

# 全息感知视域下卷烟市场监管多源数据融合与智能预警机制研究

刘彤\* 杨龙龙 周潇 孙昕

山东潍坊烟草有限公司, 中国·山东 潍坊 261205

**摘要:** 卷烟市场监管存在数据孤岛和滞后问题, 本文提出了将全息感知应用于卷烟市场监管中, 设计了四层(感知层、数据层、模型层、应用层)架构, 并对三类异构数据进行了整理、融合并建立了关联规则融合模型; 针对假私冲重点户, 开展特征工程, 建立了进货异常指数、销售偏差指数等多维预警指标体系, 采用随机森林、支持向量机以及 DBSCAN 等算法建立预警模型并优化了动态阈值。研究表明, 多源数据融合能够有效打破部门壁垒, 提高风险识别精准度, 形成的预警机制可用于卷烟市场监管工作中的技术支持, 也可为烟草行业智慧监管提供一定理论参考。  
**关键词:** 全息感知; 卷烟市场监管; 多源数据融合; 智能预警; 假私冲

## Research on Multi-source Data Fusion and Intelligent Early Warning Mechanism for Tobacco Market Regulation under the Holographic Perception Framework

Liu Tong\*, Yang Longlong, Zhou Xiao, Sun Xin

Shandong Weifang Tobacco Co., Ltd., China Shandong Weifang 261205

**Abstract:** There are issues of data silos and delays in cigarette market regulation. This article proposes applying holographic perception to cigarette market supervision, designing a four-layer architecture (perception layer, data layer, model layer, application layer), and organizing, integrating, and establishing an association rule fusion model for three types of heterogeneous data. Focusing on key accounts involved in fake or private production, feature engineering was carried out, and a multi-dimensional early warning indicator system was established, including purchase anomaly index and sales deviation index. Algorithms like random forest, support vector machine, and DBSCAN were used to build early warning models and optimize dynamic thresholds. The study shows that multi-source data integration can effectively break departmental barriers, improve risk identification accuracy, and the resulting early warning mechanism can provide technical support for cigarette market regulation and offer a theoretical reference for intelligent supervision in the tobacco industry.

**Keywords:** Holographic perception; Cigarette market regulation; Multi-source data fusion; Intelligent early warning; Counterfeit and private sales fraud detection

## 0 引言

目前卷烟市场监管存在数据壁垒高、信息融合不够、风险预警滞后等问题, 导致市场监督管理能力无法充分发挥。为此本文提出全息感知理念, 目标是打造多源数据融合和智能预警监管模式。通过搭建四层(感知层、数据层、模型层、应用层)架构, 实现多源异构数据整合, 应用机器学习算法建立预警模型, 精准识别、提前预警卷烟市场监管中的“假私冲”等重点风险, 提高卷烟市场监管的智能化、精准化水平。

## 1 全息感知卷烟市场监管的理论内涵与框架构建

全息感知重点在于运用多种手段采集、获取多维度数

据, 并进行有效融合, 从而实现对所感知对象的全方位立体化认识。将其应用于卷烟监管中, 打造全景式监管模式, 以解决传统监管方式存在的信息滞后及盲区问题。

监管体系四层逻辑架构图分为: 感知层、数据层、模型层、应用层, 四层相互作用形成的闭环, 对应的是采集—治理—分析—应用的全流程, 监管机构通过这样的方式, 进行数据收集、整合、加工、处理和预警调度。

多源数据融合是这个体系的重点, 主要作用就是: 打破部门之间的信息壁垒, 使数据互联互通; 关联分析内外数据, 提高对假私非等违法行为的发现精准度; 为智能预警模型提供高质量的基础数据, 这是达到由被动监管向主动预警监管转变的重要手段。

## 2 卷烟市场监管多源数据的采集与融合机制

### 2.1 多源异构监管数据的采集维度与类型

卷烟市场监管涉及来自不同渠道的多源异构数据,从三个方面整理数据资源:一是烟草专卖内部数据,包括零售户订货记录、许可证信息、历史违规情况、经营档案、周边群众评价等,用于呈现零售主体经营状况和信用状况;二是跨部门协同数据,包括市场监管局的营业执照年检和经营范围变更信息、公安局的涉烟违法人员名单和嫌疑车辆活动轨迹、税务部门的纳税数据、物流快递数据等,用来描述监管对象的外部关联特征;三是外部感知数据,主要是指治安监控和市政交通摄像头的高清视频流、GPS定位数据、通信运营商的通话记录和基站切换轨迹、互联网平台上的交易记录和社交账号动态等非结构化流数据。这些外部感知数据是构成全息感知能力的重要基础,在时空域对监管场景进行全方位立体化覆盖。结合具体的业务场景来看,烟草专卖内部数据主要用于零售户信用评分、订货异常预警;跨部门协同数据主要用于违法网络溯源、关联主体识别;外部感知数据主要用于空间时间特征跟踪、布控。从数据本身的特点来看,内部数据结构化程度高且有时间戳;协同数据是关系型网络;外部感知数据是非结构化流数据,三类数据相互补充、互相叠加,可以形成围绕监管对象经营状态、社会关系和社会行为的全息化描述,使得多维度数据更加完善,也为后期进行机器学习建模和智能分析提供更加充分的数据基础。

### 2.2 多源数据的预处理与标准化方法

多源异构数据接入以后需要经过系统性的预处理和标准化才能入库。数据清洗过程中,利用数据比对算法识别并修正异常记录,通过均值填充、回归插补等方式处理缺失值,通过统计异常检测和孤立森林算法等检测方法找到离群点并进行标记,将预处理后的数据输入到库中,以保证后续建模分析用的数据是真实的、可靠的。数据标准化就是统一各种不同数据源数据的格式、规范和编码,建立数据质量监控机制,使得跨部门之间的数据具有可比性和互操作性。在存储层使用 Hive、Spark 等分布式大数据技术搭建基础层,在 ETL 工具中实现数据抽取、转换、加载过程,完成多源数据的高效整合和统一管理。并且用完整性、一致性和及时性等多维度的数据评估指标来对进入数据库的数据质量打分评级,并定时分析和定位入库数据质量存在的问题并追根溯源,将数据准确率、字段填充率、更新及时率等纳入到考核指标当中来,从源头上确保数据分析和建模用的数据基础是扎实可靠的。同时还需建立数据质

量反馈闭环机制,及时将清洗过程中发现的质量问题反馈给数据采集部门,不断净化数据源质量。在标准实施层面,对于不同部门之间的数据语义区别需要制定统一的数据字典和元数据规范,明确每个字段的业务含义及取值范围,以保证在跨系统的数据融合过程中达到语义一致、口径统一的效果,从而保证关联规则挖掘所用的数据是优质的。

### 2.3 基于关联规则的数据融合模型构建

完成标准化后,需要建立融合模型,提炼多源数据内在关联,在此基础上搭建跨部门数据协同管理平台,实现数据接入、更新、存储的统一管控,打破信息孤岛。运用关联规则算法扫描多源数据集,寻找数据字段之间的隐含关系和模式,采用 Apriori 和 FP-Growth 算法挖掘零售户订货行为、物流轨迹等的频繁项集,查找零售户订货行为和物流轨迹之间的耦合规律。融合的过程以时间、地点、对象为维,关联匹配不同系统的异构数据,形成完整的证据链。对于可能产生的记录冲突问题,利用一致性校验和冲突消解机制,通过优先级规则以及可信度评估策略来保证融合结果的正确性。将融合之后的多维数据用于监管决策,例如基于融合后的数据形成零售户综合风险画像、发现跨区域违法活动网络、预测重点时段及区域的监管需求等,帮助烟草专卖局制定差异化巡查方案、精准化配置监管资源,促进由经验型向数据型转变。当融合模型建成后可以通过对接大屏或者报表系统的方式实现多维分析的功能,协助监管人员全面了解市场情况、准确找到风险点并作出正确判断和决定。

## 3 卷烟市场智能预警模型构建与算法设计

### 3.1 假私冲重点户识别的特征工程与模型构建

假私冲重点户精准识别是智能预警系统的重中之重,要实现该项功能必须建立一个高区分度的特征体系,并选择合适的分类算法。筛选出进货渠道可疑度、进销存偏差度、销售金额和纳税额匹配度、历史上是否发生过违规行为、周围群众的举报频率、经营面积与客流量匹配度以及卷烟陈列是否规范等经营行为、财务关联和环境反馈三个方面的主要特征,来对该零售户进行全面描述,在此基础上采用信息增益、卡方检验和递归特征消除相结合的方法论体系,对每个特征变量的区分能力以及相关冗余度进行分析评价,将一些信息值较低、相关性较高的低区分度特征从原来的特征集合中予以排除,仅保留对于目标变量具有较高解释力的高区分度特征子集,进而有效规避维度灾难问题并提高计算效率<sup>[1]</sup>。鉴于各类别的样本分布极不均匀,假私冲重点户属于稀有样本,在此利用 SMOTE 过采

样及代价敏感学习相结合的方式,对少数类样本适当扩充,充分挖掘假私冲重点户的潜在样本空间,提高模型对假私冲重点户的敏感性。

### 3.2 智能预警指标体系与动态阈值设定

科学的预警指标体系是实现风险精准识别的重要前提,本文构建的预警指标体系由进货异常指数、销售偏差指数、行为异常指数和关联风险指数组成,分别从货源渠道、销售波动和关联网络三个方面量化违法违规风险。针对预警阈值,传统的固定阈值无法紧跟市场环境的变化,因此本文提出的基于机器学习的自适应阈值算法结合市场季节性波动以及政策调控等实时反馈数据动态调整预警指标阈值预警规则及权重,并根据风险程度大小划分预警等级,实现风险初始出现即能及时预警的目的,也可以较好的控制误报率和漏报率。对于预警等级,将风险信号由低到高设定为蓝色关注、黄色提示、橙色警戒、红色预警4个等级,对应的响应措施分别为一般关注、重点监测、现场核查和立即处置,在预警等级判断上,结合预警等级的综合风险评分区间和触发条件,实施预警等级判定,做到预警有梯度、响应有重点。

### 3.3 基于机器学习的预警算法优化策略

预警模型的不断优化有利于保障系统长时间稳定地工作。为了及时了解各种预警模型的自身性能情况,本文提出了基于准确率、召回率以及F1值等指标建立预警模型性能监测体系,并使用AUC-ROC曲线和精确率-召回率曲线作为辅助评判手段,针对不同的预警阈值分析模型优劣,分析预警模型自身性能的变化规律,对预警算法做周期性评价;并且采用TensorFlow、PyTorch等深度学习框架搭建深度神经网络模型,抽取多源融合数据中的非线性特征关系,提高对隐秘违规行为的识别能力。同时,系统利用实时新数据反馈,采用梯度下降法等优化算法实时修正模型参数,并反复迭代优化模型,从而实现精准击中灰黑名单目标。就整个迭代优化过程而言,系统采取在线学习和离线批处理并行开展的方式:离线部分则利用每个月整月全量的历史数据,在每个预警模型下每天将该月数据集分割成27个时点,并利用此模型学习,当A/B测试证明新模型确实优于旧版时才能更新线上代码;在线部分则是通过流式处理引擎实时读取新增加的部分样本,进行模型增量学习及参数细微调节,并建立模型版本管理和快速回滚机制,在每次迭代升级过程中步步为营谨慎操作,使预警机制跟随市场环境的变化以及监管政策的要求不断发展完善,提高系统运用实践效果及监管智能化水平<sup>[3]</sup>。

## 4 全息感知智能预警系统的应用路径与保障机制

### 4.1 智能预警系统的功能架构与技术集成

全息感知智能预警系统是基于模块化设计思想研发的,综合应用了全息感知摄像头应用、GPS定位和轨迹追踪、涉烟人员轨迹还原分析、实时预警推送、执法调度指挥等功能。技术上通过多协议适配中间件实现异构厂家摄像头接入;通过HLS、RTMP流媒体技术保证视频数据高效传输;利用Ceph分布式存储满足海量视频数据存储需求;使用RabbitMQ、Kafka消息队列实现多级联控预警实时触发;将深度学习视频分析模型嵌入到系统中对异常行为进行实时智能监测,在关联周边执法力量移动终端的同时调度周边摄像头对准事发区域,形成从感知发现到处置响应的闭环管理链;通过构建数据传输-存储-访问全生命周期的安全防护体系,利用TLS加密协议保障数据链路安全,使用AES-256加密算法对存储层中的敏感数据进行加密处理,利用角色基于授权与多因素身份认证方式实现对访问层中具有不同敏感等级的信息的完整授权,并且建立完善操作日志审计系统对重要数据和信息的访问及系统操作行为全过程进行记录,保证了系统的安全可控运行以及有效保护了数据资产<sup>[3]</sup>。

### 4.2 跨部门协同联动的执法保障机制

以业务流程整合和信息共享为基础,以烟草专卖、市场监管、公安等部门业务系统为支撑,实现从线索发现、案件移送、联合执法到案件处置全过程的线上协同作业;通过建立常态化数据共享、联合研判和应急联动工作机制,结合智能调度算法实现执法资源的科学调配和快速响应;当触发预警时,通过弹窗消息、短信等方式将预警信息推送至监管人员的移动终端或指挥中心大屏上,同时联动交通部门规划最优拦截路线,利用无人机高空巡查辅助定位追踪,形成全方位、立体化、多层次的执法保障网络;加强对系统管理人员和一线执法人员的培训培养,建立分层分类培训体系,对系统管理人员重点开展技术运维和数据分析培训,对一线执法人员重点开展系统操作和智能预警响应实务培训,同时,组织跨部门联合演练和模拟执法场景训练,并将系统使用能力纳入到执法人员绩效考核体系中,确保通过不断的能力建设使得智能化监管工具在执法实践中发挥出最大效用<sup>[4]</sup>。

### 4.3 系统应用的效果评估与持续优化路径

对系统应用效果的评价从假私冲卷烟查获率、预警响应速度、监控覆盖范围、联合执法案件办结率、举报线索

转化率等方面建立一套科学的量化评价指标体系；并针对零售户信用等级评价结果，对于合规经营并且整改到位的重点监管户及时移出重点监管名单，对于新的行为异常对象及时纳入重点监管，根据新获取的线索，不断运用机器学习算法迭代修正预警模型，实现监管资源的精准投入和调剂使用。在系统推广上，计划先试点再推广、循序渐进的方式，先选择监管需求高、信息化基础好的地区作为首批试点单位，在实际工作中摸索总结出可推广、可复制的建设模式和实施办法，在全省范围内分期分批有序推开，以典型示范带动整体应用水平提升<sup>[5]</sup>。

## 5 结语

通过研究可知，运用全息感知建立的多源数据融合和智能预警机制能有效解决卷烟市场监管过程中存在的数据孤岛问题以及滞后反应问题，利用多源数据建立预警模型并采用随机森林、DBSCAN 等算法对重点风险“假私冲”等进行预警预报，能够达到较高精度并且具有一定的实时性。提出的四层结构及实施方案，是提高监管精准度和时效性的直接技术手段，同时也是推进烟草行业向智慧监管方向发展的可行的理论参考和实践路径，需要在今后的工

作中不断优化模型，并加强跨部门协同。

## 参考文献：

[1] 姜思明, 刘荣, 谭升达等. 基于零售终端数据治理的卷烟市场状态监测研究[J/OL]. 烟草科技, 1-21[2026-05-25].

[2] 王平, 廖磊, 孔庆民. 基于多维指标的区域卷烟市场容量动态预测模型[J]. 上海商业, 2026(04): 206-208.

[3] 宁涛, 高琦, 薛梓远等. 基于利益相关者理论的卷烟零售点市场容量影响因素指标体系研究[J]. 中国市场, 2026(07): 99-104.

[4] 陈雪芬, 李家成, 严杰等. 基于时序融合 Transformer 的卷烟市场信息动态还原模型研究[J]. 商场现代化, 2025(22): 1-4.

[5] 林原, 张赵成, 张晗等. 烟草制品市场管控体系研究[J]. 中国市场, 2025(31): 72-75.

作者简介：\* 通讯作者：刘彤（1991-），女，汉族，山东潍坊人，本科，研究方向：主要从事烟草专卖管理方面工作。