

信息化研究

2017年第3期 总第118期

国家信息中心信息化研究部

2017年2月27日

“一带一路”大数据分析服务体系建设

于施洋 杨道玲 王璟璇 许婷婷

【摘要】 从定位、建设内容、服务流程与方法等方面对“一带一路”大数据分析服务体系进行了阐述，指出该服务体系旨在以专题分析、指数研究为主要内容，从需求反馈、业务研究、数据准备、数据挖掘与分析、结果展现等环节建设流程模型，采用关键词分析、情感分析、多源融合分析、预测性分析等方法，为政府和企业提供大数据分析支撑和服务。最后从拓展业务服务范围、创新分析维度和分析方法、加强大数据人才队伍建设等方面提出了有关建议。

【关键词】 一带一路；大数据；数据分析；服务体系

“一带一路”大数据分析服务体系建设¹

于施洋 杨道玲 王璟璇 许婷婷²

在“一带一路”战略推进过程中，必然会产生各种类型的海量数据，如何从海量、多源、异构的数据中发掘和释放潜藏的、具有决策价值的信息，显得尤为重要。“一带一路”大数据分析服务，主要是运用大数据分析技术及时、准确地追踪我国与沿线国家的“一带一路”合作进展，为政府、企业提供实时、预测性的数据支撑和咨询服务，促进“一带一路”相关配套政策措施和投资合作行为更具针对性和科学。

一、“一带一路”大数据分析服务的定位

（一）为政府“一带一路”宏观决策提供科学支撑

在大数据时代，管理者决策越来越依靠数据分析，而不是依靠经验和直觉。“一带一路”大数据分析旨在基于“一带一路”综合数据库归集的海量数据，应用大数据技术，为政府提供常态化的数据分析和预测服务，提高决策的科学性和政策的有效性。具体而言，“一带一路”大数据分析服务体系主要支撑政府四类决策需求：一是进展分析，通过海量数据分析及时跟踪各项工作进展，客观反映“一带一路”建设的成绩；二是态势研判，对沿线国家合作环境、重大事件或重大决策的社会关切等方面的发展形势进行深入分析，为制定对策提供参考；三是政策评价，对“一带一路”重大政策实施前的公众期待以及实施后的贯彻落实情况和社会反应情况、公众政策诉求等进行分析，提升政策制定的科学性；四是潜力预测，对与沿线国家“一带一路”合作的潜力进行监测预测，为制定中长期规

1 基金项目：国家发展改革委“一带一路”建设2016年专项课题“‘一带一路’发展指数编制思路研究”。原文刊载于《电子政务》杂志2017年第1期

2 于施洋（1974—），男，管理学博士，副研究员，现任国家信息中心信息化研究部副主任，兼任国家发改委大数据分析中心执行副主任、国家信息中心“一带一路”大数据中心执行副主任，中国信息协会电子政务专业委员会会长。研究方向：大数据与政府治理创新。

杨道玲（1979—），男，管理学博士，副研究员，国家信息中心信息化研究部电子政务研究室副处长，兼任国家信息中心“一带一路”大数据中心主任助理，中国信息协会电子政务专业委员会副秘书长。研究方向：政务大数据、电子政务战略规划与绩效管理。

王璟璇（1984—），女，管理学博士，助理研究员，国家信息中心信息化研究部大数据分析处副处长。研究方向：电子政务绩效评估、电子政务顶层设计、大数据分析。

许婷婷（1990—），女，政治学硕士，国家信息中心“一带一路”大数据中心数据分析师，研究方向：“一带一路”大数据。

划、深化合作提供参考。从2015年4月至今，国家信息中心“一带一路”大数据中心基于“一带一路”综合数据库，初步建立起“一带一路”常态化大数据监测分析和报送机制，陆续向推进“一带一路”建设工作领导小组办公室报送40份大数据分析报告，基本探索出一套相对成熟的“一带一路”大数据分析服务体系。

（二）为企业参与“一带一路”建设提供投资参考

“一带一路”沿线国家市场广阔且潜力巨大，但调研发现，我国大部分“走出去”的企业存在对海外市场的认知不够、合作规划较为盲目、风险意识缺乏等问题。大数据分析服务体系旨在为企业提供更细粒度的信息咨询服务，通过综合各种相关的数据进行大数据分析，为参与“一带一路”的企业提供更为详尽、实时的调查报告以及风险规避方案，进一步提高企业的市场拓展能力。“一带一路”大数据分析服务可满足企业“走出去”的三类需求：一是帮助企业明确合作方向。通过现时数据和历史数据的综合比较，整体了解目标国行业、企业的微观动态以及发展潜力。二是帮助发掘市场需求。结合企业业务拓展目标，建立多维数据源的综合分析，及时反馈竞争企业动态以及市场价格变化情况，预知市场发展趋势。三是有助于防范市场风险。通过大数据分析发现异常趋势和异常点，及时跟进目标国的政策法规变化，对投资合作风险进行有效预测。

二、“一带一路”大数据分析服务体系的核心内容

“一带一路”大数据分析服务体系包括专题分析和指数研发两大类服务。目前，专题分析主要从重点国别、重点区域、重要行业、重点省市、重点项目等五个方面分析“一带一路”建设的有关进展情况及未来发展潜力。指数研发重点通过设计综合指数和专项指数的方式，定量化、持续性地观测和评价“一带一路”进展。其中，综合指数包含国别、省市、企业三类指数，分别从宏观、中观、微观三个层面评价“一带一路”建设成效；专项指数从国际产能合作、外贸竞争力、对外投资、智库影响力、信息化发展水平等若干专项主题角度评价“一带一路”相关领域建设情况（参见图1）。

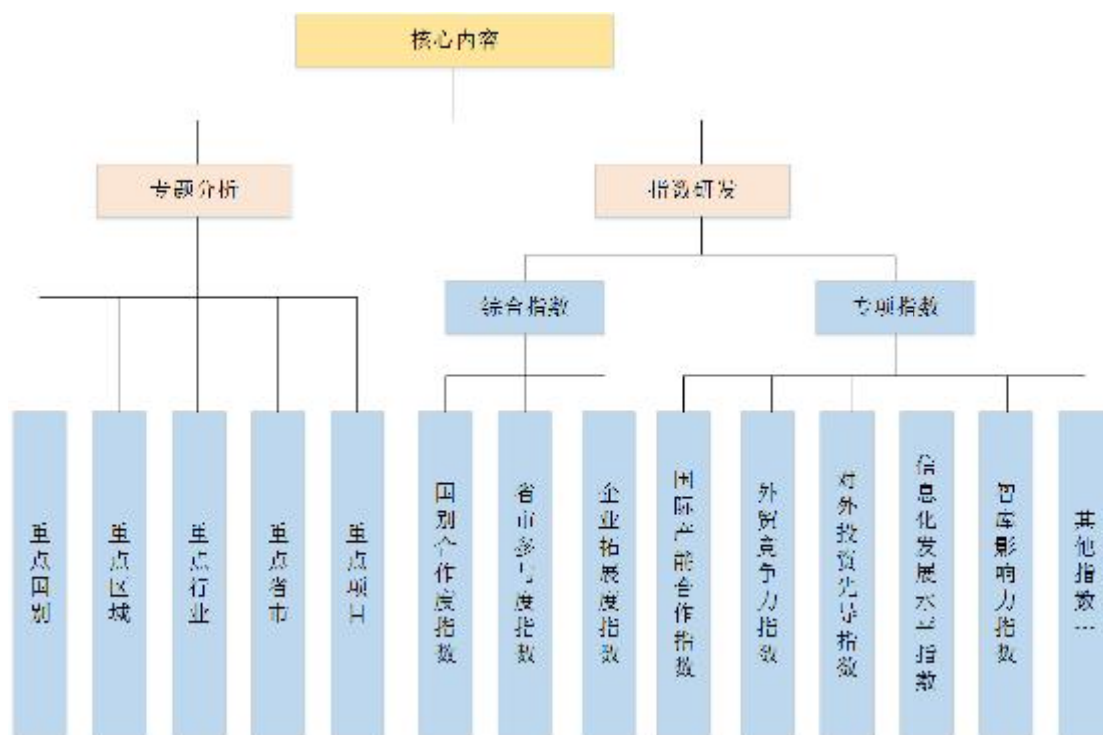


图1 “一带一路”大数据分析服务的重点

（一）专题分析

“一带一路”大数据专题分析服务基于海量数据分析，形成涵盖重点国别、重点区域、重点省市、重点行业、重点项目等五个类型的专题分析报告，确保政府和企业及时了解沿线国家和地区的合作前景、国内各省市的参与情况、产能行业的合作方向以及重大项目进展等情况，并做出有效决策。

1、重点国别、重点区域“一带一路”合作潜力分析

重点国别与重点区域分析旨在为政府“一带一路”决策部门出台相关投资政策、确定国别投资方向提供支撑，为企业“走出去”提供国别投资和风险防范指南。基于“一带一路”综合数据库，着重从如下几个方面进行数据挖掘与分析：

①重点国别与重点区域内网民对我国“一带一路”战略的关注趋势以及关注重点。一是运用大数据挖掘技术，分析监测期内国内外网民对我国与沿线国家合作的关注变化趋势；二是分析国内外媒体和网民对双方合作的情感倾向，并对网民的讨论内容进行特征抽取，发现潜在的合作价值与风险；三是将网民对合作领域的舆情数据与统计数据进行对比，寻找差距与潜力，为双方深化合作提供决策参考。

②根据我国与沿线国家经贸合作的历史发展规律，对未来投资环境进行预测。一

是分析近年来双边经贸合作的总额与增速数据，发现未来发展趋势；二是对比分析沿线国家与其他合作伙伴的合作特征，为我国对外经贸转型提供借鉴。③基于对现状和潜力的分析，提出深入推进我国与“一带一路”沿线国家合作的建议。

2、重点省市“一带一路”参与情况分析

对重点省市“一带一路”的参与情况分析，旨在为国家统筹“一带一路”建设和地方决策部门了解本地区“一带一路”进展、制定推进方案提供决策参考，同时为有关企业了解各省市“一带一路”建设现状、政策环境、资源禀赋等提供参考。主要分析内容包括：①分析省市“一带一路”战略实施的进展情况。通过分析地方政府与“一带一路”沿线国家在政策沟通、设施联通、经贸畅通、资金融通、民心相通等方面的进展情况，全面客观地评价各省市对接“一带一路”战略的成效与问题。②分析国内外媒体和网民对该省市与“一带一路”沿线国家合作的关注情况，获取网民的情感倾向以及评价观点，了解该省市参与“一带一路”建设的民情基础。③结合该省市的建设进展、民情基础、资源禀赋、发展目标等，提出该省市与沿线国家开展合作的具体建议。

3、重点行业对外合作现状与潜力分析

重点行业对外合作现状与潜力分析旨在为相关部门了解某行业领域对外合作现状、出台有针对性的经贸投资政策提供参考，为企业了解海外市场发展动态、发掘潜在市场、确定投资方向提供咨询服务。主要分析内容包括：①统计分析各行业贸易与投资数据，展现各行业对外合作发展的前景趋势；②观测分析国内外媒体和网民对各行业海外合作的关注分布特征，发掘对外合作的机遇与挑战；③比较研究各行业的国内发展与海外市场拓展，就存在问题提出针对性的解决方案。

4、重大项目海外建设进展跟踪分析

重大项目海外建设进展跟踪分析旨在为相关部门和企业就“一带一路”重点项目的总体建设情况及不同领域、不同行业的项目进展提供调研分析报告，为拓展投资方向提供参考。在分析过程中，基于“一带一路”综合数据库中的工程项目数据，根据建设时间将项目分为已建、在建和拟建三大类，结合项目金额、数量以及投资国别、投资领域等，分析我国对外投资的发展特点以及发展趋势。

（二）指数研发

“一带一路”建设是一项长期复杂的系统工程，建立科学的综合评价体系是

客观反映“一带一路”成效、指引未来发展的重要手段。随着“一带一路”战略的不断推进，相关评价体系也相继出现，经调研发现，目前国内关于“一带一路”的评价主要有两类：一类是专项领域指数，如上海航运交易所发布的“一带一路”货运贸易指数与“海上丝绸之路”运价指数、宁波航运交易所发布的海上丝路指数、中国人民大学国家发展与战略研究院发布的“一带一路”能源投资政治风险指数、中国银行发布的“一带一路”人民币汇率指数、中国电子科学研究院发布的“一带一路”信息化投资指数，等等。二是综合性评价，相比专项领域类评价，能更全面地反映“一带一路”进展。例如，北京大学海洋研究院提出的“一带一路”五通指数、中国城市竞争力研究会发布的“一带一路”最具竞争力城市评价体系等。总体而言，目前的评价体系还处于探索阶段，专项领域类评价在操作上相对容易，但其仅局限于某特定行业领域，难以反映“一带一路”建设全貌；综合性评价尽管力求覆盖全面，但由于“一带一路”建设涉及面广、数据可得性难度大，总体上还存在数据滞后、偏主观定性等问题，且主要针对沿线国家互联互通的考察，对国内各地方参与情况更是鲜有涉及。

为科学、全面地反映“一带一路”建设成绩，跟踪各项工作进展，及时发现存在的短板和问题，为未来战略持续实施提供参考，我们启动了“一带一路”系列指数研发工作，作为“一带一路”大数据分析服务体系的重要组成部分。

1、综合指数

综合指数主要是以国别、省市、企业为重点对象，从宏观、中观、微观三个层面对“一带一路”的进展进行客观、全面评价。其中，国别合作度指数围绕“一带一路”五大对外合作重点，从政策沟通度、设施联通度、贸易畅通度、资金融通度、民心相通度等五个维度，评价我国与“一带一路”沿线国家的互联互通及合作发展的成效；省市参与度指数按照“基础—行动—效果”三个层次构建评价模型，重点从政策环境、设施配套、经贸合作、人文交流、综合影响等五个维度，评价各省“一带一路”战略参与情况及实施效果；企业拓展度指数从国际合作条件、跨境运营成效、国际综合影响力等方面评价我国企业参与“一带一路”建设的情况与潜力，引导国内企业的海外拓展。2016年10月28日，国家信息中心“一带一路”大数据中心发布了“一带一路”国别合作度指数与省市参与度指数，对沿线64个国家和国内31个省区市进行了综合测评。“一带一路”综合评价指

数参见图 2。

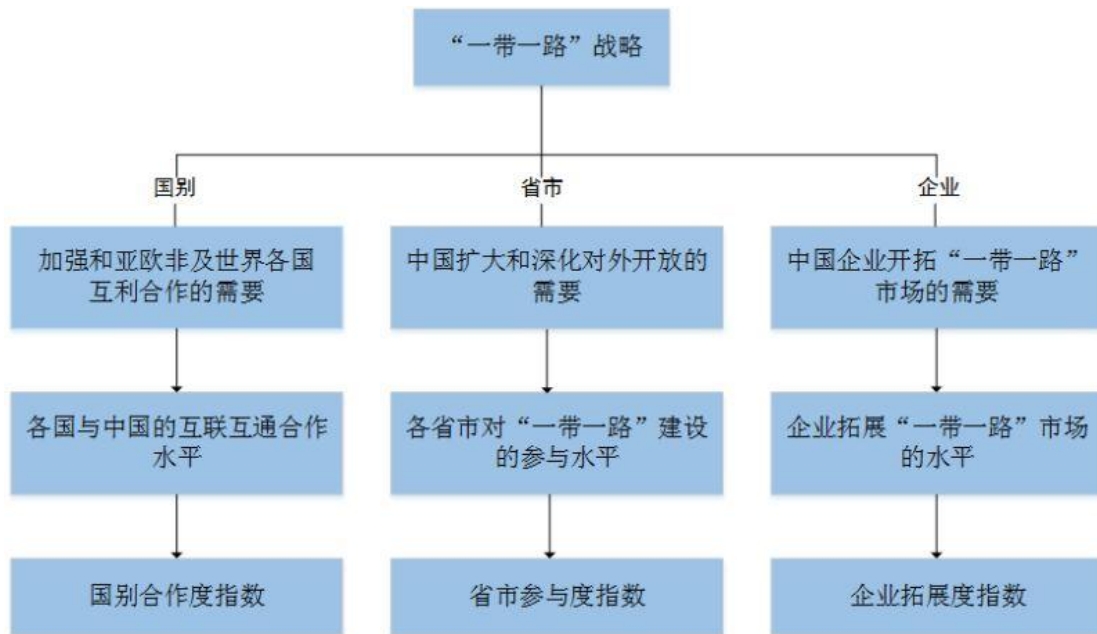


图 2 “一带一路”综合评价指数

2、专项指数

专项指数主要选取一些核心领域，评价我国与沿线国家互联互通的基础与合作情况，目前已经针对信息化发展水平、外贸竞争力、投资发展、产能合作、智库影响力等方面开发出一系列指数。例如，信息化发展水平指数是指从信息通信基础（ICT 基础）和信息通信应用（ICT 应用）两个维度对“一带一路”沿线 64 个国家进行测评，旨在了解沿线国家信息化发展水平；外贸竞争力指数，主要是运用全球海关进出口贸易数据对我国与“一带一路”沿线国家的外贸发展状况和贸易结构进行评价，以期科学推进我国外贸转型；对外投资先导指数，利用服务企业“走出去”的企业业务数据，对我国跨境企业的海外投资发展趋势进行评估和测评；国际产能合作指数，主要是国际产能合作重点领域的对外贸易与投资数据以及互联网数据，评价我国与沿线国家的国际产能合作水平与潜力；“一带一路”智库影响力指数，主要评价国内和海外“一带一路”智库的研究能力和影响水平。

三、“一带一路”大数据分析服务的流程与方法

（一）基本流程

“一带一路”大数据分析服务体系的主要流程包括需求反馈、业务研究、数据准备、数据挖掘与分析、结果展现等五个环节（参见图3）。

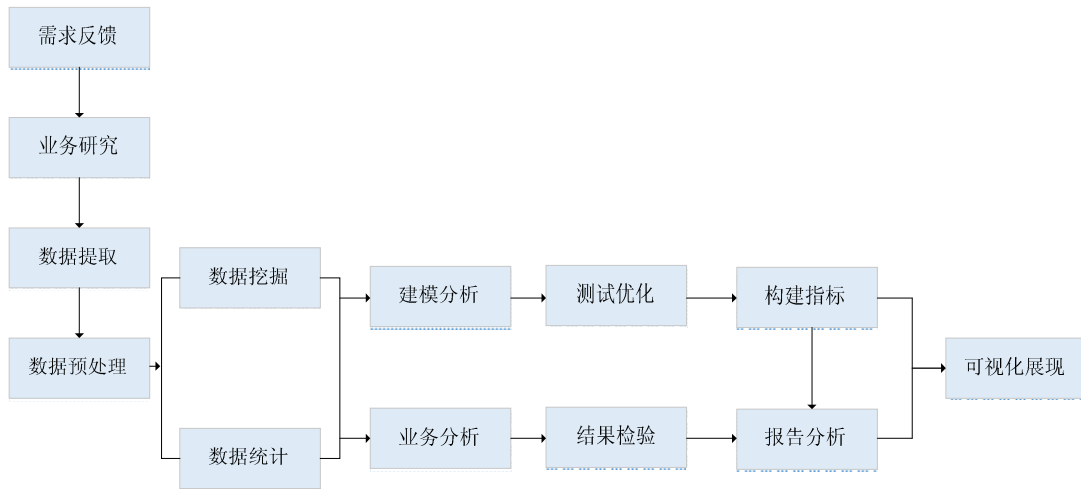


图3 “一带一路”大数据分析服务的基本流程

1、需求反馈

“一带一路”建设以政府和企业为参与主体，致力于建立和加强沿线各国间的互联互通，我国政府和企业需加强对沿线国家的深入了解，推动与沿线国家战略对接，发掘区域间的合作期待以及区域内的市场潜力。需求反馈主要通过以下两个步骤：一是积极建立与政府部门、企业的沟通交流机制，深入全面了解政府和企业的需求；二是梳理业务问题，准确定位需求本质，为业务研究和数据挖掘分析提供基础方向。

2、业务研究

针对业务需求内容和特征，对业务进行深层次研究，主要包括：①背景梳理。研究业务领域的文件，调研现状，确定需求的宏观背景以及发展环境。②确定业务分析框架。结合业务背景的分析结果，确定业务分析要件。③分解业务流程。对复杂业务流程进行逐层分解，并确定使用数据以及分析主题内容。

3、数据准备

在前期需求反馈和业务研究的基础上，从“一带一路”综合数据库中抽取需要的数据组成数据集合，为下一步数据挖掘分析做好数据准备工作。从综合数据库中抽取的数据主要包括：业务数据、国内统计数据、海外统计数据、国内互联网数据、海外互联网数据以及采购定制的数据。

4、数据挖掘与分析

对获取的数据集合进行预处理和深度分析挖掘，采用文本分析、主题聚类、关联分析、倾向性分析等方法进行数据处理；同时，结合业务研究，建立专门算法模型，挖掘数据背后反映的事件间的相关关系和规律性、趋势性；然后，将初步分析结果进行测试与检验，将业务需求与实际发展情况进行对比，保证相关规律既符合现实又具有启发意义。

5、结果展现

数据分析结果的呈现方式一般采用三种：分析报告、交互性大屏以及决策支持系统。分析报告主要以文档的形式呈现大数据分析结果，服务于“一带一路”相关政府部门和企业，简单、便捷、可操作性强；交互式大屏主要是融合图形图像、计算机图形技术和设计学等手段，以多维的形式将大数据所蕴含的有价值信息可视化展现出来，以人机交互的方式提高用户对数据的洞察能力；决策支持系统利用大数据分析和云计算技术，通过对各类“一带一路”数据信息进行集成展示和联动交互，及时响应系统终端的变量分析需求并以图形化、动态化的方式输出分析结果，展现终端方式可分为桌面终端和移动终端。

（二）主要分析方法和维度

“一带一路”大数据分析以互联网数据挖掘与分析为重心，辅之以经贸等传统统计数据的分析。在已有研究的基础上，结合政府和企业的业务需求特征，“一带一路”大数据分析服务体系采取的主要方法有关键词分析、情感分析、多源融合分析、预测性分析等。

1、关键词分析

关键词分析通过网页爬虫技术获取用户访问内容，利用网页分块技术、块和标题之间的语义分析进行正文提取，根据词表找到该主题下所有词及其频次分布，确定用户关注行为，一般包括时间搜索趋势、热度地域/行业、渠道关注等分析维度。①在时间搜索趋势方面，通过统计用户对某一关键词在搜索引擎中的搜索量以及搜索频率的时间变化趋势，获知在监测期间事件的发展态势，根据态势峰值分析引起网民关注的高峰事件，得出网民对事件的关注点随时间的变化过程。②在热度分布特征方面，主要从国家、行业等角度考察，分析国内各省市、沿线国家的媒体和网民关注热度分布特征，并就关注内容以及关注合作行业进行深入

分析。③在渠道关注方面，一方面统计新闻媒体、微博、微信、贴吧等渠道关键词的搜索量，分析各渠道的关注情况；另一方面，分析各渠道的行业、地域等属性。

2、情感分析

情感分析主要是如何识别和分析文本中包含的主观性情感信息并进行情感极性判定，通过文本特征抽取、构建情感词典，然后加总所有识别情感倾向值，从而得到整个文本的倾向。分析“一带一路”大数据的网民情感有助于促进政府决策支撑的情感认知，提高和拓展企业的营销能力。情感分析的内容以文本为主，从情感分析的流程看，主要有情感倾向的判定以及情感信息归纳。①判定情感倾向。通过自然语言处理技术对网民的评论信息进行文本预处理、情感信息提取，对情感极性进行分类。鉴于沿线国家语言种类繁多，语言的情感处理技术受限，情感极性一般分为积极、消极、中立三类。②情感信息归纳。在获得网民对“一带一路”相关专题的情感倾向之后，通过文本分析，归纳每种情绪的网民评论特征，进一步深化分析网民的情感内容。

3、多源融合分析

多源融合分析主要将不同渠道、多种抓取方式获取的不同数据结构的信息进行归纳、汇集与分析，基于多元表示原理、相关性原理、意义构建等支撑理论实现线上与线下数据的融合。“一带一路”战略的推进需要宏观与微观相结合，通过对统计数据、互联网数据以及业务数据等多源、异构的数据进行融合分析和交叉印证，得出相对全面、客观的研究结果。大数据有不同的呈现形式和不同的结构，代表不同的视角和领域，“一带一路”以政策沟通、设施联通、经贸畅通、资金融通、民心相通为五大工作重心，因此在数据分析的过程中，将不同领域、不同结构的数据汇聚进行关联分析，全面阐释“一带一路”相关因素的关联。最常见的方法为将互联网数据与统计数据相融合进行基本态势的研判。

4、预测性分析

针对“一带一路”建设，既需要及时了解“一带一路”的进展也需要对未来的发展趋势有清楚的认知，因此基于“一带一路”的属性信息和历史发展态势进行时间序列预测，涉及趋势分量预测、季节性分量预测、随机误差分量预测等。通过分析公众在网上的搜索记录以及舆情变化来洞察公众的选择倾向，预测社会

经济发展趋势以及行业合作潜力。①对于发展趋势的预测，通过抓取相关关键词在搜索引擎中的搜索情况以及舆情变化，观测人们在现实世界里的需求、兴趣和关注，例如宏观经济走势、具体经济活动运行等。②对于行业合作潜力的预测，通过观测沿线国家主要搜索引擎中对我国的行业类型的搜索变化，以及互联网文本中对期待合作的内容讨论，剖析我国与沿线国家的合作重点。

四、建议与展望

第一，深入研究服务对象需求，拓展业务服务范围。聚焦重点国别、重点区域、重点行业、重点省市的研究与分析，坚持国际国内、监测预测、统计与非统计相结合的原则，提高报告的质量和决策支持效果；积极拓展面向“一带一路”重点成员单位和地方“一带一路”决策部门的服务。同时，积极开展面向企业的决策支持服务，重点以企业“走出去”服务为主线，为大中型金融、投资类企业等提供数据和咨询服务。

第二，创新分析维度和方法，强化大数据分析的技术研发。不断归纳总结历史业务需求特征，丰富和创新多维度、多视角的数据分析方法，形成系列决策服务和咨询品牌；继续推动“一带一路”的基础性研究，构建与完善重点指标体系，科学评估“一带一路”的进展与成效；强化数据建模、数据挖掘、数据展现的技术研发，提升数据分析的准确性与高效性，提高大数据分析的处理能力。

第三，优化大数据分析团队的知识结构，加强人才队伍建设。储备计算机科学与技术、统计学、情报学、空间信息与数字技术、国际政治等方面的人才，优化具备社会科学与自然科学双重知识背景的综合人力资源结构；定期进行大数据分析方法与技术的培训，加强与专家、智库、高校的交流与合作，提高团队对“一带一路”相关大数据分析业务的分析能力。

参考文献：

- [1] Grobelink M. Big-data computing: Creating revolutionary breakthroughs in commerce, science and society [N/OL]. 2012-10-02.
- [2] 上海航运交易所发布“一带一路”航运指数[EB/OL].
http://finance.ce.cn/rolling/201507/29/t20150729_6077169.shtml, 2015-7-29.
- [3] “海上丝路指数”亮相伦敦 中国航运指数首次走出国门[EB/OL].

- http://news.ifeng.com/a/20151023/45976282_0.shtml, 2015-10-23.
- [4] 国内首个能源对外投资政治风险指数发布
[EB/OL]. http://cn.chinagate.cn/news/2016-01/18/content_37598891.htm, 2016-1-18.
- [5] 中行发布首个“一带一路”人民币汇率指[EB/OL].
http://news.xinhuanet.com/ttgg/2015-12/03/c_1117341760.htm, 2015-12-3.
- [6] “一带一路”信息化投资指数(2016版)正式发布[EB/OL].
http://news.ifeng.com/a/20160811/49758750_0.shtml, 2016-8-11.
- [7] 中国经济年鉴(一带一路卷). 北京: 中国经济年鉴社. 2015.
- [8] 2015 中国一带一路最具竞争力城市排行榜名单[EB/OL].
<http://www.askci.com/news/data/2015/06/25/21165h119.shtml>, 2015-6-25.
- [9] 罗海艳,《移动用户网络行为分析与预测方法研究》[D], 沈阳农业大学, 2015.
- [10] Das S R, Chen M Y. Yahoo! for Amazon: Sentiment extraction from small talk on the web [J]. Management Science, 2007, 53(9):1375-1388.
- [11] 李广建,《大数据分析情报关系辨析》[J], 中国图书馆学报, 2013, 213:14-21.
- [12] 化柏林、李广建,《大数据环境下多源信息融合的理论与应用探讨》[J], 图书情报工作, 2015, 59(16): 5-10.
- [13] 苏志、张骐,《基于网络大数据的社会经济预测综述》, 情报杂志, 2015, 07:18-21.

本期责编: 张岳

审稿: 于凤霞

联系地址: 北京市西城区三里河路 58 号 100045 国家信息中心信息化研究部

联系电话: 010-68557221