Model Mining Method for Collaborative Behavior of Knowledge Agent in Innovation Ecosystem

Wen Li(✉)

Business School, Suzhou University of Science and Technology,
Suzhou 215009, Jiangsu, China
cucmcum@163.com

Abstract. Conventional model of cooperative behavior mining method, can carry on the analysis, data mining to the conventional collaborative behavior but for specific subject knowledge in innovation ecosystem cooperative behavior, and analysis of the data mining results shooting low deficiencies, therefore puts forward innovation ecosystem in knowledge collaborative behavior main body model of the mining method. Based on knowledge innovation ecosystem in the main body composition analysis of collaborative behavior model, used algebraic representation, data processing design collaborative behavior model, realized the coordinated behavior model of innovation ecosystem knowledge subject data processing; According to the parameter fitting of collaborative behavior of knowledge subject in innovation ecosystem, the mining results were displayed to realize the model mining of collaborative behavior of knowledge subject in innovation ecosystem. The experimental data show that the proposed collaborative behavior model mining method is 41.84% higher than the traditional mining method, which is suitable for the model mining of collaborative behavior of knowledge subjects in the innovation ecosystem.

Keywords: Innovation ecosystems · Knowledge subject · Cooperative behavior · Model mining

1 Introduction

The model mining method of conventional cooperative behavior can analyze the data mining of conventional cooperative behavior. But when we analyze the collaborative behavior of knowledge agents in specific innovation ecosystem, because of the limitations of the collaborative behavior model data handler, there is a shortage of low hit rate of mining results. According to the given collaborative behavior pattern, the traditional collaborative behavior model mining method matches the set of supply chain instances which conform to the collaborative behavior pattern on the candidate chain set of personalized supply chain, and selects the supply chain instances which conform to the user's personalized demand according to the user's personalized demand in the set of matched supply chain instances. However, model mining that is not suitable for collaborative behavior of knowledge agents in innovation ecosystems [1], this paper presents a model mining method for collaborative behavior of knowledge agents in

G. Gui and L. Yun (Eds.): ADHIP 2019, LNICST 302, pp. 372–380, 2019.
https://doi.org/10.1007/978-3-030-36405-2_37

innovative ecosystems. Using principal component analysis, the composition of knowledge agent collaborative behavior model is analyzed. Represented by algebra, a large amount of related data information is projected into the feature subspace of low-dimensional data. Design collaborative behavior model data processing process, data processing of collaborative behavior model of knowledge agents in innovation ecosystem is completed. The least square method is used to simulate the collaborative behavior parameters of knowledge agents in innovative ecosystem. Depending on the normal distribution of discrete group points, the results of the mining are displayed, the model mining of collaborative behavior of knowledge agents in innovation ecosystem is completed. In order to ensure the effectiveness of the designed collaborative behavior model mining method, simulating the experimental environment of collaborative behavior of knowledge agents, using two different cooperative behavior model mining methods, a simulation test of the hit rate of the excavation is carried out. The results of the experiment show that the cooperative behavior model mining method is highly effective.

2 System Objective and Analysis

The model mining methods of collaborative behavior of knowledge agents in innovative ecosystem mainly include:

(1) On the basis of the data array composed of n parameters and m sample values, a small number of comprehensive variables are established to analyze the formation of collaborative behavior models of knowledge entities in innovative ecosystem. It is represented by algebra.
(2) A large number of highly relevant variables in the production process are mapped to the principal component space defined by a small number of implicit variables through multivariate statistical projection. To reveal its main structure, the input of the model and the simplification of the variables are realized.
(3) The least square method is used to fit the data. Based on the normal distribution of discrete group points, the mining results are displayed. To solve the problem of model mining method of common cooperative behavior.

3 Data Processing of Knowledge Agent Synergetic Behavior Model in Innovation Ecosystem

3.1 Analysis on the Composition of Knowledge Agent Synergetic Behavior Model in Innovation Ecosystem

An innovative ecosystem is a system of resources, talent, institutions and infrastructure that can combine new ideas or technologies with products, services and production processes. Knowledge agent cooperative behavior model is one of the commonly used methods in multivariate statistical analysis. It is based on a data array consisting of n parameters m samples with a certain dependence. By creating a smaller number of composite variables, make it more concentrated to reflect the original n parameters contained in the change information. The basic method is to determine the primary and secondary position of the direction of change according to the variance of the data. The main elements are obtained in primary and secondary order, these main elements are independent of each other. With this tool, can extract change information, reduce the complexity of data analysis. It has the ability to remove the data correlation, reduce the noise effect, reduce the data dimension and other advantages, in data compression, signal processing, pattern recognition, fault diagnosis and other fields have been widely used [2]. After the traditional principal component analysis method is used to standardize the data processing, the data show a "uniform" distribution. In this paper, the concept of relative principal component is proposed. At the same time, the concepts of the relative transformation of the data and the uniform distribution of the data are defined. A relative principal component analysis method is established. The geometric meaning of it is given.

Collaborative behavior model of knowledge subject in innovation ecosystem, the central idea is to reduce the dimensions of data sets that contain a large number of related variables. At the same time, train variables in the dataset as much as possible. Multiple regression discriminant analysis uses variable selection procedures to reduce dimension, but it results in the loss of one or more variables [3]. The CBIE method uses all the original variables to get a small set of new variables. The new variable can then be approximated to the original variable. The greater the correlation of the original variable, the smaller the number of new variables required. Principal Components, Principal Components, (PCs) are irrelevant, orderly, therefore, a few principal components can be used to train most variables that exist in the original data set.

The principal component analysis is concerned with through several linear combinations of a set of variables to explain the variance-covariance structure of this set of variables. Its general purpose is data compression and data interpretation. Although n elements are required to reproduce system-wide variability, however, most of the variability is often explained by only a few m (m < n principal components. When this is the case, these two principal components contain almost as much information as n primitive variables [4, 5]. So this m principal components can replace the original n variables, and the original data set is composed of N times measurements of n variables, compress the data set consisting of N values of M principal components (Fig. 1).

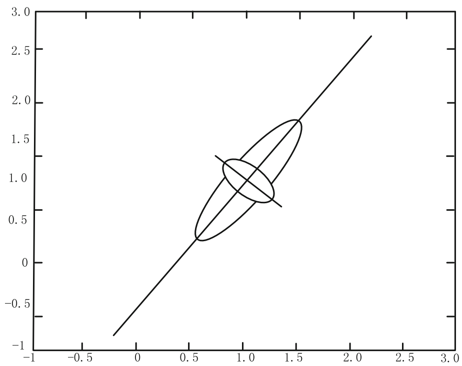

Fig. 1. Normal density ellipsoid diagram

3.2 Algebraic Representation of Cooperative Behavior Model

Suppose random vector X′ = [X1, X2, X3, …, XN] has covariance matrix ∑, its eigenvalue λ1 ≥ λ2 ≥ λ3 ≥ … ≥ λn ≥ 0. Consideration of linear combination

$$\begin{cases} Y_1 = a_1^{'}X = a_{11}X_1 + a_{21}X_2 + \ldots + a_{n1}X_n \\ Y_2 = a_2^{'}X = a_{12}X_1 + a_{22}X_2 + \ldots + a_{n2}X_n \\ \ldots \\ Y_m = a_m^{'}X = a_{1m}X_1 + a_{2m}X_2 + \ldots + a_{nm}X_n \\ Van(Y_i) = a_1^{'}\sum a_i, i = 1, 2, \ldots, n \\ Cov(Y_i, Y_m) = a_1^{'}\sum a_m, i = 1, 2, \ldots, n \end{cases} \tag{1}$$

Principal component is an unrelated linear combination Y1, Y2, …, Ym.

The first principal component is a linear combination of the maximum variance, and even Var(Y1) = a1′∑a1, maximization. Var(Y1) = a1′∑a1 It will increase by multiplying any a1 by a constant. In order to eliminate this uncertainty, focus only on coefficient vectors with unit length. So we define the first principal component and the second linear combination a1′X, a1′a1 = 1, it makes Var(aa1′X) become higher [6]. a1′X is second principal component bilinear combination, when it in a1′a1 = 1和Cov (a1′X, am′X, X) = 0(m < i), it makes Var(a1′X) become bigger.

a2′X is the second primary and second linear combination, when a2′ a2 = 1 and Cov(a1′ X, a2′ X) = 0, it makes var(a2′ X) become bigger. Conclusion random vector X′ = [X1, X2, X3, …, XN]'s covariance matrix. Feature vector (λ1, e1), (λ2, e2), (λ3, e3), …, (λn, en), λ1 ≥ λ2 ≥ λ3 ≥ … ≥ λn ≥ 0. I principal component is Yi = ei′ X = ei1′ X+ei2′ X+ … +a in Xn, i = 1, 2, 3, …, m. Therefore, $Van(Y_i) = e_1^{'}\sum e_i$, $i = 1, 2, \ldots, n$, $Cov(Y_i, Y_m) = e_1^{'}\sum e_m, e \neq m$ if the λi is equation, so the selection of the corresponding coefficient vector e_i is not unique.

3.3 Design of Data Processing Process for Collaborative Behavior Model

CBIE's main task is dimension reduction of data Matrix. A large number of highly relevant variables that exist in the production process multivariate statistical projection Map to a principal component space defined with a small number of hidden variables, so as to reveal its main structure, implementation of model input and variable simplification.

Although, sometimes a large number of multiple (n) variables are required to describe a complex system more perfectly, however, in general, most of the variability of the system is often explained by a few (m, m5n) principal components. And these m principal elements basically can contain from most of the original n variables or primary information. In other words it can be effectively represented with m principal elements or explain the characteristic structure of the data matrix formed by the sequence of variables in the original system, in order to achieve the purpose of data dimension reduction, the data matrix composed of a sequence of system variables is $X \equiv X(k, k + N - 1) = [x(k), x(k+1), \ldots, x_n(k + N - 1)]$, if the data consisting of the first component sequence of the system is represented by Xi as i $X_i = [x(k), x(k+1), \ldots, x_n(k + N - 1)]$.

Because the results of principal component analysis are affected by dimension, so in the principal component analysis, data need to be standardized, So we can construct a mean value of 0, standardized data with variance of 1 [7, 8]. First standardize x as follows: $Xs = [X - (11\ldots1)^T M_{eans}]diag(1/S_1, 1/S_2, \ldots, 1/S_n)$ Means = [m1, m2, …, mn] is The mean value of the variable X, S = [S1, S2, …, Sn] is Standard deviation of variables. If we remember the covariance matrix of the data matrix XS, which is composed of the sequence of system variables, the is:

$$\sum X_s = E\{[X_s(k, k+N-1) - E\{X_s(k, k+N-1)\}][X_s(k, k+N-1) - E\{X_s(k, k+N-1)\}]^T\}$$

then you can pass $|\lambda I - \sum Xs| = 0$ and $|\lambda I - \sum Xs| ei = 0$, i = 1, 2, …, n. Calculate the matrix separately $\sum Xs$, eigenvalues of λI and corresponding Eigenvectors ei, for ease of description, suppose $\lambda 1 \geq \lambda 2 \geq \lambda 3 \geq \ldots \geq \lambda n$. Eigenvector obtained by using the upper expression e, get the following n combinations of data

$$\begin{cases} V_1 = (e_1)^T X_s = e_{11}X_{s1} + e_{21}X_{s2} + \ldots + e_{n1}X_{sn} \\ V_2 = (e_2)^T X_s = e_{12}X_{s1} + e_{22}X_{s2} + \ldots + e_{n2}X_{sn} \\ \ldots \\ V_m = (e_n)^T X_s = e_{1n}X_{s1} + e_{2n}X_{s2} + \ldots + e_{nn}X_{sn} \end{cases} \tag{2}$$

Satisfy the property, and

$$\begin{cases} Van(V_i) = (e_i)^T \sum xe_i = \lambda, i = 1, 2, \ldots, n \\ Cov(Y_i, Y_m) = (e_i)^T \sum xe_J = 0, i \neq j \end{cases} \tag{3}$$

If the preceding m (m < n) vectors are selected in order V1, V2, V3, …, Vm, then they can be used to analyze the principal components of the system. The principal component analysis is mainly to collect a large amount of historical data. Set up the statistical database, extract the normal data set, and establish the statistical unit, according to the principal component analysis method, the process data projects a lot of related data information into the feature subspace of low-dimensional data. Identify collaborative behavior model data processing. Based on the composition analysis of collaborative behavior model of knowledge agents in innovation ecosystem, and using algebraic representation, Determine the data processing of the collaborative behavior model, the collaborative behavior model data processing of knowledge agents in innovation ecosystem is realized.

Based on the knowledge agent collaborative behavior model data processing in the innovation ecosystem, the collaborative behavior parameters of knowledge agents in innovation ecosystem are fitted. The fitting of its parameters is divided into two parts. First, the parameter model is transformed to make it a unified model parameter. Secondly, according to the regression parameters, there are many fitting methods. In this paper, the least square method is used to fit the data.

Parameter model transformation, relying on the original mathematical relations, when the data collection parameter is substituted into the mathematical relation, the mathematical relation cannot be related to the computer language. Therefore, a bridge between mathematical relation and data mining technology is built by constructing a parameter model transformation, and its variable transformation table is shown in Table 1.

Table 1. Variable transformation table

Original variable	Variable after change	Radix	Form
η	A(a,b)	2	Numeric type
σ	B(a,b)	2	Numeric type
λ	C(a,b)	2	Hollerith type
θ	D(a,b)	2	Time type

According to Richard Henderson's mathematical mining formula, the transformed parameters are replaced, and the least-square data fitting is carried out at the same time. The least square method is one of the methods to realize the data fitting. The least square method is suitable for mining calculation, the running volume is small, and the calculation is simple, so the least square method is used to fit the data.

Let a parameter be a1 for the first operation, a2 for the second operation, and so on, and an for the operation. If the data a1 is the same as a certain operation parameter an-m, then choose a1 or an-m instead. The purpose of the data is to parse hundreds of thousands of related data for the second time, and the least square data fitting first calculates the result of the first operation. And the statistics can be expressed as, $a_1 = \sum_{i=1}^{n} (\widehat{a_0} + \widehat{a_1} x_{i1} + \ldots\ldots + \widehat{a_n} x_{in} - y_i)^2$ [9, 10].

In the formula, x_{in} represents the relevant operation parameter of mathematical calculation, the eigenvalue selected when the mathematics is running at $\widehat{a}$. If the partial derivation of each data segment is obtained, then the deviation square difference of each data segment can be judged, and whether its a1 can replace an-m can be determined.

3.4 Results of Knowledge Agent Cooperative Behavior Mining in Innovation Ecosystem

Based on the least square method, the collaborative behavior information of the knowledge subjects in the innovative ecosystem is determined based on the mining results of the collaborative behavior of the knowledge subjects in the innovative ecosystem. The discrete group analysis of the relevant data such as a1/A2CU. An and so on is carried out, and the conclusion is drawn. Taking the mining theory value x = μ as the computing center, the discrete 3 s range of the data is calculated, and its S = (a1a 2. An)/n). The cooperative behavior of knowledge agents in the innovation ecosystem is calculated by monitoring data mining in the interval of [−3 s, 3 s] and mining the optimal values of the weights of the monitoring data parameters. The mining results are displayed by relying on the normal distribution of discrete group points, and the schematic diagram of normal distribution of discrete group points is shown in Fig. 2.

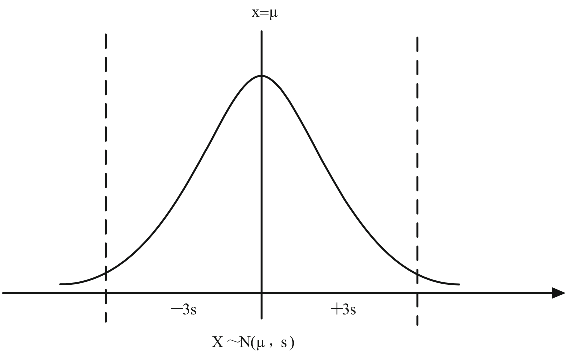

Fig. 2. Showing the normal distribution of discrete group points

Data processing based on collaborative behavior model of knowledge agents in innovation ecosystem. The least square method is used to realize the collaborative behavior parameter fitting of knowledge agents in innovative ecosystem, and the results of mining are displayed based on the normal distribution of discrete group points. The model mining of collaborative behavior of knowledge agents in innovation ecosystem is completed.

4 Test and Analysis

In order to ensure the effectiveness of the model mining method proposed in this paper for collaborative behavior of knowledge agents in the innovation ecosystem, simulation experiments are carried out. In the process of experiment, the simulation experiment of mining hit rate is carried out with different knowledge agents' cooperative behavior as the test object. The different range of independent variables and behavior structure of knowledge agent cooperative behavior are simulated. In order to ensure the effectiveness of the experiment, the conventional cooperative behavior model mining method is used as the comparison object, and the results of the two simulation experiments are compared, and the test data are presented in the same data chart.

4.1 Test Preparation

In order to ensure the accuracy of the simulation test process, the test parameters are set. In this paper, we use different collaborative behavior of knowledge agents as the experimental object, and use two different mining methods of collaborative behavior model to carry out the simulation test of mining hit rate, and analyze the results of the simulation experiment. Because the analytical results and the analytical methods obtained by different methods are different, it is necessary to ensure the consistency of the environmental parameters in the test process. The results of the test data setup in this article are shown in Table 2.

Table 2. Test parameter settings

Simulation test parameters	Scope/parameters	Remarks
Simulated independent variable	1/CBS	Complexity of collaborative behavior of knowledge subjects' count backwards
1/CBS	[0,1]	Does not contain 0
Simulation and test system	DJX-2016-3.5	Windows terrace

4.2 Test Result Analysis

In the process of experiment, two different cooperative behavior model mining methods are used to work in the simulation environment, and the variation of the hit ratio is analyzed. At the same time, because two different collaborative behavior model mining methods are adopted, the analysis results can not be directly compared. Therefore, the third party analysis record software is used to record and analyze the test process and results. The results are shown in the curve of comparison of the results of this experiment. In the simulation test result curve, the function of the third-party analysis recording software is used to eliminate the uncertainty caused by the personnel operation in the simulation laboratory and the factors generated by the simulation computer equipment, and only for the collaborative behavior of different knowledge subjects. Different mining methods of cooperative behavior model were used to simulate the hit rate of mining. The contrast curve of the test results is shown in Fig. 3. According to the test curve results and the third party analysis and record software, the mining method of cooperative behavior model is proposed. Compared with the conventional cooperative behavior model mining method, the mining hit rate of the model mining method is processed with arithmetic average value, and the result shows that the proposed collaborative behavior model mining method is better than the traditional mining method. Results the hit rate was increased by 41.84, which was suitable for the model mining of collaborative behavior of knowledge agents in the innovation ecosystem.

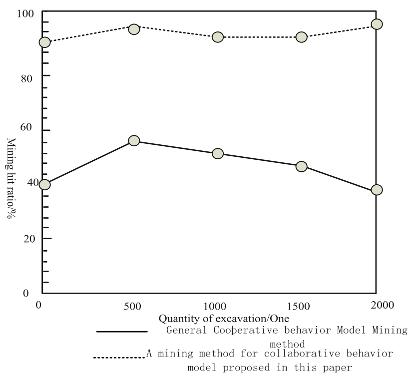

Fig. 3. Contrast curve of test results

5 Conclusion

In this paper, a model mining method for collaborative behavior of knowledge agents in innovation ecosystem is proposed, which is based on data processing of collaborative behavior model of knowledge agents in innovation ecosystem and parameter fitting of collaborative behavior of knowledge agents in innovation ecosystem. Realize the research of this paper. Experimental data show that the proposed method is highly effective. It is hoped that the research in this paper can provide theoretical basis for collaborative behavior model mining.

Acknowledgments. Supported by the National Natural Science Foundation of China (Grant No. 71771161).

Suzhou Science and Technology Program (Soft Science) Project (Grant No. SR201710).

References

1. Li, F., Chen, T.T.: Research on modeling and simulation of public network cooperative protection. Comput. Simul. **34**(6), 298–301 (2017)
2. Zheng, S.Z., Wu, Q.: Research on urban innovation capability from the perspective of intellectual property rights and construction of innovation ecosystem——taking Zhuhai as an example. Sci. Technol. Manag. Res. **36**(5), 111–116 (2017)
3. Tong, Z.H., Han, C.H.: Research on the impact of environment disturbance on NKCB and KCB in innovation activities. Sci. Technol. Prog. Policy **34**(13), 136–143 (2017)
4. Min, X.P., Shi, Y.L., Li, H., et al.: Mining collaborative behavior based on dynamic supply chain network. Comput. Integr. Manuf. Syst. **22**(2), 324–329 (2016)
5. Chen, J.B., Gao, S.L., Guo, Y.L.: Research on governance strategy of innovation ecosystem in small and medium-sized knowledge enterprises. Technoeconomics Manag. Res. **25**(10), 26–30 (2017)
6. Li, X.M., Bao, F.N.: Research on the impact of university social capital and collaborative behavior on collaborative innovation performance. Sci. Technol. Prog. Policy **34**(4), 122–128 (2017)
7. Fang, G., Zhou, Q., Yang, W.: Research context and progress from industry-university-institute cooperation to collaborative innovation: based on bibliometrics analysis. Technol. Econ. **35**(10), 26–33 (2016)
8. Zhang, Y.Y., Zhang, S.T., Peng, H.J., et al.: Cyberspace knowledge innovation behavior under innovation ecosystem perspectives——multiple case studies based on grounded theory. Sci. Technol. Prog. Policy **34**(6), 139–146 (2017)
9. Shan, M.M., You, J.X., Shao, J.: Co-evolution and optimization modes of industrial innovation ecosystem: a case study of Zhangjiang bio-pharmaceutical industry. Shanghai Manag. Sci. **39**(3), 1–7 (2017)
10. Mao, B.Q., Deng, W., Feng, S., et al.: Calculation method of vertical target dispersion based on multidisciplinary collaboration simulation. Comput. Simul. **33**(1), 20–23 (2016)