



## Using data mining technique to enhance tax evasion detection performance

Roung-Shiunn Wu<sup>a</sup>, C.S. Ou<sup>b</sup>, Hui-ying Lin<sup>b</sup>, She-I Chang<sup>b</sup>, David C. Yen<sup>c,\*</sup>

<sup>a</sup> Department of Information Management, National Chung Cheng University, 168 University Rd., Min-Hsiung, Chia-Yi, Taiwan, ROC

<sup>b</sup> Department of Accounting and Information Technology, National Chung Cheng University, 168 University Rd., Min-Hsiung, Chia-Yi, Taiwan, ROC

<sup>c</sup> Department of DSC and MIS, Miami University, 2042C, FSB, Miami University, Oxford, OH 45056, USA

### ARTICLE INFO

#### Keywords:

Data mining  
Value-added tax  
Tax evasion  
Association rule

### ABSTRACT

Currently, tax authorities face the challenge of identifying and collecting from businesses that have successfully evaded paying the proper taxes. In solving the problem of tax evaders, tax authorities are equipped with limited resources and traditional tax auditing strategies that are time-consuming and tedious. These continued practices have resulted in the loss of a substantial amount of tax revenue for the government. The objective of the current study is to apply a data mining technique to enhance tax evasion detection performance. Using a data mining technique, a screening framework is developed to filter possible non-compliant value-added tax (VAT) reports that may be subject to further auditing. The results show that the proposed data mining technique truly enhances the detection of tax evasion, and therefore can be employed to effectively reduce or minimize losses from VAT evasion.

© 2012 Elsevier Ltd. All rights reserved.

### 1. Introduction

Tax revenue is one of the most necessary financial resources of a government for accomplishing specific goals. However, some businesses often attempt to evade their payment of correct taxes. Consequently, tax evasion creates a critical impact on the budgetary income of these businesses and of the government. These businesses incur additional social costs because they spend their valuable resources in finding means to evade taxes, rather than focus on their operations. On the side of the government, tax authorities have to bear the costs of the detection and prevention of illegal tax evasion activities. As a result, effective ways to detect related tax evasion activities have always been an important and challenging issue for tax authorities in any country.

If the government cannot effectively detect illegal tax evasion activities, public investment would be negatively affected due to the budgetary shortage resulting from the loss of tax revenues. VAT (value-added tax) evasion is one of the important issues for many tax authorities. Gebauer, Nam, and Parsche (2007) report that, based on German data, the VAT revenue gap derived from the comparison of the quantified, hypothetical, and the actual collected revenues increased from 5.1% in 1995 to 7.5% in 1996. They also estimate that VAT revenue losses were approximately EURO 18 billion in 2001 for Germany alone.

Gebauer et al. (2007) also suggest that VAT evasion not only leads to significant revenue losses, but also to a considerable increase in administrative costs used to detect the illegal tax evasion activities. In addition to the decreased tax revenue and increased administrative costs, VAT evasion also has a significant negative side effect on the collection of income taxes from businesses. This occurs because VAT evasion, which implies an indirectly underreported taxable income from business, is often directly accompanied by underreported sales revenues.

In order to realize the benefits of spending valuable, albeit limited, resources to detect VAT evasion, tax authorities need to deploy their resources wisely. As such, tax authorities have often relied on the sampling method and the personal judgment of tax auditors in selecting suspicious tax reports to audit for potential tax evasion activities. Thus, the purpose of the current study is to determine a more scientific approach to improve tax auditor's productivity and performance in handling the detection tasks of VAT tax evasion.

All over the world, tax authorities are under increasing pressure to locate underreporting taxpayers, collect additional tax revenues, and predict the irregular behavior of non-paying taxpayers. Without the assistance of information technology tools, most tax authorities need to pull in tax data from a variety of independent sources or perform data matching and checking with other sources to find cases of non-compliance. As a result, tax evasion detection performance has been rather limited.

Business intelligence (BI) in general, and data mining in particular, may be effective tools for enhancing the efficiency and effectiveness of the detection of illegal tax evasion (Fadaio, Williams, Trotman, & Onyekelu-Eze, 2008). In the US, Texas was one of the

\* Corresponding author. Tel.: +1 513 529 4827; fax: +1 513 529 9689.

E-mail addresses: [roungwu@ccu.edu.tw](mailto:roungwu@ccu.edu.tw) (R.-S. Wu), [actcso@gmail.com](mailto:actcso@gmail.com) (C.S. Ou), [enya@email.ntact.gov.tw](mailto:enya@email.ntact.gov.tw) (H.-y. Lin), [actsic@ccu.edu.tw](mailto:actsic@ccu.edu.tw) (S.-I. Chang), [yendc@muohio.edu](mailto:yendc@muohio.edu) (D.C. Yen).

first states to apply data mining techniques for detecting suspicious tax evasion reports and thereby recoup unpaid taxes (Hoover, 2009). Songini (2004) reports that Texas uses a BI system that is able to flag a situation in which a business is suspected to be evading taxes. This suspicious tax report is referred to an audit staff for further investigation. Since the introduction and application of the BI system, USD 362 million of tax losses have already been recovered. The tax authority in Texas has also committed strongly to data mining for spotting suspicious tax reports. As cited by Songini (2004), Lisa McCormack, a manager in the tax audit division in Austin, Texas, claims, “We only audit 1% of the taxpayers. . . We have to try and figure out how to make the best use of the [government’s investigative] resources.”

The current study intends to utilize data mining as a tool to enhance tax evasion detection performance. Data mining is a methodology used to discover hidden information from rough data (Fayyad, Piatetsky-Shapiro, & Padhraic, 1996; Yoon, 1999). It can be applied in the process of decision support, prediction, forecasting, and estimation (Liao, 2003). Moreover, data mining techniques are able to efficiently handle a large number of records and data (Ravisankar, Ravi, Raghava Rao, & Bose, 2011). Compared to general statistics, data mining is able to identify certain patterns and match specific data via efficient computing technology. In other words, the interpretation of data allows flexibility (Liao, 2003).

This study employs the association rule of the data mining technique to the VAT database to uncover patterns and relationships among attributes that are useful for identifying problematic tax evasion reports. In this research, a screening model will be developed based on specific patterns or rules discovered from identified VAT evasion tax reports. This screening model is utilized to select the cases that are suspected to be non-compliant VAT reports for further auditing checks. In other words, the goal of using data mining as a technique in detecting VAT evasion enhances the tax auditor’s productivity in recovering tax revenue losses.

The current paper is organized as follows. After the introductory section, Section 2 provides a literature review. Section 3 illustrates the proposed framework. Sections 4 and 5 discuss the design and development of the screening model and the experimental results, respectively. Finally, Section 6 provides the conclusion, including the limitations of this study and future implications.

## 2. Literature review

### 2.1. Value-added tax evasion detection in Taiwan

Keen and Lockwood (2010) in exploring the causes and consequences of the remarkable worldwide attention given to VAT in recent years, find that more than 130 countries have implemented the VAT scheme. In addition, VAT has raised 20% or more of all tax revenues in those countries. Their estimated figures also suggest that the adoption of VAT contributes positively in the establishment of an effective tax system for most countries under this study. They argue that, “By any standards, the rise of the VAT has been the most significant development in tax policy and administration of recent decades.”

The goal of VAT is to collect taxes on the difference between receipts from sales and expenditures on purchases for each business transaction. According to the Department of Investment Services in Taiwan (2010):

“All sales of goods and provision of services in Taiwan, as well as all imports of goods into Taiwan, are subject to Business Tax (Sales Tax). . . The sale of goods is defined as the transfer of goods to another entity for consideration in Taiwan. . . The sale of services is defined as the supply of services to others or the provision of goods for the use, production of earnings

by others for a consideration. . . The import of goods includes all goods that are imported into Taiwan from a foreign country.”

In addition, tax authorities in Taiwan use Input Documentary Evidence (IDE) as a receipt from sales and Output Documentary Evidence (ODE) as an expenditure on purchase. In a business transaction, the buyer obtains the IDE and the seller possesses the ODE. According to statistics from the Ministry of Finance (2011), VAT is an important tax income among overall taxes, ranking only second to income tax from 2003 to 2010. However, according to Huang and Lin (2009), in 1999–2001, more than 35% selected cases in Taiwan involved VAT evasion. Clearly, tax evasion is a serious problem in Taiwan.

One of the techniques that tax authorities traditionally use to detect VAT evasion is to conduct a cross-matching of the IDE and the ODE. Due to upstream to downstream interdependence, the auditing of a seller’s VAT report is usually useful for auditing the validity of a buyer’s VAT reports in detecting VAT evasion. Once a mismatch between the costs in IDE and the sales in ODE is detected, penalties can be imposed on the guilty taxpayers.

The current cross-matching processes are done manually by tax auditing staff with the help of simple computer software tools. This task can be very tedious and time-consuming. In addition, the effectiveness of this auditing task heavily relies on the experience and skills of the tax auditing staff. Dealing with the huge volume of IDEs and ODEs and processing all VAT reports in an efficient and effective manner might be impossible. In reality, tax authorities are able to screen only a small percentage of tax reports for further auditing because of limited staff resources. Consequently, a systematic approach to identify the VAT reports with high tax-evasion potential among all tax reports becomes a critical and timely necessity for tax authorities.

Gebauer et al. (2007) have examined VAT evasion caused by carousel fraud. Their results indicate that firms in the EU repeatedly carry out cross-border transactions. Carousel fraud actually leads to the fraudulent retention of revenues. Their study examines three reform models of the VAT system in Germany for detecting VAT evasion. They argue that, under the new system, many small cases of fraud can cause an enormous amount of tax losses, which may actually generate a more severe damage than a few large cases. In addition, the administrative overhead, the associated personnel, and equipment costs may be unintentionally underestimated. Therefore, they suggest that, before a radical change in the VAT system is made, tax authorities should run available options to make the present VAT system more effective in the prevention and detection of tax evasion.

Some taxpayers intentionally become involved with some optimal degree of tax evasion. Gupta (2008) analyzes the relationship between tax evasion behavior and taxation policy. He further suggests that, from a policy perspective, increasing the penalty rates may turn to be the best way to reduce tax evasion when taxpayers face budgetary income pressures.

### 2.2. Data warehouse technologies

More and more companies are using BI tools to analyze sales and other related transactional data to detect fraud. Software companies, such as SAS Institute Inc., SPSS Inc., NCR’s Teradata, and IBM’s Cognos Business Intelligence all provide fraud-detection oriented BI tools. Some US government agencies are also employing BI tools to detect tax evasion. According to Lisa McCormack, a manager in the audit division in Austin, the comptroller’s office in Texas relies on a data warehouse tool to check for sales tax compliance.

A data warehouse is a subject-oriented, integrated, time-variant, and non-volatile dataset (Inmon, 1996). Data warehousing is utilized to collect data from different sources that can then be or-

متن کامل مقاله

دریافت فوری ←

**ISI**Articles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات