Zhu, Xulyu; Yan, Zheng; Ruan, Jianfei; Zheng, Qinghua; Dong, Bo

IRTED-TL An Inter-Region Tax Evasion Detection Method Based on Transfer Learning

Published in:
Proceedings - 17th IEEE International Conference on Trust, Security and Privacy in Computing and Communications and 12th IEEE International Conference on Big Data Science and Engineering, Trustcom/BigDataSE 2018

DOI:
10.1109/TrustCom/BigDataSE.2018.00169

Published: 05/09/2018

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Please cite the original version:
Zhu, X., Yan, Z., Ruan, J., Zheng, Q., & Dong, B. (2018). IRTED-TL An Inter-Region Tax Evasion Detection Method Based on Transfer Learning. In *Proceedings - 17th IEEE International Conference on Trust, Security and Privacy in Computing and Communications and 12th IEEE International Conference on Big Data Science and Engineering, Trustcom/BigDataSE 2018* (pp. 1224-1235). Article 8456038 (IEEE International Conference on Trust, Security and Privacy in Computing and Communications). IEEE.
https://doi.org/10.1109/TrustCom/BigDataSE.2018.00169

IRTED-TL: An Inter-Region Tax Evasion Detection Method based on Transfer Learning

Xulyu Zhu[1,2], Zheng Yan[3,4], Jianfei Ruan[1,2], Qinghua Zheng[1,2], Bo Dong[5,6]
[1] Shaanxi Province Key Laboratory of STN Tech. R&D, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China
[2] School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China
[3] State Key Laboratory on Integrated Services Networks, School of Cyber Engineering, Xidian University, Xi'an 710071, China
[4] Department of Communications and Networking, Aalto University, Espoo 02150, Finland
[5] School of Continuing Education, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China
[6] National Engineering Lab for Big Data Analytics, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China
zhuxvlv@163.com, zyan@xidian.edu.cn, xjtu_jfruan@163.com, qhzheng@xjtu.edu.cn, dong.bo@xjtu.edu.cn

***Abstract*—Tax evasion detection plays a crucial role in addressing tax revenue loss. Many efforts have been made to develop tax evasion detection models by leveraging machine learning techniques, but they have not constructed a uniform model for different geographical regions because an ample supply of training examples is a fundamental prerequisite for an effective detection model. When sufficient tax data are not readily available, the development of a representative detection model is more difficult due to unequal feature distributions in different regions. Existing methods face a challenge in explaining and tracing derived results. To overcome these challenges, we propose an Inter-Region Tax Evasion Detection method based on Transfer Learning (IRTED-TL), which is optimized to simultaneously augment training data and induce interpretability into the detection model. We exploit evasion-related knowledge in one region and leverage transfer learning techniques to reinforce the tax evasion detection tasks of other regions in which training examples are lacking. We provide a unified framework that takes advantage of auxiliary data using a transfer learning mechanism and builds an interpretable classifier for inter-region tax evasion detection. Experimental tests based on real-world tax data demonstrate that the IRTED-TL can detect tax evaders with higher accuracy and better interpretability than existing methods.**

Keywords—tax evasion; transfer learning; interpretability; inter-region detection

I. INTRODUCTION

Tax evasion causes a large revenue loss in China. The Chinese government reported that the rate of tax revenue loss in China was more than 22 percent [1]. Especially in recent years, tax evasion measures have become more diverse and covert in China. Many companies use advanced facilities, accounting methods, and human factors to evade taxation inspections, which makes auditing work more difficult. Meanwhile, with the rapid development of economy, tax data has been growing rapidly. The number of annual tax-related business records is up to 1 billion, and the daily peak of these records is up to ten million. Such large amount of data brings tremendous pressure to the tax audit work.

National governments have taken a series of measures to detect tax evasion. Three means for tax evasion detection have been adopted by tax authorities in their daily operation: manual case selection, computer-based case selection and whistle-blowing-based selection [1]. The computer-based case selection is primarily based on machine learning techniques, which extract evasion-related features from historical data for training and obtain a model that can be used in tax audit. Thus, it is considered a semi-automatic and labor-saving method applicable in the era of big data.

However, traditional machine learning-based methods have a concrete problem in practice, inter-region tax evasion detection. Machine learning-based tax evasion detection methods assume an ample supply of training examples as a fundamental prerequisite to construct an effective tax evasion detection model in a geographical region. However, the annotation of datasets in the taxation domain tends to be expensive and time-consuming. In addition, a tax evasion detection model trained for a specific region may have a high generalization error when applied to other regions due to different economic and social conditions. Different regions have different distributions of features. Most statistical models must be rebuilt from scratch using newly collected training data. This is a big challenge in developing a universal tax evasion detection model for different regions with varying economic and social conditions. Thus, construction of an evasion detection model for a region with the help of auxiliary data from another region has become an important and challenging issue.

Transfer learning [2] is a method to use knowledge gained while solving one problem to solve a different but related problem. It lessens the need for expert experience and greatly reduces the amount of labeled data needed in a target research domain. Transfer learning has been widely applied in document classification, image recognition, speech recognition, knowledge discovery and other fields. Applying transfer learning, we can absorb auxiliary knowledge from a source region that holds adequate training data and apply it to a label-sparse target region to augment learning in the presence of regional differences caused by economic and social disparities. It can be applied for inter-region tax evasion detection.

However, there are several challenges when applying the transfer learning to inter-region tax evasion detection.

First, no existing studies on tax evasion detection are based on transfer learning. Few works have explored how transfer learning can be used in inter-region tax evasion detection. Current studies about transfer learning are difficult to directly apply to tax evasion detection due to the high accuracy and interpretability requirements in the field of taxation.

Second, due to regional differences, there are few common features between regions, which causes difficulty in the transfer process. Even with common features, their marginal probability distributions can be quite different. For example, the average age of legal representatives in China's coastal cities is 36, but it is 47 in China's inland cities.

Third, as presented by Tian et al. [1], the results of machine learning-based methods are not explainable and counterintuitive. Almost all machine learning models and transfer learning methods are black box models due to the feature mapping operation, which are vulnerable to security attacks [3] [4] [5]. Making the model interpretable is an important issue for developing a robust and stable tax evasion detection system.

In this paper, we propose an Inter-Region Tax Evasion Detection method based on Transfer Learning (IRTED-TL) to overcome the above challenges and realize inter-region tax evasion detection with high accuracy and sound interpretability. It integrates Transfer Adaboost (TrAdaBoost) [6], Transfer Component Analysis (TCA) [7] and LightGBM [8]. Specially, TrAdaBoost is an instance-based transfer learning method, which ensures the transfer ability of the IRTED-TL. LightGBM supports the interpretability and accuracy of the IRTED-TL. TCA, a feature-based transfer learning method, reduces the difference between the regions, which further optimizes the performance of the model. Therefore, the IRTED-TL can provide an effective and explainable model for a label-sparse tax evasion detection task in a target region.

In the IRTED-TL, we extract features based on random forest and Kullback–Leibler (KL) divergence. The random forest is used for feature importance extraction and the KL divergence measures the similarity between feature distributions. Based on these extracted features, we map the features with KL divergence exceeding a threshold value using TCA. Then, LightGBM is adopted to identify whether a taxpayer has exhibited tax evasion behavior based on mapped features. We circularly reweight sample weights in the source and target regions according to the classification result by applying TrAdaBoost. To evaluate the effectiveness of the IRTED-TL, experiments based on the real-world tax data of five regions in two provinces in China were conducted. The results show that the IRTED-TL can detect tax evaders with higher accuracy and better interpretability than existing methods.

The IRTED-TL is original and differs substantially from previous methods of tax evasion detection. Those methods normally use traditional machine learning techniques and assume an ample supply of training examples as a prerequisite to construct a tax evasion detection model in a specific region. The method proposed in this paper adopts transfer learning techniques by using inter-region auxiliary data to augment learning when training examples in a target region are lacking. The IRTED-TL builds an interpretable classifier to induce interpretability into the detection model so that the derived results can be traced. The main contributions of this paper can be summarized as follows:

- We propose a novel method for inter-region tax evasion detection by considering the absence of training data in a target region and the interpretability of the derived results.
- We provide a unified framework that takes advantage of auxiliary data by applying transfer learning and builds an interpretable classifier to handle tax evasion detection issues when available training samples in a target region are lacking.
- We justify the performance of the IRTED-TL through comparison with existing work based on a large real-world dataset with thirteen transfer scenarios in five regions in two provinces in China. The results show that the IRTED-TL can greatly improve the accuracy of inter-region tax evasion detection and provide better interpretability than existing work.

The remainder of the paper is organized as follows. Section 2 provides a brief review on related work. In Section 3, we formulate the problem that we aim to address and provide key notations that are used in the paper. We propose the IRTED-TL for inter-region tax evasion detection in Section 4. We describe the experimental results and provide analysis and discussions in Section 5. Finally, a conclusion is presented in the last section.

II. Related Work

In this section, we briefly review the related work on tax evasion detection and transfer learning methods.

A. Tax Evasion Detection Methods

There are three frequently used methods in tax evasion detection: manual case selection, whistle-blowing-based selection and computer-based case selection methods. Manual case selection and the whistle-blowing-based selection methods are time-consuming and tedious, and computer-based case selection methods are considered to be the most promising and comprehensive approach used by tax administrations to detect tax evasion.

Machine learning-based tax evasion detection methods, a type of computer-based case selection, learn an automatic model without expert experience from historical tax evasion detection data. Existing methods include association analysis [9], cluster analysis [10] [11] [12] [13], classification [14] [15] [16] [17], genetic algorithm [18] [19] [20], simulation [21] [22] [23], and reinforcement learning [24] [25].

For example, Pamela Castellón González et al. [13] gave evidence that it is possible to characterize and detect potential users of false invoices by using different types of data mining techniques. Moreover, they adopted clustering algorithms, such as self-organizing map and neural gas, to identify taxpayer groups with similar behaviors. Wu et al. [9] applied a data mining technique based on association rules to enhance the

performance and productivity of value-added tax evasion detection in Taiwan. Chen and Cheng [14] proposed a hybrid model that combines a Delphi method and a rough set classifier to classify vehicle license tax payment to solve real-world problems faced by taxation agencies. Junqué de Fortuny et al. [15] applied Support Vector Machine (SVM) and Naïve Bayes to detect residence fraud of taxpayers. Goumagias et al. [25] presented a dynamic, Markov-based decision support model to predict the behaviors of a risk-neutral enterprise in Greece.

However, there are two main problems of machine learning-based tax evasion detection methods. First, they use statistical techniques to identify whether a taxpayer has evaded taxes and require a set of manually labeled data contributed by auditors. Few regions in China have enough labeled data because data annotation in the field of taxation is very time-consuming work. Therefore, tax evasion models can only be trained in the few regions with abundant labeled data, and these models are not generic to other regions due to discrepancies between different regions. Second, it is difficult to explain and trace the results obtained by applying these models.

B. Transfer learning Methods

Transfer learning uses the knowledge gained in solving one problem and applies it to fix a different but related problem [2]. There are four types of transfer learning methods [2]: instance-based transfer learning [6] [26] [27], feature-based transfer learning [7] [28] [29] [30], parameter-based transfer learning [31] [32] and relation-based transfer learning [33] [34]. Our method is related to instance-based and feature-based transfer learning. The motivation is that: (1) instance-based transfer learning can preserve the original tax features and enhance the interpretability of the tax evasion detection model; and (2) feature-based transfer learning can close the gap between the source and target regions through feature mapping.

TrAdaBoost [6] is the most classic method used to solve the transfer learning problem through instance transfer. It utilizes a small amount of labeled data in the target domain to leverage the source domain data and construct a high-quality classification model for the target domain. Shi et al. [35] noted that TrAdaBoost performs poorly given improper source data, hence Yao et al. [32] proposed a method named MultiSurceTrAdaBoost to address this limitation. TransferBoost, proposed by Eaton et al. [36], considers that each source domain can be a potential component of a target domain distribution. The source domain that is drawn from the shared component of the target distribution could be used to augment the target training data. However, instance-based transfer learning works poorly when the gap between domains is large because it assumes that the source domain has available instances for the target domain.

The feature-transfer methods aim to find "good" feature representations to minimize domain divergence and classification or reduce regression model error. Pan et al. [7] proposed TCA for domain adaptation, which learns some transfer components across domains in a Reproducing Kernel Hilbert Space (RKHS) using maximum mean discrepancy. Gong et al. [29] presented a kernel-based method that takes advantage of low-dimensional structures. It models domain shift by integrating an infinite number of subspaces that characterize changes in geometric and statistical properties from the source domain to the target domain. Long et al. [30] proposed a Transfer Kernel Learning approach to learn a domain-invariant kernel by directly matching the source and target distributions in the RKHS. However, feature-based transfer learning is not interpretable and thus cannot be directly used for tasks that require interpretability.

Although these techniques have been applied and examined in different domains, studies on tax evasion detection based on transfer learning are lacking. We can construct a tax evasion detection model in the absence of labeled data in one target region using evasion knowledge extracted from other regions by applying transfer learning. One important issue that has been neglected is the relevance and differences between the features of tax evasion in various regions. Based on our survey, the literature has not yet studied how to use the relevance between regions to eliminate the effects of regional differences for inter-region tax evasion detection.

III. Problem Statement and Notations

In this paper, we focus on three main problems that are widespread in tax evasion detection. (i) How to build an accurate tax evasion detection model for a target region lacking labeled data, (ii) How to handle the discrepancies between different regions, and (iii) How to induce interpretability in a tax evasion detection model?

To address these challenges, we focus on the problem of inter-region tax evasion detection based on transfer learning. We use a transfer learning technique to extract evasion-related taxation knowledge from one region with sufficient labeled data and apply it to build a tax evasion detection model for another label-sparse target region. When extracting knowledge, the discrepancies between different regions will be filtered, leaving the knowledge that can be commonly used across regions. We combine transfer learning with an interpretable classifier to induce interpretability in the tax evasion detection model.

Next, we provide some notations and definitions.

We formulate a source region and target region. We denote the source region as S, which has abundant labeled data $D_S = (X_S^i, Y_S^i)\,(i = 1 \ldots n)$ for training, where X_S^i is the data instance and Y_S^i is the corresponding class label. Similarly, we denote the target-domain data as T, and the quantity of its training data $D_T = (X_T^i, Y_T^i)\,(i = 1 \ldots m)$ is often inadequate to train a reliable detection model.

For the definitions of "domain" and "task" in the inter-region tax evasion detection, a domain consists of two components: a feature space χ and a marginal probability distribution $P(X)$, where $X = \{q_1, \ldots, q_j\} \in \chi$ represents the tax features in each region and q_j is the j^{th} term in the feature space. $P(X)$ indicates the distribution for each feature. In the case of inter-region tax evasion detection, two different regions share some tax features but not all because the records in different regions are not identical. Moreover, they have different distributions due to economic and social disparities.

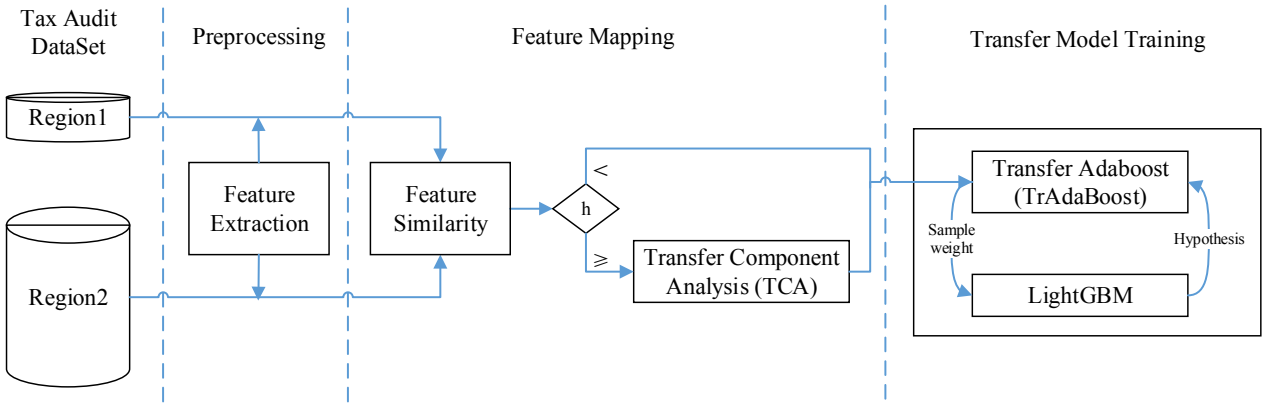


Fig. 1. The framework of the proposed IRTED-TL

Given a specific domain $\{\chi, P(X)\}$, a task expresses a label space γ and an objective function $f(\cdot)$. In the field of tax evasion detection, two regions have the same label space γ, where $y = \{0, 1\} \in \gamma$. $y = 1$ represents a taxpayer with tax evasion and $y = 0$ represents a normal taxpayer. The function $f(\cdot)$ is designed to detect whether a taxpayer evades taxes.

Our formal description of inter-region tax evasion detection is as follows. Given source region data $D_S = (X_S^i, Y_S^i)(i = 1 \ldots n)$ and target region data $D_T = (X_T^i, Y_T^i)(i = 1 \ldots m)$, where $m << n$, their domain has the following characteristics: $|X_S| \neq |X_T|$, $X_S \cap X_T \neq \emptyset$, $P(X_S) \neq P(X_T)$, and the label space γ of the source and target regions is the same. We aim to learn a detection model $h : (X_S, X_T) \rightarrow Y_T$ in the target region using training data in both the source and target regions regardless of the differences in their distributions.

We summarize the notations used in the paper in Table I.

TABLE I. NOTATIONS USED

Variable	Description
X_S	Domain of the source region
X_T	Domain of the target region
Y_S	Label set of X_S
Y_T	Label set of X_T
D_S	Data set of the source region
D_T	Data set of the target region
n	Number of instances in the source region
m	Number of instances in the target region
$P(\cdot)$	Marginal probability distribution
$\|\cdot\|$	Number of dimensions
i	Instance index
j	Feature index
x_i	i^{th} instance
y_i	Label of the i^{th} instance
w_i	Sample weight of the i^{th} instance
q_j	j^{th} term in the feature space

IV. PROPOSED SCHEME

In this section, we describe the framework of the IRTED-TL, the details of our method, and the IRTED-TL procedure.

A. *The Framework of the IRTED-TL*

The IRTED-TL is a novel inter-region tax evasion detection method based on transfer learning, which aims to construct an evasion detection model for label-sparse target region using auxiliary data from another region. It provides a unified framework that takes advantage of auxiliary data by applying transfer learning and builds an interpretable classifier to handle tax evasion detection issues in the presence of few available training samples in a target region.

The IRTED-TL integrates TrAdaBoost, TCA and LightGBM to solve the three challenges of inter-region tax evasion detection. TrAdaBoost is an instance-based transfer learning method, which ensures the transfer ability of the IRTED-TL. TCA reduces the differences between different regions, which enhances the performance of the IRTED-TL. Moreover, LightGBM provides interpretability for the IRTED-TL.

The framework of the IRTED-TL is shown in Figure 1 and is comprised of three main stages.

In the *Preprocessing* stage, we collect two tax data sets from two regions (Region1 and Region2). They are both high-dimensional semi-structured data sets and their sizes are different due to the uneven distribution. Therefore, preprocessing is required for tax data, including filling in missing values, reducing feature dimensions and extracting features from sequence data. To achieve automated tax evasion detection preprocessing, feature selection methods such as random forest [37] are adopted to automatically select the features associated with tax evasion. Then, for the convenience of training, sequence data such as transaction amounts are distilled into a static indicator.

In the *Feature Mapping* stage, features are divided into two groups according to the feature similarity between Region1 and Region2. The group with low similarity is mapped to the same feature space using TCA, which narrows the distance between the two regions.

In the *Transfer Model Training* stage, we use TrAdaBoost to circularly reweight the sample weight of each instance based

on the hypothesis calculated by LightGBM. An evasion detection model is generated when the error rate converges.

B. The IRTED-TL Framework

1) Preprocessing Stage

There are two categories of features to describe a taxpayer: static information and dynamic information. Static information indicates the properties of the taxpayer, such as the registered capital, registered address, number of employees, and age of the legal representative. Dynamic information includes time-series data in the process of interacting with other taxpayers, such as transaction amounts, transaction taxes, and billing days.

In general, tax databases have more than 10,000 features in China. However, not all these features are helpful for evasion classification. In fact, few would be helpful for tax evasion detection. Therefore, we adopt a random forest classifier to calculate the feature importance based on source data to select the most useful information from the feature space. In the random forest, we use a Gini index [38] to present the importance of features in each decision tree:

$$Gini(D,q_j) = (1-\sum_{i=1}^{n_c} p_{D_i}^2) - \sum_{v=1}^{V} \frac{|D^v|}{|D|}(1-\sum_{i=1}^{n_c} p_{D_i^v}^2) \quad (1)$$

where n_c denotes the number of classes, V denotes the set of all possible values for the feature q_j and p_{D_i} is the ratio of the i^{th} class in data D. The Gini index reflects the probability of an inconsistent category when selecting two samples from the dataset randomly. Then, we calculate the Gini Importance (GI) for each feature q_j:

$$GI = \sum_{i=1}^{n} \frac{C(i)}{|D|} Gini(q_j) \quad (2)$$

where n denotes the number of times a feature splits nodes in the random forest and $C(i)$ denotes the number of sample q_j splits. Finally, we select 200 features with the highest GI for tax evasion detection.

Note that some of these features are sequence data such as the tax amount payable, which refers to the amount of income calculated by tax authorities in accordance with the provisions of tax law and is declared once a month. We take the maximum, minimum, average and median of those features as the original data.

2) Feature Mapping Stage

Although we extracted features from the original data, the features in X_S and X_T may follow different distributions. Therefore, it is necessary to map the features into the same distribution. The Maximum Mean Discrepancy (MMD) [7] is a widely used objective in domain adaption that can dramatically minimize the distance between domain distributions by projecting the data onto the learned transfer components:

$$D(P(X_S), P(X_T)) = \left\| \frac{1}{n}\sum_{i=1}^{n} \phi(x_{s_i}) - \frac{1}{m}\sum_{j=1}^{m} \phi\left(x_{t_j}\right) \right\|_H \quad (3)$$

where $\phi(\cdot)$ denotes a mapping function that maps the original features to a high-dimensional space. The definitions of n, m, S and T are illustrated in Section III.

Next, we calculate the KL divergence between each pair of features in X_S and X_T to evaluate the similarity between features in the source and target regions:

$$KL\left(q_{s_j} \| q_{t_j}\right) = \int_{-\infty}^{\infty} s(x) \log \frac{s(x)}{t(x)} dx \quad (4)$$

where q_{s_j} denotes the j^{th} feature in the source region, q_{t_j} denotes the same feature in the target region, $s(x)$ and $t(x)$ are the probability density functions of q_{s_j} and q_{t_j}, respectively.

A low similarity feature distribution will reduce the effect of the transfer model. Thus, we select low similarity features from the feature space by setting a threshold h on the KL value and mapping them to the same distribution. If the KL of a feature is greater than h, it means that the feature has low similarity between regions. h is selected with two goals: (1) increasing h to maintain the original structure of the feature to its maximum extent, and (2) decreasing h to narrow the distance of the distribution between X_S and X_T. Our test results show that a reasonable choice of h helps improve the classifier performance.

Then, we minimize the MMD for the features with higher KL divergence than h to close the distance between the distribution of the regions. The following formula is the optimization goal for minimizing the MMD and the detailed solution is proposed in the TCA [7].

$$\min \sum_{c=0}^{C} tr\left(A^T X M_C X^T A\right) + \lambda \|A\|_F^2 \quad (5)$$

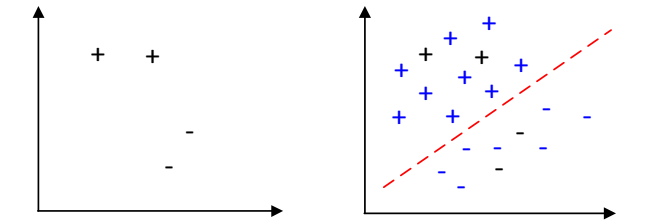
where A is the transformation matrix, X is the data that combines the source and target regions and M_C is an MMD matrix.

3) Transfer Model Training Stage

With the above steps, we closed the distance between the distribution of X_S and X_T so that auxiliary data from D_S can be reused in the classification of D_T. However, there are some instances in D_S that have a negative effect on the classification of D_T due to the regional differences. To solve this problem, we adopt TrAdaBoost to select useful instances from the source region.

If an instance from D_S is mistakenly predicted, the instance may likely conflict with D_T. TrAdaBoost decreases its training weight to reduce its negative effect. In the next round, the misclassified instances in source region, which are dissimilar to those in the target region, affect the learning process less than in the current round. After several iterations, the instances in

the source region that fit those in the target region better will have larger training weights whereas the source region training instances that are dissimilar to those in the target region will have lower weights. The instances with large training weights help the algorithm train better classifiers. TrAdaBoost can extract valuable knowledge from the source region and improve the classification performance of the target region, which dynamically adjusts the sample weights of the instances to distinguish instances that may be beneficial for the learning target classifier. Figure 2 shows an illustration of TrAdaBoost [6].



(a) Classification is very difficult when there are few labeled training data.

(b) If we have sufficient auxiliary training data (blue "+" and "-"), we may be able to estimate the classification hyperplane.

(c) Auxiliary data may also be misleading classification results; the black "-" with a green circle is misclassified.

(d) TrAdaBoost moves the classification hyperplane in the correct direction by increasing the weight of misclassified source training data and reducing the weight of misclassified target training data

Fig. 2. An illustration of TrAdaBoost

TrAdaBoost uses an SVM as a base classifier, but it is not interpretive. To make the model interpretable, a tree-based classifier is adopted. However, most single tree-based classifiers perform worse than an SVM. Tree-based ensemble learning is an intuitive method to address this problem, which improves the performance of the classifier in each round and produces a traceable result. Therefore, we adopt LightGBM [8], a newly published tree-based boosting method, to cover the interpretability deficit of TrAdaBoost. Compared with other tree-based gradient boosting frameworks, LightGBM has two novel techniques: Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB). GOSS reduces data instances with small gradients so that the remaining data with high gradients can be used to calculate the information gain. Thus, useless instances in the source region are filtered. Because the data instances with larger gradients play more important roles in the computation of information gain, GOSS can obtain an accurate estimation of the information gain with a much smaller data size. EFB can reduce the number of features by bundling mutually exclusive features. The experiment [8] explained that LightGBM is ten times faster than the previous method due to EFB and GOSS.

C. The Procedure of the IRTED-TL

Algorithm 1 IRTED-TL

Input: source region data $D_S = (X_S, Y_S)$; target region data $D_T = (X_T, Y_T)$; test data Z; threshold h; sample weight $w^l = (w_1^l, \ldots, w_{n+m}^l)$; iterations N.

1: Feature Extraction

2: Initialization

$$w_i^0 = \begin{cases} \frac{1}{n}, & 1 \le i \le n, \\ \frac{1}{m}, & n+1 \le i \le n+m, \end{cases}$$

3: Caluculate KL divergence in eq.(4)

$$v_j = KL(q_{s_j} || q_{t_j})$$

4: Feature Mapping

$$(X_S^{q_j}, X_T^{q_j}) = \begin{cases} (X_S^{q_j}, X_T^{q_j}), & v_j \le h, \\ TCA(X_S^{q_j}, X_T^{q_j}), & v_j > h, \end{cases}$$

5: Set $\beta_s = \frac{1}{1+\sqrt{2\ln\frac{n}{N}}}$

6: **for** $k = 1, \ldots, N$ **do**

7: Normalize w^l. Set $w_i^k = w_i^k / \sum_{k=1}^{n+m} w_i^k$

8: Call Learner f_k =LightGBM(D_S, D_T, S, w^k).

9: Calculate the error of f_k on D_T:

$$error_k = \sum_{i=1}^{n+m} \frac{w_i^k \cdot |h_k(x_i) - c(x_i)|}{\sum_{i=n+1}^{n+m} w_i^k}$$

10: Set $\beta_k = min(\frac{error_k}{1-error_k}, 0.5)$

11: Update the new weight vector:

$$w_i^{k+1} = \begin{cases} w_i^k \beta_s^{|h_k(x_i)-c(x_i)|}, & 1 \le i \le n, \\ w_i^k \beta_k^{-|h_k(x_i)-c(x_i)|}, & n+1 \le i \le n+m, \end{cases}$$

12: **end for**

Output: hypothesis of test data y_h:

$$y_h(x) = \begin{cases} 1, & \sum_{k=\lceil N/2 \rceil}^{N} f_k(x) \ln\beta_k \le \frac{1}{2} \sum_{k=\lceil N/2 \rceil}^{N} \ln\beta_k \\ 0, otherwise \end{cases}$$

The procedure of the IRTED-TL is given in Algorithm 1. We initialize the sample weight for each instance in D_S and D_T

and we distinguish between them by applying different sample weights at the beginning $\frac{1}{n}$ for the instances in D_S and $\frac{1}{m}$ those in D_T. Notably, $m << n$, thus D_T plays a more important role than D_S because D_S is used to assist the training of D_T. TrAdaBoost calculates the sample weight for the training of LightGBM and LightGBM provides the hypothesis for updating the sample weight. β_s and β_k are two factors for updating sample weight; β_s updates the sample weight for instances in D_S and β_k updates the sample weight for instances in D_T. Note that $\beta_s^{|h_k(x_i)-c(x_i)|} \in (0, 1]$ and $\beta_k^{-|h_k(x_i)-c(x_i)|} \geq 1$, thus the sample weight of D_T is higher than that of D_S. The error rate is a weighted average of the loss and sample weights for updating β_k. When entering a new target region taxpayer, the model derives the prediction from the post-N / 2 classifier.

V. Performance Evaluation

To evaluate the performance of the IRTED-TL, we introduce our experimental design. Then, we investigate the following four research questions:

***RQ** 1: Can our method classify tax evasion taxpayers in the target region with high accuracy?* For this purpose, we evaluate the IRTED-TL based on large real-world datasets and compare its accuracy with six baseline approaches (see Section IV-C-1 for details).

***RQ** 2: How to determine the value of the parameters in the IRTED-TL to achieve the best detection accuracy?* For this purpose, we adjust the values of parameters and test their effects on the performance of the IRTED-TL (see Section IV-C-2 for details).

***RQ** 3: Can our method be adapted to a small amount of labeled data in the target region?* For this purpose, we evaluate the IRTED-TL with the increment of training data and compare it with the baselines (see Section IV-C-3 for details).

***RQ** 4: Can our method explain the process of tax evasion detection and provide evidence of tax evasion?* For this purpose, we visualize a partial result of our method to provide an impression of the interpretability of the IRTED-TL (see Section IV-C-4 for details).

A. Datasets and Metrics

There is no public standard dataset to evaluate tax evasion detection. For our experiments, we obtained tax data from tax authorities in China. We collected the tax information of 122,047 taxpayers in the industrial categories of wholesale and retail from two provinces. The taxpayers were divided into tax evaders and non-tax evaders. Each taxpayer has two categories of data: static data and dynamic data. Static data indicates the inherent attributes of taxpayers, including legal representative information (i.e., age, gender, and region), company size, registered capital and other company-related indicators. The dynamic data comprises a series of trading and declaration data with time-varying properties, such as transaction amounts, transaction taxes, and billing days. For convenience, we named one province G and the other province S. G has sufficient labeled data for tax evasion detection and S processes little tax evasion detection data in daily audit tasks. We selected four independent regions from G, $\{G_1, G_2, G_3, G_4\}$, and each region has its own economic and social conditions. Thus, we gathered actual tax data from five regions $\{G_1, G_2, G_3, G_4, S\}$. Moreover, through feature selection, they are located in the same feature space but differ in the marginal probability distribution. The features of G_1, G_2, G_2 and G_4 are more similar than that of S because they are in the same province. We designed thirteen inter-region transfer scenarios, as shown in Table II.

In Table II, the data set $G_1 \rightarrow G_2$ indicates that G_1 is the source region and G_2 is the target region, and its task is to extract knowledge from G_1 to aid in the classification of G_2. The other scenarios are named in the same manner. $|D_S|$ denotes the amount source region data used for training. $|D_T \cup Z|$ is the amount of data in the target region, D_T denotes the 20% of target region data used for training, and Z denotes the remainder used for testing.

TABLE II. Thirteen Transfer Scenarios

Scenarios	$\|D_S\|$	$\|D_T \cup Z\|$
$G_1 \rightarrow G_2$	3388	1630
$G_1 \rightarrow G_3$	3388	1580
$G_1 \rightarrow G_4$	3388	646
$G_2 \rightarrow G_1$	1630	1630
$G_2 \rightarrow G_3$	1630	1580
$G_2 \rightarrow G_4$	1630	646
$G_3 \rightarrow G_1$	1580	1580
$G_3 \rightarrow G_2$	1580	1580
$G_3 \rightarrow G_4$	1580	646
$G_4 \rightarrow G_1$	646	646
$G_4 \rightarrow G_2$	646	646
$G_4 \rightarrow G_3$	646	646
$G \rightarrow S$	8940	2768

The metrics used in our experiments are shown in TABLE III.

TABLE III. Descriptions of The Metrics

Term	Abbr	Definition
True Positive	TP	# of correctly classified tax evaders.
True Negative	TN	# of correctly classified non-tax evaders.
False Negative	FN	# of incorrectly classified non-tax evaders.
False Positive	FP	# of incorrectly classified tax evaders.
Error Rate	ER	(FN + FP) / (TP + TN + FN + FP)
Precision	P	TP / (TP + FP)
Recall	R	TP / (TP + FN)
F-measure	F1	2PR / (P + R)
ROC Area	AUC	Area under ROC curve
Top N hit rate	TopN	(# correctly classified tax evaders in top N) / N

B. Comparison Methods

To verify the performance of the IRTED-TL, we used traditional machine learning methods and a transfer learning method as comparison methods in the experiments. Traditional machine learning methods include the Multilayer Perceptron (MLP) [39] and SVM methods [40]. An MLP is a class of feedforward artificial neural network that consists of at least three layers of nodes; each node is a neuron that uses a nonlinear activation function. An SVM is a discriminative classifier formally defined by a separating hyperplane. TrAdaBoost(SVM) was adopted as a transfer learning baseline, as described in previous work [6].

According to the source of training data, the comparisons generate six baselines that fall into three categories, as shown in Table IV. The S-classifier indicates that we use source data to train the model and apply it to the target region. The T-classifier indicates that the model is directly trained by data in the target region. ST-SVM and TrAdaBoost(SVM) use both D_S and D_T as training data.

TABLE IV. THE BASELINE METHODS

Baseline	Training Data	Test Data	Classifier
S-MLP	D_S	D_T	MLP
S-SVM	D_S	D_T	SVM
T-MLP	D_T	D_T	MLP
T-SVM	D_T	D_T	SVM
ST-SVM	$D_S \cup D_T$	D_T	SVM
TrAdaBoost(SVM)	$D_S \cup D_T$	D_T	SVM

TABLE V. ERROR RATES OF TAX EVASION DETECTION WITH DIFFERENT METHODS

Data Set	S-MLP	S-SVM	T-MLP	T-SVM	ST-SVM	TrAdaBoost(SVM)	IRTED-TL
$G_1 \rightarrow G_2$	0.172	0.260	0.150	0.317	0.187	0.122	**0.018**
$G_1 \rightarrow G_3$	0.216	0.335	0.223	0.344	0.168	0.085	**0.024**
$G_1 \rightarrow G_4$	0.157	0.108	0.133	0.235	0.127	0.074	**0.034**
$G_2 \rightarrow G_1$	0.105	0.239	0.075	0.181	0.139	0.063	**0.015**
$G_2 \rightarrow G_3$	0.259	0.392	0.222	0.346	0.261	0.102	**0.019**
$G_2 \rightarrow G_4$	0.141	0.130	0.114	0.257	0.102	0.094	**0.019**
$G_3 \rightarrow G_1$	0.152	0.348	0.070	0.191	0.146	0.076	**0.018**
$G_3 \rightarrow G_2$	0.491	0.416	0.179	0.297	0.272	0.137	**0.028**
$G_3 \rightarrow G_4$	0.110	0.393	0.126	0.464	0.186	0.088	**0.037**
$G_4 \rightarrow G_1$	0.340	0.488	0.063	0.119	0.146	0.067	**0.020**
$G_4 \rightarrow G_2$	0.314	0.483	0.170	0.312	0.308	0.164	**0.020**
$G_4 \rightarrow G_3$	0.420	0.497	0.206	0.347	0.477	0.136	**0.024**
$G \rightarrow S$	0.488	0.505	0.226	0.276	0.302	0.166	**0.136**

TABLE VI. F1 SCORE OF TAX EVASION DETECTION WITH DIFFERENT METHODS

Data Set	S-MLP	S-SVM	T-MLP	T-SVM	ST-SVM	TrAdaBoost(SVM)	IRTED-TL
$G_1 \rightarrow G_2$	0.818	0.704	0.850	0.608	0.803	0.876	**0.982**
$G_1 \rightarrow G_3$	0.770	0.566	0.769	0.681	0.848	0.912	**0.976**
$G_1 \rightarrow G_4$	0.863	0.899	0.863	0.714	0.882	0.933	**0.968**
$G_2 \rightarrow G_1$	0.900	0.728	0.926	0.813	0.862	0.939	**0.985**
$G_2 \rightarrow G_3$	0.772	0.493	0.788	0.682	0.702	0.901	**0.981**
$G_2 \rightarrow G_4$	0.873	0.871	0.884	0.691	0.897	0.909	**0.983**
$G_3 \rightarrow G_1$	0.866	0.726	0.931	0.810	0.867	0.925	**0.982**
$G_3 \rightarrow G_2$	0.056	0.669	0.817	0.630	0.716	0.860	**0.972**
$G_3 \rightarrow G_4$	0.894	0.726	0.868	0.219	0.824	0.920	**0.964**
$G_4 \rightarrow G_1$	0.487	0.066	0.936	0.876	0.842	0.932	**0.980**
$G_4 \rightarrow G_2$	0.617	0.096	0.819	0.590	0.574	0.832	**0.980**
$G_4 \rightarrow G_3$	0.319	0.062	0.806	0.702	0.121	0.872	**0.976**
$G \rightarrow S$	0.663	0.627	0.748	0.642	0.708	0.830	**0.868**

TABLE VII. AUC SCORE OF TAX EVASION DETECTION WITH DIFFERENT METHODS

Data Set	S-MLP	S-SVM	T-MLP	T-SVM	ST-SVM	TrAdaBoost(SVM)	IRTED-TL
$G_1 \rightarrow G_2$	0.944	0.871	0.936	0.783	0.906	0.950	**0.991**
$G_1 \rightarrow G_3$	0.891	0.858	0.873	0.757	0.849	0.963	**0.998**
$G_1 \rightarrow G_4$	0.957	0.972	0.942	0.921	0.956	0.981	**0.997**
$G_2 \rightarrow G_1$	0.958	0.864	0.970	0.904	0.936	0.977	**0.998**
$G_2 \rightarrow G_3$	0.877	0.718	0.872	0.736	0.819	0.949	**0.998**
$G_2 \rightarrow G_4$	0.919	0.900	0.954	0.931	0.918	0.970	**0.998**
$G_3 \rightarrow G_1$	0.953	0.779	0.973	0.896	0.934	0.969	**0.998**
$G_3 \rightarrow G_2$	0.911	0.696	0.912	0.805	0.798	0.942	**0.993**
$G_3 \rightarrow G_4$	0.959	0.835	0.934	0.928	0.902	0.973	**0.997**
$G_4 \rightarrow G_1$	0.927	0.924	0.979	0.953	0.949	0.975	**0.998**
$G_4 \rightarrow G_2$	0.837	0.910	0.927	0.783	0.889	0.930	**0.994**
$G_4 \rightarrow G_3$	0.768	0.788	0.890	0.745	0.860	0.924	**0.998**
$G \rightarrow S$	0.684	0.688	0.882	0.897	0.746	0.915	**0.940**

TABLE VIII. TOPN(N=100) SCORE OF TAX EVASION DETECTION WITH DIFFERENT METHODS

Data Set	S-MLP	S-SVM	T-MLP	T-SVM	ST-SVM	TrAdaBoost(SVM)	IRTED-TL
$G_1 \rightarrow G_2$	0.990	0.970	0.970	0.910	0.950	0.988	**1.000**
$G_1 \rightarrow G_3$	0.856	0.780	0.942	0.930	0.710	0.952	**1.000**
$G_1 \rightarrow G_4$	0.990	1.000	0.964	1.000	0.970	0.990	**1.000**
$G_2 \rightarrow G_1$	0.972	0.970	0.978	0.980	0.940	0.976	**1.000**
$G_2 \rightarrow G_3$	0.868	0.700	0.914	0.900	0.910	0.952	**1.000**
$G_2 \rightarrow G_4$	1.000	0.980	0.970	0.950	0.980	0.990	**1.000**
$G_3 \rightarrow G_1$	0.970	0.930	0.964	0.990	0.960	0.940	**1.000**
$G_3 \rightarrow G_2$	0.992	0.880	0.968	0.940	0.960	0.990	**1.000**
$G_3 \rightarrow G_4$	0.990	0.880	1.000	0.940	0.960	0.990	**1.000**
$G_4 \rightarrow G_1$	0.924	0.870	0.976	0.940	0.960	0.952	**1.000**
$G_4 \rightarrow G_2$	0.938	0.960	0.976	0.850	0.950	0.982	**1.000**
$G_4 \rightarrow G_3$	0.854	0.810	0.934	0.880	0.860	0.902	**1.000**
$G \rightarrow S$	0.946	0.886	0.954	0.984	0.914	0.930	**0.992**

C. *Experimental Results*

1) Effectiveness of the IRTED-TL

We evaluated the effectiveness of the different methods in different scenarios using four metrics. The evaluation tables are shown in Tables V to VIII.

Table V shows the error rates of the different methods. For all thirteen transfer scenarios, compared with the traditional machine learning-based methods and TrAdaBoost(SVM), the IRTED-TL performs better in terms of error rate. The error rates of the IRTED-TL are more than 4.00% lower than that of the others, an average 19.6% lower.

Table VI shows the F1 scores of different methods. The IRTED-TL achieves the best performance in all thirteen transfer scenarios. The F1 scores of the IRTED-TL are greater than 0.96 in the G_1, G_2, G_2 and G_4 transfer scenarios. Notably, when the MLP and SVM perform badly in the condition of $G_2 \rightarrow G_3$ and $G_4 \rightarrow G_3$, the transfer learning methods TrAdaBoost(SVM) and IRTED-TL maintain excellent performance, which demonstrates that a tax evasion detection model trained with single-region data may not apply to other regions, causing high generalization error.

As shown in Table VII, the IRTED-TL outperforms other methods and shows outstanding AUC scores, as the feature mapping and boosting classifier of the IRTED-TL cause performance gains.

As shown in table VIII, the proposed method can achieve almost 1.000 TOPN (N=100) in tax evasion detection. If we provided tax authorities with a list of 100 taxpayers suspected of tax evasion, all the taxpayers on the list would be verified to be correctly classified. The results justify the stability of the proposed methods against region noise and data errors because the IRTED-TL has high performance in all thirteen transfer scenarios.

We focus the ROC curve of the $G \rightarrow S$ transfer scenario, as shown in Figure 3. This scenario is representative of the thirteen transfer scenarios because its source and target regions are in different provinces. The IRTED-TL always outperforms other methods and shows outstanding detection accuracy.

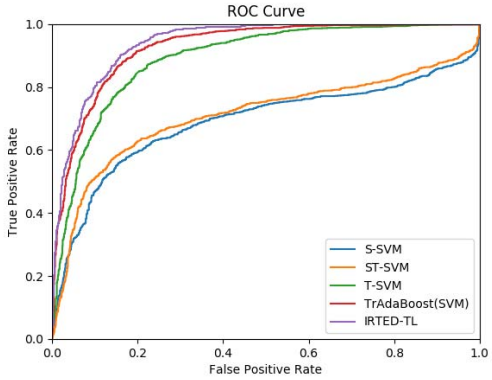


Fig. 3. The ROC Curve of different methods

In conclusion, the IRTED-TL greatly improves the accuracy of tax evasion identification according to various metrics.

2) Effect of Parameters

The threshold h plays an important role in the IRTED-TL, controlling the balance between the original structural information and the distance of the distribution.

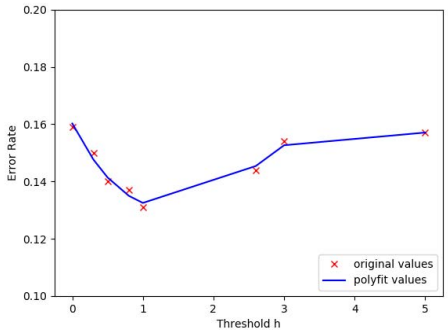


Fig. 4. Error rate of the IRTED-TL with different thresholds

Figure 4 illustrates the error rates with different values of the threshold h. We sampled h at the same density in the set of KL divergence. As most of the KL divergence between features was lower than 1, we can just sample a few thresholds

for different collections greater than 1. The error rate gradually decreases when h is in the range of 0 to 1 and increases slowly when h is greater than 1. When $h \geq 5$, none of the features undergo a mapping operation and when $h = 0$, all the features undergo the mapping operation. Our method performs the best when $h = 1$. Thus, appropriate feature mapping operations will improve the performance of our method. The smaller the value of h, the more features undergo the feature mapping operation and the less structural information is preserved. This demonstrates that an appropriate threshold setting can help close the gap between different distributions and maintain the original internal data structure during the feature mapping operation.

3) Stability of the IRTED-TL

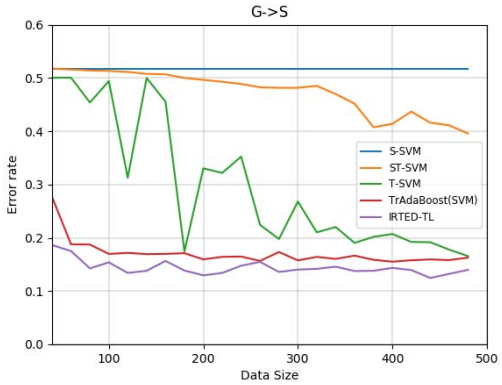


Fig. 5. The error rate of the IRTED-TL with different training data sizes for the target region

Figure 5 shows the error rates of different methods with different amounts of target region training data. We omitted the MLP-based baselines result because it was similar to the SVM. We maintained the amount of source region training data and changed the size of the target training data step-by-step. The amount of target region training instances was gradually increased from 20 to 500. The IRTED-TL always outperforms other methods with different amounts of target region data. The error rate of the S-SVM remains constant with increasing target data sets because it only uses source region data for training. The ST-SVM converges slowly because it is seriously affected by the source data and not all of data in source region are useful for target classification, thus the error rate decreases slowly with increasing target training data. The T-SVM only uses target data for training and subtle changes in the training data have a large impact on the classification, which leads to a jitter phenomenon with the growth of training data. Compared with the ST-SVM and T-SVM, the error rate of TrAdaBoost(SVM) and the IRTED-TL are quite low at the beginning and converge quickly even if provided a small amount of data. The IRTED-TL is better than TrAdaBoost(SVM) due to the advantages of the feature mapping and boosting technique. When the data size increases, the effect of the T-SVM gradually approximates the IRTED-TL because the target region has sufficient data for training. When the target region has a small amount of training data, the IRTED-TL would extract useful knowledge from the source region to improve the performance of tax evasion detection in the target region. Therefore, the IRTED-TL has excellent performance and stability in tax evasion detection in the absence of target region data.

4) Interpretability of the IRTED-TL

Most machine learning-based tax evasion detection methods cannot clearly explain the process of tax evasion when proposing a list of suspicious taxpayers. The IRTED-TL has the advantage of interpretability, thus it has the ability to interpret the detection process. Figure 6 shows a partial visualization of the training results of the IRTED-TL.

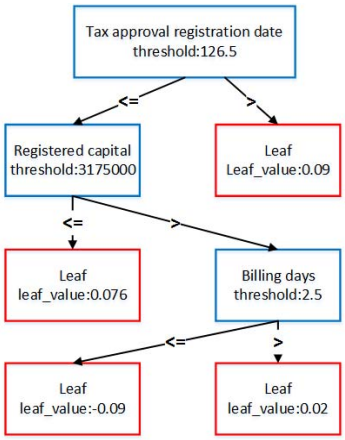


Fig. 6. Interpretability of the IRTED-TL

Figure 6 consists of two types of nodes colored in blue and red. The blue node is an attribute node, indicating the feature name and the corresponding threshold to split the node. The red node is a value node that will give a value to the sample for the next iteration. The value of the red node is meaningful. Positive values tend to be more prone to tax evasion whereas negative values tend to be normal. We can obtain the suspect value of tax evasion by adding the values of red nodes passed by a taxpayer. Therefore, we have strong evidence if the IRTED-TL detects a tax evader. The IRTED-TL can tell which indicators are abnormal using a tree structure. Moreover, the IRTED-TL explains the linkage between the various indicators. It is normal to examine the metrics alone, but there may be problems when combining them. In practice, the IRTED-TL can provide a set of tax evasion detection programs as a supplementary audit tool.

D. Further discussion

1) Advantages:

Based on the above evaluation results, we summarize the advantages of our approach below.

Hybrid solution: The proposed approach benefits from the advantages of transfer learning and a boosting classifier. The intent of this hybrid solution is to solve the problems of traditional tax evasion detection method such as poor generalization, the need for large labeled data and the lack of interpretability. The performance test shows that our approach achieved the best performance for all metrics.

Stability: The proposed method can be applied to detect tax evasion with a small amount of labeled data in a target region. Our performance evaluation tested target regions with different training data sizes, and the results imply that our approach can converge to the best performance with a small amount of training data. When the data size becomes larger, our approach still outperforms existing methods. Therefore, we can conclude that our method can stably improve the effectiveness of tax evasion detection when training examples are insufficient.

Robustness: Our method is robust even in the presence of data perturbations. The data sets of each region have strong local characteristics, which are regarded as abnormal disturbances in the transfer scenario. Our performance test showed that the proposed method can remove interference and extract common aspects of tax evasion in each region. In addition, the data set also has some perturbations due to registration error or empty data. Applying the boosting technique in our hybrid approach can overcome this problem.

Accuracy: Based on our evaluation, the proposed approach can achieve higher tax evasion detection accuracy than baselines using different groups of datasets. In addition, our method converges to a low error rate quickly with a small amount of training data.

Interpretability: The proposed approach is interpretable. When our approach detects a tax evader, it can provide a detailed tax evasion detection trail, which explains the reason for tax evasion. Auditors can deduce some information from the visualization of the model and develop specific inspection programs for taxpayers because some thresholds in the model are specific to the region.

2) Disadvantages

There are some disadvantages in our approach. We must further improve the following aspects. First, we must collect other tax evasion detection data to justify the scalability of the method. We could build a more robust model through multisource transfer learning, which extracts knowledge from at least two regions to detect the tax evasion of target region. Second, to improve the accuracy of the approach, we adopted the feature mapping technique but it increases the computation time. Thus, we are committed to finding a faster method to reduce the time complexity of our approach.

VI. Conclusion

This paper proposed the IRTED-TL, a novel inter-region tax evasion detection method. It integrates transfer learning and an interpretable classifier to augment training data for a label-sparse target region and induce interpretability in the detection model. First, by combining feature-based and instance-based transfer learning techniques, auxiliary knowledge from a source region with adequate training data is absorbed and applied to a label-sparse target region to augment the learning data in the presence of regional differences. Second, to offer a clear explanation of tax evasion, LightGBM is adopted to build an interpretable and accurate tax evasion detection model in the target region. The IRTED-TL outperforms existing methods with higher accuracy. In addition, it provides a good explanation of tax evasion behavior.

In future work, we will improve the design and seek a new method to optimize the speed of the process. Moreover, we aim to study multi-source transfer learning in tax evasion detection, which uses auxiliary training data sets from multiple source regions to achieve better tax evasion detection performance in a target region.

Acknowledgements

This work was sponsored by "The Fundamental Theory and Applications of Big Data with Knowledge Engineering" of the National Key Research and Development Program of China under Grant No. 2016YFB1000903, Innovative Research Group of the National Natural Science Foundation of China (61721002), Innovation Research Team of Ministry of Education (IRT_17R86), the National Science Foundation of China under Grant Nos. 61502379, 61532015 and 61672410, Project of China Knowledge Centre for Engineering Science and Technology, and the Academy of Finland (Grant No. 308087).

References

[1] F. Tian, T. Lan, K.-M. Chao, N. Godwin, Q. Zheng, N. Shah*, et al.*, "Mining suspicious tax evasion groups in big data," *IEEE Transactions on Knowledge and Data Engineering,* vol. 28, pp. 2651-2664, 2016.

[2] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering,* vol. 22, pp. 1345-1359, 2010.

[3] D. Wang, P. Wang, and J. Liu, "Improved privacy-preserving authentication scheme for roaming service in mobile networks," in *Wireless Communications and Networking Conference (WCNC), 2014 IEEE*, 2014, pp. 3136-3141.

[4] D. He and D. Wang, "Robust biometrics-based authentication scheme for multiserver environment," *IEEE Systems Journal,* vol. 9, pp. 816-823, 2015.

[5] D. Wang and P. Wang, "Two birds with one stone: Two-factor authentication with security beyond conventional bound," *IEEE transactions on dependable and secure computing,* 2016.

[6] W. Dai, Q. Yang, G.-R. Xue, and Y. Yu, "Boosting for transfer learning," in *Proceedings of the 24th international conference on Machine learning*, 2007, pp. 193-200.

[7] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Transactions on Neural Networks,* vol. 22, pp. 199-210, 2011.

[8] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma*, et al.*, "LightGBM: A highly efficient gradient boosting decision tree," in *Advances in Neural Information Processing Systems*, 2017, pp. 3149-3157.

[9] R.-S. Wu, C.-S. Ou, H.-y. Lin, S.-I. Chang, and D. C. Yen, "Using data mining technique to enhance tax evasion detection performance," *Expert Systems with Applications,* vol. 39, pp. 8769-8777, 2012.

[10] G. J. Williams and P. Christen, "Exploratory multilevel hot spot analysis: Australian taxation office case study," in *Proceedings of the sixth Australasian conference on Data mining and analytics-Volume 70*, 2007, pp. 77-84.

[11] Z. Assylbekov, I. Melnykov, R. Bekishev, A. Baltabayeva, D. Bissengaliyeva, and E. Mamlin, "Detecting Value-Added Tax Evasion by Business Entities of Kazakhstan," in *Intelligent Decision Technologies 2016*, ed: Springer, 2016, pp. 37-49.

[12] X. Liu, D. Pan, and S. Chen, "Application of hierarchical clustering in tax inspection case-selecting," in *Computational Intelligence and Software Engineering (CiSE), 2010 International Conference on*, 2010, pp. 1-4.

[13] P. C. González and J. D. Velásquez, "Characterization and detection of taxpayers with false invoices using data mining techniques," *Expert Systems with Applications,* vol. 40, pp. 1427-1436, 2013.

[14] 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, pp. 2161-2174, 2010.

[15] E. Junqué de Fortuny, M. Stankova, J. Moeyersoms, B. Minnaert, F. Provost, and D. Martens, "Corporate residence fraud detection," in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2014, pp. 1650-1659.

[16] D. DeBarr and Z. Eyler-Walker, "Closing the gap: automated screening of tax returns to identify egregious tax shelters," *ACM SIGKDD Explorations Newsletter,* vol. 8, pp. 11-16, 2006.

[17] M. Gupta and V. Nagadevara, "Audit selection strategy for improving tax compliance–Application of data mining techniques," in *Foundations of Risk-Based Audits. Proceedings of the eleventh International Conference on e-Governance, Hyderabad, India, December*, 2007, pp. 28-30.

[18] E. Hemberg, J. Rosen, G. Warner, S. Wijesinghe, and U.-M. O'Reilly, "Detecting tax evasion: a co-evolutionary approach," *Artificial Intelligence and Law,* vol. 24, pp. 149-182, 2016.

[19] G. Warner, S. Wijesinghe, U. Marques, O. Badar, J. Rosen, E. Hemberg*, et al.*, "Modeling tax evasion with genetic algorithms," *Economics of Governance,* vol. 16, pp. 165-178, 2015.

[20] E. Hemberg, J. Rosen, G. Warner, S. Wijesinghe, and U.-M. O'Reilly, "Tax non-compliance detection using co-evolution of tax evasion risk and audit likelihood," in *Proceedings of the 15th International Conference on Artificial Intelligence and Law*, 2015, pp. 79-88.

[21] J. A. Noguera, F. J. M. Quesada, E. Tapia, and T. Llàcer, "Tax compliance, rational choice, and social influence: An agent-based model," *Revue française de sociologie,* vol. 55, pp. 765-804, 2014.

[22] T. Llacer, F. J. Miguel, J. A. Noguera, and E. Tapia, "An agent-based model of tax compliance: an application to the Spanish case," *Advances in Complex Systems,* vol. 16, p. 1350007, 2013.

[23] L. Antunes, J. Balsa, and H. Coelho, "Agents that collude to evade taxes," in *Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems*, 2007, p. 212.

[24] N. Abe, P. Melville, C. Pendus, C. K. Reddy, D. L. Jensen, V. P. Thomas*, et al.*, "Optimizing debt collections using constrained reinforcement learning," in *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2010, pp. 75-84.

[25] N. Goumagias, D. Hristu-Varsakelis, and A. Saraidaris, "A decision support model for tax revenue collection in Greece," *Decision Support Systems,* vol. 53, pp. 76-96, 2012.

[26] A. Gretton, A. J. Smola, J. Huang, M. Schmittfull, K. M. Borgwardt, and B. Schölkopf, "Covariate shift by kernel mean matching," 2009.

[27] M. Sugiyama, T. Suzuki, and T. Kanamori, *Density ratio estimation in machine learning*: Cambridge University Press, 2012.

[28] S. J. Pan, X. Ni, J.-T. Sun, Q. Yang, and Z. Chen, "Cross-domain sentiment classification via spectral feature alignment," in *Proceedings of the 19th international conference on World wide web*, 2010, pp. 751-760.

[29] B. Gong, Y. Shi, F. Sha, and K. Grauman, "Geodesic flow kernel for unsupervised domain adaptation," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, 2012, pp. 2066-2073.

[30] M. Long, J. Wang, J. Sun, and S. Y. Philip, "Domain invariant transfer kernel learning," *IEEE Transactions on Knowledge and Data Engineering,* vol. 27, pp. 1519-1532, 2015.

[31] Z. Zhao, Y. Chen, J. Liu, Z. Shen, and M. Liu, "Cross-people mobile-phone based activity recognition," in *IJCAI*, 2011, pp. 2545-250.

[32] Y. Yao and G. Doretto, "Boosting for transfer learning with multiple sources," in *Computer vision and pattern recognition (CVPR), 2010 IEEE conference on*, 2010, pp. 1855-1862.

[33] J. Davis and P. Domingos, "Deep transfer via second-order Markov logic," in *Proceedings of the 26th annual international conference on machine learning*, 2009, pp. 217-224.

[34] L. Mihalkova, T. Huynh, and R. J. Mooney, "Mapping and revising Markov logic networks for transfer learning," in *AAAI*, 2007, pp. 608-614.

[35] X. Shi, W. Fan, and J. Ren, "Actively transfer domain knowledge," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 2008, pp. 342-357.

[36] E. Eaton and M. desJardins, "Selective Transfer Between Learning Tasks Using Task-Based Boosting," in *AAAI*, 2011.

[37] A. Liaw and M. Wiener, "Classification and regression by randomForest," *R news,* vol. 2, pp. 18-22, 2002.

[38] L. Breiman, *Classification and regression trees*: Routledge, 2017.

[39] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural networks,* vol. 2, pp. 359-366, 1989.

[40] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning,* vol. 20, pp. 273-297, 1995.