Acta Technica 62 (2017), No. 3B, 757–766

Analysis of "One Belt and One Road" construction and tax policy adjustment in Anhui Province from the perspective of tax preference and tax coordination

Yunxia Yao[2]

Abstract. When there is a lack of knowledge or integrity in the field of taxation, it is possible to use tax preference and tax coordination methods for data mining. In this method, analysis of tax policy adjustment based on the perspective of tax preference and tax coordination can achieve the "One Belt and One Road" construction of continuous eigenvalues by using tax domain knowledge. The analysis algorithm of tax policy adjustment (AATPA) is also proposed, applicable to data description problems of mixed tax types. The adjustment analysis using the algorithm can be automatically recorded and determined, and the frequent degree corresponding to its internal eigenvalue can effectively correct the tax policy adjustment center. Under the construction of "One Belt and One Road", the output of tax policy adjustment can be expressed and explained by concept combination. Finally, through the study of the real tax data in Anhui Province, it is shown that the AATPA algorithm proposed in this paper is effective.

Key words. One Belt and One Road Initiative, tax incentives, tax coordination, tax policy adjustment.

1. Introduction

At present, our country has entered the era of big data with the rise of data mining technology. Using this method can extract useful and unknown information from the mass of data. And this information is useful for decision-making[1–2]. If the tax domain knowledge is relatively little or the data set is incomplete when the mining task is carried out, the method of adjusting analysis can be used at this time[3–4]. Using this method, unidentifiable data can be classified into different classes and

[1]Acknowledgment - This research is supported by the foundation of Intension to Enhance the Whole Staff Action Plan Major Scientific Research Project of Anhui Finance & Trade Vocational College: Research on tax policy and tax planning of anhui export-oriented enterprises under the background of "One Belt and One Road" (granted by 2017nhrwa01).

[2]Workshop 1 - Anhui Finance & Trade Vocational College, Hefei, Anhui, 230601, China

implemented automatically, without the influence of prior knowledge. In this way, the relevant information in the original data set can be preserved to the maximum extent [5−6]. In the field of machine learning, the so-called tax incentive and tax coordination analysis method refers to that the description of objects' compound expressions makes the data can be divided according to this. And the measurement is based on the similarity of the data objects, rather than the geometric distance [7−8]. In tax incentive and tax coordination technology, the most prominent feature is that it is based on the corresponding class to determine the coverage, and also can explain the adjustment results. According to different treatment degrees of the concept normalization and specialization of tax incentives and tax coordination characteristics, a hierarchical conceptual description can be obtained[9]. In addition, the use of tax incentives and tax coordination can also solve the problem of incremental data mining, that is, if new data are added to the data mining, there is no need for the original data to be adjusted again[10].

This paper focuses on the related research of tax preference and tax coordination. In this paper, a method of tax policy adjustment based on tax preference and tax coordination is put forward, and the continuous characteristic value succeeds in conducting "One Belt and One Road" by using the knowledge of the tax field. By setting the similarity threshold, the number of adjusted partitions can be realized automatically, and for the similarity of data objects, it can also be judged according to the tax preference and tax coordination corresponding to the eigenvalue. And for the adjustment center, dynamic adjustment can be realized according to the frequency of eigenvalue. The adjustment output conducts "One Belt and One Road" construction, and the relative adjustment coverage description is obtained. The hierarchical concept tree can be obtained by adjusting the output. After experimental analysis, it is proved that the AATPA algorithm proposed based on the research of predecessors is effective.

2. Concept and characteristics of tax preference and tax coordination

2.1. Definition of tax preference and tax coordination

$U = D_1 \times D_2 \times ... \times D_m$Is a given discrete vector space, which is n-dimensional, in which D_ja finite set of symbols $(j = 1, 2, ..., m)$.|U|indicates the number of elements in the setU, which can describe the scale of the set.

Definition 1. $\forall_u \in U$is an instance of U, i. e., it describes $u = \langle a_1, a_2, ..., a_m \rangle$ by vector symbol, and can satisfy $a_j \in A_j \subseteq D_j$, $j = 1, 2, ..., m$, corresponding to the tuple S_i, $i = 1, 2, ..., n$.

Definition2. The description structure of the objects in the subset $T \subset U, T$is the same, which can be expressed as a set of several tuplesS_i, referred to as the database table$T = \{S_1, S_2, ..., S_n\}$, in which T refers to the subset of interest in the table.

Tuples S_i belong to the objects to be classified, and there are a total of n tuples, which can be recorded as$T = \{S_1, S_2, ..., S_n\}$. In$S_i$, the total number of features is $A_j$described to be m ones, $j = 1, 2, ..., m$. The range corresponding to

A_j is a symbolic field with a number of differences. It can be expressed in the form of$Dom\,(A_j) = \{a_1, a_2, ..., a_m\}$. $|Dom\,(A_j)|$refers to the scale of $Dom\,(A_j)$. In the feature domain, when describing objects are different, there will be different structural relationships among elements, which can be divided into four categories, namely,catego rical, linearand hierarchy. The corresponding sets' range under different structures may be disordered, ordered and graphically ordered, and described in order ofD_c,D_l andD_h.

D_l represents the ordered conceptual domain, and $A_j \subseteq D_l$ shows that all characteristic values are ordered concepts.

For example, the customer deposit balanceD_l = {none, less, relatively less, medium, relatively much, much, very much, extremely much'};

D_c refers to the concept domain of disorder. If$A_j \subseteq D_c$, it means that the characteristic values are all disordered concepts, taking the tax market performance as an example, D_c= D_c= {'Lingzhang tax', 'superdrop tax', 'Banker tax', 'speculation tax', 'informationtax'};

$$D_l A_j \subseteq D_l$$

D_h represents a characteristic that is normally a tree-like hierarchy. The parent node can generalize the child node, and at this time, in each concept layer, all the features are divided into two kinds, that is, order, disorder. Therefore, all node values correspond to an unordered set or an ordered set. For example, over a period of time, the customer's tax transactions are mostly ordered, as shown in figure 1. But relatively, newly listed securities or those circulating in the market are unordered, as shown in Fig. 1. The knowledge of the tax field is reflected by the change of characteristics, which can reveal the interest of the user.

2.2. Tax preference and tax coordination structure

It is assumed that the data object of interest meets the relevant conditions in definition 2 and the characteristics of the tuple $S_i \in T(i = 1, 2, ..., n)$ can be effectively measured by tax preference and tax coordination.

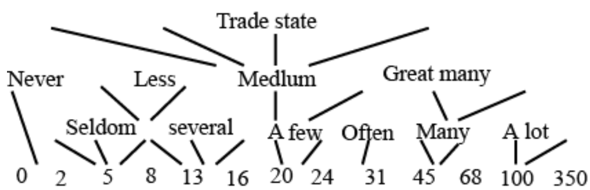

Fig. 1. Ordered hierarchical structure

(1) Tax preferences and tax coordination of symbolic eigenvalues:

$$d_c\,(a_{ik}, a_{jk}) = \begin{cases} 0 & a_{ik} = a_{jk} \\ 1 & a_{ik} \neq a_{jk} \end{cases} \quad , i \neq j \tag{1}$$

$d_c\,(a_{ik}, a_{jk})$ represents the corresponding tax preference and coordination of the element $a_{ik}, a_{jk} \in A_k (i \neq j)$ in the range of the kth eigenvalue Wie of the two tuples

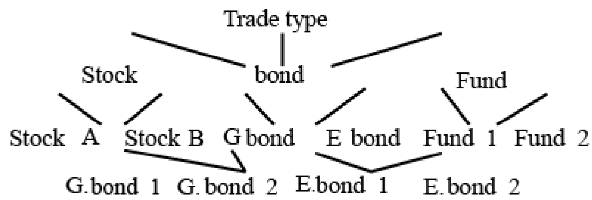

Fig. 2. Unordered hierarchical structure

S_i and S_j. If the two eigenvalues are the same, the value of $d_c\left(a_{ik},a_{jk}\right)$ is 0. Vice versa, it is 1.

(2)Tax preference and coordination of ordered eigenvalues:

$$d_l\left(a_{ik},a_{jk}\right)=\left|a_{ik}-a_{jk}\right|. \tag{2}$$

$d_l\left(a_{ik},a_{jk}\right)$ represents the value of the kth feature of both $S_i$and S_j is derived from the corresponding tax concessions and tax coordination of the set D_l: $a_{ik},a_{jk}\in A_k(i\neq j)$. Take the deposit of the customer as an example, that is, Deposit$=\{a_1,a_2,a_3,a_4,a_5,a_6\}=\{$'$few$, 'relatively few, 'medium,'' relatively many', 'many', 'a lot of The tax preference and the tax revenue coordination is $d_l\left(a_1,a_2\right)=$ 2, namely tax preference and the tax revenue coordination at the level of 'less' and 'medium' is 2. The tax preference and the tax revenue coordination $d_l\left(a_1,a_1\right)$=0, meaning that' less' tax preference and tax coordination is 0.5%.

(3) The definition of tax preference and tax coordination $d\left(S_i,S_j\right)$ between two tuples $S_i,S_j\in T$ is:

$$d\left(S_i,S_j\right)=Dis\text{−}o\left(S_i,S_j\right)+Dis\text{−}c\left(S_i,S_j\right)=\sum_{k=1}^{k_o}U_k\left|a_{ik}-a_{jk}\right|+\sum_{k=1}^{k_c}U_kd_c\left(a_{ik},a_{jk}\right). \tag{3}$$

$Dis-o\left(S_i,S_j\right)=\sum_{k=1}^{k_o}U_k\left|a_{ik}-a_{jk}\right|$ refers to the sum of all the ordered characteristic values corresponding to the tax preference and coordination between$S_i$and S_j. k_o is the number of ordered features among the characteristics contained in S. $d\left(S_i,S_j\right)=Dis-o\left(S_i,S_j\right)=\sum_{k=1}^{k_c}U_kd_c\left(a_{ik},a_{jk}\right)$ is the total sum of tax preference and tax coordination corresponding to the disordered characteristics between two tuples, in which there are k_c disordered features, and U_k represents the weight factor vector.

2.3. Guidelines for the analysis of tax policy adjustment

Suppose that $Y=\{Y_1,Y_2,...,Y_p\}$is a set of $p$adjustment centers. In$Y_i=\langle y_1,y_2,...,y_m\rangle$, there are $k_o$ordered feature fields, and the other $k_c$feature fields belong to unordered features, which jointly constitute the adjustment center, in which there are $m\left(=k_o+k_c\right)$vectors. These vectors have the same characteristic description structure as$S_i\in T$, i.e. the characteristic description structure between tuples to be classified is the same. Using the discriminant criterion $J$to measure all the$P$ adjustment centers in the distance$P$, the minimum value of tax preference and tax coordination is obtained.

$$J = \min \sum_{l=1}^{p} d\left(S_i, Y_l\right), i = 1, 2, ..., n \tag{4}$$

$d\left(S_i, Y_l\right)$ represents the sum of all the features of the tax preference and coordination betweenS_i and Y_l, which can be expressed as:

$$d\left(S_i, Y_l\right) = \sum_{k=1}^{k_o} U_k \left|a_{ik} - y_{lk}\right| + \sum_{k=1}^{k_c} U_k d_c\left(a_{ik}, y_{lk}\right). \tag{5}$$

U_k represents the weight vector, and k $= 1, 2, ...,$ mis the number of features in the tuple, so that the contribution degree of each characteristic in the adjustment of tax policy can be adjusted reasonably. This adjustment process is realized through the knowledge of the tax field.

3. Tax policy adjustment analysis algorithm

Through the study of tax preference and tax coordination, this paper describes the similarities between tuples$S_i$or customer behavior characteristics. Through this method, the closest ones will be grouped into the same category, and the adjustment center $Y = \{Y_1, Y_2, \ldots, Y_p\}$ can use the adjustment function to get the dynamic adjustment, which can then be based on tax preference and tax coordination to automatically classify tuples in T into the nearest class. Moreover, the number of tuples is changing in the process of adjusting and dividing, which will make the adjustment center's concept value have the corresponding adjustment. This adjustment will continue until the system has stabilized. Most of data sets, in the process of dynamic data mining, need to be summed up.

3.1. Conceptualized segmentation of tax eigenvalues

There is also a need for conceptual division of tax policy adjustments in the database to form an orderly conceptual value to ensure full interpretation of the output results and ensure that the adjustment space can be fully compressed. According to the data distribution form, one can determine the division method of the tax characteristics. In this paper, it is based on the data distribution density to make the tax characteristic value with discontinuous distribution get the corresponding concept value algorithm after mapping, that isCGA(conceptual generalization algorithm). Because the number of the characteristic $A_j = \{j = 1, 2 \ldots, m\}$contained in $T$is$m$, and the range of the part A_j is a real number field, using the CGAalgorithm, the corresponding values of a particular featureA_j can be mapped to obtain a relative concept value.

Through Algorithm 2.1, the CGA algorithm can be described simply:

(1)Take samples from the table$T$in a random way to get the table$T_s \subseteq T$; take into account the number of samples and the random characteristics when sampling;

(2)Characteristics A_j to be dealt with: the number G of concepts desired and the corresponding accuracy factor T can be defined in conjunction with tax domain knowledge, and usually the factor is between 5 and 10, and from this to determine the $\text{partitioninterval} = (\text{upmost} - \text{lowest})/(G \times T)$. Through the segmentation, the corresponding range of $A_j$is divided into several small intervals, and the total number of intervals is$n\ (= G \times T)$; the number of data available in each interval is counted, denoted as$\text{s} - \text{countk}$.

(3)Computation of concept segmentation: according to the order, $\text{s} - \text{countk}$ accumulates and eventually getssumj. When the value of sumjis close to the set threshold $\text{total} - \text{cnt}/G(\text{total} - \text{cnt}{=}|T_s|$, the concept segmentation will be generated. The upper and lower limits of the segment are the upper bound corresponding to the last interval and the lower bound corresponding to the first interval insumj, respectively. In this order, the segment is incorporated into the concept table for storage. This process will continue throughout the statistical interval and all characteristic values of A_j will be processed through this process.

Tax policy adjustment analysis algorithm

The implementation process of the whole algorithm can be divided into two stages. First of all, it is required to be able to clearly adjust and analyze$p$. This stage corresponds to the AATPA algorithm. The algorithm can describe the AATPA algorithm simply: select a tuple $S_i \in T$ from T, which is not yet class-divided, and construct a hyper sphere, whose radius is a the similarity threshold Simth, and count the number of tuples falling into the ball, denoted as count. If the construction meets the count threshold Nth and maintains a far distance with the center of other classes, $p$will add 1at this time.

Theorem 1: When the relevant parameters have been determined, including Simth, Nth and so on, the algorithm AATPA is able to determine the tuple similarity in the table T under the action of adjusting criterion J. Moreover, the time required will be kept within the limit of $O\ (n)$, and p adjustment centers $\text{Cluster} - \text{C}$.will be obtained.

Verification: Suppose that there are n tuples in the table$T$, and the eigenvalues corresponding to each tuple S_i are finite. If the threshold valuesSimth, Lth are set at this time, the output of AATPA algorithm can be determined because the number of different distributed classes is judged by similarity.

3.2. Generation of tax policy adjustment analysis algorithm

The AATPA algorithm ensures that the data objects that need to be processed are substantially reduced. The algorithm can be implemented in memory. Suppose that$T_1 \subset T$, in which there are $p$adjustments, denoted ask_p. Moreover, each tuple corresponds to the atomic rule and has m characteristics. In AATPA algorithm, the table $T_1$is represented as a rule set $\text{R} - \text{set0}$.

Table 1 shows part of the rule table outputted by the AATPA algorithm, which represents the law of the customer's trading behavior within a given time range.

If the similarity threshold k_1 is 5, one can obtain the corresponding inductive results:$(1,2),(3,4),(5,6,8)$and$(7,8)$; when k_2=3, the corresponding results are $(1,2,3,4)$ and $(5,6,7,8$. Figure 2 shows the tax policy adjustment analysis structure that can be obtained through the AATPA algorithm. The analysis shows that this structure can be used to obtain the characteristics of the transaction behavior of tax customers.

Table 1. Adjustment output R-set0 generated by AATPA algorithm

#	Stk−number	Trd−mode	Trd−types	Strength	Frequency	Assets	Custm−type	Loss	Samples
1	3. . 5	A	A∨Z	1. . 5	1. . 3	few	A	5. . 15	11
2	3. . 5	B	A∨G	1. . 5	1. . 3	*few*	A	5. . 15	18
3	3. . 5	A	A	5. . 12	3. . 6	many	A	5. . 15	23
4	3. . 5	B	A	5. . 12	3. . 6	medium	A	5. . 15	15
5	5. . 8	A	A	1. . 5	6. . 10	medium	A	15. . 25	8
6	5. . 8	C	A	5. . 12	6. . 10	medium	A	15. . 25	28
7	5. . 8	C	G	1. . 5	1. . 3	few	B	5. . 15	31
8	5. . 8	A	G	1. . 5	1. . 3	medium	B	15. . 25	16

(1)[custom*ertype* = *retailinvestor*][number of Transaction tax payment= 3∼5][frequency is smaller than 3][Items are less than 5]⇒[*A*ssest = small][Credibility = 30%];

(2)[*A*ssest = *samll*][cust*omertype* = *retailinvestor*][Type of transaction A or B]⇒[Loss rate of 5 to 15%][Credibility =90%];

(3)[Clients involved in treasury bond trading]⇒[Customer class=B]∧[Trading frequency = less][Credibility=34%].

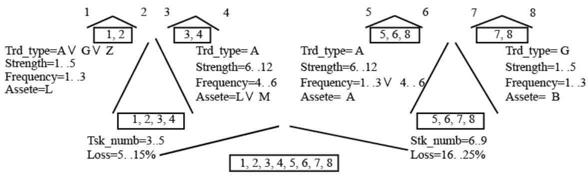

Fig. 3. hierarchical concept of the AATPA algorithm output

4. Experimental analysis

In order to analyze the feasibility of the AATPA algorithm proposed in this paper, the real tax data and simulation data of Anhui Province are used in this paper, mainly to test the execution time required by the algorithm, and the corresponding adjustment output when certain thresholds are selected differently.

Thus, 100Krecorded data are selected from it, which become the sample of this paper. Then, by random sampling, the extraction of data subsamples are conducted. The number of subsamples extracted from the sample is as follows: 5K, 10K, 50K, 100K, 500K and 1M, respectively. However, each sub-sample runs the algorithm ten times for recording and analyzing the time required by the algorithm, and calculating the mean value. Table 2 shows the scale of tax policy adjustment and its corresponding time variance. It can be obtained by analysis that the efficiency of the algorithm meets the requirements of mining.

4.1. Relationship between the execution time of the algorithm and the scale of database table $|T|$

The experimental results show that if the table $T$is increased by 10 times, the corresponding execution time will be increased by the same multiple. While generating simulated data, the basic principle is that the dimension of vector space is 20 dimensions, in which 50 adjustment points are generated randomly, and there are 20 coordinates corresponding to each adjustment point, requiring that at each point it can be generated in a positive or too distributed manner, with a value ranging from $[1, 10000]$, and all of them are integers.

Table 2. Comparison of scale of tax policy adjustment and time of implementation

Record size	5K	10K	50K	100K	500K	1000K
Average running time(s)	14.9	34.82	141.2	341.8	1524	3752
Covariance of time	3.62	3.9	4.41	7.9	11.5	24.8

Number record, average execution time average execution time difference

4.2. Influence of "One Belt and One Road" construction on the adjustment and analysis of tax policy in Anhui Province

The prominent feature of the algorithm AATPA is that, through the definition of tax preference and tax coordination, the adjustment goal is accomplished. In the whole process, the adjustment results will be significantly affected by the size of the feature range. As shown in figure 3, when there are differences in the number of conceptual segments and Simthtaken for the same data sample, the final adjustment number $p$and the diagram of the relationship between each other will be changed. Thereof, the x axis represents the number of conceptual segments$G$, and the $y$axisSimth. In addition, the $z$axis represents the influence of Gon p due to the adjustment error. If the database table$T$is given, when Gincreases, the number of elements of the resulting tax preference and tax coordination concept features range will also increase, which indicates that the user's concern about the tax preference is relatively small. Therefore, the adjustment analysis will increase at this time. On the contrary, the adjustment analysis will decrease. When Sim_{th}decreases, it indicates that there are strict similarity requirements between tuples, and then the number of categories will be increased.

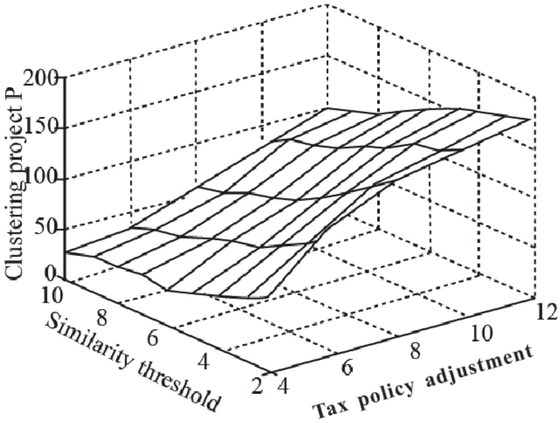

Fig. 4. Impact of tax preference and tax coordination on tax policy adjustment

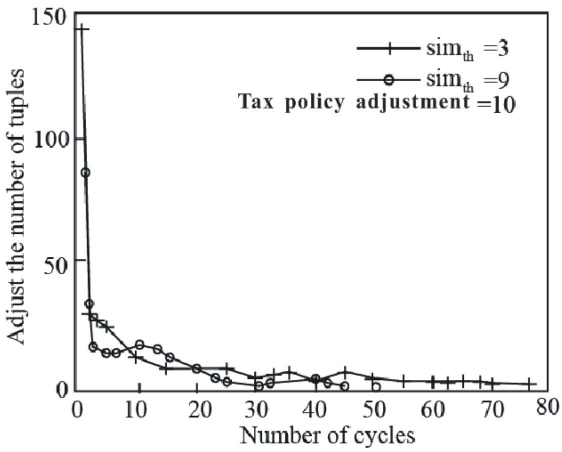

Fig. 5. Impact of tax coordination on the number of times of dynamic adjustment

5. Conclusion

At present, in order to effectively deal with the current big data environment, there is a need to strengthen the research and analysis of the data mining technology, especially the adjustment analysis method which has become an important research field. And the tax incentive and tax coordination method has relatively good application for incomplete data information or the information lacking the tax field knowledge. This paper from the perspective of tax incentives and tax coordination, analyzes the influence degree of the relevant threshold on the performance of the algorithm under the knowledge in tax field. Through the analysis of Anhui Province's real experimental data of tax, it can be obtained that: the algorithm designed in this paper is able to effectively solve this problem with a strong feasibility. In the following research, the algorithm will be improved in order to improve its versatility and make it be applied to a wider field.

References

[1] G. Brosio: *Personal taxation incentives and tax reform.* Journal of Mass Spectrometry *48* (2013), No. 9, 1015–1018.

[2] I. Yang: *The effectiveness of tax credits as a policy tool to promote r & d activity: a systematic review and meta-regression analysis.* Arthritis Research & Therapy *17* (2011), No. 1, 1–7.

[3] J. E. John: *Medicines for tropical diseases; support through tax-incentives.* Drug Discovery Today *16* (2011), No. 3, 91–92.

[4] Y. U. Jianqiu, Z. Cao: *The analysis and forecast of rmb internationalization on one belt and one road.* Applied Energy *88* (2015), No. 7, 2480–2489.

[5] T. Kuennen: *One and one is one [road construction].* Journal of Organometallic Chemistry *99* (2010), No. 3, 397–405.

[6] P. Lignier: *The role of taxation in addressing the externalities of road transport: a critical analysis of australian tax policy.* Journal of the American College of Cardiology *53* (2011) 2754–2760.

[7] L. D. Cesare, M. Sportelli: *Fiscal policy lags and income adjustment processes.* Chaos Solitons & Fractals *45* (2012), No. 4, 433–438.

[8] D. H. NgOngOla: *Analysis of policy reforms and structural adjustment programs in malawi with emphasis on agriculture and trade.* Science *161* (2010), No. 3840, 445–446.

[9] A. B. Gilmore, B. Tavakoly, G. Taylor: *Understanding tobacco industry pricing strategy and whether it undermines tobacco tax policy: the example of the uk cigarette market.* Addiction *108* (2013), No. 7, 1317–1326.

[10] G. Navratil: *Adjustment computations: spatial data analysis.* International Journal of Geographical Information Science *25*, (2010), No. 25, 326–327.

Received November 16, 2017