

大数据技术在乡村画像中的应用研究

李望月^{1,2}, 刘瑾^{1,2}, 陈娜^{1,2}

1. 北京国研网信息股份有限公司, 北京 100010; 2. 国务院发展研究中心信息网, 北京 100010

摘要

在国家大数据战略和乡村振兴战略的大背景下,创新性地将大数据技术引入乡村画像中。在系统研究乡村大数据源、大数据画像技术水平和乡村振兴研究的热点、重点和难点问题的基础上,从乡村发展基础、乡村发展状态和乡村发展行为3个方面构建了大数据乡村画像的概念模型。借助原始指标、知识图谱、政策文本3类标签化方法,将概念模型具体化为可进行实际操作的3套标签体系,在此基础上,对3套标签体系的大数据计算方法、可视化方法进行阐述,并进行画像举例。最后探索了大数据乡村画像未来的研究方向。

关键词

乡村画像;大数据技术;概念模型;标签体系;乡村振兴;知识图谱

中图分类号:TP301

文献标识码:A

doi: 10.11959/j.issn.2096-0271.2020008

Application research of big data technology in rural portrait

LI Wangyue^{1,2}, LIU Jin^{1,2}, CHEN Na^{1,2}

1. Beijing DRCNET Information Co., Ltd., Beijing 100010, China

2. State Council Development Research Center Information Network, Beijing 100010, China

Abstract

Under the background of the national big data strategy and the rural revitalization strategy, big data technology was innovatively introduced into rural portrait. On the basis of systematically studying the research on rural big data sources, big data portrait technical level and the hotspots, key points and difficult problems of rural revitalization research, the conceptual model of the big data rural portraits were constructed from three aspects: rural development foundation, rural development status and rural development behavior. With the help of three types of labeling methods: original index, knowledge map and policy text, the conceptual model was concretized as three sets of label systems that can be used in practice, and on this basis, the big data calculation methods and visualization methods of the three sets of label systems were expounded, and taking examples of portraits. Finally, the future research direction of big data rural portrait was explored.

Key words

rural portrait, big data technology, conceptual model, label system, rural revitalization, knowledge map

1 引言

随着国家大数据战略的实施推进,大数据技术蓬勃发展,正加速成为发现新知识、创造新价值的新一代生产力,其应用已渗入经济社会的各个微观单元,并逐渐成为各行各业寻求突破、创新发展的关键力量。其中,大数据技术中的用户画像技术在电子商务、金融等领域的成功应用,为这些行业带来了巨大的经济效益。在最早应用用户画像技术的电子商务领域,目前已实现对上亿用户的智能化画像,为电商企业快速、便捷、精准地了解用户消费行为、兴趣偏好等特征提供了新的途径,同时也为营销策略的精准制定和实施提供了科学依据。

大数据技术同样已经渗入乡村发展的某些领域,其中,应用比较广泛的是农业领域。通过对气候、土壤、农作物基因库、农业传感器返回的数据等农业大数据进行挖掘、分析发现,大数据技术在推动精准农业实施建设方面做出了重要贡献,在农业产量预测、农业自然灾害的监测预警、农产品价格预测、农产品营销等方面的应用也取得了良好的效果。但到目前为止,无论是理论层面,还是实践层面,尚未有文献对大数据技术在乡村层面的系统性应用进行研究。

党的十九大报告确立了“乡村振兴”战略的重要地位,实施乡村振兴战略的第一步是立足乡情,全面、准确地了解当前“乡村”这一微观单元的经济社会运行全貌和特征,即首先要开展“乡村画像”工作。据2016年第三次全国农业普查数据显示,截止到2016年底,全国约有4万个乡镇、60万个行政村,对这些乡镇、村的发展全貌和特征属性进行系统性分析与挖掘需要耗费

大量的人力、物力、财力,如果不借助大数据画像技术,将很难实现。大数据乡村画像可以实现对数十万乡村的智能化管理,是现代农村建设发展的智能型管理工具,为快速了解乡村发展全貌与现代化建设趋势,挖掘乡村优势特征与不足,制定政策支持方向、科研方向、投资决策方向,提供了有力的理论依据和数据支撑。根据乡村画像的结果,政府部门可以明确乡村产业发展方向、乡村治理领域,同时监测乡村振兴实施效果、制定乡村现代化发展规划;科研机构可以在乡村画像系统的海量数据基础上,使用科学方法构建乡村发展模型,为解决我国“三农”问题提出建议;商业机构可以使用乡村画像系统,根据自身的投资方向,确定最佳投资对象、投资方式和投资期限。

综上所述,大数据技术在多个领域的应用已较为成熟,但在乡村领域的应用大多聚焦于农业领域,对于其在电商等领域成功应用的画像技术,在乡村领域尚未进行应用与实践,而乡村画像又是一项重要且必要的工作。鉴于此,本文将对大数据技术在乡村画像中的应用开展前瞻性研究,在借鉴用户画像方法的同时,基于乡村大数据情况,建立大数据乡村画像的概念模型、标签体系以及实施算法。

2 大数据乡村画像的相关研究现状

鉴于当前尚未检索到以大数据乡村画像为主题的直接文献,本文将沿着“以数据为基础,以技术为手段,以助推乡村振兴发展为目的”的思路,从乡村大数据源、大数据画像技术和乡村振兴领域研究的热点、重点和难点问题3个方面出发,分别进行研究。

2.1 乡村大数据源的研究现状

乡村大数据源是大数据技术开展乡村画像的基础。在对现有文献进行检索研究时,并未发现有关“乡村大数据”“农村大数据”的研究,但近两年关于“农业大数据”^[1-2]的研究逐渐增加,赵瑞雪等人^[1]从数据来源上梳理了国内外在农业科技活动中长期积累和整编的海量科学数据,即农业科学大数据,通过对其列出的数据源进行梳理发现,数量上汇集了12个国外数据源和30个国内数据源(包括6个地方政府门户网站开放数据),内容上大致可分为4类:自然条件与资源数据,如气象数据、地质数据、再生资源数据、空间数据等;农业经济数据,如宏观农业数据、生态农业、中国饲料数据库等;生命科学数据,如生物信息数据、细胞数据、微生物病毒数据、基因数据等;农业行政机构数据,主要来自于政府部门开放门户。姜侯等人^[2]从农业大数据产生的角度,提出了农业物联网数据、农业遥感和农业无人机数据、农业网络数据、科研及农户生产经验数据4类农业大数据,设计了农业大数据平台整体架构,并对农业大数据的应用方向进行了论述。除上述微观层面的农业大数据外,中华人民共和国农业农村部、国家统计局、中华人民共和国海关总署等部门拥有的农业统计数据同样是农业大数据。省、市级别宏观汇总数据可从《中国统计年鉴》《中国农业年鉴》等公开的年鉴资料中获取,但微观数据较难获得。

农业大数据是乡村大数据的一部分,除农业之外的其他乡村大数据主要集中在国家统计局、地方统计局以及其他政府部门。其中,省、市级别的乡村数据可在《中国农村统计年鉴》《中国农业机械工业年鉴》《中国环境统计年鉴》等公开年鉴中

获取,而微观数据由各部门保存,一般不对外公开,外部组织机构也无法通过网络爬取获得。在国家或地方各部门拥有的乡村数据中,以农业普查数据中包含的农村信息最为全面,普查表中的行政村综合表和乡镇综合表囊括了乡村基本信息和特征信息;普查表中的农户基层表包括了村内农户生活、生产经营活动的方方面面,进一步丰富了乡村数据。农业普查数据可提供全面的乡村数据信息,局限性为每十年开展一次,最近一次是2016年开展的第三次全国农业普查。但随着近两年地方乡村振兴战略统计工作的开展,乡村振兴战略统计的调查指标与普查指标具有较高的相似度,可对相同的指标进行不同年度的连续分析。除农业普查数据外,乡村层面还存在着数量可观的非传统数据,如中国精神文明网网站(中国共产党中央委员会宣传部、中央精神文明建设指导委员会办公室)发布的5个批次共计4 717个全国文明村镇的名单数据,中华人民共和国建设部、中华人民共和国文化和旅游部、中华人民共和国财政部等7个部门发布的5个批次共计6 819个中国传统村落的名单数据,中华人民共和国农业农村部发布的两个批次共计300个中国美丽休闲乡村的名单数据。此外,随着《乡村振兴战略规划(2018—2022年)》和《数字乡村发展战略纲要》^[3]的推进实施,数字乡村的建设速度加快,乡村大数据将会得到快速沉淀。

2.2 大数据画像技术的研究现状

大数据画像技术是大数据乡村画像的核心技术手段。大数据画像的概念来源于用户画像,用户画像的英文概念“user persona”最早由艾伦·库珀提出,其认为user persona是真实用户的虚拟代表,根据用户行为、动机等将用户分为不同的类

型,从中抽取每类用户的社会属性、生活习惯、消费行为和兴趣偏好等信息的共同特征,并设定名字、照片、场景等要素对其进行描述^[4]。用户画像又称为人群画像,是根据用户信息抽象出的一个标签化的用户模型,即用高度精练的特征描述一类人,例如年龄、性别、职业、兴趣偏好等。

大数据画像技术是用户画像在大数据环境下的延伸和发展,是基于大量传统数据和非传统数据、静态数据和动态数据、结构化数据和非结构化数据,使用统计分析、机器学习和人工智能等算法,构建模型、挖掘数据、提取特征并进行动态可视化展示的一门数据分析与展示技术。在中国知网文献数据库中,以“大数据画像”“乡村画像”“画像”等关键词进行检索,对现有文献从画像对象、画像技术方法两个方面进行归纳总结。

从画像对象来看,当前文献库中可检索到以“用户画像”“产品画像”“企业画像”“城市画像”为标题的研究,其中,“用户画像”作为画像技术的初始应用领域,研究的文献最多,应用也最为广泛,尤其是在互联网电商、电信、金融等领域^[5-7];研究“产品画像^[8]”和“企业画像^[9]”的文献数量较少,研究内容和视角基本沿用“用户画像”的思路;而“城市画像”相关文献中对画像的理解和研究差异较大,一些文献局限于城市传统数据的传统图表的画像,还有一些文献使用GPS数据、互联网等新兴大数据进行可视化分析^[10-11]。但以“乡村画像”为标题的文献尚未发现。

从画像技术方法来看^[9-23],特征值(标签值)的提取是画像技术的关键。当前画像技术中特征值的提取方法众多,主要可以分为两大类:一类是人工抽取特征值的方法,即通过查阅文献、调研、访谈并在专家的建议下提取用户特征,构建用户画像,此种方法适用于数据表达清晰、数据

量小的场景,主要应用于经济社会宏观和中观领域;另一类是数据挖掘的方法,主要的特征提取大数据技术方法有 k 均值聚类(k -means)^[15]、线性判别分析(linear discriminant analysis, LDA)^[16]、朴素贝叶斯(naive bayesian, NB)^[17]、卷积神经网络(convolutional neural networks, CNN)^[18]、决策树^[19-20]等,主要的大数据关联规则方法有APRORI算法^[21]、FP-Growth关联规则算法^[22],这些方法主要应用于电商、金融等行业的用户画像^[23]。虽然当前关于特征值提取的方法很多,但是特征值挖掘和提取的准确性仍是要持续攻关的难点。

2.3 乡村振兴领域研究的热点、重点及难点问题

乡村振兴领域的热点、重点及难点问题是大数据乡村画像的重要应用方向。自2017年党的十九大报告中提出乡村振兴战略,到2018年《乡村振兴战略规划(2018—2022年)》正式出台,再到2019年《数字乡村发展战略纲要》的提出,学术界关于乡村振兴领域研究的热点、重点及难点问题的研究大多围绕着乡村振兴发展而开展,或是聚焦乡村振兴发展的产业兴旺、生态宜居、乡村文明、治理有效、生活富裕5个维度整体,或是聚焦某一个局部展开研究。通过对乡村领域相关文献进行研究和分析,可以总结为:乡村振兴领域研究的热点是乡村振兴实施效果评价^[24-26],文献以乡村振兴评价指标体系构建为主,部分文献选取对象进行实证分析;乡村振兴领域研究的重点是乡村振兴实施路径^[27-29],当前文献大多对乡村振兴的某一方面进行重点研究,从问题导向给出乡村振兴的实施路径;乡村振兴领域研究的难点是乡村振兴差异化的前期

规划^[30-33],如黄祖辉提出要从区域新型城镇化战略和乡村差异化发展的实际出发,落实乡村振兴战略。其中,除乡村振兴的实施效果可以进行定量分析外,其他研究大多为定性分析。

大数据乡村画像系统的成果可以为乡村振兴领域热点、重点及难点问题的研究提供科学有力的支撑,在实际应用过程中,大数据乡村画像系统主要有三大作用。第一,充分了解乡村全貌。传统的乡村发展评价模型一般使用年鉴数据和调查数据,对互联网上大量可用数据的利用率不高,维度通常也局限于年鉴数据提供的信息,无法全面刻画乡村面貌。大数据乡村画像系统可以充分利用年鉴数据、政府网站公开数据等结构化数据,同时也可以利用地理位置信息、气候信息、政策文件信息、政策称号信息等非结构化数据,多维度、全方位、立体化地对乡村发展情况进行描述,帮助政府根据乡村存在的现实问题,有针对性地制定政策。第二,寻找差异,为因地制宜发展乡村振兴提供理论支持。利用大数据乡村画像系统,可以对省份层面、地市层面、县域层面、乡镇层面以及行政村层面等各级行政区划的乡村振兴发展情况进行分析,应用范围广泛。通过对比,可以发现研究主体的优势和劣势,帮助当地政府因地制宜,开展“千村千面”个性化、特色化规划,解决规划的同质性和落地难等问题。同时,大数据乡村画像系统可智能化匹配生产条件、自然条件或者社会条件相似的标杆性乡村,帮助落后乡村找到适合当地的乡村振兴发展路径和实施路径。第三,监测乡村振兴政策实施效果。大数据乡村画像系统可以对乡村振兴发展情况进行连续时间的监测,通过对多年的数据和信息进行对比,系统可使用标签对分析结果进行展示,例如,某省份人均农林牧渔产值连续5年增加,乡村画像标签可展示

为“连续5年人均农林牧渔产值增加”。同时,大数据乡村画像系统也可设置预警机制,在指标值出现异常波动或者下降趋势时(针对正向指标),发出预警信号,帮助政府部门及时做出应对。

综上所述,本文结合已有研究成果,以乡村大数据为基础,以大数据画像技术为手段,以辅助解决乡村振兴发展中的热点、重点和难点问题为导向,开展乡村画像概念模型和标签体系的构建工作。

3 乡村画像概念模型构建及数据支撑

概念模型来源于数据库设计,是对现实世界中问题域内的事物进行描述的模型,是现实世界到信息世界的第一层抽象,是用户与数据库设计人员进行交流的语言,因此,概念模型一般具有较强的语义表达能力,能够方便、直接地表达应用中的各种语义知识,且简单、清晰、易于用户理解。乡村画像概念模型是对乡村微观单元中的经济和社会发展进行描述的模型,是对真实乡村的经济社会运行状态和行为的刻画,是大数据乡村画像的核心内容,是影响乡村画像结果的关键。本文将在确立乡村画像构建原则的基础上,从基本属性、状态属性和行为属性3个方面构建乡村画像概念模型,并在概念模型的基础上,系统性梳理所需的乡村画像数据支撑。

3.1 乡村画像概念模型的构建原则

乡村画像概念模型是大数据乡村画像的基础,构建科学合理的大数据乡村画像概念模型应遵循全面性、系统性、特征性、可行性和智能性的原则。

(1) 全面性原则

乡村画像概念模型在构建时应遵循全

面性原则,即应充分考虑各类乡村的属性,构建的乡村画像模型应覆盖全部村的基本属性,可以实现对任何一个乡村进行画像,不应存在不能进行乡村画像的情况。

(2) 系统性原则

乡村画像的系统性是指在乡村画像概念模型的构建中,坚持全局意识、整体观念。概念模型构建时应系统地考虑各要素的层次性,各要素由外至内层层深入,构建一个乡村经济社会内各子系统、各要素相互独立又相互联系的有机整体。

(3) 特征性原则

乡村画像概念模型的构建要突出特征性,特征属性是乡村画像概念模型的重要组成部分,是彰显乡村特色、挖掘乡村价值、因地制宜地推进乡村振兴的重要工作抓手。

(4) 可行性原则

乡村大数据是大数据乡村画像落地实施的前提,乡村画像是在乡村大数据分析、计算和研究的基础上开展的,如果乡村画像的数据不可获得,那么乡村画像概念模型的构建将会停留在理论研究层面,无法走向实际操作层面。

(5) 智能性原则

概念模型的构建应清晰、明确,能够被计算机理解,便于进行程序化实施,进而保障后续对4万个乡镇、60万个行政村的大数据画像的自动化、智能化、动态化实现。

3.2 乡村画像概念模型的建立

基于乡村画像概念模型的构建原则,在乡村画像概念模型构建的过程中,综合考量乡村经济社会的基本属性和特征属性、静态属性和动态属性、微观属性和宏观属性,最终确定乡村画像概念模型从“基本属性”“状态属性”“行为属性”3个方面进行构建。在构建概念模型时,将按照

“345”的原则构建,即乡村画像基本属性由3个维度构成,分别为“地理区位”“自然条件”“政策称号”;乡村画像状态属性由4个维度构成,分别为“规模”“速度”“结构”“效率”;乡村画像行为属性由5个维度构成,分别为“发展产业”“保护生态”“延续文明”“加强治理”“推动共同富裕”。乡村画像概念模型如图1所示。

3.3 乡村画像概念模型的数据支撑

乡村画像概念模型及标签体系的设立是基于广泛的、多维度的乡村大数据源进行的,包括传统数据和非传统数据、结构化数据和非结构化数据等多种类型。本文使用的主要数据来源及相应用途见表1。

4 大数据乡村画像标签体系的构建

在乡村画像概念模型的框架下,以乡村大数据为基础,以展示乡村全貌、挖掘乡村特征、助力乡村振兴发展为目标,构建乡村画像标签体系。相对于概念模型而言,乡村画像标签体系更加具体和形象,在获取数据的基础上,可直接进行可视化标签分析。

4.1 标签体系的构建方法

与指标体系不同,标签体系不再受各指标层级必须相同、指标值必须为数值型数据的限制,是一种更为灵活、更适合大数据生态的模型体系。本文将从指标数据、知识图谱、政策规划文本3个方面,介绍5类标签的构建方法。其中,从指标数据出发,可以构建2类原始标签;从知识图谱出发,可以通过对原始标签进行扩展和泛化处理,生成2类新的标签;从政策规划文本出发,提取关键词,可以生成宏观属性的乡

村行为标签。

(1) 基于指标名称的原始标签提取

从乡村数据指标出发,对数据指标名称进行转换,在保持指标含义不变的条件下,将原始指标转换为以短词语或短词组形式为主、指标名称更为精简的标签,并将其标记为原始标签。如第三次全国农业普查中行政村普查表中的“生活污水是否经过集中处理”这一指标,在提取标签时,可以提取为“生活污水集中处理”。此外,对于指标中存在的“个数”“人数”“户数”等内容,在进行标签提取时,一般进行删除处理,如“村集体创办的互助型养老服务设施个数”指标,在提取标签时,转化为“村集体创办互助型养老服务设施”。

(2) 基于指标数据计算的标签提取

从乡村数据出发,借助传统描述性统

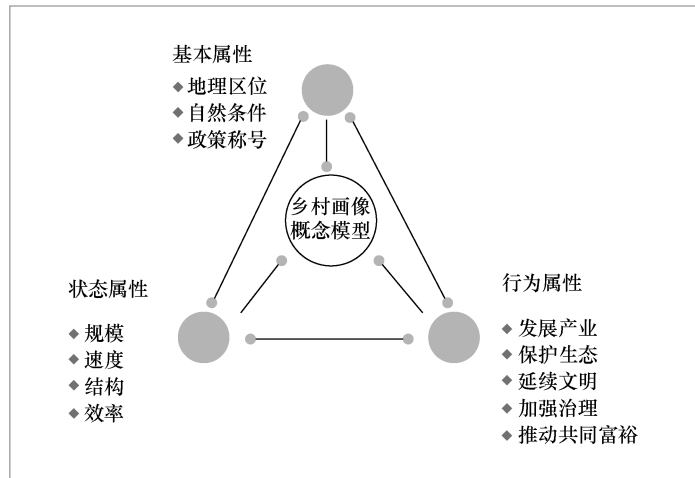


图1 乡村画像概念模型

表1 乡村画像概念模型数据源汇总

属性类型	一级标签	数据来源	数据用途
基本属性	地理区位	(1) 中华人民共和国民政部网站提供的《中华人民共和国县级以上行政区划代码》 (2) 国家各类经济区域划分标准 (3) 《中国统计年鉴》《中国农村统计年鉴》等年鉴中涉及行政区划的信息	利用行政区划信息、年鉴信息和地理位置信息,构建包含省、市、县、乡、村名称、归属关系、地理位置等信息在内的数据库
	自然条件	(1) 百度地图、Google地图等地图信息 (2) 国家气象信息中心提供的地面气象、高空气象、卫星探测、气象灾害等数据	利用地图信息、自然资源信息、气象信息,构建能够识别行政村自然条件的数据库
	政策称号	(1) 中华人民共和国农业农村部、中华人民共和国科学技术部、中华人民共和国生态环境部、中华人民共和国工业和信息化部、中华人民共和国国家发展和改革委员会、中华人民共和国财政部、中华人民共和国水利部、中华人民共和国文化和旅游部、中华人民共和国住房和城乡建设部等国家部委公布的多个政策称号的各个批次名单 (2) 省级、市级、县级政府公布的农业农村相关政策称号名单	利用公开政策称号名单,构建能够识别行政村是否有某个称号、是否在某个政策区域范围内、是否曾经有称号而后来被去除的数据库
状态属性	规模	(1) 《中国统计年鉴》《中国农村统计年鉴》《中国农业年鉴》《中国休闲农业年鉴》《中国品牌农业年鉴》《中国农产品加工业年鉴》《中国智能农业发展报告》《全国农产品成本收益资料汇编》《中国农业机械工业年鉴》《中国农村贫困监测报告》《中国农垦统计年鉴》《中国环境统计年鉴》《中国教育统计年鉴》《中国社会统计年鉴》《中国民政统计年鉴》《中国人口和就业统计年鉴》等年鉴数据,以及国民经济与社会发展统计公报、民政事业发展统计公报等公报数据 (2) 各省份单独发布的农业农村方面的年鉴 (3) 各省份单独发布的第三次全国农业普查主要数据公报 (4) 地方政府构建的乡村大数据平台,如云南的数字乡村平台,公布了云南省每个行政村的特色产业、基层组织、自然资源等信息,属于微观数据库	利用宏观数据和微观数据,可构建从上自下的包含农业生产、农民生活、农村环境等信息的全方位、多维度的农业农村数据库
	速度		
	结构		
	速率		
行为属性	发展产业	(1) 《中国统计年鉴》《中国农村统计年鉴》《中国农业年鉴》《中国休闲农业年鉴》《中国品牌农业年鉴》《中国农产品加工业年鉴》《中国智能农业发展报告》《全国农产品成本收益资料汇编》《中国农业机械工业年鉴》《中国农村贫困监测报告》《中国农垦统计年鉴》《中国环境统计年鉴》《中国教育统计年鉴》《中国社会统计年鉴》《中国民政统计年鉴》《中国人口和就业统计年鉴》等年鉴数据,以及国民经济与社会发展统计公报、民政事业发展统计公报等公报数据 (2) 各省份单独发布的农业农村方面的年鉴 (3) 各省份单独发布的第三次全国农业普查主要数据公报 (4) 地方政府构建的乡村大数据平台,如云南的数字乡村平台,公布了云南省每个行政村的特色产业、基层组织、自然资源等信息,属于微观数据库	利用宏观数据和微观数据,可构建从上自下的包含农业生产、农民生活、农村环境等信息的全方位、多维度的农业农村数据库
	保护生态		
	延续文明		
	加强治理		
	推动共同富裕		

计分析,计算某一指标或原始标签的基本统计量,对单时点数据计算最大值、最小值、均值、秩(排序),对连续时点数据计算增长率等,并对计算后的统计量中的突出特征进行提取。以单时点第三次全国农业普查数据为例,针对行政村“外来人口”这一指标,在全省样本层面计算指标排名,针对排名提取最具特征的标签,如“外来人口省内Top3”“外来人口省内前10%”等。对于非传统数据的标签化处理,视数据情况而定,其中,政府部门公开评选的乡村称号数据,如“乡村振兴示范村”“生态村”等文本型数据,无需转换,可直接作为乡村画像标签。

(3) 基于知识图谱的原始标签扩展

基本知识图谱的原始标签扩展是指从原始指标出发,借助相关知识图谱,引入外部相关信息,在原始指标的基础上,扩展出新的标签。关于扩展型标签的知识图谱的构建,应将乡村的地理区位等基本属性与国家政策、本区域的特征、相邻区域的特征建立直接或间接的关系,借助相关知识图谱,可以扩展出是否为“两省边界村”、是否“与一线城市相邻”、是否“与经济大省相邻”等扩展性标签。

(4) 基于知识图谱的原始标签泛化

基本知识图谱的原始标签泛化是指从原始指标出发,借助知识图谱技术,挖掘出多个指标的共同属性,并将其泛化成一个共性标签。对于泛化型标签的知识图谱的构建,应将乡村的各项经济和社会活动进行系统化处理,并建立乡村各项经济和社会活动的层级关系,进而基于建立的层次关系,实现原始标签准确、高效的泛化。如某个乡村在“设施农业、循环农业、工厂化生产”“高标准农田”“按测土配方施肥”方面开展了大量工作,则基于相关的知识图谱,可以泛化出“开展新型生产模式”这一标签。

(5) 基于政策规划文本的标签提取

从国家级政策规划文本出发,提取其中的关键词,对关键词进行转换,生成新的标签,一般适用于宏观属性一、二级标签的提取,对于三级以上标签的提取仍需依赖相关的知识图谱。这种标签提取方法紧跟国家政策形势,提取的标签可直接用于政策实施进度、实施效果的监测,受到政府相关部门的高度关注。如从《乡村振兴战略规划(2018—2022年)》中提取“构建现代农业产业体系”一词,可以将其标签化为“推行现代农业产业化发展”。

4.2 乡村画像基本属性标签体系

在乡村画像概念模型中,乡村画像基本属性包含“地理区位”“自然条件”“政策称号”3个维度。地理区位从行政区划出发,借助基于地理、经济等知识图谱体系,生成行政区划、区域扩展、临近区域等新型标签;自然条件将地形、气候、土壤、资源、自然灾害作为画像标签;政策称号将乡村、乡村所属市县园区、乡村下属企业在涉农领域获取的国家部委机构公开评选的政策称号,直接作为乡村标签。这3个维度的标签值在短期内相对稳定,除行政村区划调整和滚动评审的政策称号之外,绝大多数乡村的基本属性在短期内不会变动。乡村画像基本属性标签体系如图2所示。

乡村画像基本属性标签体系构建中,地理区位标签的初始数据一般仅为乡村所属的行政区划这一个数据值的信息,对于其他地理区位标签,需借助知识图谱进行标签扩展;自然条件的数据支撑以农业部信息中心、国家气象信息中心等政府部门的公开数据为主,进行标签构建;政策称号数据以中华人民共和国农业农村部、中华人民共和国国家发展和改革委员会、中华人民共和国科学技术部、中华人民

共和国财政部、中华人民共和国生态环境部、中华人民共和国水利部、中华人民共和国文化和旅游部、中华人民共和国住房和城乡建设部、中华人民共和国民政部等政府部门联合或单独公开评审的称号名单为主,进行标签构建,其中“农业产业化国家重点龙头企业”和“国家自然保护区”的评审为滚动式评选,当年评审后,之前发布的名单作废;国家有机食品生产基地会定期对前面审批的基地进行复核,公布审核后的名单。最终确立一级标签3类,二级标签12类,三级标签29类,四级标签35类,五级标签35类。具体标签体系见表2。

4.3 乡村画像状态属性标签体系

乡村画像状态属性标签体系,主要用来标识当前乡村发展的基础性、代表性指



图2 乡村画像基本属性标签体系

标的发展状态,反映乡村发展的基本水平。通过状态属性可以对当前乡村的发展水平有一个概括性的了解,一般地,状态标签多为长期以来政府部门关注的指标。对于状态属性标签化体系的构建,将从“规

表2 乡村画像基本属性标签体系

一级标签	二级标签	三级标签	四级标签	五级标签	
地理 区位	行政 区划	省-市-区县-乡镇	/	/	
	区域 扩展	两类划分	沿海区域、内陆区域	/	
		三类划分	东、中、西部地区	/	
		六类划分	东北、华北、华中、华东、华南、华西地区	/	
		三大经济圈	环渤海经济圈、长三角经济圈、珠三角经济圈	/	
		三大超级城市群	京津冀、长三角、粤港澳	/	
		所属园区	位于国家级/省级开发区、位于自贸区、位于综合保税区	/	
	临近 区域	相邻区域	相邻的行政区	与直辖市相邻、与省会城市相邻、与一线城市相邻、与经济强市/县相邻、与人口大市/县相邻	与国家级/省级开发区相邻、与自贸区相邻、与综合保税区相邻 与少数民族聚居村相邻、与中国传统村落相邻、与特色景观旅游名村相邻
			相邻的园区	与国家级/省级开发区相邻、与自贸区相邻、与综合保税区相邻	
		相邻特征村	与少数民族聚居村相邻、与中国传统村落相邻、与特色景观旅游名村相邻		
	边界村/边境村	2个行政机构边界村/边境村	两国边境村、两省边界村、两市边界村、两县边界村、两乡镇边界村		
		3个行政机构边界村	三省边界村、三市边界村、三县边界村、三乡镇边界村		
		多个行政机构边界村	多省边界村、多市边界村、多县边界村、多乡镇边界村		

(续表)

一级标签	二级标签	三级标签	四级标签	五级标签
自然条件	地形	高原、山地、平原、丘陵、盆地	/	/
		气候	降水量	/
	日照强度		/	/
	温度		/	/
	湿度		/	/
	温差		/	/
	土壤		土地肥沃度	红土地、黑土地
		土壤酸碱度	盐碱地	/
	资源	土地资源	村占地面积	/
			耕地资源	/
		水资源	充足、不充足/海水、江水	/
		矿产资源	/	/
		森林资源	/	/
		渔业资源	/	/
畜牧业资源		/	/	
自然灾害		干旱、冰雹、地质灾害	/	/
政策称号	村庄政策称号	村庄政策称号	乡村振兴示范村	全国乡村振兴示范村
			生态村	国家级生态村
			人居环境示范村	改善农村人居环境示范村
			休闲旅游村	全国乡村旅游重点村
				中国美丽休闲乡村
				中国优秀国际乡村旅游目的地
				特色景观旅游名村
				中国传统村落
				少数民族聚居村
				国家级贫困村
		脱贫摘帽村		
	村庄上级行政机构政策称号	乡镇政策称号	产业强镇	农业产业强镇
			环境优美小镇	全国环境优美乡镇
			特色小镇	特色小镇
文明村镇			全国文明村镇	
园区政策称号	县级政策称号	国家自然保护区	/	
		全国“平安农机”示范县	/	
	示范区政策称号	可持续/绿色/创业创新	国家农业可持续发展实验示范区(暨绿色农业发展试点先行区)	
			全国农村创业创新园区(基地)	
		农产品加工示范基地	全国农产品加工业示范基地	
		有机/特色农产品	国家有机食品生产基地	
			特色农产品优势区	
		产业融合	国家农村产业融合发展示范园	
			全国农村一二三产业融合发展先导区	
		现代/科技/高新技术农业	国家农业科技园	
	现代农业产业园			
企业政策称号	农业企业称号	龙头企业	农业产业化国家重点龙头企业	
		加工企业	全国农产品加工业示范企业	
			全国主食加工示范企业	

模”“速度”“结构”“效率”4个维度开展。在规模和速度维度下, 主要对乡村产业、乡村人口、村集体收支的规模和速度进行考量; 在结构维度下, 对乡村非农收入、老龄化人口、农民受教育、医疗资源配置、教育资源配置的结构进行考量; 在效率维度下, 对人均产值、农业机械化效率进行考量。乡村画像状态属性标签体系如图3所示。

乡村画像状态属性标签体系中的数据一般为常见性数据, 对于省、市层面的数据, 可以在《中国农业年鉴》《中国农村统计年鉴》等公开年鉴中获取; 对于乡村层级的状态数据, 一般保存在各县级以上的统计部门, 不对外公开。全国乡村的状态标签数据可从农业普查数据中获取, 但一般乡村级别底层微观数据的获取存在一定困难。对于非政府部门开展的状态标签的画像, 可以考虑借助非传统数据, 如借助遥感数据获取农业产业数据, 借用手机信令数据获取乡村人口流动的信息。最终确立一级标签4类, 二级标签13类, 三级标签20类, 具体状态标签体系见表3。

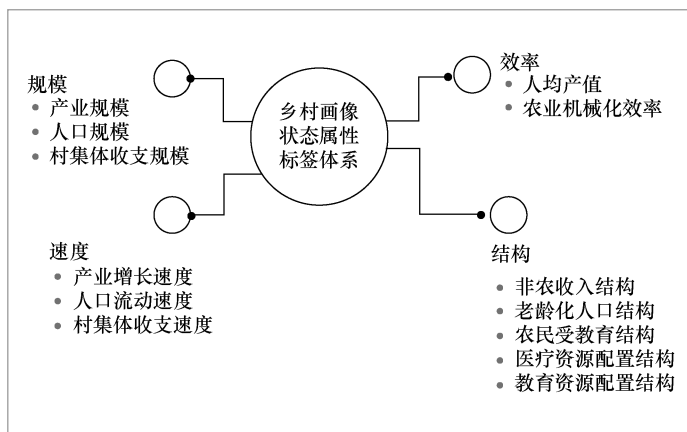


图3 乡村画像状态属性标签体系

4.4 乡村画像行为属性标签体系

乡村画像行为属性标签体系, 将从乡村微观行为标签和乡村宏观行为标签两个方面开展, 微观行为标签主要反映被刻画的对象的单一行为, 较为具体, 而宏观行为标签则为多种行为的综合反映, 较为抽象。微观行为标签有2类: 一是某些单时点原始指标可直接被转换为微观行为标签, 如集中处理生活垃圾、村集体创办卫生室、施用农家肥; 二是对连续时点的状态性指

表3 乡村画像状态属性标签体系

一级标签	二级标签	三级标签	一级标签	二级标签	三级标签	
规模	产业规模	农作物规模	速度	产业增长速度	农业增长速度	
		林作物规模			非农业产业增长速度	
		畜牧业规模			人口流动速度	外出人口增长率
		渔业规模			村集体收支速度	外来人口增长率
		特色种植业规模			村集体收入增长速率	
	非农业规模	村集体办公支出增长速度				
	人口规模	户籍人口规模		结构	非农业收入结构	/
		常住人口规模			老龄化人口结构	/
		外来人口规模			农民受教育结构	/
	村集体收支规模	外出人口规模		医疗资源配置结构	教育资源配置结构	每千人拥有的医生数
		村集体收入规模		教育资源配置结构		每千人拥有的教师数
		村集体支出规模		效率	人均产值	/
					农业机械化效率	/

标变动进行加工计算,将其转换为微观行为标签,如可以通过计算“村集体支出”两年间的变动,生成“缩减村集体支出”这一标签。宏观行为标签可借助政策文本进行提取,同时也可借助相关知识图谱进行扩展和泛化。

基于《国家乡村振兴发展战略规划(2018—2022)年》(以下简称战略规划)和第三次全国农业普查的数据指标,从“发展产业”“保护生态”“延续文明”“加强治理”“推动共同富裕”5个维度,构建了乡村画像行为属性标签体系。其中,从战略规划原文提炼了乡村画像宏观行为标签,标识为“乡村画像一级标签和部分二级标签”,同时从第三次全国农业普查指标数据中提取了乡村画像的微观行为标签,最终构建乡村画像行为属性标签体

系如图4所示。

在乡村画像行为属性标签体系的构建中,一级标签是基于战略规划原文信息直接提取的,二级和三级标签以第三次全国农业普查的调查表指标为基础,以农业普查中的行政村、农业经营户、规模以上农业经营户3类主体的调查信息表为依据,构建乡村画像行为属性标签体系。乡村画像行为属性包含正向属性和负向属性,本文的行为标签体系多以正向属性的形式给出,以期引导乡村振兴正向发展,在应用中视乡村各标签值的大小,加入相应的程度副词或否定词进行负向刻画。最终确定一级标签5类、二级标签23类、三级标签44类、四级标签31类,具体行为属性标签体系见表4。

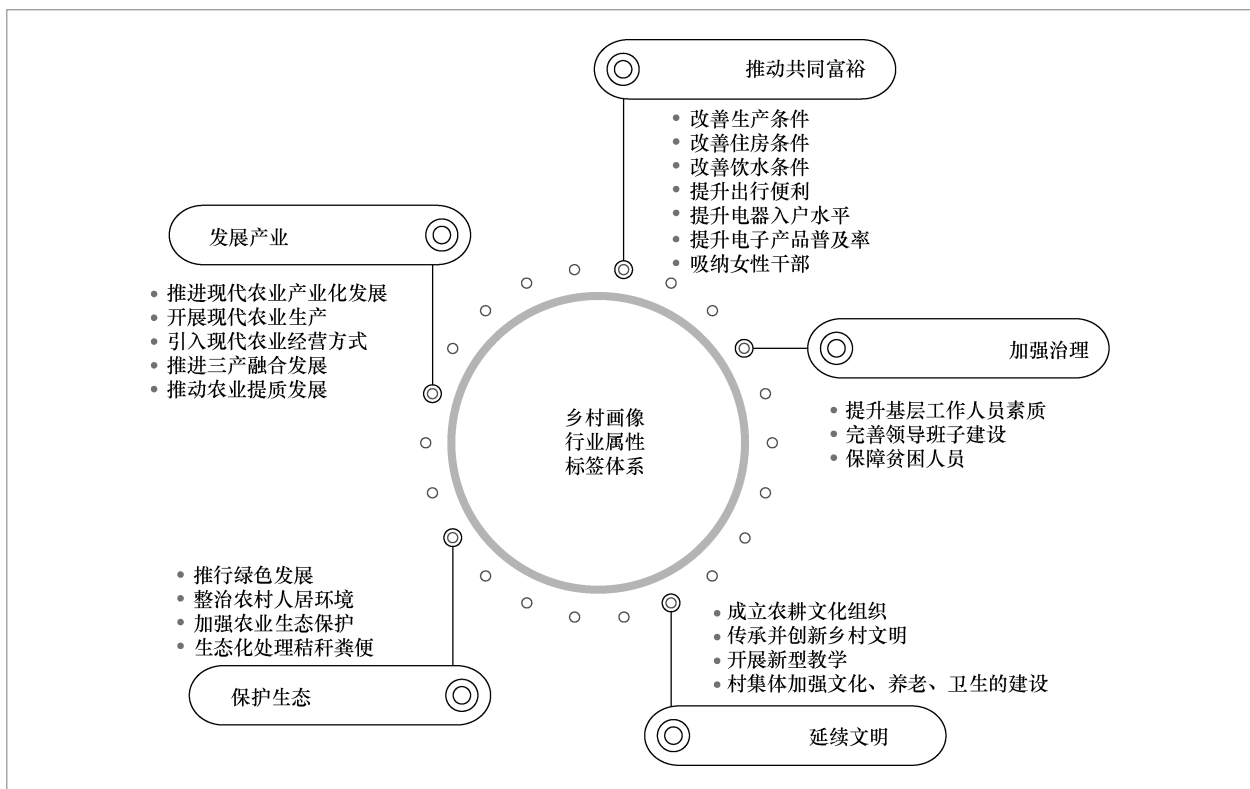


图4 乡村画像行为属性标签体系

表4 乡村画像行为属性标签体系

一级标签	战略规划原文	二级标签	三级标签	四级标签
发展产业	构建现代农业产业体系	推进现代农业产业化发展	进行工商登记	注册个人工商营业执照 注册家庭农场
			开展农业技术培训	/
	构建现代农业生产体系	开展现代农业生产	参加农业保险	参加政策性保险、参加商业保险
			转变农业生产方式	以生产为主、生产与加工一体化、生产加工销售一体化
			开展新型生产模式	设施农业、循环农业、工厂化生产 拥有高标准农田 按测土配方施肥
	构建现代农业经营体系	引入现代农业经营方式	引入先进生产设施	拥有滴灌、喷灌、渗灌设施
			参加经营组织形式	公司+农户、农民合作社、专业协会、土地托管
	农村一二三产业深度融合发展	推进三产融合发展	转变农产品销售方式	自销、中间商经销、按生产订单销售
			开展新型经济活动	餐饮住宿、采摘、垂钓、农事体验
			推进农业与旅游业	特色景观旅游名村
推进农业与餐饮业			拥有营业执照的餐馆	
推动农业从增产导向转向提质导向	推动农业提质增效	生产高品质农产品	无公害农产品、绿色食品、有机食品	
保护生态	推行乡村绿色发展方式	推行绿色发展	推行绿色生活能源	电、沼气、太阳能
	加强农村人居环境整治	整治农村人居环境	集中处理生活垃圾	/
	人与自然和谐共生	加强农业生态保护	集中处理生活污水	/
			开展农业生态保护	生态公益林补偿 保护井水和泉水
			降低农业生态危害	降低化肥施用量 降低农药喷洒量
	生态化处理秸秆粪便	拥有畜禽粪便无害化处理设施 施用农家肥 秸秆粉碎还田	/	
延续文明	深入挖掘农耕文化	成立农耕文化组织	农民业余文化组织	/
	保护传承的基础上创造性转化、创新发展	传承并创新乡村文明	推行文体卫发展	创办幼儿园、托儿所 创办图书室、文化站 创办体育健身场所 创办卫生室
			推行先进零售模式	营业面积50m ² 以上的综合商店和超市
	焕发出乡风文明新气象	开展新型教学	小学教学点	/
	村集体加强文化、养老、卫生的建设	村集体创办幼儿园、托儿所 村集体创办到助型养老服务设施个数 村集体创办卫生室 村集体创办变电站	/	

(续表)

一级标签	战略规划原文	二级标签	三级标签	四级标签
加强治理	加强农村基层基础工作	提升基层工作人员素质	引入大学生村干部	/
			提升领导班子的文化水平	党支部书记受教育程度 村委会主任受教育程度
			健全乡村治理体系	完善领导班子建设
确保农村社会安定有序		保障贫困人员	扶助低保户	农村最低生活保障户数 农村最低生活保障人数
			供养五保户	农村五保户供养户数 农村五保户供养人数
			推动共同富裕	改善了农村生产条件
改善了农村生活条件	改善了农村生活条件	改善住房条件	引入专业农业机械	耕整机、旋耕机、播种机、水稻插秧机、排灌动力机械、联合收获机、机动脱粒机
			推进畜牧业机械	饲草料加工机械、挤奶机、剪毛机、增氧机
			引进林业机械	果树修剪机
提升出行便利	改善饮水条件	提升饮用水源品质	钢筋混凝土、砖混、砖石木、竹草土坯	/
			降低取水困难	经过净化处理的自来水、受保护的井水和泉水、不受保护的井水和泉水、江河湖泊水、收集雨水、桶装水
			提升出行便利	家用交通工具
提升电器入户水平	提升电子产品普及率	提升电子产品普及率	普及制冷电器	拥有小汽车 拥有摩托车、电瓶车
			普及传媒电器	拥有空调 拥有电冰箱
			普及卫生电器	拥有彩色电视机 拥有有线彩色电视机
提升电子产品普及率	提升电子产品普及率	提升电子产品普及率	提升电子产品的拥有率	拥有沐浴热水器 拥有电脑 拥有手机
			提升电子产品的联网率	电脑联网 手机联网 互联网购物
			促进社会公平正义	吸纳女性干部

5 大数据乡村画像的实施算法及画像举例

5.1 大数据乡村画像的标签计算方法

(1) 乡村画像基本属性标签的计算方法

一是地理区位标签计算。从乡村所属行政区划出发,借助大数据工具,识别乡村行政区划边界,借助知识图谱,通过建立区域特征判别算法、邻近村特征识别算法、边界村识别算法,对临近区域中的相邻区域、边界村、边境村进行计算识别。

二是自然属性标签计算。在乡村数据较难获得的情况下,可借助所属县、市自然属性

的数据,进行标签化处理。

三是政策称号标签计算。通过大数据技术查找比对政府部门公开评选的各类涉农名单,政府公开评选的涉农的各种名称头衔可直接用于乡村画像,直接生成乡村画像的特征标签。此类标签的生成以收集整理政府部门公开评审的涉农名单为前提,在收集、积累涉农数据的基础上,分类建立“村”“镇”“县”“产业园”等名录。在具体开展乡村画像工作时,可对被画像的乡村名称与名录库中公开评选的名称进行查找、比对,若能在名录库中查找到乡村名称,则记录此乡村对应的被评选的称号,可直接将被评的称号作为标签;若查找不到,则不生成标签,不进行处理。对于“有进有出”型的评选,在生成标签时,不仅对当年入选获取的称号生成标签进行标识,对于曾被评选上但当年未被评选上的对象,同样可粘贴标签,如标签化处理可为“某年被评选为某称号,但在某年被摘掉称号”。

(2) 乡村画像状态标签的计算方法

乡村画像状态属性标签的计算,主要对乡村的“规模”“速度”“结构”“效率”等数据进行加工计算,计算其在“本乡镇”“本县”“本市”“本省”以及“全国”层面中的“位次”,并筛选特征指标,进行标签化处理,在具体特征标签计算时,可采用“市内Top3”“省内Top10”“省内前10%”等规则进行标签化。

(3) 乡村画像行为标签的计算方法

从标签体系出发,将原始数据按照标签体系进行模型化处理,最常见的是指数合成方法,将初始标签值加工计算为各级标签值。根据计算的各级标签值的大小,选定相关的程度副词或否定词,生成具体的行为标签。如经指数合成计算,得出“改善生产条件”这一标签值为零,那么在具体生成行为属性标签时生成为“未改善生产条件”。

5.2 大数据乡村画像的可视化方法

(1) 乡村画像的地图可视化方法

在对乡村进行画像时,首先应将乡村的基本情况可视化展示,其中包括地理区位、自然条件、政策称号等基本属性标签和产业、人口等部分状态属性标签的内容。对于地理区位标签,对乡村行政区划的边界、行政区划的面积以及在上级行政区划中的位置进行展示,借助地图对乡村地理位置进行明确展示。同时在地图中对乡村人口、用地、政策称号等内容进行文字说明。此外,对于乡村发展重要的、特殊的自然条件,如江河、煤矿、气候等标签,也可在地图上标识展示。

(2) 乡村画像之大数据词云图方法

“词云”这一概念由美国西北大学新闻学副教授、新媒体专业主任里奇·戈登提出,是对网络文本中出现频率较高的“关键词”予以视觉上的突出,形成“关键词云层”或“关键词渲染”,从而过滤掉大量的文本信息,使浏览网页者通过词云就可以领略文本的主旨。当前词云图被大量应用在对政策要闻等词频的统计中。

在乡村画像领域,将引入词云图对乡村的诸多标签进行词云可视化展示。与当前词频统计画像不同,乡村画像词云图以各标签无量纲化处理后的标签值进行画像。在无量纲化处理时,采用“累积分布函数 $\times 100$ ”的方法进行无量纲化处理,即将原始数据减去均值后除以标准差后的值,放入正态分布的累积分布函数中计算函数值,并将函数值乘以100。这种处理方法将原始数据映射在区间(0,100),能较好地保持各乡村标签值之间的差距,其中得分接近100代表标签属性具有突出性特征,得分接近0代表标签特征不明显。经过无量纲化处理后的得分数据,可实现对乡村

上百个标签的同时展示。

(3) 乡村画像之大数据综合图方法

乡村画像大数据综合图将乡村地图、乡村画像词云图、数据表、统计图等多种图表分析的可视化方法进行综合运用,基于乡村大数据条件,充分发挥各类统计图表的优势,科学设计综合图的可视化展示内容、展示形式和展示布局,最终提供大数据乡村画像的综合图。

5.3 大数据乡村画像举例

借助第三次全国农业普查的普查表和非传统的乡村数据,对乡村微观的各项指标数据进行模拟,并以此作为大数据画像可视化基础。在模拟数据的基础上,依据乡村基本属性标签体系、状态属性标签体系和行为属性标签体系,对数据进行加工处理,绘制乡村地图、词云图以及各项统计图,最终确定大数据乡村画像,如图5所示。

乡村地图的可视化部分布置在大数据画像图的左上角,采用地图与实景图相结合的方式,在构建系统时可通过鼠标点击实现地图和实景图之间的切换,其中,在地图层面,对乡村的基本属性标签和部分状态属性标签进行标识。乡村画像标签词云图部分布置在大数据画像图的左下角两块区域:一是大数据词云图,将对原始数据、状态标签、行为标签中的定量数据进行加工处理,生成数百个乡村画像标签,进行词云展示;二是将词云图中标签值小、肉眼难以识别的标签提取出来,单独列为“乡村振兴有待提升标签”,以数据表格的形式进行展示。状态属性标签可视化部分布置在大数据画像图的右上角,分别对乡村规模、速度、结构、效率采用“标签(省内Top10等)+传统统计图”的方式进行可视化展现。行为标签可视化部分布置在大数据画像图的右下角,对于行为标签各维度的标签值进行可视



图5 大数据乡村画像举例

化显示。最后,在右侧中间位置,基于3套标签体系,合成乡村振兴推进水平指数,从5个维度对本乡村与全省水平进行可视化比较展示。

6 结束语

大数据画像技术是一种描述事物全貌、挖掘事物特征、发现事物发展规律的重要技术,在乡村画像领域具有广阔的应用前景。本文创新性地构建了乡村画像概念模型、标签体系,明确了大数据乡村画像的实施算法,设计了乡村画像的可视化形式,为大数据乡村画像奠定了理论与实践的基础。大数据画像的结果可直接用于评价乡村振兴效果为乡村振兴前期发展规划、发展路径的制定和实施提供决策参考。但是,大数据画像工作仍然面临着诸多挑战。未来,大数据乡村画像实践工作的研究,可以从以下几个方面展开。

一是系统性梳理乡村大数据源,对数字乡村建设中产生的乡村大数据进行前瞻性设计。乡村大数据是乡村画像的基础。突破当前以农业科学大数据为核心,以辅助农业科研、生产、加工、销售为目的建设农业大数据库的思想,将农业大数据的建设扩展到乡村大数据建设,对农业之外的其他乡村大数据源进行系统性梳理,预判和规划数字乡村建设的数据内容,深入研究乡村大数据的应用方向,推进乡村大数据的应用方向从农业精准生产到乡村振兴精准管理与规划的跨越式发展。

二是构建乡村画像知识图谱体系,进一步扩充乡村画像的标签体系。知识图谱是智能化开展大数据乡村画像的关键。对于标签扩展和标签泛化,均依赖于知识图谱,而当前乡村研究领域尚未发现相关

知识图谱的构建。乡村画像知识图谱的构建是一项系统工程,应在对乡村大数据、自然资源、乡村经济和社会活动、国家政策之间的关系进行系统性梳理的基础上展开。

三是建立大数据乡村画像的标签筛选机制。标签筛选机制是乡村画像工作顺利进行的必要保障,从初始的乡村大数据中可以提取大量的原始标签,而借助知识图谱技术又可以扩展和泛化出新的标签,标签数量可达数百个,且随着数字乡村工作的开展,乡村大数据数据量不断增加,标签的数量也将不断增加,可达上千个,如不建立标签筛选机制,直接将全部标签进行画像展示,不仅不能通过画像刻画乡村的特征,而且会陷入标签灾难,鉴于此,标签筛选机制在未来也将是一项重要的工作。

四是探索大数据乡村画像的监测预警体系。监测预警是乡村画像的重要应用方向。将电商、金融领域成功运用的用户画像预测分析方法,尤其是把 k -means、决策树、支持向量机等机器学习方法引入乡村画像体系之中,通过构建相应的模型算法,实现对乡村经济社会发展的监测预警,为政府进行政策调整提供理论依据。

参考文献:

- [1] 赵瑞雪,赵华,朱亮.国内外农业科学大数据建设与共享进展[J].农业大数据学报,2019,1(1):24-37.
ZHAO R X, ZHAO H, ZHU L. Progress in the development and sharing of big data in agricultural science between China and foreign countries[J]. Journal of Agricultural Big Data, 2019, 1(1): 24-37.
- [2] 姜侯,杨雅萍,孙九林.农业大数据研究与应

- 用[J]. 农业大数据学报, 2019, 1(1): 5-15.
JIANG H, YANG Y P, SUN J L. Research and application of big data in agriculture[J]. Journal of Agricultural Big Data, 2019, 1(1): 5-15.
- [3] 中国共产党中央委员会办公厅, 中华人民共和国国务院办公厅. 数字乡村发展战略纲要[Z]. 2019-05-16.
General Office of the CPC Central Committee, General Office of the State Council of the People's Republic of China. Digital rural development strategy outline[Z]. 2019-05-16.
- [4] 阿兰·库珀. 交互设计之路[M]. 北京: 电子工业出版社, 2006: 115-135.
ALAN C. The road of interaction[M]. Beijing: Publishing House of Electronics Industry, 2006: 115-135.
- [5] 李佳慧, 赵刚. 基于大数据的电子商务用户画像构建研究[J]. 电子商务, 2019(1): 46-49.
LI H J, ZHAO G. Research on the construction of e-business user portrait based on big data[J]. E-Business Journal, 2019(1): 46-49.
- [6] 汪丹萍. 基于电信多维数据的房地产业用户分析及行业挖掘[D]. 北京: 北京邮电大学, 2018.
WANG D P. User analysis and industry mining of real estate industry based on Telecom multidimensional data[D]. Beijing: Beijing University of Posts and Telecommunications, 2018.
- [7] 单晓红, 张晓月, 刘晓燕. 基于在线评论的用户画像研究——以携程酒店为例[J]. 情报理论与实践, 2018, 41(4): 1-10.
SHAN X H, ZHANG X Y, LIU X Y. Research on user portrait based on online review: taking ctrip hotel as an example[J]. Information Studies: Theory & Application, 2018, 41(4): 1-10.
- [8] 王颖. 基于知识图谱的产品画像构建研究[D]. 南京: 南京理工大学, 2018.
WANG Y. Research on the construction of product portrait based on knowledge map[D]. Nanjing: Nanjing University of Technology, 2018.
- [9] 田娟, 朱定局, 杨文翰. 基于大数据平台的企业画像研究综述[J]. 计算机科学, 2018, 45(Z2): 58-62.
TIAN J, ZHU D J, YANG W H. Research on enterprise portraits based on big data platforms[J]. Computer Science, 2018, 45(Z2): 58-62.
- [10] 滕致遥. 基于时空数据的城市行为可视画像[D]. 成都: 电子科技大学, 2018.
TENG Z Y. Visual image of urban behavior based on spatiotemporal data[D]. Chengdu: University of Electronic Science and Technology, 2018.
- [11] 马亚雪, 李纲, 谢辉, 等. 数字空间视角下的城市数据画像理论思考[J]. 情报学报, 2019, 38(1): 58-67.
MA Y X, LI G, XIE H, et al. Theoretical thinking on city profile from the perspective of digital space[J]. Journal of the China Society for Scientific and Technical Information, 2019, 38(1): 58-67.
- [12] 刘海鸥, 孙晶晶, 苏妍娜, 等. 国内外用户画像研究综述[J]. 情报理论与实践, 2018, 41(11): 155-160.
LIU H O, SUN J J, SU Y Y, et al. Literature review of persona at home and abroad[J]. Information Studies: Theory & Application, 2018, 41(11): 155-160.
- [13] 张小可, 沈文明, 杜翠凤. 贝叶斯网络在用户画像构建中的研究[J]. 移动通信, 2016(22): 22-26.
ZHANG X K, SHEN W M, DU C F. Research on bayesian network in user portrait construction[J]. Mobile Communications, 2016(22): 22-26.
- [14] GAUCH S, SPERETTA M, CHANDRAMOULI A, et al. User profiles for personalized information access[J]. The Adaptive Web, 2007: 54-89.
- [15] HUANG Z. Extensions to the k-means algorithm for clustering large data sets with categorical values[J]. Data Mining & Knowledge Discovery, 1998, 2(3): 283-304.
- [16] ZHANG H Z, QIU B J, GILES C L, et al. An LDA-based community structure discovery approach for large-scale social

- networks[C]// 2007 IEEE Intelligence and Security Informatics, May 23-24, 2007, New Brunswick, USA. New York: IEEE, 2007.
- [17] RISH I. An empirical study of the naive bayes classifier[C]// IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence, August 4-10, 2001, Seattle, USA. San Francisco: Morgan Kaufmann, 2001, 3(22): 41-46.
- [18] SEVERYN A, MOSCHITTI A. Learning to rank short text pairs with convolutional deep neural networks[C]// The 38th International ACM SIGIR Conference, August 9-13, 2015, Santiago, Chile. New York: ACM Press, 2015: 373-382.
- [19] QUINLAN J R. Induction of decision trees[J]. Machine learning, 1986, 1(1): 81-106.
- [20] QUINLAN J R. C4.5: programs for machine learning[M]. San Francisco: Morgan Kaufmann, 1993.
- [21] AGRAWAL R S R. Fast algorithm for mining association rules[C]// The 20th International Conference on Very Large Databases (VLDB), September 12-15, 1994, Santiago, Chile. San Francisco: Morgan Kaufmann, 1994: 487-499.
- [22] HAN J, PEI J, YIN Y. Mining frequent patterns without candidate generation[C]// The 2000 ACM SIGMOD International Conference on Management of Data, May 15-18, 2000, Dallas, USA. New York: ACM Press, 2000: 1-12.
- [23] 高广尚. 用户画像构建方法研究综述[J]. 数据分析与知识发现, 2019, 3(3): 25-35.
- GAO G S. A survey of user profiles methods[J]. Data Analysis and Knowledge Discovery, 2019, 3(3): 25-35.
- [24] 闫周府, 吴方卫. 从二元分割走向融合发展——乡村振兴评价指标体系研究[J]. 经济学家, 2019(6): 90-103.
- YAN Z F, WU F W. From binary segmentation to convergence development: a study on the evaluation index system of rural revitalization[J]. Economist, 2019(6): 90-103.
- [25] 贾晋, 李雪峰, 申云. 乡村振兴战略的指标体系构建与实证分析[J]. 财经科学, 2018(11): 70-82.
- JIA J, LI X F, SHEN Y. Indicator system construction and empirical analysis for the strategy of rural vitalization[J]. Finance & Economics, 2018(11): 70-82.
- [26] 张挺, 李闽榕, 徐艳梅. 乡村振兴评价指标体系构建与实证研究[J]. 管理世界, 2018, 34(8): 99-105.
- ZHANG T, LI M R, XU Y M. The construction and empirical study of the rural revitalization evaluation index system[J]. Management World, 2018, 34(8): 99-105.
- [27] 廖彩荣, 陈美球. 乡村振兴战略的理论逻辑、科学内涵与实现路径[J]. 农林经济管理学报, 2017, 16(6): 795-802.
- LIAO C R, CHEN M Q. The theoretical logic, scientific connotation and achieving methods of rural revitalization strategy[J]. Journal of Agro-Forestry Economics and Management, 2017, 16(6): 795-802.
- [28] 刘合光. 乡村振兴战略的关键点、发展路径与风险规避[J]. 新疆师范大学学报(哲学社会科学版), 2018, 39(3): 25-33.
- LIU H G. The key points, development path and potential traps in promoting rural vitalization strategy[J]. Journal of Xinjiang Normal University (Edition of Philosophy and Social Sciences), 2018, 39(3): 25-33.
- [29] 周思悦, 申明锐, 罗震东. 路径依赖与多重锁定: 经济地理学视角下的乡村建设解析[J]. 经济地理, 2019(6): 183-190.
- ZHOU S Y, SHEN M R, LUO Z D. Path dependence and multiple lock-in process in rural construction: an economic geography perspective[J]. Economic Geography, 2019(6): 183-190.
- [30] 郭晓鸣, 张克俊, 虞洪, 等. 实施乡村振兴战略的系统认识与道路选择[J]. 农村经济, 2018(1): 11-20.
- GUO X M, ZHANG K J, YU H, et al. Systematic understanding and road choice of implementing the strategy of rural

- revitalization[J]. Rural Economy, 2018(1): 11-20.
- [31] 张强, 张怀超, 刘占芳. 乡村振兴: 从衰落走向复兴的战略选择[J]. 经济与管理, 2018, 32(1): 6-11.
ZHANG Q, ZHANG H C, LIU Z F. Systematic understanding and road choice of implementing the strategy of rural revitalization[J]. Rural Economy, 2018, 32(1): 6-11.
- [32] 陈国磊, 张春燕, 曾菊新, 等. 中国“一村一品”的空间分异格局研究[J]. 经济地理, 2019, 已录用.
CHEN G L, ZHANG C Y, ZENG J X, et al. Spatieal differentiation patterns of “one village one product” in China[J]. Economic Geography, 2019, Accepted.
- [33] 黄祖辉. 准确把握中国乡村振兴战略[J]. 中国农村经济, 2018(4): 2-12.
HUANG Z H. On the strategy of rural revitalization in China[J]. Chinese Rural Economy, 2018(4): 2-12.

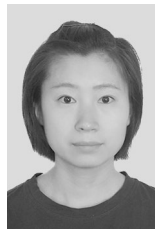
作者简介



李望月(1987-), 女, 博士, 国务院发展研究中心信息网高级项目咨询师, 主要研究方向为微观数据、非传统数据、数据挖掘、大数据商业价值、企业创新、产业竞争力、产业集聚、区域规划等。



刘瑾(1993-), 女, 国务院发展研究中心信息网金融部研究员, 主要研究方向为金融监管、资产定价、量化投资、金融科技、监管科技、企业画像、大数据和机器学习等。



陈娜(1989-), 女, 国务院发展研究中心信息网数据工程师, 主要研究方向为数据库建设、更新和维护、数据分析、数据质量管理、数据处理、数据清洗、数据产品策划。

收稿日期: 2019-08-28