

# 大数据技术在经济运行监测分析中的应用

● 戎 天

**摘 要:**从省级政府部门角度,对大数据技术在经济运行监测分析工作中的应用场景进行了探讨,分析了当前工作应用中存在的缺乏专业人才支撑、涉企政务数据共享有待进一步加强、成熟适用的分析模型及应用场景比较少等问题,提出强化专业人才培养、加快推进政务数据信息共享、深挖经济运行监测分析业务与大数据技术契合点等建议。

**关键词:**大数据 数据挖掘 经济运行 Python

## 一、引言

全球范围内,研究发展大数据技术并运用大数据推动经济发展、完善社会治理、提升政府服务和监管能力正成为趋势。习近平总书记指出“要运用大数据提升国家治理现代化水平。要建立健全大数据辅助科学决策和社会治理的机制,推进政府管理和社会治理模式创新,实现政府决策科学化、社会治理精准化、公共服务高效化”;借助大数据推动政府职能转变,利用大数据提升政府治理能力,构建服务型政府,是本届政府始终关心并重视的问题。就地方政府及工作部门而言,提高大数据应用水平对更好地贯彻落实国家宏观调控政策、提升经济调控能力和决策水平至关重要。在此背景下,本文拟从省级部门工作实际出发,围绕经济运行监测分析工作中数据采集、数据分析、数据呈现等业务需求,分析探讨网络爬虫、文本挖掘、分类预测、可视化分析等大数据技术在实际工作的具体应用,为更好地开展经济运行监测分析提供新的视角。

## 二、经济运行监测分析及大数据简介

### (一) 经济运行监测分析简介

经济运行监测分析是对经济运行动态指标进行连续性的观测及其规律性的揭示,需抓住关系经济运行的全局性问题进行深入分析,提出政策建议。以工业为核心的实体经济是国民经济的基础,工业与其他产业之间的关联紧密,是经济运行监测分析的主要着眼对象。因此,经济运行监测分析主要是从工业生产、效益、投资、外贸、价格等方面,以及与工业上下关联、影响并制约发展的其他因素(如与工业生产高度关联的能源、交通、信贷等主要生产要素情况等),来描述工业经济运行的状态,并结合横向部门、各地工信部门、行业协会、重点企业等多方面信息,综合研判分析经济运行态势。经济运行监测分析不同于一般的统计分析,统计分析把数据的准确及时作为主要目标,而经济运行监测分析除了对数据的严格要求外,更加注重于对动态、活情况的把握,重点反映经济运行中出现的新情况、新问题,并对苗头性、倾向性问题进行深入分析,准确把握经济运行的走势,并把对经济运行工作进行指导作为主要目标。

### (二) 大数据及相关分析工具简介

大数据是商业智能化的一个新的发展阶段,目前人们对于大数据还缺乏一个权威的说法,不同的机构对其认知各不相同。当前,人们在讨论大数据时,更多的是通过其若干基本特征去认识。IBM 公司把大数据的特征概括为三个“V”,也就是规模(Volume)、快速(Velocity)、和多样(Variety),但是更多的人则将其概括为四个“V”,即规模(Volume)、快速(Velocity)、多样(Variety)、价值(Value)。大数据技术,是指从各种类型的数据中快速获得有价值信息的一种技术,其水平反映了提取有用信息的能力,主要包括数据采集、数据清理、数据存储及管理、数据挖掘、数据可视化等。其中与经济运行监测分析工作密切相关的大数据技术主要是数据采集、数据挖掘及数据可视化。近年来,以 Python 为代表的大数据采集、数据分析工具日趋完善,在 Python 的集成开发环境(IDE)中有大量优秀的开发工具包,其胶水语言的特性能够将这些工具包无缝衔接,为数据分析人员提供网络爬虫、自然语言处理、机器学习、可视化分析等常用工具。

指从各种类型的数据中快速获得有价值信息的一种技术,其水平反映了提取有用信息的能力,主要包括数据采集、数据清理、数据存储及管理、数据挖掘、数据可视化等。其中与经济运行监测分析工作密切相关的大数据技术主要是数据采集、数据挖掘及数据可视化。近年来,以 Python 为代表的大数据采集、数据分析工具日趋完善,在 Python 的集成开发环境(IDE)中有大量优秀的开发工具包,其胶水语言的特性能够将这些工具包无缝衔接,为数据分析人员提供网络爬虫、自然语言处理、机器学习、可视化分析等常用工具。

## 三、大数据技术在经济运行监测分析工作中的应用场景探索

### (一) 运用网络爬虫技术扩展经济运行数据采集范围

网络爬虫技术又被称为网络机器人、网路蜘蛛,是一种按照规则,自动抓取信息的程序或者脚本。如果把互联网比喻成蜘蛛网,那么网络爬虫就是在网上爬来爬去的蜘蛛,根据既定目标选择性地访问互联网上的网页与相关链接,循环读取网页内容,获取相关信息并存储。经济运行数据的获取和开发,是经济运行监测分析工作的关键。在大数据环境下,经济运行监测分析工作不仅需要了解掌握本部门及相关部门的内部统计数据,也需要并应当从互联网公开数据中获取与经济运行息息相关的更多数据,如:大宗产品价格信息、土地市场供应信息、上市公司企业财务报表和经营状况数据等。但这些信息大多以网页等格式存在,难以直接获取下载,可在遵守相关法律法规的前提下,利用网络爬虫工具批量抓取网页上的有用信息,形成可用于分析的数据集,拓展经济运行信息采集的广度和深度。比如:为了解掌握近年来江苏全省工业企业土地要素保供情况,可设计网络爬虫工具抓取江苏土地市场网中 13 个设区市及下辖县(市、区)工业用地成交信息(除了行政区、地块编号、宗地位置、土地用途、成交时间、面积、金额等字段外,还可进一步抓取地块编号链接详情页中的竞得单位、出让年限等信息),用于监测分析各地工业用地供应量、供应地价变化等。

### (二) 应用文本挖掘技术分析企业反馈信息主题及情绪变化

文本挖掘(Text Mining, TM)是指从大规模文本数据库中提取隐含的、以前未知的、潜在有用的信息和知识的过程,已广泛应用于网络舆情分析、商品评价分析、微博热点分析、学科前沿热点分析和有关政策文本分析等领域。在相关政府部门履职尽责过程中,以企业为主要用户或服务对象的相关政府部门业务系统内,积累了大量反映企业生产经营情况以及体现企业有关政策诉求和建议的文本数据。相较于数值型数据,文本

数据作为企业反应情况的主要形式,蕴含着更为丰富的信息,但受数据量大、记录粗糙以及相应信息处理能力不足等多种因素限制,对文本数据系统挖掘相对缺乏,制约了对企业反馈信息的深入研究利用。就经济运行监测分析工作而言,可将基于LDA主题模型的文本主题抽取、基于朴素贝叶斯算法的文本情感分析等文本挖掘技术引入到对苗头性、倾向性、潜在性问题以及企业诉求建议的研判分析中来,定期对政府部门企业监测直报系统、问卷调查系统以及工作调研中企业反馈的文本信息进行挖掘分析,精准反映不同类型、不同行业的企业困难及诉求主题变化、情绪倾向,以获取企业真实准确的看法,为研究提出有关政策建议提供依据。

### (三) 应用机器学习算法识别预警企业生产经营风险

机器学习,也就是计算机从大量数据中学习特定规律的过程,这些规律包括分类、关联分析、回归、聚类等,涉及到决策树、随机森林、逻辑回归、支持向量机、朴素贝叶斯、K近邻算法、K均值算法、AdaBoost、神经网络、马尔科夫等十几种算法。在运用大数据提升政府治理能力的时代背景下,以机器学习算法为代表的数据挖掘技术已经被广泛应用于各个领域,综合利用机器学习算法对经济运行数据进行深入挖掘分析,从中发现归纳有规律、有共性的问题和趋向,提炼出有价值的信息、建议,可以更好地服务于政府决策。比如:在企业生产经营风险监测预警工作中,可综合应用LM神经网络算法、CART决策树算法,对来自市场监管、税务、人社、供电等部门的企业登记、纳税申报销售、缴纳税金、职工人数、社保缴费、用电量等企业生产经营指标,以及纳税状态是否正常、社保费缴费是否正常、电费缴纳是否正常等信息,建立对高生产经营风险企业和正常生产经营企业具有较强分类能力的风险识别模型,及时识别生产经营风险较大的企业,为预警和工作处置提供支持。

### (四) 应用可视化技术呈现数据分析结果

数据可视化,是指将相对晦涩的数据通过可视的、交互的方式进行展示,从而形象、直观地表达数据蕴含的信息和规律。数据可视化是大数据应用的“最后一公里”。一方面,可视化技术可以帮助数据分析人员一眼洞悉数据背后隐藏的信息,不再受制于枯燥晦涩的分析算法;另一方面,可视化技术能够非常有效地传达信息中各种经济指标之间的复杂模式、趋势和相关性,更好地展现分析发现和结论,从而增进阅读者的理解。除了使用常用的柱状图、折线图、饼图、散点图、雷达图等图表呈现经济运行状况之外,还可以运用词云图呈现大量企业反馈信息文本的分析结论;运用动态图更加清晰地反映某种特征的变化趋势;运用GIS地图呈现产(行)业经济数据地理空间分布规律等,往往可以起到“一图胜千言”的效果。比如:应用动态气泡图展现近几年来江苏13个设区市工业增加值、工业用电量及高耗能行业占比等关联指标的变化情况,X轴和Y轴分别描绘工业增加值与工业用电量,不同颜色标识不同的设区市,气泡大小则反映高耗能行业占比情况,近几年的时间变化由气泡位置的不断变化体现,最终形成具有动画视觉效果的气泡图,表达出需要呈现的三维信息。

### (五) 数据报表自动化处理

在日常经济运行监测分析工作中需要定期分析处理来自相关部门的分行业、分地区工业生产、效益、投资、价格报表,以及几百家重点监测企业生产经营指标数据,这些报表数据量大、指标多、需要计算分析的维度也多,如果依靠手工计算将

占用分析人员大量时间精力。而利用大数据分析工具python开发相应的数据分析脚本,即使数据总量巨大、指标众多,也可以轻松实现对报表数据的自动化处理。比如:工业企业主要经济指标表中共有包括企业数、营业收入、营业成本、利润总额在内的17项指标,可以按照常用的规模大小、行业、所有制类型等分析维度,一键计算生成重点分析指标的同比、环比、占比、贡献拉动等数据,并同步绘制形成反映规模、走势、结构的柱状图、折线图、饼图等相关分析图表,大大提高基础数据分析处理工作效率。

## 四、大数据技术在经济运行监测分析应用中存在的困难问题

### (一) 缺乏专业人才支撑

一直以来政府相关职能部门中承担经济运行监测分析的工作人员主要以经济学、统计学、以及中文等学科背景的人才为主,而数据挖掘需要综合运用计算机、数学以及统计学等相关知识,目前既懂数据分析和挖掘技术又具备经济运行业务知识的复合型人才较为匮乏,难以适应新形势下经济运行大数据分析和挖掘的要求。

### (二) 涉企政务数据共享有待进一步加强

虽然近年来江苏出台了一系列政策文件,并以“不见面”审批服务为牵引,大力推进政务数据资源的整合共享。但受制于多种因素的制约,政务数据资源的整合与共享仍面临诸多困难。据不完全统计,江苏省级政务部门无条件共享数据量占部门掌握总数据量的比例不足10%,信息系统数据的共享比例不足50%,政务数据尤其是涉企政务数据共享不足,一定程度上制约了大数据技术在经济运行监测分析工作上的深化应用。

### (三) 成熟适用的分析模型及应用场景比较少

就现阶段经济运行监测分析工作而言,相关应用场景和分析模型较少,数据挖掘的深度、综合利用的程度还有待提高。如:在工业生产、效益等主要经济指标预测方面,如何选用合适的特征指标、应用哪几种机器学习算法,怎么优化算法模型达到最优预测效果,都需要在工作实践中进行深入的研究探索。

## 五、加快利用大数据技术提升经济运行分析水平的对策建议

### (一) 强化专业人才培养

一是加强对现有人员的技术培训。积极引导经济运行分析工作人员树立大数据分析理念,定期组织业务骨干参加前沿的、专业的技术培训,提升数据挖掘分析能力,打造经济运行大数据分析团队。二是加大对复合型人才招录力度。在新进人员招考和引进方面,设置相关专业需求,加大对具有统计分析、计算机技术(软件工程)、经济管理等多学科知识的复合型人才招录,引进新生力量,提高分析团队大数据素养。三是引入外部专业力量。数据分析人才的培养需要一定的条件和较长的周期,因此相关政府部门也可以通过购买服务、与高校院所开展课题合作等方式,借助专业机构技术力量开展经济运行大数据分析。

### (二) 加快推进政务数据信息共享

进一步深化改革各自为政的数据资源管理模式,明确数据资源采集、存储、管理、使用等各环节的责任分工,理清权属关系,广泛汇聚、整合数据,并通过建设完善省级大数据共享交换平台,构建全省一体化的共享交换体系,实现全省政务大数据资源的全面汇聚、共享和应用,为包括经济运行监测分析

# 区域营商环境优化探讨

## ——以苏州市吴中区为例

● 姚 晶 陈 果 李若愚

**摘 要:** 随着《优化营商环境条例》的实施,我国营商环境优化的法律保障得到完善。如何利用好政策保障发挥已有工作优势,促进区域产业转型升级,为当地经济社会发展寻找新动力,成为很多地区研究的重要课题。文章通过调研近年来苏州市吴中区营商环境优化现状,运用 SWOT 理论进行分析,并指出了要进一步优化当地营商环境的政策建议。

**关键词:** 营商环境优化 SWOT 分析 苏州市吴中区

### 一、引言

近年来,按照江苏省和苏州市深化“放管服”改革,优化营商环境工作的总体安排和工作部署,苏州市吴中区持续深化“放管服”改革,加快推动“3550”改革<sup>①</sup>走向常态化,努力构建快捷、高效、规范、透明的营商环境,营商环境指数位居江苏省前列。2019 年吴中区实现地区生产总值 1176 亿元,增长 6%;一般公共预算收入 175.6 亿元,增长 9.3%;全社会固定资产投资 554.6 亿元,增长 14%;新兴产业产值占规上工业总产值比重达到 59.5%<sup>[1]</sup>。2019 年,吴中区位列全国综合实力百强区第九位、绿色发展百强区第九位、全国科技创新百强区第十位<sup>[2]</sup>、全国投资潜力百强区第十六位<sup>[3]</sup>。但与苏州的其他各市、区比较,吴中区营商环境优化工作仍存在一定差距。本文基于 SWOT 分析,就进一步优化营商环境和促进经济发展提出相应的政策建议。

### 二、苏州市吴中区营商环境 SWOT 分析

#### (一) 优势分析

1. 生态区位优势明显。一是生态环境优势明显。吴中区拥有太湖五分之三水域、五分之四峰峦、184 公里湖岸线,全区拥有 1 个国家级旅游度假区、6 个太湖国家风景名胜、1 个

国家 5A 级景区,5 个国家 4A 级景区、2 个国家森林公园、1 个国家地质公园、1 个国家现代农业示范园区、2 个国家湿地公园,被誉为太湖最美的地方<sup>[4]</sup>,近年来旅游的人次保持年均 7—8% 的增长。二是区位优势凸显。近年来,吴中区累计投入 200 亿元,建成了中环快速路吴中区段等 18 条主干道,公路总里程突破 1300 公里(不含市政道路),公路密度 167 公里/百平方公里,区内任何地方 15 分钟即可驶上高速公路,2 小时交通圈覆盖长三角 5 座国际机场和 2 个港口,5 条地铁交织成网,10 大交通工程畅通城乡,构架起立体大通道<sup>②</sup>,凸显吴中多元化大交通格局。

① 江苏省政府办公厅 2017 年 6 月 20 日发布的《关于全省推行“3550”改革的意见》提出“全面推行开办企业 3 个工作日内完成、不动产登记 5 个工作日内完成、工业建设项目施工许可 50 个工作日内完成的改革目标”。

② 5 座国际机场分别是上海浦东国际机场、上海虹桥国际机场、苏南硕放国际机场、南京禄口国际机场和杭州萧山国际机场,2 个港口分别是上海港和苏州港。数据来源 《吴中区十三五综合交通运输发展规划》。

在内的各项专题应用、基础应用奠定工作基础。

#### (三) 深挖经济运行监测分析业务与大数据技术契合点

在应用场景方面,围绕日常工作需要和业务痛点,探索运行数据快速采集、精准分析、直观展现等多种应用场景,积极尝试构建基于大数据基础的经济运行监测分析体系。在分析模型方面,加强相关工作研究,找准经济运行监测分析业务场景和数据挖掘技术的结合点,设计开发固定套路的分析模型,将成熟的、行之有效的数据分析方法、数据挖掘技术以模板的形式固定下来,形成拿来即可用的应用模型,大大降低大数据分析难度。

### 六、结语

虽然大数据技术在经济运行监测分析工作中的应用面临着一些困难和挑战,但更有广阔应用前景。如:网络爬虫技术可以扩展经济运行监测分析数据采集来源,文本挖掘技术将经济运行分析中的结构化数据延伸拓展至无结构化数据,机器学习技术能够分类识别运行风险,可视化分析技术可以形象便捷地展示经济运行动态全貌,而基于 python 的自动化数据处理技术能够极大提升基础数据处理效率等。在新时代新形势下,经济运行监测分析工作需加强对大数据技术的应用与创新,找准分

析技术与分析业务更多的切合点,促进大数据技术与经济运行工作的深度融合,不断提升经济运行监测分析水平,及时、精准分析研判经济运行态势,为地方党委政府当好参谋助手。

### 参考文献:

- [1] 朱宏任. 如何做好经济运行监测分析工作 [J]. 中国经贸导刊, 2002 (22).
- [2] 大数据的基本概念及相关问题 [R]. 中国电子商务企业发展报告, 2013.
- [3] 潘巧智, 张磊. 浅谈大数据环境下基于 python 的网络爬虫技术 [J]. 网络安全技术与应用, 2018 (05).
- [4] 湛志群, 张国焯. 文本挖掘与中文文本挖掘模型研究 [J]. 情报科学, 2007 (07).
- [5] 李苍舒. 基于机器学习理论的中国新金融业态风险预判与防范 [J]. 学术论坛, 2019 (04).
- [6] 罗东玲, 刘瑛. 加快推进江苏数字政府建设的对策研究 [J]. 江苏科技信息, 2019 (26).

(戒天, 江苏省工业和信息化厅)