

基于大数据架构的国家海洋创新指标及其预测方法研究

刘大海, 李晓璇, 王春娟, 李先杰

(国家海洋局第一海洋研究所 青岛 266061)

摘要: 基于国家创新体系理论和大数据思想, 探讨国家海洋创新体系定义和范围; 根据国家海洋创新体系的具体内容, 选取对外海洋技术依存度、海洋仪器设备国产化率、海洋关键技术自主化率、海洋科技进步贡献率、海洋科技成果转化等作为国家海洋创新预测指标; 在 Bigtable+MapReduce 框架内讨论适用于海洋创新指标的预测方法, 探索国家海洋创新指标预测技术思路, 为建立支撑海洋创新战略制定的预测体系做了有益尝试。

关键词: 海洋创新指标; 大数据架构; 预测方法

中图分类号: F204 ; P7 文献标志码: A 文章编号: 1005-9857(2016)03-0009-05

On the National Marine Innovation Index and Marine Innovation Forecasting Method Based on Big Data Architecture

LIU Dahai, LI Xiaoxuan, WANG Chunjuan, LI Xianjie

(The First Institute of Oceanography, SOA, Qingdao 266061, China)

Abstract: Based on the theory of National Innovation System and the theory of big data, this paper explored the definition and scope of national marine innovation system. According to the specific contents of national marine innovation system, it put forward the national marine innovation forecasting indexes including foreign dependence of marine technology, marine equipment nationalization rate, the autonomization rate of marine key technology, the contribution rate of marine science and technology progress, marine science and technology transfer rate and so on. In addition, in the framework of Bigtable+MapReduce, it discussed the forecasting methods that were suitable for marine innovation index and explored the thought of national marine innovation index forecasting technology, which made attempts to establish the forecasting system for better marine innovation strategy-making.

Key words: Marine innovation index, Big data architecture, Forecasting method

基金项目: 海洋公益性行业科研专项经费项目“海洋强国建设的评价体系研究及应用”(2014418029); 国家海洋局项目“海洋科技创新评估与预测研究”(A201547); 基本科研业务费专项资金项目“海洋科技创新战略研究”(GY0214T08); 基本科研业务费专项资金项目“我国海洋创新评估体系及预测技术研究与应用”(2015T09); 国家海洋局项目“海洋科技创新指数及企业创新能力研究”。

作者简介: 刘大海, 助理研究员, 博士, 研究方向为海洋创新政策研究, 电子信箱: liudahai@fio.org.cn

通信作者: 李晓璇, 硕士研究生, 研究方向为海洋创新政策研究, 电子信箱: lixiaoxuan@fio.org.cn

党的十八大将创新驱动作为国家发展的主要目标,提出“实施创新驱动发展战略,到2020年进入创新型国家行列”,《中共中央关于全面深化改革若干重大问题的决定》明确提出要“建设国家创新体系”。海洋创新是建设创新型国家的关键领域,也是国家创新体系的重要组成部分。习近平总书记在中共中央政治局第八次集体学习时强调要“着力推动海洋科技向创新引领型转变”,突出体现了中央对海洋创新领域的高度重视。

为贯彻落实党和国家的重要战略部署,国家海洋局积极推进国家海洋创新工作。在工作扎实推进的同时,也遇到一些困难:①尚未形成系统的海洋创新体系,难以深入分析我国海洋创新发展的现状和问题;②尚难实现对我国海洋创新领域发展趋势的定量化预测;③缺少对未来海洋创新战略重点的布局。出现问题的原因在于:海量多源的海洋创新数据、难以预测的海洋创新行为、高度复杂的海洋创新环境,使得传统海洋创新预测体系面临严峻挑战。当前,我国海洋创新数据应用局限于简单的报表,海洋创新领域预测仍主要依赖“专家咨询”和“问卷调查”等定性方法,缺乏对海洋创新大数据的深度挖掘,更缺乏系统的指标预测研究,给我国海洋创新领域发展趋势的定量化预测和未来海洋创新战略重点的有效布局造成一定的困难。

因此,有必要基于大数据思想,尽快探索构建海洋创新指标预测方法体系:以分布式数据存储系统Bigtable支持下的海量多源海洋创新数据为基础,以基于国家海洋创新体系理论确定的海洋创新指标为对象,在MapReduce计算框架下探索预测海洋创新领域未来发展的整体趋势、重点领域和关键技术的方法体系,为海洋创新总体规划和海洋科技政策中的预测信息提供方法支撑。

1 国家海洋创新体系的构建

国外对国家创新体系理论的研究始于20世纪80年代中期,随后,该理论在发达国家的政策部门和学术界的影响迅速扩大。Niosi等^[1]认为国家创新体系是指以促进本土科学技术创造为目标,由企业、大学、政府机关等主体相互作用构成的一个体

系。我国2006年出台的《国家中长期科学和技术发展规划纲要(2006—2020年)》将国家创新体系界定为:以政府为主导、充分发挥市场配置资源的基础性作用、各类科技创新主体紧密联系和有效互动的社会系统。

将国家创新体系理论引入海洋领域,国家海洋创新体系^[2]是指通过政府、科研机构、高校、企业等创新主体相互合作,进行促进全社会海洋创新资源合理配置和有效利用的活动,涵盖涉海的科学研究、人才培养、产业发展、创新服务与管理等一系列内容的系统。也就是说,其范围主要包括:①创新主体,包括涉海科研机构、高等学校和企业;②创新活动,包含海洋科技成果、结题课题、论文、专利、企业效益、社会效益、环境效益等;③创新环境,涵盖宏观社会制度环境与微观个体观念环境。

2 国家海洋创新指标研究

海洋创新是国家创新的关键领域,有效评估海洋创新以反映国家海洋创新的发展问题,对于实现创新型海洋强国具有重要的战略意义和现实意义。海洋创新评估指标的选取是海洋创新评估工作的关键,对海洋创新评估的结果影响重大。

结合定义,为全面有效地反映海洋创新领域发展状况,根据国家海洋创新体系的具体内涵,选取拟预测的海洋创新指标见表1。

3 国家海洋创新指标预测方法研究

可将国家海洋创新指标预测方法分成传统海洋创新预测方法和基于大数据思想的预测方法。

3.1 传统海洋创新预测方法研究

传统海洋创新预测方法主要应用于传统指标预测和技术预见两方面^[3]。

在传统指标预测方面,通常运用趋势外推法进行预测^[4]。其预测需基于指标测算结果,指标具体包括海洋科技进步贡献率^[5]、海洋科技成果转化率^[6]和海洋科技投入产出效率^[7]。在测算方法上,科技进步贡献率主要运用时滞灰色生产函数^[5]和索洛余值法等^[6];科技成果转化率主要运用综合评价法和主成分分析法等^[7-8];科技投入产出效率主要运用数据包络分析和柯布—道格拉斯生产函

数等^[9-10]。

表1 拟预测的海洋创新指标

海洋创新指标		指标解释
主体指标	对外海洋技术依存度	海洋技术引进经费与海洋研究与发展经费支出的比值
	海洋仪器设备国产化率	海洋技术引进后,国内该项技术生产件的数量与所有生产件数量的比值
	海洋关键技术自主化率	海洋领域具有自主知识产权设备的价值与设备总投资的比例
活动指标	海洋科技进步贡献率	在海洋经济各行业中,海洋科技进步的增长率在整个海洋经济的增长率中所占的比重
	海洋科技成果转化率	一定时期内成熟应用的海洋科技成果占全部海洋科技应用技术成果的百分率
	海洋科技投入产出效率	包括海洋科技综合效率、技术效率、规模效率和规模报酬等一系列的内容
环境指标	海洋技术成熟度	评估高新海洋技术可见度
	海洋知识产权保护力度	对涉海科研机构、高等学校和企业等海洋创新主体的海洋知识产权的保护力度
	当地研究与培训专业服务状况	能否从本地的世界级或国家级机构中获得专业研究和培训服务
	海洋战略性新兴产业集群发展状况	海洋战略性新兴产业相关活动和服务的集聚程度
新型指标	海洋创新研究前沿	海洋创新领域研究过程中出现的潜在研究方向,是研究的新趋势和新动态
	海洋创新研究热点	一定时期内,有内在联系的、数量相对较多的一组海洋创新领域论文所探讨的科学问题或专题
	海洋创新社会关注度	一定时期内,海洋创新相关关键词的网络搜索频率

在技术预见方面,主要方法是德尔菲法、情景分析和专家会议等^[11]。

传统海洋创新预测方法在实际工作中的局限性包括:

(1)定量难。我国海洋经济统计数据在完备性和衔接性上存在不足,无法为海洋创新指标的定量测算提供结构化数据支撑。

(2)契合度低。由于创新领域的特殊性,海洋创新数据大多呈暴发性无序增长,而传统预测方法只适用于有明显趋势性的数据。

(3)研究重点定位难。传统预测得出的结果仅显示相应数据的趋势性增长规律,无法确定海洋创新的研究前沿和研究热点,在对海洋创新发展战略提供决策服务方面有效性不足。

3.2 基于大数据思想的预测方法研究

与传统数据模式不同,大数据不再采用随机样本数据,而是全体数据。也就是说,大数据不用抽样调查方法,而是对所有数据进行分析处理,大量、高速、多样和价值被公认为大数据的四大特点。随着大数据时代的到来,海洋创新数据类型和数量量级均发生巨大变化,对传统的预测方法提出了挑战。为解决此类问题,一些新技术、新方法应运而生,其中 MapReduce 以其良好的扩展性、容错性和大规模并行处理的优势成为大数据处理领域的代表技术。MapReduce 可以定义为一种用于大规模数据集并行运算的编程模型,分为 Map(映射)和 Reduce(规约)两步。

(1)关于 Map 函数算法的研究。Map 函数将数据映射到预先定义好的群组成类,主流算法有:决策树, Bayes 分析、神经网络、支持向量机等^[12-15]。其中,神经网络主要处理数值型数据, Bayes 分析主要用于基于已知先验概率的情况下进行决策和推理,支持向量机更适用于小样本。决策树算法相比于以上算法,在处理非数值型数据上拥有显著优势,可以在相对短的时间内对大型数据源计算出可行且效果良好的结果。

(2)关于 Reduce 函数算法的研究。Reduce 的算法主要有聚类分析、主成分分析、BP 神经网络等^[16-18]。其中,主成分分析是将多个变量通过线性变换以选出较少个数重要变量, BP 神经网络能学习和存贮大量的输入-输出模式映射关系,聚类分析

在处理大量的、不完全的、含有噪声的数据中具有强大优势。

需要说明的是,实现 MapReduce 还需要 HBase、Cassandra、Bigtable 等非关系型数据库支持。其中, Bigtable 是 Google 开发的分布式数据存储系统,在适用性、扩展性和计算性能方面具有较强优势。

4 国家海洋创新指标预测技术思路

在对我国海洋创新发展状况深入分析的基础上,确定拟预测的海洋创新指标,收集整理海洋创新数据,构建 Bigtable 分布式数据存储系统+MapReduce 技术的大数据架构,对海洋创新指标的预测方法及其应用进行研究(图 1)。

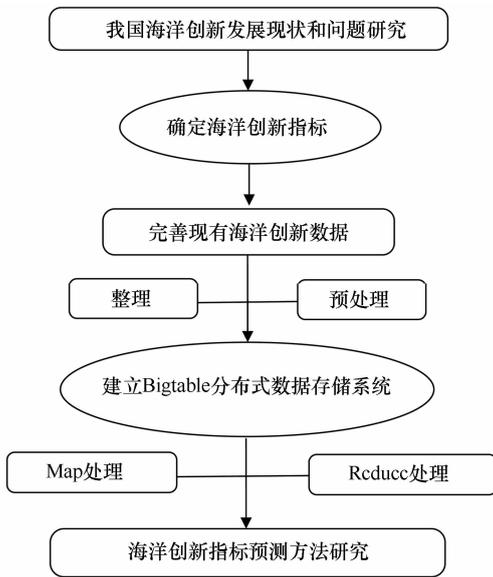


图 1 技术路线图

技术路线可分为 4 个阶段:①进行我国海洋创新发展现状和问题研究,定义国家海洋创新体系,确定拟预测的海洋创新指标;②收集海洋创新指标所需数据,完善健全现有数据;③对数据进行整理和预处理,建立 Bigtable 分布式数据存储系统;④综合运用合适算法对海洋创新数据进行 Map 处理和 Reduce 处理,形成完善的海洋创新指标预测方法体系。

技术路线的核心在基于大数据思想的海洋创新指标预测方法研究部分,具体分成两个方面:①数据收集整理与 Bigtable 数据库构建。收集海洋创

新指标所需数据,并对数据进行整理和预处理,建立 Bigtable 分布式数据存储系统,对海洋创新数据进行存储和管理;②基于 MapReduce 计算框架的指标预测方法研究。针对各类非结构化、非线性、无因果关系数据,基于 MapReduce 计算框架,形成一套系统的海洋创新指标预测方法,具体包括 Map 处理和 Reduce 处理两方面。运用设定决策树的最大高度来限制树的生长或者设定每个节点必须包含的最少记录数对决策树算法进行改进,并通过改进的决策树算法对海洋创新数据进行 Map 处理。根据数据类型不同,综合运用 BP 神经网络和不同类别的聚类算法,对海洋创新数据进行 Reduce 处理,为海洋创新指标的预测奠定基础。

Bigtable+MapReduce 架构下海洋创新指标预测最优方法的实现是研究的难点。具体表现为:如何基于大数据思想,运用 Bigtable 分布式数据存储系统有效地管理海量多源海洋创新数据?如何在 MapReduce 计算架构下综合运用合适算法对海洋创新数据进行 Map 处理和 Reduce 处理,实现并寻找到最优算法?若能实现,将能为海洋创新总体规划编制和海洋科技政策制定提供全面准确的预测信息,实现对我国海洋创新领域发展趋势的定量化预测和未来海洋创新战略重点的有效布局。

5 结束语

基于大数据架构开展国家海洋创新指标预测有双重意义。一是在学术方面,将 MapReduce 技术与海洋创新数据结合,探索海洋创新指标预测方法,可以建立起一套支撑海洋创新战略制定的预测体系。二是在应用方面,通过 Bigtable 分布式数据存储系统对海洋创新数据进行存储和管理,深入分析多源数据,进行科学合理的海洋创新指标预测方法研究,实现对我国海洋创新领域发展趋势的定量化预测,可以为海洋创新战略和政策的制定提供决策辅助服务。

参考文献

- [1] JORGE N, SAVIOTTI P, BELLON B, et al. National systems of innovation: in search of a workable concept[J]. *Technology in Society*, 1993(15): 207-227.
- [2] 刘曙光,丁丽君.海洋创新体系建设国际经验与借鉴[J]. *海洋*

- 开发与管理,2012,29(3):63—66.
- [3] Hajime Eto. The suitability of technology forecasting/foresight methods for decisionsystems and strategy A Japanese view [J]. Technological Forecasting & Social Change, 2003:231—249.
- [4] 王元地,潘雄峰,刘凤朝. 科技进步贡献率测算及预测实证研究[J]. 商业研究,2005(5):28—31.
- [5] 鲁亚运. 基于时滞灰色生产函数的我国海洋科技进步贡献率研究[J]. 科技管理研究,2014,32(12):55—59.
- [6] 赵蕾,林连升,杨宁生,等. 综合评价方法在中国水产科学研究院科技成果转化研究中的应用构想[J]. 科技管理研究,2011,29(6):42—45.
- [7] 吕晨,曾明彬. 基于 DEA 的中国区域科技投入产出相对效率研究[J]. 科学管理研究,2014,32(2):101—104.
- [8] 徐士元,何宽,樊在虎. 基于浙江面板数据的海洋科技进步贡献率研究[J]. 海洋开发与管理,2013,30(11):111—116.
- [9] 王黎明. 福建省科技成果转化评价研究[D]. 福州:福建农林大学,2014:33—34.
- [10] 孟庆军,许莲艳. 基于 C—D 函数的高新技术产业科技投入产出效率分析[J]. 河北工业科技,2015,32(1):17—21.
- [11] 中国未来 20 年技术预见研究组. 中国未来 20 年技术预见(续)[M]. 北京:科学出版社,2007:27.
- [12] 曹宁,高莹,徐根祺. 决策树方法的研究[J]. 科技视野,2014(20):72.
- [13] 祝翠,刘焕彬. Bayes 方法在学习效果评价上的应用[J]. 科技论坛,2014(1):138—139.
- [14] 孔玉静,华尔天. 基于神经网络的无线传感器网络数据融合算法研究[D]. 杭州:浙江工商大学,2013.
- [15] 丁世飞,齐丙娟,谭红艳. 支持向量机理论与算法研究综述[J]. 电子科技大学学报,2011,40(1):3—10.
- [16] 郭红建,陈一飞. 采用 K-means 聚类算法提高审计分析质量[J]. 中国管理信息化,2015(1):9—10.
- [17] 张杰. 基于主成分—聚类分析法的管道风险评价方法[J]. 油气储运,2014(2):139—143.
- [18] 崔东文. 多隐层 BP 神经网络模型在径流预测中的应用[J]. 水文,2013(1):68—73.