

# 门控变关系图卷积网络的涉烟案件当事人预警

冯鹏程<sup>1</sup>, 张高豪<sup>2</sup>, 谢刚<sup>1</sup>

1. 贵州师范大学大数据与计算机科学学院, 贵州 贵阳 550025;

2. 贵州省烟草公司贵阳市公司, 贵州 贵阳 550002

## 摘要

为了落实“精准监管”方针, 烟草公司需要提高涉烟案件命中率。过往的方法缺少了对涉烟案件高危当事人的研究, 阻碍了案件命中率的提高。基于烟草公司存有的大量历史数据, 挖掘出准确的预警名单是提高案件命中率的有效途径。进行高危当事人特征分析后, 提出门控变关系图卷积网络, 以得到准确的高危当事人预警名单。首先, 门控变关系图卷积网络使用变关系图卷积网络, 捕捉当事人的关系与关键特征。然后, 门控层进一步提取特征。最后, 把特征输入Softmax层得到分类结果, 进而得到预警名单。通过对比实验, 证明构建的模型效果更佳。某市专卖局应用本项目的系列成果后, 其案件命中率从约0.01%提升到了约0.5%, 这证明构建的预警模型能满足真实监管的需求。

## 关键词

烟草行业; 涉烟案件; 高危当事人识别; 门控层; 图卷积网络

中图分类号: TP391.4

文献标志码: A

doi: 10.11959/j.issn.2096-0271.2024040

## *Gated variable graph convolutional network for warning of parties in tobacco-related cases*

FENG Pengcheng<sup>1</sup>, ZHANG Gaohao<sup>2</sup>, XIE Gang<sup>1</sup>

1. School of Big Data and Computer Science, Guizhou Normal University, Guiyang 550025, China

2. Guiyang Branch, Guizhou Tobacco Company, Guiyang 550002, China

## Abstract

In order to implement the "precision supervision" policy, tobacco companies need to increase the hit rate of tobacco-related cases. Past approaches had lacked research on high-risk parties in tobacco-related cases, which had hindered the improvement of case hit rates. Based on the large amount of historical data stored in tobacco companies, mining accurate warning lists is an effective way to improve case hit rates. After conducting the analysis of high-risk parties features, a gated variable relationship graph convolutional network was proposed to obtain an accurate high-risk parties warning list.

Firstly, the gated variable relation graph convolutional network used variable relation graph convolutional network to capture the relationship and key features of the parties. Then, the gated layer was applied to further learn the features. Finally, the learned features were inputted to the Softmax layer to get the classification results, and then an alert list is obtained. Through comparison experiments, the constructed model is proved to be more effective. After a municipal monopoly bureau applied the results of this project, its case hit rate improved from about 0.01% to about 0.5%, which proves that the early warning model can meet the needs of real regulation.

### Key words

Chinese tobacco industry, cigarette case, high-risk parties recognition, gated layer, graph convolutional network

## 0 引言

烟草行业是我国税收的重要来源。据中国烟草总公司、国家烟草专卖局数据,2022年烟草行业工商税利总额达14 413亿元,财政总额14 416亿元,为国家和地方财政做出积极贡献<sup>[1]</sup>。然而,不法分子通过售假走私等违法行为严重危害国家利益和消费者权益。在福建省的“2101”专案中,专案组成功并联总涉案金额高达21.9亿元的40余起案件<sup>[2]</sup>。从2022年10月至2023年2月,全国各地烟草部门累计查处“可乐罐”等非法电子烟案件593起,涉案金额总计5.07亿元<sup>[3]</sup>。近年来,随着互联网发展,涉烟违法犯罪越发猖獗,有关部门要迅速精准打击。

目前,中国烟草的数字化之路存在模型算法简单、数据孤岛等问题<sup>[4]</sup>。如何将数据挖掘等技术应用在烟草市场监管领域已成为当前研究的热点。例如,乔浪超等<sup>[5]</sup>对涉烟寄递数据进行时空特征分析,以识别涉烟邮递。该研究的缺点是数据太少,且没考虑类不平衡问题。针对涉烟车辆小数据集,王贞等<sup>[6]</sup>对车辆数据提取特征,并使用标签传播算法对异常车辆分类。

传统的高危当事人识别由办案人员依据办事经验确定,但人工找出高危当事

人耗时耗力,还可能存在疏漏。这个情形急需合适的技术解决。识别高危涉案当事人,获取高危预警名单,稽查人员可以进行重点监管。然而,目前缺少涉及烟草案件的前沿研究,因此涉烟案件当事人的高危预警研究很有价值。

参照其他领域的分类应用<sup>[7-9]</sup>,解决问题的关键是当事人之间的关系与涉案金额等关键特征的提取。其中,当事人的关系可以被视为图结构数据,由此得出解决思路,即把原数据转化成图,然后用图模型提取特征,最后进行分类识别。

为挖取图结构信息,图嵌入领域诞生了DeepWalk、图因式分解、SDNE (structural deep network embedding) 等方法,但这些经典方法不适合用于高危当事人识别。因为它们多是基于图结构的图表示学习的方法,极少利用节点特征。图神经网络 (graph neural network, GNN) 能捕捉图结构与特征信息,更适合用于高危当事人识别。

在此背景下,本文用子关系矩阵代替完整的标准化邻接矩阵,把图卷积网络改进为变关系图卷积网络 (variable relation graph convolutional network, VRGCN), 并且加上门控层和Softmax层,提出基于门控变关系图卷积网络 (gated variable relation graph convolutional network, GVRGCN) 的高危当事

人识别模型。最后,实验证明GVRGCN能较好地满足监管要求。

## 1 相关工作

图卷积网络(graph convolution network, GCN)是GNN的经典模型之一,其特点是通过每层网络乘以邻接矩阵来融合节点数据。由于GCN简洁高效,本文对GCN进行改进,从而获得编码部分。

为达成归纳式的节点分类目标,SPN(structured proxy network)结合了图神经网络、结构化预测的优势,在条件随机场框架中引入了由GNN表征的灵活势函数来代替原势函数<sup>[10]</sup>。SPN最大的问题是条件随机场对节点和边信息的融合机制并不适用于高危当事人识别场景。

为提高无监督图嵌入准确性与可扩展性,文献[11]提出图融合,利用节点特征的 $k$ 个最近邻关系构造权重邻接矩阵。文献[12]提出了基于GCN的图修正模块,学习如何修正边。为解决在错乱的图上难以实现GCN的问题,文献[13]利用相似节点动态调整图邻接关系,提出图学习卷积网络,并采用池化和蒸馏减少过拟合。

本文认为高危预警名单可以通过高危当事人识别获取,并进一步把高危当事人识别视作一个分类任务。周纯英等<sup>[14]</sup>提出的基于图模型的GKCI(GNN-based key class identification)算法启发笔者使用交叉熵损失函数训练模型,然而,该研究使用DeepWalk改进的Node2Vec学习嵌入。如果将Node2Vec应用于高危当事人识别,很可能引起引言叙述不适用的问题。因此,在杨晓晖等<sup>[15]</sup>提出的模型的启发下,本文提出VRGCN,并在其基础上加上一层门控层,提出GVRGCN的高危当事人识别模型。

## 2 高危涉案当事人的特征分析

高危当事人是指已经或将要犯下危害案件、造成高度危险、值得高度关注的涉案人员。参照人工识别经验,归纳高危当事人的识别规则,能更好地优化识别模型。判断当事人是否高危有以下参考标准。

第一,涉案金额。涉案金额高的比涉案金额低的更危险。当涉案金额超过规定数额就会被列为特大案件,涉案当事人极大概率被标记为高危人员。

第二,涉案人数。涉案人数多的案件比涉案人数少的更值得注意,其中主犯比其他涉案人员更危险。

第三,案件性质。案件性质反映当事人的险情级别。涉烟案件分为行政案件和刑事案件。一般情况下,刑事案件的当事人要比行政案件的危险。并且,行政案件和刑事案件也存在不同级别,例如行政违法行为有无证经营和销售非法产品的区别,刑事案件有省标、国标的分级。当事人的险情级别应该具体情况具体判别。

最后,涉案人员关系。涉案人员的关系影响其危险程度。当事人多次涉案会留下多次记录,这些记录属于一个人,是同人关系。显然,累犯比初犯危害更大。根据办案经验,涉案两次及以上的涉案人员往往会继续犯案,且涉案方式更隐蔽。而所属案件相同的涉案人员,属于同案关系。与高危人员同案的比与其他涉案人员同案的更高危。比较复杂的情况是,多个涉案人员不直接同案,却与同一当事人同案。本文称这种关系为二阶同案。二阶同案里难以决断的是不直接同案的涉案人员之间的关系,他们可能有隐秘联系没被探明。这使得他们比无关系人员更危险,但不如其他关系高危。而比二阶同案关系更琐碎或复杂的

关系,一律当作二阶同案处理。

根据上述识别规则,本文编排出数据预处理手段,把案件转换成图数据。实验部分可以使用识别规则进行对比。

### 3 门控变关系图卷积网络预警模型

GVRGCN预警流程如图1所示,首先运用专门的预处理手段把真实案件数据构造成图,接着使用GVRGCN将图转化为低维嵌入,最后通过Softmax层预测当事人是否高危。

模型编码部分改进自GCN,因此GVRGCN存在性能随深度增加而下降<sup>[16]</sup>的问题。此外,GCN直接应用在高危当事人识别会有很多问题。具体如下。

#### 3.1 门控变关系图卷积网络

GCN数学描述如式(1)所示,其中邻接矩阵 $A$ 与节点特征矩阵 $X$ 输入编码器GCN得出嵌入向量 $Z$ 。 $\sigma()$ 是激活函数。 $W_0$ 、 $W_1$ 是对应网络层的参数。 $D$ 是 $A$ 的度矩阵,经过处理,邻接矩阵 $A$ 就能标准化为 $\tilde{A}$ 。

$$\begin{aligned} GCN(A, X) &= \tilde{A}\sigma(\tilde{A}XW_0)W_1 = Z, \\ \tilde{A} &= D^{-1/2}AD^{-1/2} \end{aligned} \quad (1)$$

随着编码器层数的加深,由于节点的同人关系和同案关系,原本的GCN无法满足高危当事人识别的要求,存在以下两个问题。

问题1:数据平均。标准化的邻接矩阵与节点特征矩阵相乘,会导致同人关系的节点数据被平均。属于同一当事人的记录,存在一

些特征(如是否为主犯)不能直接合并成一个节点。本文希望GCN融合同人关系的节点以共享数据,使它们方便识别。然而,标准化让它们加权平均,跟这一目标截然相反。

问题2:数据同化。如果节点在邻接矩阵中对应的行相等,经过GCN融合后,它们的数据会变得一样。处理同人关系时,这个特性导致同属一人的节点数据相同,这意味着任意数量的同人节点的识别结果都无差别,这将完成实质上的节点合并。而处理同案关系时,这个特性导致同案节点的输出相同,这种数据同化干扰了识别结果,因为这意味着同案关系同等危险。图2举例描述这些问题。

为解决这些问题,本文提出VRGCN,相对于GCN,VRGCN每层不再使用完整的标准化邻接矩阵,而是根据处理流程使用其子图的关系矩阵。

根据第2节的关系分析,完整的邻接矩阵可以分成同案子关系矩阵和同人子关系矩阵。子关系矩阵在运算中仍然保有邻接矩阵的行相等输出结果相同的性质。如果采用第一层同案子关系矩阵、第二层同人子关系矩阵的设计,就能在第一层传递同案当事人信息,而在第二层进行同人节点数据融合。这样就解决了问题2。针对问题1,只要放弃矩阵的标准化就能解决。更进一步,还可以根据需求,修改子关系矩阵中相关元素的值。例如,传递二阶同案关系时,除主对角线外的元素乘以0.5,就能达到二阶同案关系不如其他关系危险的效果。VRGCN的数学描述如式(2)所示:

$$\begin{aligned} VRGCN(A, X) &= \\ A_d\sigma(\sigma(A_{d-1}XW_{d-1}))W_d &= Z \end{aligned} \quad (2)$$

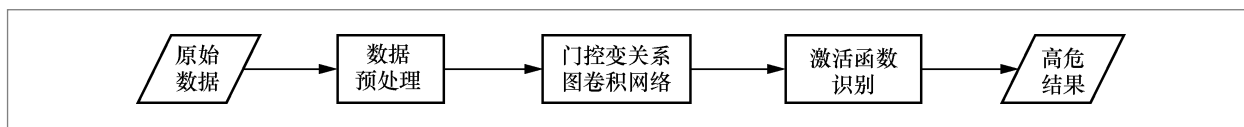


图1 GVRGCN 预警流程

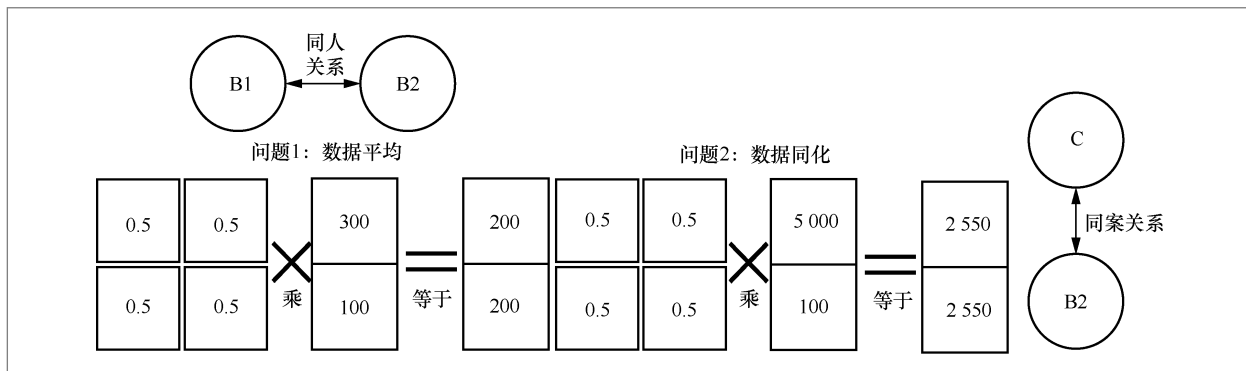


图 2 原始的 GCN 遇到的问题

其中,  $A_0, \dots, A_{d-1}, A_d$  分别是  $0 \sim d$  层网络的子关系矩阵,  $W_0, \dots, W_{d-1}, W_d$  是对应网络层参数,  $Z$  是 VRGCN 学习到的嵌入。VRGCN 能够提取较好的特征。然而, 案件数据还有时间序列信息, 这个是 VRGCN 难以学习的。为解决这个问题, 受门控循环单元 (gate recurrent unit, GRU) 架构的启发, 本文在 VRGCN 的基础上添加了一层门控层, 提出结合式 (3) ~ 式 (6) 的 GVRGCN。

$$R_t = \sigma(A_t Z W_{xr} + H_{t-1} W_{hr} + b_r) \quad (3)$$

$$U_t = \sigma(A_t Z W_{xz} + H_{t-1} W_{hz} + b_z) \quad (4)$$

$$\hat{H}_t = \tanh(A_t Z W_{xh} + (R_t * H_{t-1}) W_{hh} + b_h) \quad (5)$$

$$H_t = U_t * H_{t-1} + (J - U_t) * \hat{H}_t \quad (6)$$

其中,  $W$  与  $b$  是网络参数矩阵,  $Z$  是 VRGCN 输出的嵌入,  $H_t$  是  $t$  时刻的隐状态,  $\sigma()$  是激活函数,  $\hat{H}_t$  是  $t$  时刻的候选隐状态,  $*$  是哈达玛积运算符,  $J$  是大小与  $U_t$  相同的矩阵。式 (3) 对应着 GRU 的重置门, 式 (4) 对应着 GRU 的更新门。重置门控制着过去隐状态信息的保留情况, 更新门控制着新隐状态中的过去隐状态信息和新输入的比例。相较于 GRU, GVRGCN 的门控层的改进之处是使用若干子关系矩阵  $A_t$  控制完整的特征矩阵  $Z$  的计算, 而不是时间序列的输入。本文使用同人子邻接矩阵控制门控层的计

算, 其数量等于各当事人涉案的时间序列中的最大长度。

### 3.2 分类识别输出层

GVRGCN 的识别部分可以使用聚类。而聚类的方法会引起两个问题, 一是聚类需要一定的数据集, 这样消耗的资源较大, 二是聚类结果难以控制。由于烟草数据变化不大, 可以使用 Sigmoid 或 Softmax 进行识别。

如图 3 所示, 本文添加 Softmax 层作为模型的输出层。GVRGCN 块由第一层的 VRGCN 和第二层的门控层组成, 接着若干 GVRGCN 块和最后的 Softmax 层组成 GVRGCN。GVRGCN 的编码器改进自 GCN, 这使得 GVRGCN 的训练依靠标签训练集, 主要是标记为高危的标签数据, 所以 GVRGCN 既可以是有监督的, 也可以是半监督的, 其中 GVRGCN 主要适用于半监督学习。

## 4 实验与分析

### 4.1 实验环境

本实验在 Intel(R) i7-8700K 的 CPU、4 GB 内存、64 位 Windows 10 的计算机上运行。

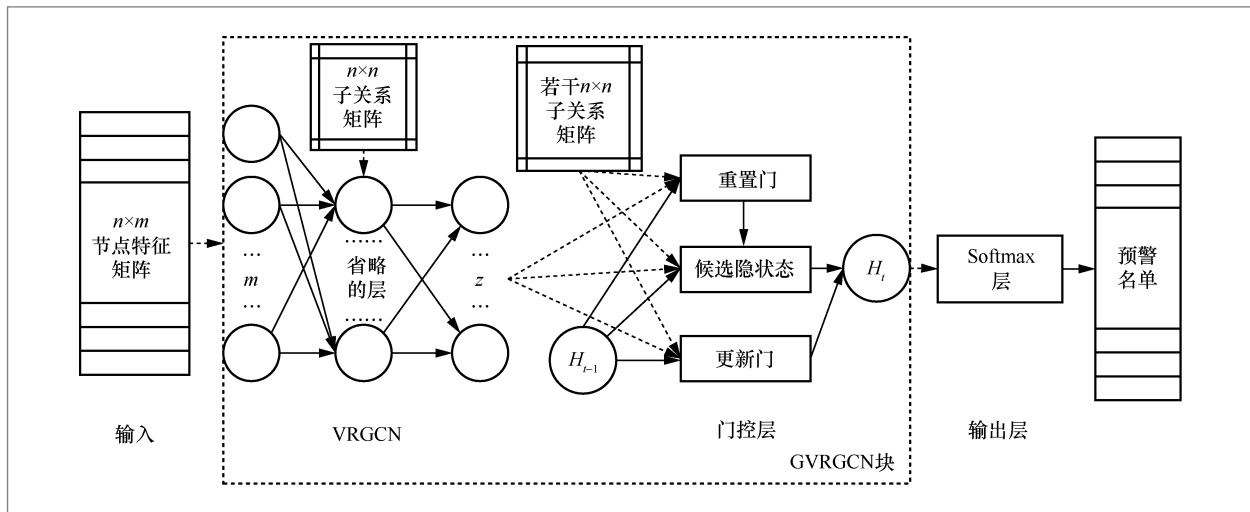


图3 门控变关系图卷积网络的网络结构

代码使用python3.9及numpy1.19.2、pandas1.1.3、statsmodels0.12.0、torch1.11.0、matplotlib3.3.2编写。

## 4.2 模型评价指标

高危当事人识别这种二分类问题，一般使用混淆矩阵衡量指标，包括TP、TN、FP、FN。

由于总样本中非高危样本占绝大多数，关注TN没有多大意义，并且在实际工作中，预测的结果还需要进一步核实。其中，精确率的高低体现了核实工作的工作量。而召回率体现了漏识别多少，直接反映预测集价值，召回率比精确率更能体现方法的好坏，因此增加召回率的权重。实验使用式(7)精确率、式(8)召回率与式(9)按照实际需求修改的F1分数评价实验方法。

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}) \quad (7)$$

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN}) \quad (8)$$

$$\text{F1} = 3 \times \text{Precision} \times \text{Recall} / (2 \times \text{Precision} + \text{Recall}) \quad (9)$$

## 4.3 数据集

本实验的数据源自某烟草公司近7年来的历史数据，总共包含3种表。根据保密条例，本文只能列举仿写样例。第一种表是案件表，包含案件名称、违法行为、刑事罪名等17个数据项，见表1。

第二种表是涉案当事人表，包含所属案件、姓名、身份证号等17个数据项。样例见表2。

第三种表是存有高危当事人信息的高危表。高危当事人的数量远少于一般人，因此识别中存在类不平衡情况。

## 4.4 数据预处理

为获得图数据，对原始数据进行预处理。

### (1) 特征工程

删除重复、绝大部分值为空、相关性高、对识别无用的特征。留下案件名称、案件类型、案发时间、涉案金额、涉案人数、违法行为、刑事罪名、所属案件、姓名、身

份证号等数据项。

### (2) 数据清洗

首先,为重名的列起别名。接着,删除案件名称、所属案件、姓名、身份证号为空的记录,删除涉案人数为0的记录。然后,进行去重、异常值处理。若数据项是否为主犯为空,则查看所属案件涉案人数,若人数是1,就认为其是主犯,否则一律判断不是主犯。

### (3) 构造图数据

图数据包含节点邻接矩阵与节点特征矩阵。

构造邻接矩阵。首先按照身份证号和案发时间对当事人数据进行排序。接着,生成若干个节点数大小乘节点数大小的单位矩阵,包括同案子邻接矩阵和若干同人子邻接矩阵。按同案关系,将同案子邻接矩阵对应位置赋值为0.5。同人子邻接矩阵的数量等于同人关系边数最多的同人节点数。为了控制门控层的计算,同人子邻接矩阵要按照同人关系里的案发时间的排序来赋值。根据这个思路,本文将在各同人子邻接矩阵的各当事人节点的第一行,按照同人关系的时间序列赋值1。为了保证同人关系节点的相同危险性,门控层最后一个参与计算的同人子邻接矩阵要按照同人关系,将同一列赋为1。

构造特征矩阵。首先生成节点数乘3大小的特征矩阵。矩阵第一列填充涉案金额,第二列填充是否为主犯,若对主犯赋值为1,其余赋为0;第三列填充案件级数,此项根据案件类型、违法行为、刑事罪名综合判定。原则上,刑事案件的案件级数要远大于行政案件的,见表3。若违法行为、刑事罪名等特征为空值,则默认行政案件的案件级数为4、刑事案件的案件级数为10。

图数据样例如图4所示。其中同人子关系邻接矩阵的元素只有1或0。

### (4) 划分训练集与测试集

本文的任务目标是验证GVRGCN的高

表1 案件记录样例(经过省略处理)

案件名称	类型	涉案金额	立案日期	违法行为	刑事罪名
贵州6号	行政	1 000	2020/1/1	无证经营	—
贵州15号	刑事	110 000	2020/2/2	无证经营	非法经营
神州9号	刑事	99 999 999	2019/9/9	无证经营	非法经营

表2 涉案当事人记录样例(经过省略处理)

所属案件	姓名	身份证号	是否主犯	违法行为	刑事罪名
贵州6号	甲	11111111	是	无证经营	—
贵州15号	甲	11111111	是	无证经营	非法经营
神州9号	乙	99999999	是	无证经营	非法经营

危当事人识别效果,进行半监督实验。实验的测试集选用2021年的数据,训练集选用其他年份的数据,并根据高危表为样本生成标签。

## 4.5 模型设置和实现方法

根据高危当事人特征分析,最多只考虑二阶同案关系,因此GVRGCN训练只有两层GVRGCN块,且GVRGCN块中的同案子关系矩阵参与第一层的计算,同人子关系矩阵参与第二层的计算。这样,模型也能避免网络变深而性能退化。两个GVRGCN块的第一层的输入、输出与第二层的输入的维度都是3,不同的是第一个GVRGCN块的第二层的输出维度是3,第二个GVRGCN块的是1。GVRGCN的激活函数是ReLU。GVRGCN的损失函数采用

表3 案件级数(有脱敏处理)

类型	违法行为/刑事罪名	案件级数
行政	无证邮寄	1
行政	未在当地进货	2
行政	擅收烟叶	3
行政	无证运输、无证经营、无证生产、其他	4
行政	销售非法烟、销售回流烟、销售走私烟	5
行政	毒物超标	6
刑事	销售伪劣产品、非法经营	11

同案子关系邻接矩阵			同人员关系邻接矩阵			完整关系邻接矩阵			特征矩阵		
0	1	2	0	1	2	0	1	2	涉案金额	是否主犯	案件级数
1	0	0.5	1	1	0	1	1	0.5	1000	1	4
0	1	0	1	1	0	1	1	0	110000	1	16
0.5	0	1	0	0	1	0.5	0	1	999999	1	16

图4 图数据样例

交叉熵函数。模型选用Adam优化器进行训练,学习率为0.01,迭代训练200轮。

图数据的构造通过循环和切片的方式实现。

为了缓解类不平衡问题,本文提出在训练过程中进行掩码迭代,但它会增加总训练轮次。实现方法是只迭代高危节点与随机选择同等数量的非高危节点,从而平衡迭代时高危节点和非高危节点的数量。本实验为测试GVRGCN的半监督效果,因此没有开启掩码迭代。

#### 4.6 实验结果与分析

为验证GVRGCN的效果,选取以下方法进行对比实验,实验结果见表4。

(1) 两层GVRGCN块的GVRGCN的Softmax分类识别。

(2) GCN的Softmax分类识别,网络结构和去除门控层的(1)相同。

(3) 使用SPN预测。

(4) 用第2节的高危当事人识别规则筛选。

如表4所示,分类识别的精确率要比规则识别的高,然而其召回率却更低。这是因为识别规则筛选出的样本数量太多。基于VRGCN的GVRGCN的表现最好,这验证了变关系与门控层的有效性。

表4还给出了各个方法运行1 000次所需的平均计算时间。这个时间除了方法的计算时间,还包含数据预处理时间。GVRGCN的计算时间较长,约为识别规则筛选所需时间的两倍,但在实际运用中,这些都是可以容忍的。

大数据稽查系统架构如图5所示,它旨在花费最少的资源、跨越最少的部门、尽可能地提升监管水平。专卖管理科的预警系统利用大数据技术和预警模型挖掘信息中心存储的历史数据,包含营销部门和稽查员日常维护的数据。得出预警名单后,预警系统发送给各个分局的稽查员,稽查员再通过预警名单进行巡查,并维护稽查数据。

案件命中率指的是市场巡查转化的案件数与商户检查次数的比率,反映了监管水平的高低。为验证有效性选取某市烟草专卖局应用本文的研究成果前的某2个月与应用后对应的月份的数据统计市场巡查涉烟案件

表4 实验结果

方法	Precision	Recall	F1	计算时间/s
GVRGCN	10.71%	100%	0.2646	5.9734
GCN+Softmax	9.68%	33.33%	0.1837	4.5887
SPN	12.5%	33.33%	0.2143	6.2075
识别规则筛选	2.7%	100%	0.0769	2.6328

命中率,结果如图6所示。需要说明的是,根据保密要求,各单位名称经过脱敏处理。

从图6可以看到,除了少数县级分局,大部分县级分局的月份市场巡查涉烟案件命中率有巨大的提升。在累计所有县级分局的命中率并除以县级分局数量后,发现总计案件命中率从原来的0.01%提升到了约0.5%,提升了接近50倍。这表明本文的研究成果完成了提升烟草市场监管能力和水平的目标,并验证了本文思路和方法的有效性。

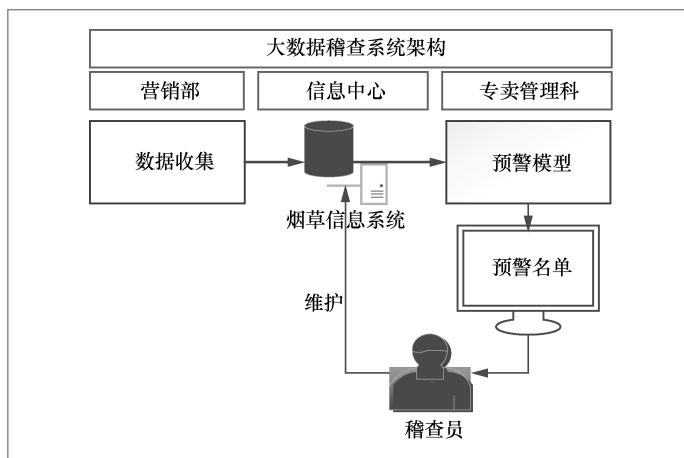


图5 大数据稽查系统架构

## 5 结束语

为提升烟草专卖监管水平,本文提出了适用于涉烟案件特定格式数据的门控变关系图卷积网络,用于获取预警名单。本文基于某烟草公司7年涉烟案件的历史数据,设计了对比实验,证明了本文方法的有效性,达到了监管需求目标。

实验发现,模型表现较差的主要原因是数据。本文研究了一套专门的数据预处理手段,但从采样结果上看还是有很多脏

数据,这要求研发出更准确的处理技术。

另外,预警方法还需要继续改进。从精确率来说,模型还需要加强。未来的研究方向是将图神经网络与准确的聚类方法结合。

## 参考文献:

- [1] 毕延浩. 2022年烟草行业实现税利总额和财政总额创历史新高[N]. 东方烟草报, 2023-03-07(001).

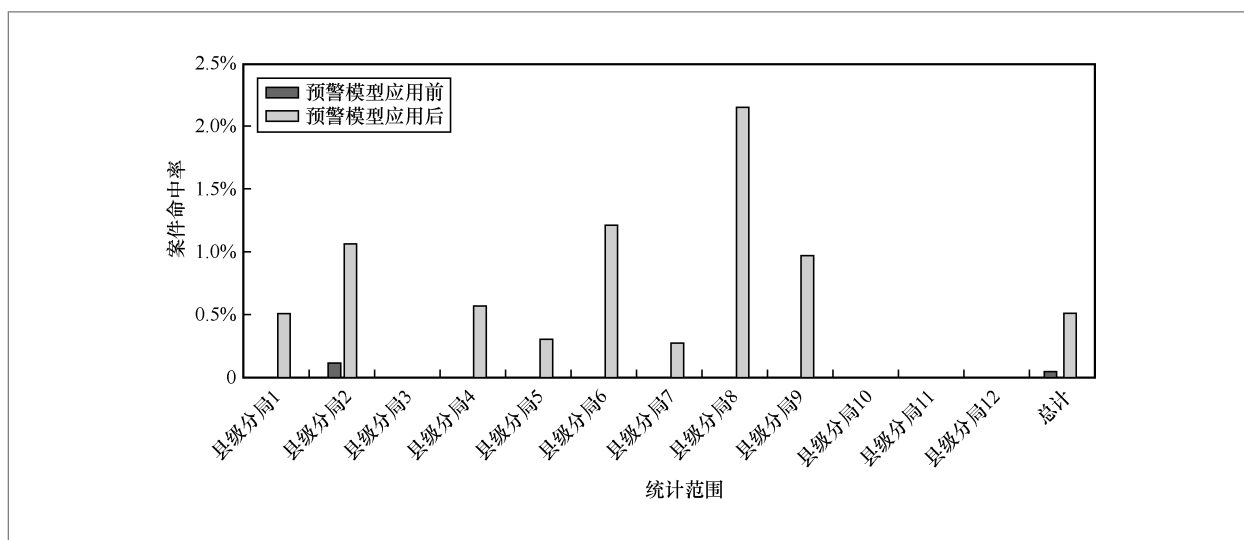


图6 某市烟草专卖局涉烟案件命中率

- BI Y H. 2022 Tobacco industry achieves record total tax profits and total finances[N]. Eastobacco News, 2023-03-07(001).
- [2] 刘娜, 陈博. 铸“数”为剑提战力[N]. 东方烟草报, 2023-01-19(001).  
LIU N, CHEN B. Casting “numbers” as a sword to enhance combat power[N]. Eastobacco News, 2023-01-19(001).
- [3] 胡敏. 持续保持监管高压态势织密未成年人“保护网” [N]. 东方烟草报, 2023-02-09(001).  
HU M. Continuing to maintain a high-pressure situation of supervision and organizing a close “protection network” for minors[N]. Eastobacco News, 2023-02-09(001).
- [4] 李中皓, 毛健, 曹培健, 等. 数字化转型: 中国烟草科技创新发展的必然选择[J]. 烟草科技, 2022, 55(1): 1-8.  
LI Z H, MAO J, CAO P J, et al. Digital transformation: an inevitable choice for innovation and development of tobacco science and technology in China[J]. Tobacco Science & Technology, 2022, 55(1): 1-8.
- [5] 乔浪超, 王进录, 高宝红, 等. 基于时空数据特征的寄递涉烟犯罪分析方法[J]. 中国烟草学报, 2023, 29(1): 116-126.  
QIAO L C, WANG J L, GAO B H, et al. Express-related counterfeit cigarette criminality analysis based on spatio-temporal data features[J]. Acta Tabacaria Sinica, 2023, 29(1): 116-126.
- [6] 王贞, 尤梓荃, 张锦程, 等. 基于标签传播的涉烟车辆异常检测[J]. 仪器仪表学报, 2021, 42(3): 161-167.  
WANG Z, YOU Z Q, ZHANG J C, et al. Anomaly detection of cigarette-smuggling vehicles based on label propagation[J]. Chinese Journal of Scientific Instrument, 2021, 42(3): 161-167.
- [7] CAO P C, LIU Q, LIAO N H, et al. A catalog of collected debris disks: properties, classifications and correlations between disks and stars/planets[J]. Research in Astronomy and Astrophysics, 2023, 23(8): 67-83.
- [8] YU B H, DENG C, BU L P. Policy text classification algorithm based on bert[C]// Proceedings of 2022 11th International Conference of Information and Communication Technology. Piscataway: IEEE, 2022: 488-491.
- [9] ZHAO G, QIU B, LUO A, et al. Deep learning applications based on wise infrared data: classification of stars, galaxies and quasars[J]. Research in Astronomy and Astrophysics, 2023, 23(8): 166-179.
- [10] MENG Q, CAI H Y, TANG J. Neural structured prediction for inductive node classification[C]//Proceedings of the Tenth International Conference on Learning Representations. Amherst: OpenReview.net, 2022: 1-21.
- [11] DENG C H, ZHAO Z Q, WANG Y Y, et al. GraphZoom: a multi-level spectral approach for accurate and scalable graph embedding[C]//Proceedings of International Conference on Learning Representations. Amherst: OpenReview.net, 2020: 1-20.
- [12] YU D H, ZHANG R H, JIANG Z B, et al. Graph-revised convolutional network[C]// Proceedings of European Conference on Machine Learning and Knowledge Discovery in Databases. Cham: Springer, 2021: 378-393.
- [13] LI K J, YE W J. Semi-supervised node classification via graph learning convolutional neural network[J]. Applied Intelligence, 2022, 52(11): 12724-12736.
- [14] 周纯英, 曾诚, 何鹏, 等. GKCI: 改进的基于图神经网络的关键类识别方法[J]. 软件学报, 2023, 34(6): 2509-2525.  
ZHOU C Y, ZENG C, HE P, et al. GKCI: an improved GNN-based key class identification method[J]. Journal of Software, 2023, 34(6): 2509-2525.
- [15] 杨晓晖, 王卫宾. 基于门控图神经网络的谣言检测模型[J]. 燕山大学学报, 2023, 47(1): 73-81.  
YANG X H, WANG W B. Rumor detection

model based on gated graph neural network[J]. Journal of Yanshan University, 2023, 47(1): 73–81.

[16] LI G H, MATTHIAS M, QIAN G C, et al.

DeepGCNs: making GCNs go as deep as CNNs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2023, 45(6): 6923–6939.

#### 作者简介



冯鹏程(1998- ),男,贵州师范大学大数据与计算机科学学院硕士生,主要研究方向为数据挖掘。



张高豪(1984- ),男,贵州省烟草公司贵阳市公司烟草专卖管理师,主要研究方向为烟草专卖。



谢刚(1980- ),男,博士,贵州师范大学教授,主要研究方向为数据挖掘。

收稿日期: 2023-09-22

通信作者: 谢刚, 48263091@qq.com

基金项目: 贵州省烟草公司贵阳市公司科技项目(黔烟筑科[2020]3号)

Foundation Item: Tobacco Company Guiyang Company Science and Technology Project, Guizhou Tobacco Building Department (2020 No. 3)