

Tax Assessment Warning Model Based on Recurrent Neural Network

Liang Hu, Wei Li, Yong Xie

School of Computer and Information Engineering, Xiamen University of Technology, Xiamen Fujian
Email: lhu@xmut.edu.cn, weili@xmut.edu.cn, 2013110701@xmut.edu.cn

Received: Oct. 1st, 2018; accepted: Oct. 15th, 2018; published: Oct. 23rd, 2018

Abstract

Tax assessment is an important and complicated task. Under the situation that there isn't a very effective warning model for tax assessment at present, this paper points out a way of using recurrent neural network (RNN) to establish warning model of tax assessment, and select the enterprises which have some tax suspects. This can solve the problem that warning model has no directive property of suspects. Through establishing warning model of tax assessment in industries, and verifying the analysis, this paper shows that this method is feasible.

Keywords

Recurrent Neural Network, Warning Model, Tax Assessment

基于递归神经网络的纳税评估预警模型

胡 亮, 李 伟, 谢 勇

厦门理工学院, 计算机与信息工程学院, 福建 厦门
Email: lhu@xmut.edu.cn, weili@xmut.edu.cn, 2013110701@xmut.edu.cn

收稿日期: 2018年10月1日; 录用日期: 2018年10月15日; 发布日期: 2018年10月23日

摘 要

纳税评估是一项重要而复杂的工作。针对目前尚无十分有效的纳税评估预警模型的情况, 提出利用递归神经网络(RNN)建立纳税评估预警模型的方法, 利用RNN的方法选出有涉税疑点的企业, 解决了预警模型无疑点指向性的问题。通过建立行业的纳税评估预警模型, 并进行验证分析, 表明该方法可行。

关键词

递归神经网络, 预警模型, 纳税评估

Copyright © 2018 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

纳税评估是对纳税人纳税能力的估测, 并据此对纳税人实际缴纳税款的真实性做出评估。虽然, 纳税人的真实纳税能力很难准确计算出来, 但其纳税异常虚假情况却能通过纳税人相关纳税指标分析评估出来[1]。纳税评估是一种在信息化条件下的风险审查纠错机制, 是申报征收与税务稽查之间的一道屏障, 也是构建税收征管体系架构之关键。纳税评估已经成为税务机关新时期强化税源管理、为纳税人提供深层次的服务、促进诚信纳税的一种具有综合效益的征管手段, 运用科学、合理的评估手段, 降低评估风险和成本, 提高纳税评估工作的整体效率和效果, 成为了纳税评估工作的关键问题[2]。

纳税评估在发达国家已形成成熟的系统理论, Allingham 和 Sandmo 首次提出了“主动纳税行为评估”问题, 建立了规范性计量经济模型[3], 其研究结果从理论上证明逃税规模与稽查处罚的概率存在负相关等有意义的结论, 然而工作远未结束, 这个数学模型方法的先行者受到很多批评, 比如 Graetz 等[4], Alm 等[5]以及 Frey 和 Feld [6]等, 其主要原因是依法纳税在理论上的评估和实际情况下的事实相差很大, 所以这种经济模型对逃税的预测能力较弱, 不具有实用价值。根据 Graetz 和 Wilde, Alm, McClelland 和 Schulze 的研究结果, 美国“风险厌恶 Arrow-Pratt 测度”基本上只介于 1 和 2 中, 但是实际观察到的依法纳税率却高达 30%, 存在很大的差距; 而且, 根据实证结果显示, 多数情况下纳税人的营业收入申报额要高于模型的预测值, 所以为了解决这个问题, 很多专家把道德因素添加到此模型中, 希望能得出更为合理的结论, 但是在 Andreoni [7]的文中, 他指出这个问题没有得到很好的解决, 仍然需要研究。

能对纳税评估问题进行定量分析首先要得益于计算机数据库的建立和计量经济学的发展, 在文献[8]中, Erard 对来自 IRS 的数据分析, 建立回归方程来研究违规行为、纳税准备方式和稽查效率之间的关系, 得到了边际税率、税务中介及遵从度之间的相关度。在文献[9]中, Frey 等人用加权有序 Probit 模型给依法纳税相关变量排序, 并可解释变量之间的相互影响程度, 测算出边缘效果。这一指标能定量反映纳税人依法纳税程度变化的趋势与规模。但在文献[10]中, Frey 指出模型中的威慑因素还是起的作用不大, 这一问题在学术界还有争论。

在我国, 政府部门积极与科研院所和高等院校合作, 相继推出了一系列理论和应用成果。值得一提的是国家自然科学基金课题“金融数学、金融工程及金融管理”的子课题“税收系统工程研究” [11], 从理论和应用层面上加强了对纳税评估的深入研究。李选举用判别分析法建立了逃税额估测模型, 这是国内首次应用数学统计的方法来建立税务稽查模型, 具有很好的借鉴意义[12]。钱澄鉴等学者在研制的税务稽查决策支持中, 运用峰值分析和推断方法, 建立了税负模型、偷漏增值税模型[13]。上海财经大学“税收征管系统工程研究”课题组则运用贝叶斯、函数变换方法建立税收诚实申报识别指标体系[14]。中国台湾大学管理学院会计系在“台湾收入署”协助下, 基于神经网络方法进行诚实申报的判别研究, 取得较好效果[15]。马庆国、王卫红、陈健等通过建立网络模型对企业纳税人申报情况的诚实度进行识别和预测, 效果显著[16]。

在进行纳税评估时，有一个很重要的环节，即纳税评估指标的选取，选择的指标合理才能公正地对企业进行评估。吴曲峰将财务报表与纳税评估相结合，将纳税指标进行分类，从财务指标的异常变动中，发现企业存在的税务风险[17]。崔丽对整个指标体系进行了完善，不仅考虑了财务指标，还添加了非财务指标，对本文的指标选取有很重要的借鉴意义[18]。李奇以超级市场零售行业为例进行了纳税评估研究，分析了该行业的行业特点，结合税源监控和纳税评估模型对具体案例进行了应用分析[19]，对本文选取的厦门地区超市零售行业的纳税评估研究有指导作用。

黄哲指出健全的纳税评估法制是进行纳税评估工作的基础，所以国家颁布的《税收征管法》应严格规定纳税评估的法律地位，健全评估制度，使各部门之间的工作能够统一衔接[20]。尽管目前纳税评估实践在慢慢趋于成熟，但是要解决的问题还有很多，洪江指出了当前亟需解决的四个主要问题，分别是纳税评估的信息化建设、评估工作的方式方法、评估机构设置和评估队伍的整体素质的提升[21]，本文在建立模型时也会参考这几方面突出问题。

2. 递归神经网络

Jordan 在 1986 年第一次提出了递归神经网络，其网络结构如图 1 所示。

在这个网络结构中，我们以 4 个神经元为例，神经元 4 是输入神经元，神经元 1、2 和 3 是全互联神经元，同时神经元 3 是输出神经元。输入神经元不进行任何运算，只负责把输入的信号传送到各互联神经元；而互联神经元具有自身反馈的特点，不仅接受来自输入神经元和其他互联神经元的输出，还要接受自身的输出，形成自环；输出神经元可以是互联神经元的一部分或者是全部。

设递归神经网络有 m 个输入神经元， n 个全互联神经元，而输出神经元是 n 个神经元中一部分。则 RNN 网络的动态特性可由下式表示：

$$h_k(t+1) = \left[1 + \frac{d}{T_k} \right] h_k(t) + \frac{d}{T_k} f(Y_k(t)) \quad (1)$$

$$Y_k(t) = \sum_{l=1}^n w_{k,l} h_l(t) + \sum_{i=1}^m w_{k,i+n} a_i(t) \quad (2)$$

式中： $h_k(t)$ 是神经元 k 在 t 时刻的输出； T_k 为时间常数； $w_{k,l}$ 是神经元 l 指向神经元 k 的权值； d 是计算步长。

误差准则函数：

$$E = \frac{1}{2} \sum_{k=1}^n (h_k(t) - b_k(t))^2 \quad (3)$$

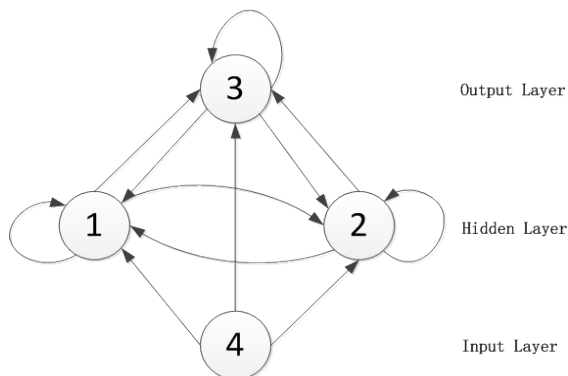


Figure 1. RNN schematic diagram

图 1. 递归神经网络示意图

可根据最速下降法来修正权值, 即有:

$$\Delta w_{ij}(t+1) = Z \sum_{k=1}^N e_k(t+1) \frac{L_{h_k}(t+1)}{Lw_{ij}} \quad (4)$$

$$\frac{L_{h_k}(t+1)}{Lw_{ij}} = \left[1 + \frac{d}{T_k} \right] \frac{L_{h_k}(t)}{Lw_{ij}} + \frac{d}{T_k} \frac{L_f(S_k(t))}{LS_k} \times \left[\sum_{l=1}^n w_{kl}(t) \frac{L_{h_l}(t)}{Lw_{ij}} + \delta_{kz_j}(t) \right] \quad (5)$$

此处, $i=1, 2, \dots, n$; $j=1, 2, \dots, n+m$

$$e(t+1) = \begin{cases} b_k(t+1) - h_k(t+1) & \text{若 } k \text{ 为输出神经元} \\ 0 & \text{若 } k \text{ 不为输出神经元} \end{cases} \quad (6)$$

$$z_j(t) = \begin{cases} h_j(t), & j=1, 2, \dots, n \\ a_j(t), & j=n+1, n+2, \dots, n+m \end{cases} \quad (7)$$

W_{ik} 为 Kronecker delta 函数

$$W_{ik} = \begin{cases} 0 & i \neq k \\ 1 & i = k \end{cases} \quad (8)$$

递归神经网络(RNN)的学习步骤如下:

- 1) 给定初始权值, 学习速度 Z , 惯量因子 T , 误差限值 A 和最大学习次数 N ;
- 2) 读入样本集 $\{A(t), H(t)\}$;
- 3) 计算每一个互联神经元在 $(t+1)$ 时刻的输出;
- 4) 修正权值;
- 5) 计算误差 E , 若 E 小于给定误差限值 A , 则退出计算; 否则进入步骤 3 重新学习。

3. 基于 RNN 的纳税评估模型

本文将厦门市某区超市零售企业的相关纳税数据作为数据样本, 建立了此行业的纳税评估预警模型。

3.1. 纳税评估的背景

纳税评估的工作处于前台征收和后台稽查之间, 是税收管理的一个基本环节, 通过纳税评估可以发现纳税疑点的企业, 为税务稽查提供稽查案例, 使稽查工作目标明确, 有针对性, 避免了以往的随意性和盲目性, 从而降低税务成本。税务工作的主要内容有信息登记、申报征收、纳税评估和税务稽查等, 具体的信息流程和业务流程如图 2 和图 3 所示。

3.2. 预警指标选取

本文根据超市零售业的产业结构特点, 结合数据来源的准确性和可获得性, 选取了 8 个主要指标作为评估预警模型的输入并对各指标进行编码, 如表 1 所示。

3.3. RNN 模型构建

3.3.1. 网络结构设计

本文选取了 8 个纳税指标, 所以将输入层的神经元节点设为 8 个, 为了更好的表示企业存在纳税问题的情况, 可以对疑点的确定度采用三类模式进行分类, 所以输出层的神经元节点为 3 个, 具体的模式分类如下。

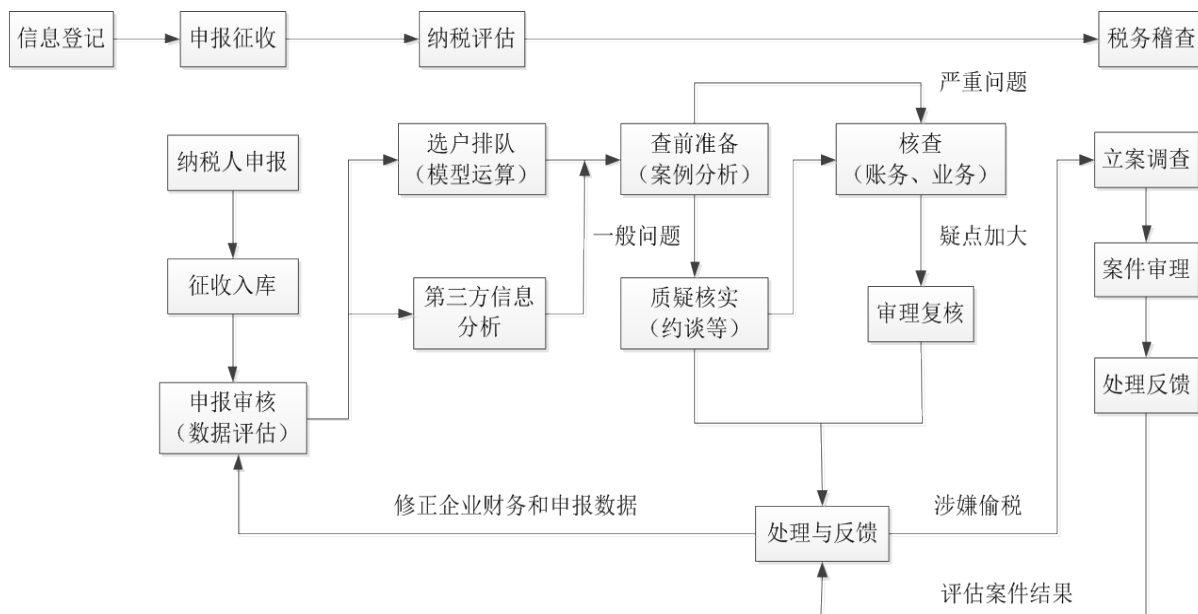


Figure 2. Work information flow chart
图 2. 工作信息流程图

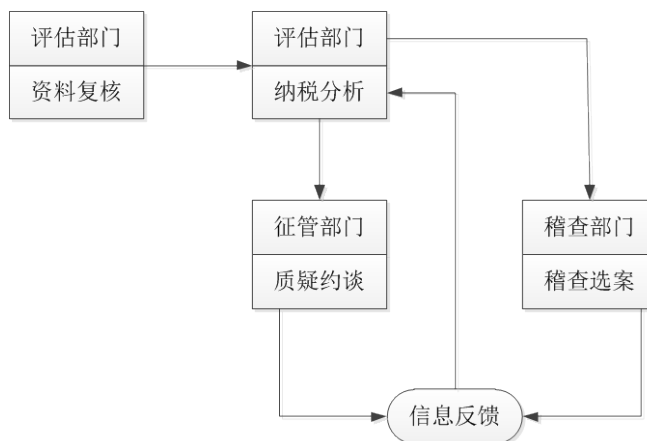


Figure 3. Business processes
图 3. 业务流程

Table 1. The index data and coding
表 1. 指标数据及编码

A1	增值税税负率(累计)
A2	增值税税负差异率(累计)
A3	增值税税负变动率(累计同比)
A4	销售毛利率(累计)
A5	主营业务成本变动率(累计同比)
A6	主营业务收入变动率(累计同比)
A7	增值税税收贡献增长率(累计同比)
A8	存货月周转率(累计平均)

将输出层对样本的输出向量定义为 $h = [h_1 h_2 h_3]$ ，当输出向量是 $h = [1 0 0]$ 时，表示企业存在纳税问题的可能性较大，当输出向量是 $h = [0 1 0]$ 时，表示企业存在纳税问题的可能性一般，当输出向量是 $h = [0 0 1]$ 时，表示企业存在纳税问题的可能性很小。

中间隐含层的神经元节点数，用下面的公式确定：

$$n = (n_t + n_0)^{\frac{2}{3}} + b \tag{9}$$

式中： n —隐层节点数； n_0 —输入节点数； n_t —输出节点数； b —1~10 之间的常数。

通过多次反复的试验和计算，将隐含层的神将元节点定为 6 个，可以更好的对企业的纳税情况进行评估，最终建立的 RNN 网络模型结构是 $8 \times 6 \times 3$ 。模型的表达式为：

$$h_n = f \left[\sum_{j=1}^6 v_j f \left(\sum_{i=1}^8 w_{ji} + s_j \right) + s_n \right], n = 1, 2, 3 \tag{10}$$

传递函数 $f(n)$ 采用 Logistic 函数，如下：

$$\text{logsig}(n) = \frac{1}{1 + \exp(-n)} \tag{11}$$

3.3.2. 样本训练及结果检验

根据已建立的 RNN 网络结构的特性再结合相关领域专家们的建议，本文选取了 150 个具有较高可信度的企业数据样本作为训练和测试对象。这些企业的数据接受过年度税务稽查，所以将纳税数据作为输入，稽查结果转化为输出向量后作为输出层的输出目标值。具体输入如表 2 所示。

在 Matlab7.0 环境下选取 RNN 学习方法进行训练。选取误差精度为 $e = 10^{-3}$ ，性能函数为 mse， $\mu = 0.01$ ，经过 458 步迭代后，网络达到稳定状态并收敛。

通过对训练样本进行模拟仿真，可以检验实际输出与期望输出之间的拟合度，根据仿真结果显示，本文的纳税评估预警模型对训练样本可以实现正确率高达 92% 的分类。为进一步进行验证，将作为测试对象的数据代入此模型，利用仿真对测试样本进行分类。结果如表 3 所示。

由表 3 可知，对测试样本的分类基本上是正确的，其中的两个误判是将在稽查结果中没有纳税问题和纳税问题一般的企业误判为纳税问题大的企业。这可能是由于存在纳税问题的企业在稽查过程中被遗漏造成的，所以产生误判也是有可能的。

Table 2. The index data

表 2. 指标数据

编号	A1	A2	A3	A4	A5	A6	A7	A8
1	0.0316	0.0398	0.1395	0.0836	0.0514	0.0326	0.1663	37.8562
2	0.0356	0.1025	-0.0802	0.0815	-0.3264	-0.3776	-0.4273	16.2031
3	0.0234	0.0728	-0.0326	0.0591	-0.3026	-0.3006	-0.3145	5.0216
4	0.0725	1.4225	-0.3526	0.0789	-0.4571	-0.4487	-0.5412	208072
5	0.0485	0.2748	0.0356	0.1824	-0.4562	-0.4132	-0.3745	4.2356
6	0.0370	0.2252	-0.0773	0.0125	-0.0920	-0.0854	-0.1562	46.5692
7	0.0705	1.4433	-0.0208	0.1588	-0.0245	-0.0482	-0.0689	14.1985
8	0.0312	0.0562	-0.0312	0.0945	0.1699	0.1931	0.1601	8.3256
9	0.1562	4.5231	4.1265	-0.1732	-0.5236	-0.6541	1.1139	3.2546
10	0.0308	0.0199	0.0356	0.0905	0.2654	0.2318	0.2568	13.4806

Table 3. The contrast on inspection results and samples outputs
表 3. 稽查结果与样本输出对比

稽查结果	模型输出	预警结果	对比
(0, 0, 1)	(0.0001, 0.0001, 0.9992)	可能性小	
(0, 1, 0)	(0.0011, 0.9856, 0.0000)	可能性一般	
(0, 1, 0)	(0.0002, 0.9992, 0.0001)	可能性一般	
(0, 0, 1)	(0.9879, 0.0002, 0.0001)	可能性大	误判
(0, 0, 1)	(0.0086, 0.0001, 0.9968)	可能性小	
(0, 0, 1)	(0.0011, 0.0020, 0.9945)	可能性小	
(1, 0, 0)	(1.0000, 0.0008, 0.0025)	可能性大	
(1, 0, 0)	(1.0000, 0.0006, 0.0023)	可能性大	
(0, 1, 0)	(1.0000, 0.0032, 0.0001)	可能性大	误判
(1, 0, 0)	(1.0000, 0.0001, 0.0007)	可能性大	
(1, 0, 0)	(1.0000, 0.0023, 0.0002)	可能性大	
(1, 0, 0)	(1.0000, 0.0001, 0.0026)	可能性大	

纳税评估预警模型的建立是为了筛选出有可能存在纳税问题的企业，从对训练样本和测试样本的模拟仿真结果来看，没有将存在大的纳税问题的企业误判为没有纳税问题或者纳税问题一般的企业，基本上避免了逃漏税的情况，所以本文建立的预警模型的效果应该是比较好的。

4. 实验分析

本文选取了增值税税负率、销售毛利率和主营业务收入变动率等多个指标进行分析。

以增值税税负率为例，经过 SPSS 软件的检验，可以发现增值税税负率基本上趋于正态分布，均值为 0.0537，标准差为 0.0316。经检验作为训练和测试样本需要补缴企业所得税的企业中，有 75% 增值税税负率在置信度为 65% 的置信区间的下方，如果企业隐瞒了营业收入的具体情况，那么其增值税税负率很可能不在此范围。当然只靠增值税税负率来判断是比较片面的，所以同时利用销售毛利率、主营业务收入变动率、主营业务成本变动率、增值税税收贡献增长率等指标，可以帮助发现企业是否有隐瞒营业收入的问题。

本文还对主营业务收入变动率和主营业务成本变动率进行了回归分析，得到模型显著性检验的 F 值为 302.0681，拒绝模型非显著的概率小于 0.0010，因此回归模型是显著的，主营业务收入变动率和主营业务成本变动率的相关系数为 0.7645，拟合的效果比较好，系数的 t 检验均拒绝等于零的假设。

利用主营业务收入变动率和主营业务成本变动率的回归模型，可以帮助确定企业隐瞒营业收入的指向性。因此，在进行纳税预警模型对企业进行分类后，在存在纳税问题的企业中再利用统计学的理论进行详细分析，可以更好的解决疑点指向性的问题。

5. 结语

本文根据超市零售业的产业结构特点，结合数据来源的准确性和可获得性，选取了 8 个主要指标作为评估预警模型的输入，预测采用递归神经网络模型。对各指标进行编码，通过递归神经网络的设计，输出预警结果，以辅助税务工作者进行正确决策。实验证明，我们的方法准确有效。下一步，数据规模比较大，我们拟考虑在大数据基础上，采用深度学习等方法来进一步提升模型的准确性和适用性。

参考文献

- [1] 财政部《税收制度国际比较》课题组. 加拿大税制[M]. 北京: 中国财政经济出版社, 2000.
- [2] 叶艺勇. 基于支持向量机和领域知识的纳税评估预警模型[J]. 数学的实践与认识, 2014, 44(1): 72-77.
- [3] Allingham, M.G. and Sandmo, A. (1972) Income Tax Evasion: A Theoretical Analysis. *Journal of Public Economics*, **1**, 323-338. [https://doi.org/10.1016/0047-2727\(72\)90010-2](https://doi.org/10.1016/0047-2727(72)90010-2)
- [4] Graetz, M.J. and Wilde, L.L. (1985) The Economics of Tax Compliance: Facts and Fantasy. *National Tax Journal*, **38**, 355-363.
- [5] Alm, J., McClelland, G.H. and Schulze, W.D. (1992) Why Do People Pay Taxes? *Journal of Public Economics*, **48**, 21-48. [https://doi.org/10.1016/0047-2727\(92\)90040-M](https://doi.org/10.1016/0047-2727(92)90040-M)
- [6] Frey, B.S. and Stutzer, A. (2002) Happiness and Economics. Princeton University Press, Princeton, 90-96.
- [7] Andreoni, J., Erard, B. and Feinstein, J. (1998) Tax Compliance. *Journal of Economic Literature*, **36**, 818-860.
- [8] Erard, B., Chih-Chin, and Ho (2001) Searching for Ghosts: Who Are Nonfilers and How Much Tax Do They Owe? *Journal of Public Economics*, **81**, 25-50. [https://doi.org/10.1016/S0047-2727\(00\)00132-8](https://doi.org/10.1016/S0047-2727(00)00132-8)
- [9] Frey, B.S. and Feld, L.P. (2002) Deterrence and Morale in Taxation: An Empirical Analysis. CESifo Working Paper No. 760, 8-20.
- [10] Frey, B.S. (2003) The Role of Deterrence and Tax Morale in Taxation in the European Union. JelleZijlstra Lecture, Netherlands Institute for Advanced Study in the Humanities and Social Sciences (NIAS).
- [11] 马国庆. 我国税收制度的改革和税收征管问题研究报告[R]. 国家自然科学基金课题“金融数学、金融工程及金融管理”之子课题“税收系统工程研究”结题报告, 2001: 30-36.
- [12] 李选举. 税收稽查选案统计研究(修改稿) [Z]. 武汉: 中南财经大学, 1997: 10-15.
- [13] 钱澄鉴, 王健, 金静学. 税务稽查决策支持系统[J]. 运筹学杂志, 1997(1): 33-37.
- [14] “税收征管系统工程研究”课题组. 税收征管中诚实申报识别研究[J]. 上海财经大学学报, 1992(12): 11-17.
- [15] Wu, R.C. (1997) Neural Network Models: Foundation and Application to an Audit Decision Problem. *Annals of Operation Research*, **75**, 291-301. <https://doi.org/10.1023/A:1018915714606>
- [16] 马庆国, 王卫红, 陈健, 等. 神经网络在税务稽查选案中的应用研究[J]. 数量经济技术经济研究, 2002, 19(8): 98-101.
- [17] 吴曲峰. 财务报表分析在纳税评估中的应用研究[D]: [硕士学位论文]. 厦门: 厦门大学, 2014.
- [18] 崔丽. 纳税评估指标体系的完善[D]: [硕士学位论文]. 呼和浩特: 内蒙古大学, 2014.
- [19] 李奇. 纳税评估研究一以零售行业为例[D]: [硕士学位论文]. 成都: 西南财经大学, 2013.
- [20] 黄哲. 《税收征管法》对纳税评估制度的规制[J]. 税务研究, 2013(5): 47-50.
- [21] 洪江. 当前纳税评估实践亟需解决的四大难题[J]. 山东科技大学学报(社会科学版), 2015(2): 85-89.

知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2161-8801, 即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: csa@hanspub.org