interquartile range (IQR) is the difference between the 75th percentile and the 25th percentile [14]. To minimize J in Eq. (6) subject to Eq. (7), the Lagrange multiplier method is applied. The Lagrange function is defined as

$$L = \sum_{i=1}^{c}\sum_{j=1}^{N} u_{ij}^{2}\rho_i(e_{ij}^{2}) \quad \sum_{i=1}^{c}\left[\sum_{j=1}^{N} w_{ij}u_{ij}\right] \quad \sum_{j=1}^{N}\lambda_j\left(\sum_{i=1}^{c} u_{ij} - 1\right) \tag{11}$$

For minimizing J, the parameter vector $\vec{a}^{i}$ for the consequent part of the i^{th} rule is obtained as:

$$\vec{a}^{i} = \left[X^{T}D_{i}X\right]^{-1} X^{T}D_{i}Y, i = 1,2,\cdots C \tag{12}$$

Where $X \subset R^{N\times(n+1)}$ is matrix with $\vec{x}(k)$as its $(k+1)^{th}$ row and the elements in the first row are all 1, $Y \in R^{N}$is a vector with $y_k$as its kth element and $D_i \in R^{N\times N}$is a diagonal matrix with $u_{ij}^{2}w_{ik}$ as its k^{th} diagonal element. Let $\rho(e_{ij}^{2}) = e_{ij}^{2}$ be defined, then it is easy to derive $w_{ij} = 1$ and D_i is a diagonal matrix with u_{ik}^{2}. The values of μ_{ik} are set by using subtractive clustering [15,16]. Subtractive clustering, a modified approach of Mountain Method, is used to partition the input space [16]. It is a density-based algorithm which reduces the number of training data based on the density of surrounding data points.

In the proposed algorithm, Gaussian membership functions are used in the premise parts, (i.e.), $A_j^{i}\left(\Omega_j^{i}, x_j(k)\right) = \exp\left\{-\frac{\left(x_j(k) - \Omega_{j1}^{i}\right)^{2}}{2\left(\Omega_{j2}^{i}\right)^{2}}\right\}$ where Ω_{j1}^{i} and Ω_{j2}^{i} are two adjustable parameters of the j^{th} membership function of the i^{th} fuzzy rules. Both parameters can be obtained from u_{ij} as follow:

$$\Omega_{j1}^{i} = \frac{\sum_{k=1}^{N}(u_{ik})^{2}x_j(k)}{\sum_{k=1}^{N}(u_{ik})^{2}}, \quad \Omega_{j2}^{i} = \sqrt{\frac{\sum_{k=1}^{N}(u_{ik})^{2}\left(x_j(k) - \Omega_{j1}^{i}\right)^{2}}{\sum_{k=1}^{N}(u_{ik})^{2}}} \tag{13}$$

In the proposed algorithm, the *I*-index is used to determine the optimal number of fuzzy rules. The *I*-index proposed by Maulik and Bandyopahhyay is defined as follow [15]:

$$I(C) = \left(\frac{1}{C} \times \frac{E_1}{E_C} \times D_C\right)^{2} \tag{14}$$

Where:

$$E_1C = \sum_{k=1}^{N}\sum_{i=1}^{C} \text{EMBED Equation.3 } \|x_k - z_i\|,$$
$$D_1C = \max_{i,k} \text{EMBED Equation.3 } \|z_i - z_k\| \tag{15}$$

In Eq.(14), $\vec{\Omega}^{i} = \left(\Omega_{11}^{i}, \Omega_{21}^{i}, \cdots, \Omega_{m1}^{i}\right)$ and the index $I(C)$ is a composition of three factors, namely, $\frac{1}{C}$, $\frac{E_1}{E_C}$, and D_C. The first factor ($\frac{1}{C}$) will reduce the values of index $I(C)$ as *C* is increased. The second factor ($\frac{E_1}{E_C}$) consists of the ratio of E_1, which is constant for a given data set, and E_C, which decreases with increase in *C*. Hence, because of this term, index *I*(C) increases as E_C decreases. Finally, the third factor (D_C) will increase with the value of *C*. It is noted that this value is upper bounded by the maximum separation between two points in the data set [15]. Thus, the value of *C* for which $I(C)$ is maximized when considering the best

number of clusters over $C = 2,3,\cdots,N$.

4. Simulation Results

The risk evaluation consisted mainly of information value of asset, level of threat and level of vulnerability grading. These input variables are interrelated with VOA, LOO and LOI, and the output variable is risk evaluation. In this study, the VOA is obtained from Eq. (1). The attribute of LOO is considered as Table 4. The attribute of LOI is considered as Table 5. There are 5 attributes for LOO, and 16 attributes for LOI. Each attribute of LOO and LOI could be with noise or outliers, and its value is a real variable between 1 and 4.

In this paper, the proposed algorithm is applied to calculate the risk evaluation. For fair comparisons, three learning algorithms of the robust fuzzy regression agglomeration (RFRA), fuzzy C-regression model (FCRM), and self-constructing neural fuzzy inference network (SONFIN) are implemented in this paper. RFRA is an algorithm for function approximation with outliers [26]. In the learning algorithms, FCRM is a fuzzy approach with general purposes and SONFIN is one kind of neural network/fuzzy modeling without specially designed for dealing with noise and outliers [27,28]. The root mean square error (RMSE) is used as the fitness function and can be calculated as

$$RMSE = \sqrt{\frac{\sum_{k=1}^{N} (y_k - \hat{y}_k)^2}{N}} \tag{16}$$

Where N is the number of training data, y_k is the kth desired output, and $\hat{y}_k$ is the output for the k^{th} training data. There are 60 input-output training data and 89 input-output data in this study. The simulation results are compared within above learning algorithms and shown in Table 6. From Table 6, The RMSE of risk evaluation for the proposed algorithm is 0.081. The proposed algorithm has the best value of RMSE among these compared approaches.

5. Conclusion

In this paper, a novel robust algorithm for information security risk evaluation is proposed. In the proposed algorithm, a modified TSK fuzzy model with the robust loss function is used to diminish the influence of noises or outlier first. Furthermore, the I-index is used to determine the optimal number of fuzzy rules in the proposed algorithm. From simulation results, the proposed algorithm outperforms other existing approaches. For SONFIN and FCRM, cannot ameliorate the performance of noise or outliers. For RFRA, the effect of robust is still not good enough. The proposed algorithm can simultaneously identify fuzzy subsets in the premise parts and parameters in the consequent parts, and it also can have robust learning effects when noise or outliers exist. From simulation results, it is shown that the proposed approach can find the best value of RMSE

for information security risk evaluation.

6. References

[1] Gao, G. H., Li, X. Y., Zhang, B. J., and Xiao, W. X. Information security risk assessment based on information measure and fuzzy clustering. Journal of Software, Vol.6, No.11, pp. 2159-2166, (2011)

[2] Fu, S., and Xiao Y. An effective process of information security risk assessment, 3rd International Conference on Computer and Automation Engineering (ICCAE 2011), pp. 124-128, (2011)

[3] Zhiwei, Y., and Zhongyuan, J. A survey on the evolution of risk evaluation for information systems security. Energy Procedia, Vol. 17, pp. 1288-1294. (2012)

[4] Feng, D. G., Zhang, Y., and Zhang, Y. Q. Survey of information security risk assessment. Journal-China Institute of Communications, Vol. 25, No. 7, pp. 10-18. (2004)

[5] Yang. Y., and Yao, S. Z. Risk assessment method of information security based on threat analysis. Computer Engineering and Applications, Vol. 45, No.3, pp. 94-96. (2009)

[6] Liu, F., Dai, K., Wang, Z. Y., and CAI, Z. P. Research on the technology of quantitative security evaluation based on fuzzy number arithmetic operations [J]. Fuzzy Systems and Mathematics, Vol. 4, No. 20. (2004)

[7] Lo, C. C., and Chen, W. J. A hybrid information security risk assessment procedure considering interdependences between controls. Expert Systems with Applications, Vol. 39, No. 1, pp. 247-257. (2012)

[8] Liu, F., Dai, K., Wang, Z., and Ma, J. Research on fuzzy group decision making in security risk assessment. In Networking-ICN 2005, pp. 1114-1121. Springer Berlin Heidelberg. (2005)

[9] Wang, P., Chao, K. M., Lo, C. C., Huang, C. L., and Younas, M. A fuzzy outranking approach in risk analysis of web service security. Cluster Computing, Vol. 10, No. 1, pp. 47-55. (2007)

[10] Takagi, T., and Sugeno, M. Fuzzy identification of systems and its applications to modeling and control. IEEE Transactions on Systems, Man and Cybernetics, Vol. 1, pp. 116-132. (1985)

[11] Ying, K. C., Lin, S. W., Lee, Z. J., and Lee, I. A novel function approximation based on robust fuzzy regression algorithm model and particle swarm optimization. Applied Soft Computing, Vol. 11, No. 2, pp. 1820-1826. (2011)

[12] Chen, D. S. and Jain, R. C. A robust back-propagation learning algorithm for function approximation, IEEE Trans. on Neural Networks, Vol. 5, No. 3, pp. 467-479. (1994)

[13] Chuang, C. C., and Lee, Z. J. Hybrid robust support vector machines for regression with outliers, Applied Soft Computing, Vol. 11, No.1, pp. 64-72. (2011)

[14] Suykens, J. A. K., Brabanter, J. De, Lukas, L., and Vandewalle, J. Weighted least squares support vector machines: robustness and sparse approximation, Neurocomputing, Special issue on fundamental and information processing aspects of neurocomputing, Vol. 48, No.1, pp. 85-105. (2002)

[15] Maulik, U., and Bandyopadhyay, S. Performance evaluation of some clustering algorithms and validity indices, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, No. 12, pp. 1650-1654. (2002)

[16] Rizman Žalik, K. Cluster validity index for estimation of fuzzy clusters of different sizes and densities. Pattern Recognition, Vol. 43, No. 10, pp. 3374-3390. (2010)

[17] Wu, K. L., and Yang, M. S. A cluster validity index for fuzzy clustering. Pattern Recognition Letters, Vol. 26, No. 9, pp. 1275-1291. (2005)

[18] Kim, D. W., Lee, K., Lee, D., and Lee, K. H. A kernel-based subtractive clustering method. Pattern Recognition Letters, Vol. 26, No. 7, pp. 879-891. (2005)

Acknowledgment

The authors would like to thank the National Science Council of the Republic of China, Taiwan for financially supporting this research under Contract No. MOST 103-2221-E-211-009 and NSC 102-2632-E-211-001-MY3.

Table 1: The attribute of confidentiality and its value

Item	Attribute	Value
1	General: information assets without the requirement of confidentiality	1
2	Restricted: information assets containing sensitive information, but there is no requirement of confidentiality, and limited use of internal staff	2
3	Sensitive: information assets within the relevant department staff in limited use	3
4	Confidential: information assets within the information, including regulated by relevant laws or regulations of confidential information	4

Table 2: The attribute of integrity and its value

Item	Attribute	Value
1	The integrity of information assets requires very low	1
2	Information assets with integrity requirements, but will not harm the integrity of the destruction and then	2
3	Information assets with integrity requirements, will be destroyed because of the integrity and cause harm.	3
4	Information assets with integrity requirements, will be destroyed because of the integrity of business interruption	4

Table 3: The attribute of availability and its value

Item	Attribute	Value
1	Failure of information assets to allow more than 3 working days, and does not require immediate repair or to find alternatives	1
2	Failure of information assets to allow more than 8 hours of work, 3 days following, do not need immediate repair or to find alternatives	2
3	Failure of information assets to allow more than 4 hours of work, 8 hours of work less, do not need immediate repair or to find alternatives	3
4	Failure of information assets to allow 4 working hours, do not need immediately repair or to find alternatives	4

Table 4: The attribute of LOO

Item	Attribute
1	Maintenance Error
2	Hardware failures
3	Theft
4	Misuse of resources
5	Operational staff error

Table 5: The attribute of LOI

Item	Attribute
1	Incorrect use of software and hardware
2	Lack of documentation
3	Lack of efficient configuration change control
4	Lack of audit-trail
5	Insufficiency security training
6	Complicated use interface
7	Insufficient maintenance
8	Lack of periodic replacement schemes
9	Insufficiency professional training
10	Inadequate service maintenance response
11	Inadequate recruitment procedures

12	Unsupervised work by outside or cleaning staff
13	Lack of security awareness
14	Wrong allocation of access rights
15	Inadequate or careless use of physical access control to buildings and rooms
16	lack of strict operation process procedure

Table 6: The RMSE of the proposed approach and other learning algorithms for the risk evaluation

Algorithm	RMSE
The proposed approach	0.081
RFRA algorithm	0.183
FCRM algorithm	0.785
SONFIN algorithm	0.819