ISSN: 2455 - 9172

EIJO Journal of Engineering, Technology And Innovative Research (EIJO–JETIR)

Einstein International Journal Organization (EIJO)

Available Online at: www.eijo.in

Volume – 2, Issue – 2, March - April 2017, Page No. : 60 - 65

Weighted Edge Computation for Mining Tax Evasion

Shivani Korde, Pratiksha Wasekar, Rohini Patil, Amruta Kale, Rajesh H Kulkarni

Pune University, Rajashri Shahu School of Engineering Narhe, Maharashtra, India

Abstract

Tax evasion is the illegal activity done by individuals, corporations, and trusts to avoid tax. There is an evidence that an increasing the property to avoid tax in an unobserved way. At the same time, taxation information related data is classic kind of big data. The issue challenges the effective solution of traditional tax evasion detection method. To avoid this, we first investigate the tax evasion cases and apply a graph-based method to characterize their property that describes two suspicious relationship trails with a same antecedent node behind an Interest Affiliated Transaction (IAT).After finding the tax avoidance cases then heterogeneous information network is applied and using this network properties are analyzed. After that CNBN (colored network based model) is used for characterizing economic behavior and social relationship IATs between the taxpayers. This method is beneficial to improve the efficiency of tax evasion detection**.**

Keywords: Weighted edge, Graph mining, tax evasion, interest affiliated transaction, heterogeneous information network.

1. Introduction

In every country the tax collection is considered topmost priority. Many people avoid paying tax and declare wrong information about taxpaying. Tax evasion is illegal evasion of taxes which is done by individuals, corporations and trusts. .Now a day's corruption and tax evasion take place in practically every country in the world [3], and should be considered a problem everywhere. It was reported that the India losses 14 trillion rupees from tax evasion yearly. Taxation information includes large amount of data with large number of tax payers. The data mining provide powerful techniques for tax administration to extract useful knowledge from huge database. A tax administration core business is, to get the right tax at the right time from the right taxpayers, and to make the funds timely available for the right tax recipients.

The national tax information collection system (NTICS) deployed in China deals with a huge volume of transactions and the related data involved.

For example, there are more than 31,910,000 taxpayers and 48,000 taxation administration offices all over the country. The number of annual tax-related business records is up to 1 billion, the daily peak of these records is up to ten million, and the volume of annual data aggregated is 12 TB, which is self-confirmed as big data [1]. All over the country, tax authorities are under pressure to locate underreporting taxpayers, and prediction of irregular behavior of non-paying taxpayers.

The tax data is collected from a many independent sources to and perform data matching the information technology tools are been used .There are many issues in traditional data mining based methods for detection of tax evasion. The reasons are: 1) the manual labeled should be given to training data set, 2) the most important issue is that some of the covert relationships are not recorded in the database[1][3]. The tax authorities have limited resources, and traditional tax auditing

methods are time-consuming .There is a need to have further inputs to a tax avoidance database and additional information resources from big data ,the hidden relation is encountered from big data base.

After analyzing classic tax evasion cases the heterogeneous information network is form and their properties are analyzed by colored network based model for characterizing social relationship, economic behavior and IAT between tax payers. The tax evasion detection process is two phase.

The first phase is mining suspicious groups it in order to identify the doubtful trading relationships from the suspicious groups which is built from the heterogeneous information network on the bases of the CNBM, The second phase is identifying tax evasion in this traditional methods can be used on all transactions related to the suspicious trading relationships to detect tax evasion within the set of suspicious groups [1]. The objective of using data mining technique in detection of tax evasion enhances the tax productivity in collection of tax.

2. Literature Survey

2.1. Mining Suspicious Tax Evasion Groups in Big Data

In this paper they have deployed the system in taxation offices. This system Can accesses national taxation information system, manages the industry and company list, collects the data from NTCIS (national tax information collection System), analyzes them to monitor the calculation of tax model and tracks the tendency of tax index [1].

2.2. Characterization and detection of taxpayers with false invoices using data mining techniques

In this is Paper is Related to detection of taxpayers and in this paper Using the various types of data mining method. First is clustering Algorithm and neural, this two types are used to identify group of Universe of taxpayer [3].

2.3. A Delphi-based rough sets fusion model for extracting payment rules of vehicle license tax in the government sector

In this paper use the Vehicle license tax payment to solve the problem that are faced by taxation agents .The taxation is consist of two Types National Tax and Local tax. The local tax problem consisting the vehicle license tax, House tax, land value increment tax, and some special taxes [5].

2.4. Destination Taxation and Evasion: Evidence from U.S.Inter-State Commodity Flows

Traders are Variety of taxes, two of which are levied at destination - sales taxes and corporate income taxes. Sales taxes are collected by the seller, if the seller has the destination state, or it not collected by the seller are reportable and payable to the buyer .then seller not planning to avoid the legal sales tax And Income tax are apply to taxable income.

3. Existing System

3.1. Database

Firstly, the information is collected from a huge database that is required for analyzing the tax evasion cases .Data is like collection of financial, receipt, stock, police related data, admin.Then this data is given to the TPIIN (Taxpayer Interested Interacted Network).

3.2. Taxpayer Interested Interacted Network (TPIIN):

In this method the relationship is extracted between different edges which affiliated to one another. For finding the relationship trails among different properties edge computation is applied. After that the algorithms for constructing the pattern tree on the database to find the suspicious tax evasion groups.

© 2017 EIJO, All Rights Reserved

3.3. Heterogeneous Information Network

For example, an information network based on relationships between legal persons and taxpayers is a homogeneous network. These mined homogeneous networks are merged into a heterogeneous information network that is modeled according to CNBM.

3.4. Detecting Suspicious Relationships

Then, a suspicious group detection method is applied to discover all relevant patterns in the heterogeneous information network. Finally, the tax evader is identified after tax evasion judgment methods.

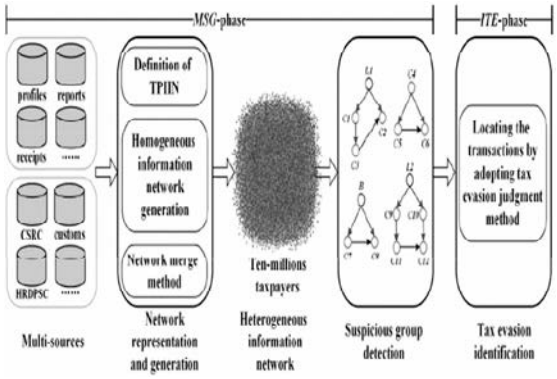

Fig.1: Existing System architecture

3.5. Clustering Of Similar Groups Of Taxpayers

Clustering is nothing but the grouping of similar type of elements. In this taxpayers clusters are formed which are related to or are affiliated to one another.

3.6. Identifying The Tax Cheater

From different clusters the taxpayers are identified.

4. Proposed System

4.1. Tax Oriented Data

Tax oriented data include the information related to tax. It contains the information of particular person or user. Personal details like the identity of person, real name, data of birth, financial details like monthly or yearly income, and property details like. In tax oriented data the overall information about user is gathered.

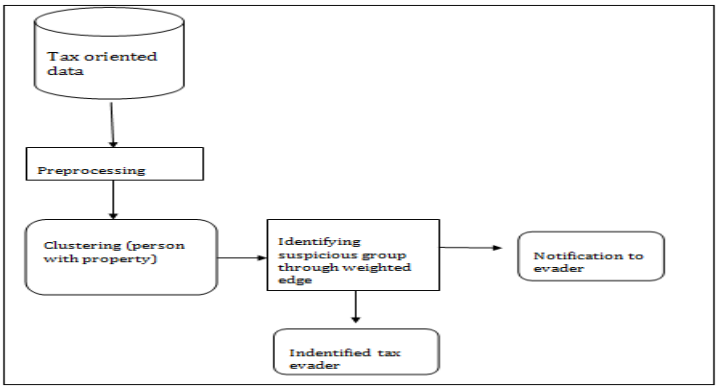

Fig.2: System architecture

© 2017 EIJO, All Rights Reserved

4.2. Preprocessing

In preprocessing knowledge discovery is applied to extract the useful knowledge from large data. In Data preprocessing contains various task like data cleaning means correcting the corrupted data , data normalization means removing redundant data, transformation means mapping of data, feature extraction reducing the amount of resources required to describe a large set of data and selection means selecting the relevant features.

Here, in preprocessing various operations are performed on collected data to avoid the anomalies like error removal, removal of noisy data, removal of redundant data.

4.3. Clustering

Clustering means grouping of data with similar properties in one cluster. In our system clustering is done on the basis of persons property .Here clustering is done to validate the property means it is illegal or legal, tax is paid or not paid.

4.4. Identifying suspicious groups through weighted edge computation

Weighted edge computation shows the relationship between one or more nodes means they are related or affiliated with each other or not. Doubtful data is extracted using weighted edge computation. Here the relationship between the affiliated properties is found.

4.5. Identified Tax Evader

In this the users who avoid the tax are detected on the basis of patterns evaluated using clustering and weighted edge.

4.6. Notification to Evader:

Here tax is calculated and notification is send to the users who had not paid tax.

5 Result

The experimental result shows the mapping of user with their property and total tax calculation of user as shown in fig 5.1 after that notify user about tax to be paid as shown in fig 5.2 .graphical representation is done with graph of users tax .

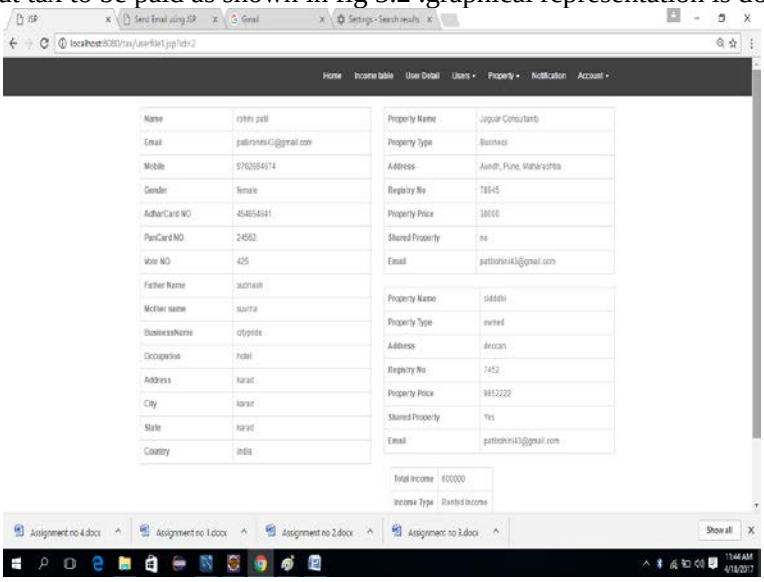

Fig 3 :user is cluster with property

© 2017 EIJO, All Rights Reserved

Fig 4: showing results of notification

6. Conclusion

In This paper the efficiency is improved of detecting the IATs based tax .Weighted edge computation is used to uncover complex relationship. It not only utilizes multiple homogeneous relationships in big data to form the heterogeneous information network, but also maximally utilizes the advantage of trail-based pattern recognition to select the suspicious groups. we simplify the heterogeneous information network into a colored model with two node colors and two edge colors. Moreover, after investigating three cases, we conclude that to identify two suspicious relationships.

7. References

[1]. Mining Suspicious Tax Evasion Groups in Big Data Feng Tian, Tian Lan, Kuo-Ming Chao, Nick Godwin, Qinghua Zheng, Nazaraf Shah, Fan Zhang.

[2]. M. J. Ferrantino , Z. Wang and X. Liu, “Evasion Behaviors of Exporters and Importers” Journal of International Economics, vol. 86, no. 1, pp. 141–157, Jan. 2012.

[3]. J. D. Vel´l´csquez and P. C. Gonz´l´cleza, “Characterization and Detection of Taxpayers with False Invoices Using Data Mining Techniques,” Expert Systems with Applications, vol. 40, no. 5, pp. 1427–1436, Apr. 2013

[4]. D. Hristu-Varsakelis, A. Saraidaris and N. Goumagias, “A Decision Support Model for Tax Revenue Collection in Greece,” Decision Support Systems, vol. 53, no. 1, pp. 76–96, Apr. 2012.

[5]. C.-H. Cheng and Y. S. Chen, “A Delphi-Based Rough Sets Fusion Model for Extracting Payment Rules of Vehicle License Tax in the Government Sector,” Expert Systems with Applications, vol. 37, no. 3, pp. 2161–2174, Mar. 2010.

[6]. D. Hristu-Varsakelis, N. Goumagias and A. Saraidaris, “A Decision Support Model for Tax Revenue Collection in Greece,” Decision Support Systems, vol. 53, no. 1, pp. 76–96, Apr2012.

[7]. Z. Gao, “Transfer Price-Based Money Laundering: A Transit Trader’s Perspective,” in Proc. 4th Int. Conf. Wireless Communications, Networking and Mobile Computing (WiCOM), pp. 1–5, Oct. 2008.

[8]. M. Kamber, J. Han, and J. Pei, Data Mining: Concepts and Techniques, Burlington, Massachusetts: Morgan Kaufmann, 2011.

[9]. R.-S. Wu, C. Ou, H. ying Lin, S.-I. Chang, and D. C. Yen, “Using Data Mining Technique to Enhance Tax Evasion Detection Performance,” Expert Systems with Applications, vol. 39, no. 10, pp. 8769–8777, Aug. 2012

© 2017 EIJO, All Rights Reserved

[10]. Yizhou Sun and Jiawei Han, "Mining Heterogeneous Information Networks: A Structural Analysis Approach," SIGKDD Explorations, vol. 14, no. 2, pp. 20-28, Apr. 2013

[11]. S. Xu, "An Application on Association Rules Data Mining in the Tax Audit System," Economic Supervision, vol. 13, pp. 43–44, Nov. 2011.

[12]. Tarjan Robert, "Depth First Search and Linear Graph Algorithms," SIAM Journal on Computing, vol. 1, no. 2, pp. 146– 160, Jun. 1972.

[13]. Yizhou Sun and Jiawei Han, "Meta-Path-Based Search and Mining in Heterogeneous Information Networks," Tsinghua.

[14]. Science and Technology, vol. 18, no. 4, pp. 329-338, Aug. 2013.

© 2017 EIJO, All Rights Reserved