

More Focus on Tax Evasion Detection with Graph Based Approach

Komal Rokade¹, Rashmi Mane²

¹(CST, Department of technology, Shivaji University, India)

²(CST, Department of technology, Shivaji University, India)

Abstract: A tax is the source of government funding. The purpose of tax is to increase revenue to fund government. The money paid by taxpayers in taxes goes to many places. It is used to paying the salaries of government workers, tax money also help to support common resources, such as police and firefighters. Tax money helps to ensure the roads you travel on are safe and well-maintained. Taxes fund public libraries and parks. Tax evasion is increased so tax evasion detection is very important in current status to avoid loss of government funding. Taxpayers are required to store and update, on an annual basis, a set of documents and information relating to international transactions or specified domestic transactions. In recent work on tax evasion detection is done but it is not effective some drawbacks are there. This article gives an introduction to related work done in tax evasion detection and describes the methods of tax evasion. Auditing is very important to find out tax evasion, and data mining techniques are applied to select record for audit, also data mining techniques are applied in tax evasion detection.

Keywords: Group mining, Tax evasion, TPIIN, trading data

I. Introduction

Introduction

The main reason that data mining has attracted a great deal of attention in information industry in recent years is due to the wide availability of huge amounts of data and the need for turning such data into useful information and knowledge. Data mining is nothing but the extracting or mining knowledge from large amounts of data. Many people uses data mining as an alternate for term, Knowledge Discovery in Databases", or KDD. Alternatively, data mining is essential and important step in the process of knowledge discovery in databases.

Tax evasion and tax fraud have been a constant issue for tax administrations, especially when pertaining to developing countries. While it is true that taxes are the source of government earning, the reality is that they send a very important signal about the commitment and effectiveness with which the State can carry out its functions and restrict access to other sources of income.

Tax evasion is illegal evasion of taxes by individuals and corporations. The number of annual tax related business records is up to 1 billion, the daily peak of these records is up to ten million. This volume of data challenges traditional data mining based methods of tax evasion. The results of the clustering based and neural network based methods are not explainable and their tax evasion identification efficiency is low. When talking about the properties of big data, traditional data mining-based methods have their limitations. The classification-based methods need a set of sample data for training, which means the data need to be manually labeled before training takes place. Moreover, the trained model is sensitive to the sample data and will be out-of-date if behaviors in tax evasion change. In addition, the results derived from clustering-based methods and neural network-based methods are difficult to explain and trace. The worse thing is that the above mentioned data-mining-based methods need to search and evaluate each transaction in the tax-oriented big data before reliable outcomes can be derived.

The proposed method is more effective and efficient than the existing approaches, as it aims to select the suspicious relations first via other related data sources and then identify those suspicious transactions.

II. Related Work

“Destination taxation and evasion: Evidence from U.S. inter-state commodity flows” in 2014 by W. F. Fox, L. Lunab, and G. Schau,[1] in that they developed a new way to examine tax evasion that focuses on corporate transactions, rather than corporate profits. Specifically, they examine how commodity flows respond to destination sales taxes, allowing for tax evasion as a function of distance between trade partners. After accounting for transportation costs, they find that the effect of taxes decreases as distance increases. Due to this way longer distances between trade partners avoid smooth government supervision and increase the

chances of tax evasion. Our results are robust with respect to outliers, strategic neighbor effects, information sharing agreements and other re-specifications.

In Apr 2013 Pamela Castellón González a, Juan D. Velásquez b, purposed “Characterization and detection of taxpayers with false invoices using data mining techniques”,[2] in that the classification and clustering methods used to characterize the taxpayers who have good or bad financial behavior associated with the use of false invoices (bill of sale).It is possible to identify some distinguishing characteristics between one group and another, according to what happens in reality. Particularly the neural gas method found that it was possible to identify some relevant variables used for differentiating behavior is good or bad, not necessarily associated with the use and sale of false invoices. Clustering algorithms like SOM and neural gas are used to identify groups of similar behavior in the universe of taxpayers. But the result of clustering based and neural network based methods are not explainable and they are not sensitive. Computer based case selection means data mining techniques are useful in tax evasion detection.

Roung-Shiunn Wua, C.S. Ou b, Hui-ying Lin b, She-I Chang b, David C. Yen c, Purposed “Using data mining technique to enhance tax evasion detection performance” in august 2012.[3] In that the operation of tax evasion detection is divided into two sub functions tax evasion prevention and tax evasion detection. The goal of tax detection is to locate defective tax reports. However, the current study has some limitations. Due to budget limitation, the current study used IBM DBMiner 2.0 as the data mining tool, rather than more advanced software. Moreover, this study only filtered out suspicious tax evasion case without processing real auditing. Actually here is limited resources and traditional tax auditing methods are time consuming and respective.

You-Shyang Chen a, Ching-Hsue Cheng b purposed “A Delphi-based rough sets fusion model for extracting payment rules of vehicle license tax in the government sector” in 2010.[4] One of the most commonly used qualitative forecasting techniques is the Delphi method that is proposed by Anderson, Sweeney, & Williams in 1998. The Delphi method used to tackle with real world decision- making problems is a systematic, interactive forecasting method, which relies on a panel of independent experts without face-to-face meetings taking place. The proposed hybrid model provides an alternative approach for those who intend to use intelligent technology in the government sector, because it effectively discovers hidden information from data and improves some service qualities for achieving better performance to unfold the derivative problems of fast-growing demand of vehicles.

III. Proposed Graph based approach

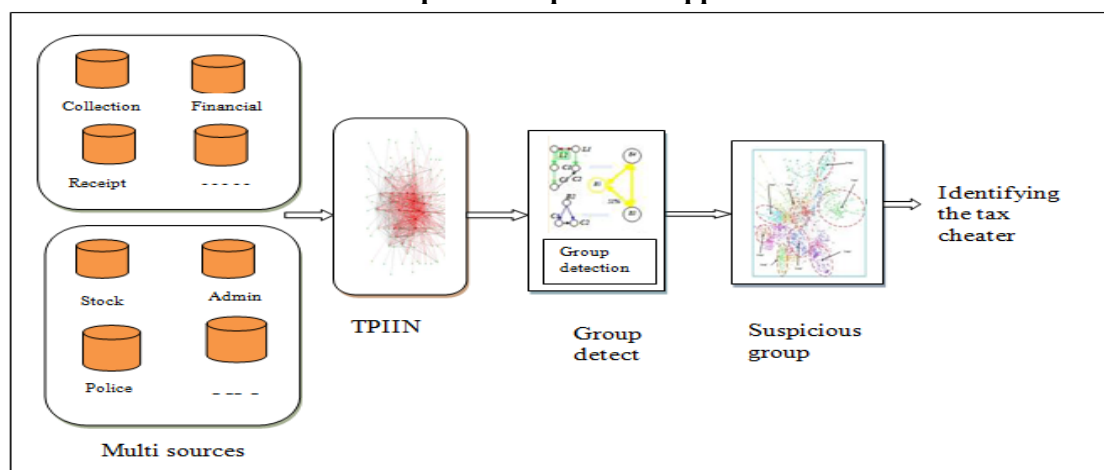


Fig 1: System architecture for proposed system

Proposed work of finding suspicious group is based on graph theory

It contains two main phases

1) Building TPIIN

Using database that contains company, director and transaction database graph will be generated. Graph shows nodes with different colors. Companies and directors will show with nodes and edges shows relation between nodes.

2) Finding suspicious groups

In this phase graph will be taken as input. That graph is called taxpayers interest interacted network. Using that heterogeneous network patterns will be generated and matched to find out suspicious tax evasion groups.

Analysis of algorithm

INPUT: Database (company, directors, transaction data).

OUTPUT: Suspicious groups

Step 1: Create graph from database.

Step 2: abstract trading relationship and saving in trading list.

Step 3: abstract antecedent relationship save as a antecedent list data.

Step 4: Generate patterns from graph

Step 5: pattern matching algorithm applied on patterns to find out suspicious groups.

Step 6: Return suspicious groups

Fig 2: mining suspicious group’s algorithm for proposed approach

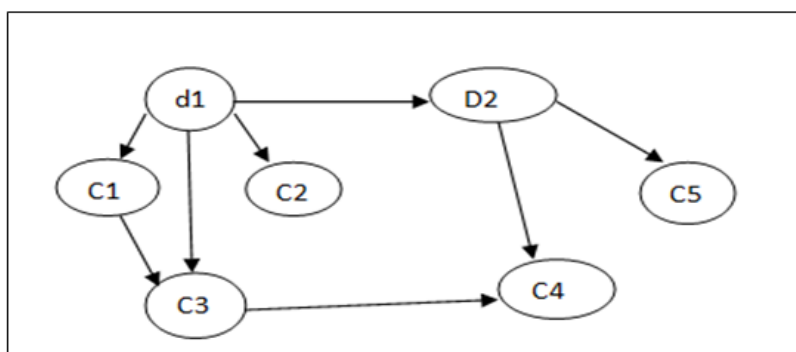


Fig 3. TPIIN graph

Step2. By using above generated graph trading data and antecedent data is generated.

Trading data

1	C1	C3
2	C3	C4

Antecedent data

1	D1	C1
2	D1	C2
3	D1	C3
4	D1	D2
5	D2	C4
6	D2	C5

Step3. Patterns generated using Trading and Antecedent

- d1->c1->c3
- d1->c3->c4
- d1->c2
- d1->d2->c4
- d1->c1->c3->c4

Step4. Patterns d1->d2->c4 and d1->c1->c3->c4 are matched. It has suspicious group.

IV. Conclusion

Tax evasion is illegal evasion of payable taxes. Due to large database finding tax evasion is difficult for tax administrator. Auditing and Tax inspection is important and effective but checking all records is time consuming. Large volume of data is challenge for traditional data mining methods Graph based approach helps to find out tax evasion. Graph based approach is based on graph theory. It finds out patterns from data and suspicious groups of tax evasion using pattern matching. So to find out tax evasion it is very useful and reduces time with increased accuracy.

References

- [1]. W. F. Fox, L. Lunab, and G. Schau, "Destination Taxation and Evasion: Evidence from U.S. Inter-State Commodity Flows," *Journal of Accounting and Economics*, vol. 57, no. 1, pp. 43–57, Feb. 2014.
- [2]. P. C. Gonzalez and J. D. Velázquez, "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.
- [3]. 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.
- [4]. Y.S. Chen and C.-H. Cheng, "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.
- [5]. J. Han, M. Kamber, and J. Pei, *data mining: concepts and techniques*, Burlington, Massachusetts: Morgan Kaufmann, 2011.
- [6]. J. Hasseldine and G. Morris, "Corporate Social Responsibility and Tax Avoidance: A Comment and Reflection," *Accounting Forum*, vol. 37, no. 1, pp. 1–14, Mar. 2013.
- [7]. *Graph theory* by Keijo Ruohonen (Translation by Janne Tamminen, Kung-Chung Lee and Robert Piché) in 2013.
- [8]. Han, J., Pei, J., Yin, Y., & Mao, R. (2004). Mining frequent patterns without candidate generation: A frequent-pattern tree approach. *Data Mining and Knowledge Discovery*, 8, 53–87.
- [9]. Yizhou Sun and Jiawei Han, "Mining Heterogeneous Information Networks: A Structural Analysis Approach," *SIGKDD Explorations*, vol. 14, no. 2, pp. 20–28, Apr. 2013.
- [10]. R. Torrini, "Cross-Country Differences in Self-Employment Rates: the Role of Institutions," *Labour Economics*, vol. 12, no. 5, pp. 661–683, Oct. 2005.
- [11]. P. Sikka and H. Wilmott, "The Dark Side of Transfer Pricing: Its Role in Tax Avoidance and Wealth Retentiveness," Apr. 2010.
- [12]. *Tax Evasion and Equity Theory: An Investigative Approach* by Sharmila King Steven M. Sheffrin in aug 2002.
- [13]. *Graph theory* by Keijo Ruohonen (Translation by Janne Tamminen, Kung-Chung Lee and Robert Piché) in 2013.
- [14]. J. Han, M. Kamber, and J. Pei, *data mining: concepts and techniques*, Burlington, Massachusetts: Morgan Kaufmann, 2011.