大数据技术在高速公路收费审计中的应用

杨世威

高速公路收费数据总量大、更新快，总量金额高，高速公路企业对收费精准度、公正性以及对逃费稽查及时性和处置高效性提出较高要求，高速公路收费审计的质量也受到企业高度关注。大数据技术具有海量数据收集、实时处理等优势，可以实现实时、全范围审计，有利于提高内部审计的效率和效果，进而提升收费业务的监管力度。本文以J公司为例，探讨高速公路收费业务大数据审计模型构建方法及应用。

一、J公司高速公路收费大数据审计模型的开发过程

（一）审计模型的设计

模型设计人员通过以下五步去设计审计相关模型。第一，确定模型构建的目标。第二，数据清洗。确认收费原始数据信息全面、准确；确保车辆通行费应收尽收；确认车辆通行费拆分的正确性。第三，确定关键数据。审计人员确定数据来源、数据类型及分析数据的方法。高速公路车辆收费大数据包含200多个字段，数据结构多样，设计人员构建模型时，根据车牌、车行、路径等核心指标设置字段，提取相关数据。第四，构建模型。主要从两个维度出发：一是业务侧驱动模型构建；二是数据侧驱动模型构建。

基于业务侧驱动的模型构建：运用数据增加交叉分析法，J企业模型设计人员首先根据数据的业务、自然属性将其分为一类基本指标、二类细化指标、三类计算衍生指标。一类基本指标主要包括通行车辆的车型、载重、营运种类等；二类细化指标是在一类指标的基础上细化分类，如货车分为4轴、5轴货车等；三类计算衍生指标是指对一类或二类指标进行分类、分组、排序、计算处理后的突显特征的指标。在对指标数据进行归类、处理后，模型设计人员会进行PDCA（计划、执行、检查、处理）闭环测试，采用管理改进措施，分析指标变化情况，确定基本模型方案。基本模型经过短期使用后，由专家对模型的有效性进行评估，若有效性达到80%以上则正式发布使用。发布后模型的有效性由系统自动进行评估，若结果呈现特征指标下降情形，当有效性降低到50%附近时，模型开发人员需要对模型进行迭代升级或开发新模型。审计人员一般3个月进行一次复核检查。

以业务驱动方式开发模型需注意：模型设计人员必须深挖业务，通过业务找数据、通过问题找数据。如数据的积累程度不够，业务场景的多元性便会收到影响。设计人员也需不断关注新技术、新资源，这些有可能对模型和数据产生重大影响。构建模型是不断迭代更新的过程，模型投入使用后，并非一成不变，设计人员需要根据实时动态的数据变化、业务变化不断对模型进行调整，以使模型可以适用于各种场景，使之成为有效的、有价值的审计工具。

基于数据侧驱动的模型构建：设计人员需了解和掌握公司已有的数据资源，寻找公司没有掌握的数据资源，尽可能多地收集与本公司行业、业务、上下游相关的数据资源。大数据分析是一个迭代的过程，很多的分析场景都是在不断的探索中找出来的。发现问题之后需要分析人员将问题带到业务具体场景中进行分析，得出结论，再优化产品、流程，然后再分析、预测、验证，由以往的经验判断过渡到数据判断。

设计人员开发模型时业务驱动方式和数据驱动方式相辅相成。审计人员可对业务信息和数据信息进行验证，验证结果有助于审计人员发现并判断问题。比如，某一车辆通过不同高速路段，一方面收费设备拍摄到了该车部分图片信息，即获得了业务信息；另一方面系统提取到该车辆所有门架信息，即获得了数据信息。审计人员将业务和数据缴费信息进行比对，查看其金额是否一致，判断是否存在问题。

模型初步构建成功后，审计专家需核验模型。通过提取样本数据，将提取到的数据导入模型，进行模型的评估和优化，确认模型是否达到可使用状态。审计专家同业务人员一起针对模型中的目标用户，检查模型是否完成预期目标，是否验证出审计问题，以及是否真的存在审计疑点。之后由计算机专业人员拟定算法，给出相应策略。如：当模型规则不清晰导致模型脚本编写有误时，专家需要就规则进行进一步沟通，调整模型脚本。当模型影响因子考虑不全面导致模型结果的实用性低时，设计人员需要加入

中国财政杂志社

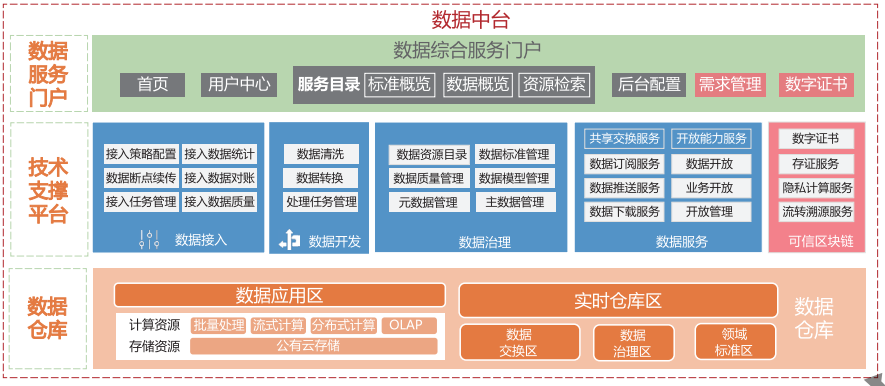

图1　J公司数据中台总体架构

新的模型因子，进行数据筛查，提高模型数据的可用性。

模型通过专家核验后，将集成到大数据审计平台进行发布供审计人员使用。后续审计人员也可对审计结果进行统计分析，建立疑点数据库，以备后期用于发现疑点问题。审计人员可对各个模型的每月疑点数据增长趋势进行分析，分析结果供审计人员确定现场核查审计重点。后期也可以通过该增长趋势查看专项审计的成效。

（二）获取通行费收缴相关数据

审计人员发布大数据审计模型之后，需要获取高速公路收费的相应数据导入模型，进行大数据分析：

首先，根据审计业务规则，梳理高速公路车辆收费大数据审计所需要的数据。数据信息包括路段门架ETC感应日志数据（车牌号、车型、时间戳），行驶过程中监控设备抓拍的日志数据（车型、车牌号、时间戳），收费站车道收费数据（车牌号、车型、出入站信息、计费信息、时间戳）和收费站日常运营日志数据（日过车数、日收费数据等）。正常的收费数据会自动传输到联网中心的数据中心，同时审计人员会去收费站站点采样，将每个站点采样的数据跟联网中心的数据进行对比，核查其数据是否一致，以此确认收费的数据信息的全面性、准确性。

其次，制定上述数据的获取方式以及同步规则。获取方式包括将企业信息系统内的数据导出或导入、从互联网上进行相关数据的下载等，同步规则可以定义为全量同步或者增量同步，同步频率可以根据时间需求定义成一天同步一次或者一月同步一次。

最后，进行数据搜集和核验：第一，确定高速公路车辆的真实行驶轨迹，统计起始地点和终止地点。第二，在收费站提取车辆通过时的收费金额记录，进行统计并上传至联网系统。第三，为确保系统内数据真实可靠，校验数据：按照车辆真实的行驶轨迹，对比应缴纳的通行金额与收费站提取的实际收费金额，确认通行与实际收费金额是否一致；确定省内通行费拆分总金额，依据省内拆分文件，确认每个社会单元应被拆分到的金额，且依据高速公路车辆实际通行情况，对比被拆分到的金额与车辆实际通行应收取的金额是否一致；对于外省通行费拆分情况，外省只提供总金额，没有具体拆分细则，本省将根据自己的社会单元进行拆分，对比不同社会单元被拆分到的金额与车辆实际通行应收取的金额是否一致。

企业将大数据分析技术运用到审计业务中去，将审计时获取的文字、图像等信息转化为数据，通过前期构建审计模型，分析、运算出数据背后蕴含的意义及信息，从而找到审计异常，现场复核，得出审计意见。

二、J公司高速公路收费审计模型成果

为应对在大数据场景下开展审计业务面临的新挑战，J公司以中台系统（见图1）呈现高速公路大数据结果，对数据持续积蓄和演化（业务数据化）。

J公司数据中台功能分为一个数据仓库和两个平台（包括技术支撑平台和数据服务门户）。数据仓库基于“公有云”建立，是数据中台的核心，一方面为高速公路运行管理提供准确全面的基础数据服务（如路网、车辆、人员、设施），另一方面通过不同业务的叠加以及部门间信息系统共享交换，融合形成专题数据（如收费、审计、营运、流量、养护等）。技术支撑平台提供数据接入、开发、治理、服务等方面的技术驱动，全面支撑数据中台的全生命周期数据管理。数据服务门户将内部与第三方数据资源整合后，集中向各类用户提供数据，实现内部各单位共享资源的统一呈现、检索、定位、使用，形成一站式信息资源门户。

在业务数据化的基础上，J公司中台进一步优化数据服务。大数据场景下，数据中台使得原有的业务工作得以解构与重塑，审计人员可以对算法服务进行一站式调用。这就在拓宽审计覆盖面的同时，做到了对于数据信息的关联分析，有利于审计人员有针对性地防范高速公路收费错漏风险，从而优化了高速公路企业的管理效能。

（本文系国家自然科学基金面上项目“政府规制、产品市场竞争网络与企业财务风险衍化”<71872040>及中央高校基本科研业务费项目<2242021S30014>的阶段性研究成果）

（作者单位：江苏交通控股有限公司审计部）

责任编辑　姜雪

中国财政杂志社