

Research on the Extension Mining Model of Implication Type Data with Multi Factors Based on Extension Theory

Xianmin Wei

School of Computer Engineering, Weifang University, 5147 Eastern Dongfeng Street, Weifang, 261061, China
wfxuweixm@126.com

Abstract

The author conducts research on the mining problem of implication type data aiming at the characteristics of implication, uncertainty, nonlinearity, dynamic, complexity existing in the process of data mining and establishes an extension mining model of implication type data with multi factors based on extension theory. This model, first of all, carries out implication analysis on the implication type data and builds corresponding implication set; then, the author conducts extension classification on the hypogynous factor in the implication set and builds classical field and segment field of the epigynous factor in the implication set based on extension type divided; the author also respectively builds the correlation function and extension goodness-of-fit model between targeted mining object and classical field of epigynous factor in the implication set, acquires comprehensive extension goodness-of-fit considering the weight of epigynous factors, which, in other words, determines the degree of closeness between targeted mining object and extension type and thus achieves the mining of implication type data. Finally, the author demonstrates the feasibility of this model by explaining and verifying the venture capital case of an enterprise.

Keywords: *Extension mining, extension theory, implication type data, multi factors, model and algorithm*

1. Introduction

The data mining of complex system is a very complicated system engineering and intelligent design engineering because of the multi-type, multi-factor, multi-data, multi-constraint, fuzziness and coupling characteristics of data in complex system, which has wide application prospect. In view of the difference of data type, many different types of data mining methods and technology come forth, like decision tree, neural network, genetic algorithm and association rules, which have achieved good application effect in engineering application. However, the association and implication of data in complex system sometimes show dominance but sometimes show potentiality. Therefore, effective mathematical model of data mining in complex system often cannot be established, which poses limitation to the data mining method based on data or model driven. To solve this problem, this paper proposes an extension mining model of implication type data with multi factors based on extension theory aiming at the problems existing in the above-mentioned mining method and model. This model can achieve effective mining of implication type data by the formal modeling conducted by matter element on data and information and establishing data mining process model blending extension qualitative analysis and extension quantitative analysis together.

2. Extension Data Mining Process Model in Complex System

2.1. The Generation of Implication Set

The factors can be divided into epigynous factor and hypogynous factor based on the implication relation between factors in complex system. The implication form of epigynous factor and hypogynous factor can be one-to-one, one-to-many, many-to-one and many-to-many, the general expression of which is:

$$C_1 \wedge \dots \wedge C_i \wedge \dots \wedge C_n \Rightarrow S_1 \wedge \dots \wedge S_j \wedge \dots \wedge S_m \quad (1)$$

Considering the dynamic characteristics of the development of different factors, there should be corresponding value $v(t)$ at t time and then the expression of value implication of corresponding factor is:

$$v_t^{C_1} \wedge \dots \wedge v_t^{C_i} \wedge \dots \wedge v_t^{C_n} \Rightarrow v_t^{S_1} \wedge \dots \wedge v_t^j \wedge \dots \wedge v_t^{S_m} \quad (2)$$

The implication characteristics C_i of epigynous factor and the implication characteristics S_j of hypogynous factor and corresponding value $v(t)$ constitute the implication set Γ , which is

$$\Gamma = \{(C_i, S_j) | \wedge C_i \Rightarrow \wedge S_j, \wedge v_t^{C_i} \Rightarrow \wedge v_t^j\} \quad (3)$$

2.2. Matter Element Modeling

In extension theory, we adopt matter element to conduct formal modeling, using N to represent the name of the object, c to represent the characteristics and v to represent the value of characteristics c , and then this object can be described with an ordered triple, which is:

$$R = (N, c, v) \quad (4)$$

If this object has q characteristics, then we can use multidimensional matter element to describe it, which is:

$$R = (N, C, V) = \begin{bmatrix} N & c_1 & v_1 \\ & \vdots & \vdots \\ & c_j & v_j \\ & \vdots & \vdots \\ & c_q & v_q \end{bmatrix} \quad (5)$$

Therefore, we take the extension type of hypogynous factor as the name of matter element, the epigynous factor as the characteristics and the dynamic value of epigynous factor as the characteristics value of matter element, so the matter-element model of the above-mentioned implication set R_s is:

$$R_s = (N, C, V) = \begin{bmatrix} S & C_1 & v_t^{C_1} \\ & \vdots & \vdots \\ & C_j & v_t^{C_j} \\ & \vdots & \vdots \\ & C_n & v_t^{C_n} \end{bmatrix} \quad (6)$$

2.3. Classical Field and Segment Field

In the process of implication type data mining of hypogynous factor, we need to classify the hypogynous factor first. After the classification, we can acquire the value

range of different epigynous factor under certain classification, which is the corresponding classical field, and we can build the classical field and segment field matter element of hypogynous factor under certain classification.

Assuming the number of classification of hypogynous factor is p , if the mining sub-classification of hypogynous factor is k , then the corresponding classical field matter

element model R_{SC}^k of classification k is:

$$R_{SC}^k = \begin{bmatrix} SC^k & c_1 & [v_k^{c_1}(t), v_k^{c_1}(r)] \\ \vdots & \vdots & \vdots \\ c_j & [v_k^{c_j}(t), v_k^{c_j}(r)] \\ \vdots & \vdots & \vdots \\ c_n & [v_k^{c_n}(t), v_k^{c_n}(r)] \end{bmatrix} \quad (7)$$

The corresponding segment field matter element model of hypogynous factor R_{SO} is:

$$R_{SO} = \begin{bmatrix} SO & c_1 & [\min_{1 \leq k \leq p} v_k^{c_1}(t), \max_{1 \leq k \leq p} v_k^{c_1}(r)] \\ \vdots & \vdots & \vdots \\ c_j & [\min_{1 \leq k \leq p} v_k^{c_j}(t), \max_{1 \leq k \leq p} v_k^{c_j}(r)] \\ \vdots & \vdots & \vdots \\ c_n & [\min_{1 \leq k \leq p} v_k^{c_n}(t), \max_{1 \leq k \leq p} v_k^{c_n}(r)] \end{bmatrix} \quad (8)$$

2.4. Implication Index Weight

This paper adopts AHP Method to acquire different implication index weight. First of all, invite evaluation experts to evaluate and give score on the importance of different index j based on the evaluation standard in Table 1 and obtain corresponding evaluation matrix $A_{n \times n}$, which can be seen in Table 2.

Table 1. Evaluation Standard of AHP Method

a_{ij} Evaluation Standard	Meaning
1	Equal Importance
3	Weak Importance
5	Importance
7	Relatively Important
9	Extreme Importance
2,4,6,8	Between the Above-mentioned Adjacent States
Reciprocal of Above-mentioned Value	$a_{ij} = 1/a_{ji}$

Table 2. Evaluation Matrix Based on AHP Method

$A_{n \times n}$	c_1	c_2	\dots	c_n
c_1	a_{11}	a_{12}	\dots	a_{1n}
c_2	a_{21}	a_{22}	\dots	a_{2n}
\vdots	\vdots	\vdots	\vdots	\vdots
c_n	a_{n1}	a_{n2}	\dots	a_{nn}

Adopt sum and product method to process the evaluation matrix and we can get:

$$\bar{a}_{ij} = a_{ij} / \sum_{i=1}^n a_{ij} \quad (9)$$

Then we can acquire the weight w_i of influencing factors i :

$$w_i = \sum_{j=1}^n \bar{a}_{ij} / \sum_{i=1}^n \sum_{j=1}^n \bar{a}_{ij} \quad (10)$$

To judge the rationality and consistency of factor weight acquired, we need to conduct consistency detection, namely the consistency ratio parameter CR needs to satisfy the following equation:

$$CR = (\lambda_{\max} - n) / (RI * (n - 1)) \leq 0.1 \quad (11)$$

In this equation: λ_{\max} is the maximum eigenvalue of evaluation matrix $A_{n \times n}$, RI is the random one-time index and can be acquired by referring to the table.

2.5. Extension Goodness-of-Fit

The extension theory gives extension distance $\rho(x, X)$ of the description point value x and interval value $X = [x_1, x_2]$, which is:

$$\rho(x, X) = \left| x - \frac{x_1 + x_2}{2} \right| + \frac{x_1 - x_2}{2} \quad (12)$$

Based on this, if the value of mining object G regarding matter element characteristics C_j is $v_g^{C_j}$, then the extension distance between it and the corresponding classical field of mining sub-classification k of hypogynous factor $\rho(v_g^{C_j}, v_k^{C_j})$ is:

$$\rho(v_g^{C_j}, v_k^{C_j}) = \left| v_g^{C_j} - \frac{v_k^{C_j}(l) + v_k^{C_j}(r)}{2} \right| + \frac{v_k^{C_j}(l) - v_k^{C_j}(r)}{2} \quad (13)$$

The extension distance between the corresponding segment field of mining sub-classification of hypogynous factor $\rho(v_g^{C_j}, v_o^{C_j})$ is:

$$\rho(v_g^{C_j}, v_o^{C_j}) = \left| v_g^{C_j} - \frac{v_o^{C_j}(l) + v_o^{C_j}(r)}{2} \right| + \frac{v_o^{C_j}(l) - v_o^{C_j}(r)}{2} \quad (14)$$

In particular, if the the value of mining object G regarding matter element characteristics C_j is $v_g^{C_j} = [v_g^{C_j}(l), v_g^{C_j}(r)]$, then the extension distance between it and the corresponding classical field of mining sub-classification k of hypogynous factor $\rho(v_g^{C_j}, v_k^{C_j})$ is:

$$\rho(v_g^{C_j}, v_k^{C_j}) = \frac{\rho(v_g^{C_j}(l), v_k^{C_j}) + \rho(v_g^{C_j}(r), v_k^{C_j})}{2} \quad (15)$$

The extension distance between the corresponding segment field of mining sub-classification of hypogynous factor $\rho(v_g^{C_j}, v_o^{C_j})$ is:

$$\rho(v_g^{C_j}, v_o^{C_j}) = \frac{\rho(v_g^{C_j}(l), v_o^{C_j}) + \rho(v_g^{C_j}(r), v_o^{C_j})}{2} \quad (16)$$

Then the extension correlation function regarding matter element characteristics C_j of mining object G and mining sub-classification k of hypogynous factor $k(v_g^{C_j}, v_k^{C_j})$ is:

$$k(v_g^{C_j}, v_k^{C_j}) = \begin{cases} \frac{\rho(v_g^{C_j}, v_k^{C_j})}{|v_k^{C_j}|}, & v_g^{C_j} \in v_k^{C_j} \\ \frac{\rho(v_g^{C_j}, v_k^{C_j})}{\rho(v_g^{C_j}, v_o^{C_j}) - \rho(v_g^{C_j}, v_k^{C_j})}, & v_g^{C_j} \notin v_k^{C_j} \end{cases} \quad (17)$$

Thus, the extension good-of-fit of mining object G and mining sub-classification k of hypogynous factor is:

$$\zeta(k) = \sum_{j=1}^n (w_j * k(v_g^{C_j}, v_k^{C_j})) \quad (18)$$

2.6. The Realization of Data Mining Algorithm

In conclusion, the description of the realization steps of the extension mining model of implication type data with multi factors based on extension theory is as follow:

(1) Acquire mining object and implication factors based on application background and confirm epigynous factor and hypogynous factor to establish implication set.

(2) Conduct extension classification on the hypogynous factor based on mining precision.

(3) Build classical field and segment field of epigynous factor aiming at different extension types.

(4) Regard hypogynous factor as the name of matter element and build matter element model of mining object that takes hypogynous factor as matter element characteristics.

(5) Conduct weight distribution of different matter element characteristics based on AHP Method and acquire corresponding weight value.

(6) Build extension distance calculation model between mining object and extension type.

(7) Build extension correlation function calculation model between mining object and extension type.

(8) Build extension good-of-fit calculation model between mining object and extension type

(9) Acquire the classification of mining object based on extension good-of-fit, namely the bigger the extension good-of-fit between mining object and certain extension classification, the more likely belonging to this extension classification.

(10) Judge whether the mining result meets the requirement. If not, execute step (11), or else, execute step (12).

(11) Conduct sub-classification with finer precision of the extension classification of hypogynous factor based on the classification done before, and repeatedly execute step (3) to step (10).

(12) End of mining process.

3. Application Analysis

The development of enterprises suffers the comprehensive influence exerted by many different factors, which means we need to consider the internal situation of enterprises as well as the external condition. There often exist complicated association and implication

between these influencing factors, which poses great uncertainty for the operation and formulation of development planning strategy of enterprises, especially for the enterprises in the rising period and growing period. The software and hardware platform of these enterprises have great development potential and mining space, which means that they need to weigh the internal and external factors of enterprise development comprehensively to determine the optimal capital input of venture capital in the process of venture capital development. To improve the development competitiveness of enterprises and win huge profit, enterprises need to consider factors like capital input, market share, profit increasing rate, production capability and financing ability in their venture capital. Based on analysis, we can know that the hypogynous factor of venture capital is capital input increasing rate S , the epigynous factor is financing ability C_1 , profit increasing rate C_2 , market share C_3 and production capability C_4 . Based on real situation of enterprise development, analysis and verification need to be conducted on the basis of the extension mining model of implication type data with multi factors based on extension theory to effectively determine the capital input of venture capital. Through the investigation, analysis of statistical data and opinion collection of the enterprise development condition, we have acquired the enterprise data in the corresponding stages. Concrete data can be seen in Table 3.

Table 3. Initial Mining Data

	epigynous factor				hypogynous factor
	C_1	C_2	C_3	C_4	S
1	0.40	0.152	0.050	0.60	0.203
2	0.45	0.109	0.032	0.75	0.265
3	0.40	0.231	0.105	0.75	0.324
4	0.65	0.362	0.176	0.85	0.566
5	0.60	0.224	0.151	0.80	0.489
6	0.50	0.198	0.128	0.85	0.327
7	0.80	0.317	0.209	0.90	0.635
8	0.70	0.216	0.185	0.85	0.416
9	0.75	0.255	0.265	0.90	0.384

We can classify the hypogynous factor into three classifications, among which the corresponding mining range of classification S_1 is 0.200 to 0.350, the corresponding mining range of classification S_2 is 0.350 to 0.500 and the corresponding mining range of classification S_3 is 0.500 to 0.650. From this, we can establish the matter element model of different extension classifications, which is:

$$R_{sc}^1 = \begin{bmatrix} SC^1 & C_1 & [0.400, 0.450] \\ & C_2 & [0.109, 0.152] \\ & C_3 & [0.032, 0.105] \\ & C_4 & [0.600, 0.700] \end{bmatrix}, \quad R_{sc}^2 = \begin{bmatrix} SC^2 & C_1 & [0.500, 0.750] \\ & C_2 & [0.198, 0.255] \\ & C_3 & [0.128, 0.265] \\ & C_4 & [0.800, 0.900] \end{bmatrix},$$

$$R_{sc}^3 = \begin{bmatrix} SC^3 & C_1 & [0.650, 0.800] \\ & C_2 & [0.317, 0.362] \\ & C_3 & [0.176, 0.209] \\ & C_4 & [0.850, 0.900] \end{bmatrix}, \quad R_{so} = \begin{bmatrix} SO & C_1 & [0.400, 0.800] \\ & C_2 & [0.109, 0.362] \\ & C_3 & [0.032, 0.265] \\ & C_4 & [0.600, 0.900] \end{bmatrix}.$$

If the matter element model of awaiting mining object is R_{sd} , which is

$$R_{SD} = \begin{bmatrix} SD & C_1 & 0.750 \\ & C_2 & 0.300 \\ & C_3 & 0.220 \\ & C_4 & 0.800 \end{bmatrix}$$

From this, we can obtain the extension distance of the awaiting mining object and the classical field and segment field of extension classification and the result can be seen in Table 4.

Table 4. Calculation Result of Extension Distance

	R_{SC}^1	R_{SC}^2	R_{SC}^3	R_{SO}
C_1	0.300	0	-0.050	-0.050
C_2	0.148	0.045	0.017	-0.062
C_3	0.115	-0.045	0.011	-0.045
C_4	0.100	0	0.050	-0.100

The extension correlation function of the awaiting mining object and the classical field of extension classification can be seen in Table 5.

Table 5. Calculation Result of Extension Correlation Function

	R_{SC}^1	R_{SC}^2	R_{SC}^3
C_1	-0.857	0	0.333
C_2	-0.705	-0.421	-0.215
C_3	-0.719	0.328	-0.196
C_4	-0.500	0	-0.333

We also conduct weight distribution of epigynous factor based on AHP Method and the evaluation matrix A built is:

$$A = \begin{bmatrix} & C_1 & C_2 & C_3 & C_4 \\ C_1 & 1 & 4/5 & 4/3 & 2 \\ C_2 & 5/4 & 1 & 5/3 & 5/2 \\ C_3 & 3/4 & 3/5 & 1 & 3/2 \\ C_4 & 1/2 & 2/5 & 2/3 & 1 \end{bmatrix}$$

Then we have weight vector W , which is:

$$W = \{0.286, 0.357, 0.214, 0.143\}$$

The extension good-of-fit between the awaiting mining object and the classical field of extension classification acquired can be seen in Table 6:

Table 6. Calculation Result of Extension Good-of-Fit

	R_{SC}^1	R_{SC}^2	R_{SC}^3
C_1	- 0.245	0	0.095
C_2	- 0.252	- 0.151	- 0.077
C_3	- 0.154	0.070	- 0.042
C_4	- 0.072	0	- 0.048
Comprehensive Good-of-fit	-0.723	-0.081	-0.072

We can see from the result in Table 4 that the extension classification of awaiting mining object is R_{sc}^3 . Then in the next stage of venture capital, the capital input should increase 50% to 65% on the basis of original capital, which can win relatively huge profit for enterprises. The result of mining is consistent with the investment amount of enterprises, which indicates the effectiveness and operability of the model.

4. Conclusions

This paper discusses the mining model and algorithm of implication relationship with multi factors aiming at mining issue of the multi-data in complex system and proposes an extension mining model of implication type data with multi factors based on extension theory. This model has achieved the formal modeling of mining data through establishing the implication set and extension type between multi factors in complex system and building matter element model of the epigynous factor and hypogynous factor in the implication set. At the same time, the applicability of the mining model is improved by establishing the extension distance, extension correlation function and extension good-of-fit model between mining object and extension classification, which also makes it possible to conduct data mining from the angle of blending qualitative and quantitative analysis. In addition, the precision of mining is effectively improved by the detailed sub-classification of extension classification, making the mining result accurate and reliable. Finally, the effectiveness and operability are demonstrated by the analysis and verification of specific case, thus providing a new solution and solving approach for multi-data mining in complex system.

Acknowledgements

This work is partly supported by National Natural Science Foundation of China (No. 61471269), Shandong Spark Program (2012XH06005), Weifang municipal Science and Technology Development Program (201301050).

References

- [1] W. H. Zhong and P. A. Qun, "Existing Situation of Data Mining Research and Its Development Tendency", *Industry and Mine Automation*, no. 2, (2011), pp. 29-32.
- [2] Q. H. Sheng and O. Y. Qun, "Application of data mining technology in enterprise human resource management", *Enterprise Economy*, no. 1, (2012), pp. 78-80.
- [3] X. Liang, "Data mining; algorithms; applications; software systems", *Computer Technology and Development*, vol. 16, no. 1, (2006), pp. 1-5.
- [4] T. Lajnef, S. Chaibi and P. Ruby, "Learning machines and sleeping brains: Automatic sleep stage classification using decision-tree multi-class support vector machines", *Journal of Neuroscience Methods*, vol. 250, (2015), pp. 94-105.
- [5] Y. Kokkinos and K. G. Margaritis, "Confidence ratio affinity propagation in ensemble selection of neural network classifiers for distributed privacy-preserving data mining", *Neurocomputing*, vol. 150, (2015), pp. 513-528.
- [6] N. Nezamoddini and S. S. Lam, "Reliability and topology based network design using pattern mining guided genetic algorithm", *Expert Systems with Applications*, vol. 42, no. 21, (2015), pp. 7483-7492.
- [7] A. Nanopoulos, A. N. Papadopoulos and Y. Manolopoulos, "Mining association rules in very large clustered domains", *Information Systems*, vol. 32, no. 5, (2007), pp. 649-669.
- [8] Y. Xu and Q. Zhu, "A New Extension Theory-based Production Operation Method in Industrial Process", *Chinese Journal of Chemical Engineering*, vol. 21, no. 1, (2013), pp. 44- 54.
- [9] T. C. Wang, A. J. Yang and S. S. Zhong, "Multi-Attribute Extension Fuzzy Optimized Decision-Making Model of Scheme Design", *Tehnički vjesnik/Technical Gazette*, vol. 21, no. 2, (2014), pp. 239-247.
- [10] T. C. Wang, A. J. Yang, S. S. Zhong and Z. D. Zhao, "Extension Adaptive Design Model of Scheme Design for Complex Mechanical Products", *Tehnički vjesnik/Technical Gazette*, vol. 21, no. 1, (2014), pp. 123-133.

Author



Xianmin Wei, he received the M. Sc. degree in computer applications from Shandong Science and Technology University (2005). He is currently an associate professor in school of computer engineering at Weifang University, China. He has published over 30 papers and 3 books in professional fields. Since 2011, he has been a member of IEEE-CS, ACM and CCCF, respectively. His fields of research are focused on swarm intelligent, intelligent sensor networks.

