

大数据、人工智能与财税服务创新

周俊亭¹, 席彦群², 周媛媛², 邱涛³, 翁安栋³

(1. 厦门大学 管理学院, 福建 厦门 361005;

2. 安徽工业大学 商学院, 安徽 马鞍山 243032; 3. 唯你网, 福建 厦门 361008)

摘要: 传统的财税工作模式呈现出重复、繁杂、低效, 耗费大量的时间与人力特点。基于大数据、人工智能的智能财税服务, 实现了票据自动识别、语义解析、自动记账、机器审核、自动对账甚至一键报税等财税新模式。财税领域的大数据平台与财税智能机器人, 根据具体规则, 持续、高效地运作, 操作简单, 极大地提升财税工作的效率, 降低人为失误。介绍了大数据、人工智能在财税实务领域创新的初步成果, 依次论述了大数据技术, 人工智能, 财务领域大数据平台, 财税智能机器人, 大数据、人工智能在财税领域的关键技术问题。其中, 财税智能机器人具体包括: 费用稽核机器人、应收应付机器人、行政事业单位核算机器人、合同比对机器人、纳税服务机器人。关键技术包括: 票据识别技术, 语义识别技术, 自动记账, 一键报税。以达到对大数据、人工智能对财税领域的创新有全面而深刻认识的目的。

关键词: 大数据; 人工智能; 财税服务; 云平台; 智能机器人

中图分类号: F275.2 **文献标识码:** A **文章编号:** 1005-0566(2020)08-0069-09

Big Data, Artificial Intelligence and Innovation of Fiscal and Taxation Service

ZHOU Jun-ting¹, XI Yan-qun², ZHOU Yuan-yuan², QIU Tao³, WENG An-dong³

(1. School of Management, Xiamen University, Xiamen 361005, China;

2. School of Business, Anhui University of Technology, Ma'anshan 243032, China;

3. Onlyou.com Company Limited, Xiamen 361008, China)

Abstract: The traditional work mode of fiscal and taxation service manifests the characteristic of repeatability, complexity, inefficiency, time consuming and manpower consuming. Based on big data, the artificial intelligence fiscal and taxation service has implemented the new work patterns which contain invoice recognition automation, semantics analytic, bookkeeping automation, robot approval examine and verify, automatically reconciliation, even one key tax declaration and so on. According to concrete regulation, the big data platform and intelligent robots of fiscal and taxation service demonstrate that they can operate persistently, efficiently and easily, which can greatly improve efficiency and reduce man-made faults. In this paper, we introduce the preliminary achievement of artificial intelligence fiscal and taxation service. Most important, we successively describe the big data technology, artificial intelligence, finance big data platform, fiscal and taxation service intelligent robots and critical technical issues. Among which, the fiscal and taxation service intelligent robots consist of expenses approval examine and verify robot, receivables and payables robot, administrative institution accounting robot, contract alignment robot, taxation service robot. The critical technical issues

收稿日期: 2019-12-16 修回日期: 2020-07-22

基金项目: 安徽省自然科学基金青年项目“基于产业链连锁反应的企业并购整合策略研究”(1408085QG138)。

作者简介: 周俊亭(1986—), 女, 河南洛阳人, 厦门大学管理学院博士研究生。研究方向: 资本市场会计与财务问题。

consist of invoice recognition automation technology, semantics analytic technology, bookkeeping automation technology, one key tax declaration technology.

Key words: big data; artificial intelligence; fiscal and taxation service; cloud platform; intelligent robots

一、引言

财税工作经常遇到这样的场景: 票据数量和种类多, 经常加班干重复的工作, 报销程序复杂, 合规风险高, 稽核耗时, 需要反复核对数据, 大量时间在查找问题, 财税工作人员花费的大多数时间是在作重复、缓慢、低效的工作, 非信息化处理的成本高, 不能为企业提供及时的反馈和决策支持。企业决策多来源于纸质材料, 财税工作表现出复杂、低效、成本高、人为失误多等特点。在企业中, 财务部门最难提效。

大数据技术、人工智能的出现, 正在改变人们的生活方式^[1]、科研方式、企业运营模式与财税实务。随着德勤、安永财务机器人的出现, 四大会计师事务所纷纷推出财税智能机器人。以德勤和普华永道为代表的财税咨询已运用人工智能。面对不断扩大的社会需求, 更多的公司加入到智能财税的研发中, 已经出现了基于大数据的功能应用的人工智能云平台与财税机器人。基于大数据、人工智能的智能财税服务实现了全数据化的工作环境, 机器学习自动记账, 甚至自动报税, 将会逐步改变传统工作模式复杂、低效、成本高、人为失误多等问题, 引领财税智能化创新。

基于大数据、人工智能的智能财税服务包括平台服务与智能财税机器人, 能够实现票据、凭证、账簿、报表等各环节的流程化、透明化、自动化的过程管理。以原始票据为切入点, 通过对原始票据扫描、文本自动识别和人工校验方式采集票据影像的电子数据, 运用深度学习自动生成会计凭证, 提供凭证制单、审核、冲正、过账、结账等基础会计核算功能, 以及外币核算、辅助核算、汇兑损益自动计算与结转的会计核算功能。同时提供账务管理解决方案、自动生成审计工作底稿、税审工作底稿、财务数据分析等增值服务。

二、大数据、人工智能在财税实务领域创新的应用方向、初步成果

智能财税应用方向主要包括: 记账: 票据智能

识别, 自动记账; 财务分析: 机器自动分析; 税务: 自动报税, 税务筹划; 审计: 自动审计, 智能分析; 财务反欺诈: 智能分析, 实现财务反欺诈; 信贷服务: 数据信用, 金融征信, 企业征信; 投资管理: 智能预警, 事前投后管理; 可视化分析; 行业报告; 企业财务管理等。

大数据、人工智能在会计实务领域的初步创新成果包括: 全数据化工作环境、机器学习自动记账和自动报税等。实现了“经济业务—单据—账务—报表—财务分析”全数据处理环境, 智能机器记账, 首先将单据影像化、电子化, 实现电子化自动采集, 省去录入环节, 单据自动汇总计算; 通过语义理解深度学习, 实现财务的自动归类、自动记账。同时快速生成可视化报表, 根据财务数据进行财务分析的大数据、全样本分析, 同时实现全样本的审计自动化。

大数据、人工智能的智能财税已使多项财税业务实现自动化。如: 利用云平台实现自动银行对账。利用业务引擎和财务引擎实现了流程自动化, 包括: 应付自动化、应收自动化、核算自动化、结算自动化、报审智能化。利用审核引擎, 实现了稽核智能化, 包括: 费用稽核、小票稽核、合同稽核。外部信息数据化是自动化的前提, 需要将软件/系统与企业信息系统对接实现数据化。同时将票据识别、用户体系、安全体系、云联接口群、场景识别等有机结合, 不断优化。

智能财税实现突破条件为, 需要利用大数据平台集合财税相关基础数据, 采取影像扫描、数据爬虫技术等进行数据输入, 利用票据识别、语义识别等技术, 同时对接企业 ERP、OA、财务软件、MIS、CRM 等系统, 识别业务事项, 利用人工智能深度学习并不断反馈优化, 根据智能应用算法模型, 应用于各类场景。智能财税的基础来源于: (1) 全数字化作业环境, 人工智能运作的前提; (2) 解决非结构化数据结构化问题; (3) 众包、闭环、机器学习, 实现自动记账。企业财税业务数据来源于大量原

始凭证与外部数据,传统财税工作中,外部数据的大部分需手工采集,大量原始凭证需手工采集,数据化传统模式依赖于大量人力手工录入。实现全数据作业环境,需要运用票据智能识别,将应用软件与信息系统中的财务软件、ERP、CIA、MIS 等系统对接,通过软件/系统对接可以实现数据化。利用票据识别技术,进行模版识别、文字识别、利用大数据进行智能修正、人工补录等循环实现票据字段和数据的提取。以智能识别替代人工录入,利用规则引擎和深度学习实现自动记账。实现机器自动会计核算,门店销售数据稽核,费用审核,付款审批,应收账款核销,应付账款核销对账,自动记账,销售数据自动抽样稽核,AR、AP 自动核销,银行、供应商自动对账等,智能财税将会以人工智能替代低端、重复的规则性工作,优化简单、重复,具备明确稳定的量化规则的传统财税工作。

三、大数据、人工智能

大数据是一种规模大到在获取、存储、管理、分析方面大大超出了传统数据库软件工具能力范围的数据集合^[2],具有海量的数据规模、快速的数据流转、多样的数据类型和价值密度低四大特征。大数据技术包含数据存储技术、数据采集技术、数据计算技术、数据挖掘技术、数据可视化等技术。大数据平台基于 Hadoop/Spark 计算平台等多种技术^[3],使用开源的系统,全球生态,相关参数可根据需要调整,灵活多变^[4-5]。大数据、云计算等显著提升了计算能力,奠定海量数据及业务活动的场景化和数据化。基于大数据平台的数据整合及数据挖掘工作,可以实现低成本数据处理^[6],具有通用型技术架构、强大数据计算能力、线性扩展能力。大数据技术与数据仓库融合,建设成为企业统一数据平台。非结构化数据、结构化数据、流数据的分析能够通过大数据技术实现^[7],使得“分析一切可分析的数据”成为可能。

大数据技术组件体系包括: Hue(图形化数据查询、监控管理工具平台), Nutch(数据搜索引擎应用), Spark GraphX(Spark 图计算), RHadoop(R 和 Hadoop 接口工具), Spark Streaming(流处理工具), Zookeeper(分布式协调服务系统), Oozie(作

业流调度系统), HBase(分布式实时列式存储数据), Spark MLlib(Spark 机器学习算法库), Mahout(Hadoop 机器学习算法库), Storm(流式实时计算框架), Hive(SQL 数据仓库系统), Pig(流式处理数据仓库系统), Impala(大数据分析查询系统), Sqoop(数据库词 ETL 工具), MapReduce(分布式离线计算框架), Spark(分布式内存计算[实时]框架), Tex(DAG 计算), Kafka(分布式发布订阅消息系统), Yarn(集群资源管理和调度系统), HDFS(分布式文件存储系统), Flume(分布式日志采集工具), Ambari(安装、部署、配置和管理工具)……

以大数据为基础的人工智能已应用在经济预测、商业零售、交通、治安等多个领域^[8]。例如,利用 Twitter 情绪词^[9]预测道琼斯指数;零售商对销售额、定价以及经济学、人口统计学和天气数据进行分析,选择合适的上架产品,判定减价的时机。货运公司对交货时间和交通模式等数据进行分析,以此对其运输路线进行微调。人工智能的出现,首先得益于三大因素的助推:大数据积累、高性能计算和深度学习算法。一是大数据积累,随着全球设备、机器和系统的连接,大量的无结构数据被创造出来。随着数据量的级数增长,历史数据量变到质变,更多的规律在小数据规模无法发现的统计规律被揭示,大量的历史数据帮助人工智能能够解决的问题也增加^[10]。二是高性能计算, GPU 要比 CPU 能更快的训练机器学习系统; FPGA/芯片发展:用 FPGA 能够更快训练出的深度学习系统; Hadoop/Spark 计算^[11]平台:例如图计算平台单台机器相当于 1000 台机器运算能力, Spark、Graphlab、Parameter Server 以及 GPU 异构计算(Caffe、torch、tensorflow)等面向大规模机器学习、人工智能的建模平台^[12-13]。三是深度学习算法,深度学习的是人工智能近年发展的驱动力,是机器学习的一个新的领域。深度学习、语音识别、图像识别技术等研究快速突破,推动人工智能发展。

四、财务领域大数据平台概述

财务领域的大数据平台的目标是构建集操作、管理、决策于一体的高效统一的智能合同管控

数据化平台,实现纵向管控、横向协同、智能决策、效率提升、规范流程、降低风险,平台支撑、灵活扩展。财务领域大数据平台能够帮助企业的不同层级实现管控目标,在帮助企业进行财务管理的同时实现管理效能的提升。财务领域大数据平台为财税智能机器人提供技术支持和平台支撑。

表 1 财务领域大数据平台下的各层级管控目标

企业层级	管控目标	管控职能
战略决策层	智能化	智能化决策支持
运营管控层	管控一体化	强化管理部门对业务的协调、管理、监督和控制能力;合同履行端到端可视平台加强合同履行过程控制及预警,降低风险和强化合同履行绩效考核
计划协同层	规范化、精细化	规范业务操作过程,明确流程各点的关键作业指标,提高作业效率;实现精细化操作管理,实现合同履行各环节上的成员能进行实时的信息共享
作业执行层	业务协同、信息整合、透明化、可视化	合同履行端到端可视平台与财经和供应链、物流信息整合;执行过程透明化、大数据可视化、全程跟踪
基础技术层	供应链技术创新、系统架构创新	大数据技术的全面集成: hadoop、hive、Spark、云计算、SaaS……;灵活、易扩展的一体化开发平台,支撑集团多元化经营和兼并扩张

大数据平台帮助管理人员了解企业经营的各项可得的内外部信息,从渠道到社交媒体,通过分析所有数据来源帮助企业经营决策;分析所有可用的操作数据并且实时做出反应,进行零延迟操作,从而优化流程;通过成本效益技术降低 IT 成本;在速度和规模上实现流程改进、新产品创新,捕获所有来源的反馈,分析庞大的市场环境,研究大量的数据,从而推动创新;通过分析所有可用数据来打造更好的欺诈/风险模型,即时进行风险控制,用流数据交易分析来实时监测欺诈;利用仪表化资产,监控资产并通过实时数据反馈来预测和预防维修问题,并且开发新产品和新服务。

五、财税智能机器人

财税服务场景中,有大量的重复工作,比如票据稽核、报销场景的审计等。财税智能机器人建立在大数据平台基础上,基于行业专家的最佳实践,积累样板,采用机器学习的方法提炼语义分析模型核心价值,借助大数据平台分布式的高效计算,通过“关键词提取、语义距离分析、主题提取/

预测—机器学习分类算法—识别算法模型—财税、金融、场景化应用”提供实时高效的服务。财税智能机器人,结合财税日常业务流水数据(财务明细、票据明细、合同明细等),补充外部信息(电商网站、国标网站上公布的行业标准等)形成巨大的语料/样本库、利用大数据高效计算提炼专家模型,实现自动票据识别、语义理解、自动记账、一键报税等服务,推动业务低成本高速发展。财税智能机器人有多种类型,目前已经出现了费用稽核机器人、应收应付机器人、行政事业单位核算机器人、合同比对机器人。纳税服务机器人处于研发阶段,反舞弊机器人仍处于下一阶段研发产品。

费用稽核机器人外挂于财务系统,不改变工作流程,配置简单,操作简单,智能易用,能够解决企业费用报销遇到的种种问题。费用稽核机器人用以解决企业在费用报销存在的诸多难题:金税三期之后,发票检查点太多,涉税风险高;纯人工审核,效率不高,审核质量不稳定;费用稽核工作简单重复枯燥,员工成就感、价值感低;费用稽核单据占会计凭证总量 70% 以上,审核量大;审核点/规则太多,难以全面把控;沟通繁琐,占费用会计大量时间等等。费用稽核机器人通过“报销单据—影像智能识别形成数据—数据集合与匹配—审核引擎—审核结果—OA/MIS—财务引擎—会计凭证—ERP、财务软件—抽取基础数据(会计科目/物料/人力/部门)”的业务流程,根据审核规则智能审核,使审核过程可视化,智能提示,避免繁琐沟通,减少约 50% - 80% 工作量,智能高效,能够显著提升企业费用管控。费用稽核机器人智能可靠,降低错误率,机器严格执行各项规则,审核步骤均被记录和监控,能够解决肉眼无法解决的难题,包括:电子发票重复报销难题;专用发票、大额发票批量验证真伪问题;普通发票纳税人识别号真伪校验问题。费用稽核机器人审核无死角,避免涉税风险。通过费用稽核机器人,生成费用大数据,能够指导业务发展,解放费用会计,有助于将人力投入转移到高附加值的工作,实现财务团队由核算会计向管理会计转型。

应收应付机器人属于非侵入式应用,嵌入在企业系统中具有高度可扩展性,遵从合规要求,质量可靠,能够实现全天候的高效低成本运作,有效释放生产力。应收应付机器人能够实现:应收应付自动化、订单处理自动化、协同对账智能化、发票管理自动化、审单核销自动化、税务处理自动化、财务核算自动化。与传统工作模式对比,优势具体表现为:影像、电子发票、少量纸质单据甚至全电子化的发票接收,自动索引知识管理,智能数据采集、自动三单匹配,机器智能审核,异常人工处理,自动服务门户在线争端解决,能够实时自动编制及大数据分析,实现业务财税一体化流程。

行政事业单位核算机器人,根据行政事业单位财务收付实现制的要求,通过“单据扫描识别—业务场景识别—会计核算方式匹配—通过接口送入系统—自动生成两大账套”的流程方法,将感知识别(识别票据种类)、认知学习(识别业务类型)、决策分析(可疑报销,是否继续生成凭证)有机结合,生成收付实现制账套。同时,行政事业单位核算机器人从识别上传影像数据开始,将单据扫描识别所获得的归类影像数据,匹配合适场景,通过自动匹配核算规则,应用模板取值规则,将经过会计核算方式匹配生成的数据转入至原有系统(如OA/网报/MIS),生成权责发生制账套。行政事业单位核算机器人解决了会计部门轮岗中遇到的知识技能衔接问题,实现行政事业单位会计核算工作的标准化,提升内控水平,实现高效集中核算,无需传递原始凭证,保证核算时效,高效应对政府会计准则的实施,解放核算人员,提升会计部门价值。

合同比对机器人能够有效降低时间成本,节约人力成本。一份50页合同的审核对比需要人工约1小时,而合同比对机器人2分钟即可完成;人工比对错误率、遗漏率高,而合同比对机器人的差异识别率100%;合同比对需要的人员人力成本高,但是创造价值低,而合同比对机器人能够替代低端劳动,使得原有人员可从事更高附加值的事务。合同比对机器人广泛适用于:信托、银行、证券、电讯、基金等合同量大的企业。

纳税服务机器人的设计目标是实现智能税务稽核与智能纳税预警。纳税服务机器人设计的核心功能包括:影像取证、企业经营大数据、重构账务、智能稽查。纳税服务机器人能够进行纳税评估、智能预警,并进行便捷取证,减少纠纷,大幅提高稽查效率、效果。

六、大数据、人工智能在财税领域的关键技术问题

1. 票据识别技术

票据识别技术,最初源于大数据图像识别技术ImageNet,其特点是利用海量的图片,以标签化技术标注特征,然后利用个体的差异,收集全视角,需要海量标本标签化,训练识别模型。

票据识别技术首先是进行模版识别,为实现版面自动识别,需要获取模版,以及大量的票据,从种类到数量,尽量地覆盖到具体财税业务中所需的各种票据,目前已达万余种票据版面,含发票、银行账单、企业内部单据等,覆盖财税、保险领域的票据数量。采用深度学习及海量图像匹配技术,以人工录入,标记票面,建立深度学习模型,对万余种票据版面进行学习 and 版面识别。

票据版面识别完成后,需要进行字段碎片切割,根据业务需求将需要识别的字段切割,切割后的字段便于精确识别,并能去除周边信息的干扰。

然后是文字识别,文字识别非常关键,利用基于深度学习技术的识别引擎,通过深度学习的方式,对GB2312库中的6763个汉字进行训练,为每个汉字自动生成各种字体、各种干扰下样本,并结合真实生产环节下汉字需要找到1万左右字体样本进行训练。利用双识别引擎交叉验证,使用2种算法引擎识别,自动获取识别错误字段。

文字识别完成后还需要智能修正,根据业务逻辑进行验证,利用深度学习及数据爬虫技术进行业务逻辑校验,大数据技术自动积累票据及客户的相关信息,利用数据爬虫获取外部相关数据,采用深度学习的方式获取票据的业务逻辑并进行验证,利用大数据进行智能纠偏,查找错误并自动修正,准确率可达90% - 97%。

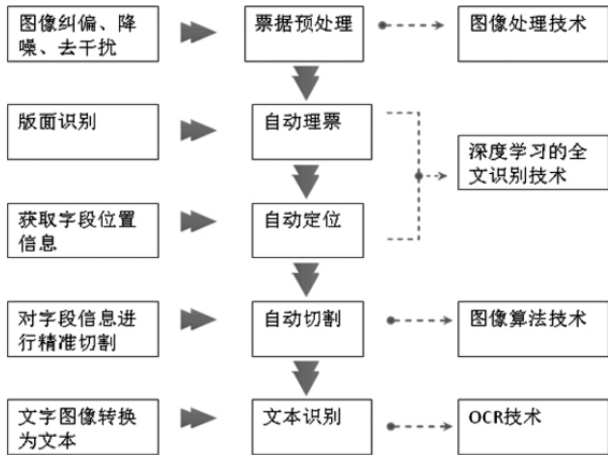


图1 票据识别流程及技术

深度学习的全文识别技术,采用人工智能的深度学习算法及全文识别算法结合预设模版,区分不同版面类别,为快速精准识别票据信息奠定基础模版。识别率可达98%,仅2%的需要人工干预理票;机器自动定位获取字段位置,自动定位准确率可达95%。图像算法技术运用人工智能的深度学习算法自动进行精确切割,切割准确率高达100%。OCR技术通过OCR技术和定位信息将图片信息转化为文本信息,OCR识别率水平可达70%~80%。

为防止错漏发生,还需要人工补录,对错误字段进行修正,并抽检。错误及无法识别的字段用

人工进行补录,对于关键字段按重要性比例进行抽检,保证发票金额等重要字段100%准确。在人工补录、抽检完成后方可以按照要求的格式交付到应用端口。

2. 语义识别技术

财税服务场景中,有大量的重复工作,比如生成凭证、记入账户、票据稽核、报销场景的审计等,在具体业务场景的辨析中,一项耗费是记入固定资产还是无形资产,是记入存货还是费用?合同规定的关键信息,如何有效地反映?需要运用什么样的会计政策?等等都需要应用语义识别进行辨析。利用大数据技术在财税领域运用语义识别,是给机器人装上语义理解的引擎,利用语义识别进行关键词提取,主题提取/预测,语义距离分析,可以极大的提高业务效率。采用机器学习的方法提炼语义分析模型核心价值,需要基于行业专家的最佳实践,积累样板,借助电商网站、国标网站上公布的行业标准,作为体系化学习的语料/样本;将种类繁多的信息通过分词技术打散,为语义分析、情感分析提供样本支持;基于分布式大数据平台的高效计算能力,提供实时的服务;结合财税日常业务流水、补充外部信息形成巨大的语料/样本、利用大数据高效计算提炼专家模型,反过来推动业务低成本高速发展。

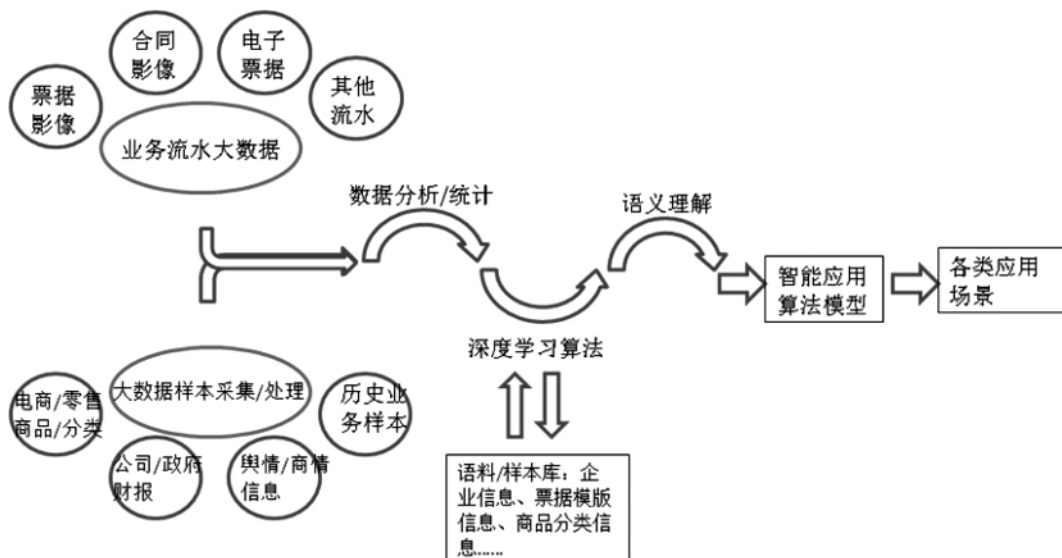


图2 财税机器人语义理解引擎

财税机器人语义理解引擎的数据是基于对业务流水数据和公司内外信息,业务流水大数据是针对企业财务管理过程中的票据、合同、电子化凭证以及其他纸质凭证、记录等传统模型的信息进行影像化,形成业务流水大数据。同时,将财务业务场景过程中的数据信息接口对接形成大数据流。企业内外信息是通过爬虫采集技术,收集大量的电商/商品明细、零售商品明细、行业标准规范、互联网舆情/商情信息以及历史业务样本数据。并且,利用强大的数据处理技术,对数据进行清洗、整合,形成高质量的样本数据。结合专家的经验,对样本进行批量分类。

通过企业对业务流水数据以及收集和采集的企业内外部信息进行数据分析和统计,进而利用深度学习算法达到语义理解,数据经过语义理解后,代入智能应用算法模型,可以应用到基于智能算法模型各类场景,如业务场景辨析、合同关键信息提取、年报数据与政策差异分析等。借助于大数据平台和深度学习,语义识别可以极大的提高业务效率,是财税智能化得到进一步提升。

3. 自动记账

财税智能机器人实现自动记账与自动驾驶原理相似,自动驾驶首先需要拍摄记录路况,将路况电子化,使得机器理解路况,将驾驶行为数据化,通过收集分析数百万车辆的实况行驶情况,最终使机器学会了驾驶。自动记账的数据基础是原始票据的各项数据,票据的智能识别奠定了数据基础;票据识别技术将原始票据的数据准确提取后

还需要进行语义分析,通过模型算法和深度学习使得机器能够辨析各类业务场景,判断不同业务场景中,具体发生项目的记账科目;随后将提取的具体票据的数据记入不同的会计科目中实现会计核算行为的数据化;大量会计的持续操作利用大数据云平台积累海量的数据,运用机器自主学习,实现自动记账,同时将结果反馈到语义分析的业务场景辨析的深度学习过程中,持续改进,不断优化。

表2 语义识别模型及应用方向

语义识别模型类别	语义识别依据基础	语义识别应用方向	语义识别应用举例
业务分类语义识别模型	商品明细分类库、企业行业知识模型、商品明细与行业语义距离分析	业务场景的辨析	判断不同业务场景中,具体发生项目的记账科目:固定资产 or 无形资产、存货 or 费用
合同语义识别模型	商业活动场景关键词库、关键词语义距离分析库	合同关键信息提取	自动准确提取合同文本中的标的、交付、开票、收款、验收、违约责任、有效期等关键信息
会计政策语义模型	会计政策识别词库、不同政策的会计处理方式	年报数据与政策差异分析	准确理解公司年报信息,并自动分析会计政策差异对报表数据的影响范围

实现自动记账需要多种数据来源,目前智能财税行业已经开发出互联网云平台加免费使用,众包模式,可以采用免费的票据识别,获取大量企业的记账数据。大规模机器学习、人工智能的建模平台已经应用 Spark、Graphlab、Parameter Server 以及 GPU 异构计算(Caffe、torch、tensorflow) 等的多种技术,包含用于图像的卷积神经网络,用于视频/文本的循环神经网络,基于大数据关联分析的

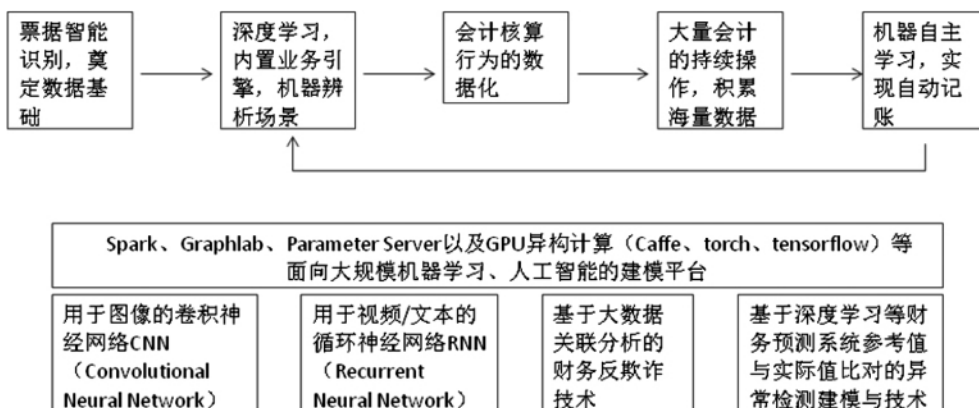


图3 自动记账与技术应用示意图

财务反欺诈技术,基于深度学习等财务预测系统参考值与实际值比对的异常检测建模与技术等。建模云平台数字化所有的信息并进行集成,同时将业务票据凭证和企业 ERP、OA、财务软件、MIS、CRM 等系统对接,识别出常规会计事项,预配置或机器学习。其中,会计事项识别包括差旅费报销,医药费报销,采购固定资产,采购原材料,在建工程,租赁办公场所,确认产成品成本等在内的各种会计事项,根据会计事项,结合信息配置科目,预配置或机器学习。识别会计事项后,需要基于深度学习,用数据标记对应场景,生成业务场景数据,根据单据和其他信息形成会计凭证(分录)的数据来源,与业务场景对应的记账对应的规则进行会计凭证配置与记账。在整个过程中不断反馈,深度学习不断优化,最终实现单据数据自动采集、机器学习自动记账。并为智能机器对账分录、自动报税等奠定基础。

4. 一键报税

传统报税模式操作复杂,重复劳动,工作量巨大。财税机器人的设计目标是希望能够改变报税繁琐的工作,极大的提升效率,节约人力成本的投入。目前,云模式下,已经出现了一键报税机器人,重复工作被机器人替代,能够提高效率,减少出错几率。在云模式下的新模式一键报税,同时连接了国税系统与地税系统,能够实现“登录—选择报税项目—报表校验—提交报税报表—完成报税”的过程。一键报税机器人,将报税的整个流程赋予机器人执行,机器人自己判断,智能筹划,实现报税过程。

应用报税机器人实现一键报税,并非一蹴而就。首先要熟悉税收法律法规和税收实践规则。每个省份和地区的税收规则在一些操作细节上会各有不同,并且税收中涉及大量的纳税调整事项和不同的税收确认基础,税收法规也会随时间而有不同的规定。因为以上情况的存在,在智能机器人报税实务中,很容易出现错报的情况。所以,报税计算模型不仅要参考企业以往的报税规则,更要考虑法规的变化和影响计税的新出现纳税事项等因素。在云模式下,需要不断调整税收计算

模型,及时处理变更事项,更新平台系统。

七、智能财税服务的宏观影响

云模式下的财税智能机器人不知疲倦,能够长时间工作,稳定性强,差错率基本为零,能够忠实执行业务,避免了传统师徒制下的不确定性和不稳定性,能够实现业务一次定义,永久生效(量化、有规则、重复性的工作)。并且,该智能财税服务对财税业务通过配置即可调整,且是全链条调整,避免了传递过程的人为失误。智能财税服务,使用方法简单明了,易于业务人员掌握、快速上手,这将极大的提升效率,减少错漏。智能财税服务在提供各种便利和增效的同时,也意味着传统财税工作模式即将被替代,未来将有大量的财税工作人员可能会失去原有的工作岗位,为政府治理与社会治理带来新的挑战。

财税智能机器人、大数据云平台已经可以完成收票/理票、记账、做报表、审核报销、资金记录等规则明确的重复性工作。短期之内,初级职称、无职称人员将受到冲击,简单、重复具有明确稳定的量化规则工作将逐步被人工智能取代。该类人员数量占比大,将会有大量的相关人员失业,同时在智能财税服务基础上,社会将产生的新的职务需求,面临结构化调整的问题。由于大量低端财会失业人员会对社会稳定的潜在不利影响,政府治理也将面临扶助与引导这些失业人员的再培训、再就业的问题。这就需要政府、社会、企业多方统筹协调,帮助财务人员转型从事分析优化流程、寻找关联关系、制定标准、从事管理会计、支持管理决策等高附加值的职务。

智能财税的发展极大地提升了效率,降低了成本,同时也存在着数据资源^[14]、前沿技术、资金等多方面的制约,这就需要根据智能财税的发展特点,制定具有一定前瞻性的政策,完善相关法律制度,加强各地区、部门、层级、系统的协调,推动业务、数据、技术的融合,打通数据、信息孤岛障碍,推进数据平台对接。同时引导产业、政府和市场多渠道资金投入,鼓励本土智能财税服务商的发展,加强前沿技术的研发,有力地推动财税服务的创新。

结束语

传统的财税工作模式呈现出重复、繁杂、低效等特点,财税工作重要但成本高,耗费大量的时间与人力。基于大数据、人工智能的智能财税服务,实现了票据自动识别、语义解析、自动记账、机器审核、自动对账甚至一键报税等财税新模式。其根据具体规则,一次定义,持续、高效地运作,操作简单,极大地提升财税工作的效率,降低人为失误。但是智能财税服务并不能在所有方面取代人的工作,财税工作的高端领域依然是人工智能所不能替代的。智能财税的最终目标是服务于人,将人类从低端、繁琐、机械的财税工作中解放出来,转而从事更有价值的工作和领域。

参考文献:

- [1] CHOUARD T, VENEMA L. Machine intelligence [J]. Nature, 2015, 521(7553): 435-435.
- [2] MCAFEE A, BRYNJOLFSSON E. Big data: The management revolution [J]. Harvard Business Review, 2012, 90(10): 60-68.
- [3] VORA M N. Hadoop-HBase for large-scale data [C]. International conference on computer science and network technology, 2011(1): 601-605.
- [4] MAVRIDIS I, KARATZA H D. Performance evaluation of cloud-based log file analysis with Apache Hadoop and Apache Spark [J]. Journal of Systems and Software, 2017: 133-151.
- [5] DING M, ZHENG L, LU Y, et al. More convenient more overhead: The performance evaluation of Hadoop streaming [C]. Research in applied computation symposium, 2011: 307-313.
- [6] FAN W, BIFET A. Mining big data: current status, and forecast to the future [J]. Sigkdd Explorations, 2013, 14(2): 1-5.
- [7] ZIKOPOULOS P, EATON C. Understanding big data: Analytics for enterprise class hadoop and streaming data [M]. McGraw-Hill Osborne Media, 2011.
- [8] BUTLER D. Tomorrow's world: Technological change is accelerating today at an unprecedented speed and could create a world we can barely begin to imagine [J]. Nature, 2016, 530(7591): 398-402.
- [9] ABAYOMI A, OLUGBARA O O, ADETIBA E, et al. Training pattern classifiers with physiological cepstral features to recognise human emotion [C]. Nature and biologically inspired computing, 2016: 271-280.
- [10] GHARAMANI Z. Probabilistic machine learning and artificial intelligence [J]. Nature, 2015, 521(7553): 452-459.
- [11] ZAHARIA M, CHOWDHURY M, DAS T, et al. Fast and interactive analytics over Hadoop data with Spark [J]. USENIX Login, 2012, 37(4): 45-51.
- [12] GU L, LI H. Memory or time: Performance evaluation for iterative operation on hadoop and spark [C]. High performance computing and communications, 2013: 721-727.
- [13] LIN X, WANG P, WU B. Log analysis in cloud computing environment with Hadoop and Spark [C]. In Proceedings of the IEEE 5th International Conference on Broadband Network & Multimedia Technology (IC-BNMT2013), 2013: 273-276.
- [14] HELBING D, POURNARAS E. Build digital democracy: Open sharing of data that are collected with smart devices would empower citizens and create jobs [J]. Nature, 2015, 527(7576): 33-35.

(本文责编:王廷芳)