

UN Comtrade 数据共享平台设计与实现^{*}

叶思菁¹⁾ 宋长青^{1,2)†} 程昌秀^{1,2)} 高剑波^{1,2)} 高培超¹⁾ 胡洛铭²⁾

(1) 北京师范大学地表过程与资源生态国家重点实验室, 100875, 北京; (2) 北京师范大学地理科学学部, 100875, 北京)

摘要 UN Comtrade(United Nations international trade statistics database)是全球最大且应用最为广泛的国际贸易数据库,具有高权威性与完整性. 本文在框架体系结构设计 with 数据表结构设计基础上构建 UN Comtrade 数据共享平台,以期为地理学研究提供数据与工具支撑. 在数据聚合策略方面,平台通过综合集成数据爬取、加载模块并嵌套多种错误修正方法,实现 5 亿多条商品贸易记录的动态高容错聚合. 在检索策略方面,平台通过分区复合索引提升数据检索指令执行效率与可扩展性. 检索试验表明,平台能够在 80 用户并发模式下稳定执行不同类型检索指令,并且通过调用 ODBC/JDBC 接口将计算过程融入检索任务,可以更加有效利用服务器端资源并节省数据传输与读写耗时,具有效率更高、简化数据处理过程等优点. 2017 年平台被应用于中-美商品显性比较优势特征检索-计算-格网化表达-对比分析,案例表明平台具有高效、稳定的并发检索效率,以及高可扩展性等优点,可以为贸易特征计算与分析提供便捷快速、形式多样的数据共享服务.

关键词 国际贸易; 大数据; 格网; 显示性比较优势指数; 地理信息系统

中图分类号 K902

DOI: 10.12202/j.0476-0301.2021279

0 引言

当今全球化时代,各国通过签署双边贸易协定,以不同形式参与经济一体化,贸易成为影响现代社会政治、经济、科技、文化、民生与国家安全的重要因素. 双边商品贸易数据在分析国家间贸易竞争与合作态势以及解析全球生态系统服务流动格局等方面发挥越来越重要的作用. UN Comtrade 数据库(United Nations international trade statistics database)是全球最大且应用最为广泛的国际贸易数据库,具有高权威性与完整性. 每年全球超过 200 多个国家或地区分别以记录国(reporter)的角色向该数据库汇交它们与其他贸易伙伴国家(partner)的年度商品(或服务)贸易额分类统计数据. 汇交过程中,这些数据记录被基于联合国统计司统一标准规范执行分类编码(例如 HS、SITC、BEC 等)与核算. 发展至今,UN Comtrade 数据库存储记录总量超过 30 亿,记录日期可以追溯到 1962 年.

UN Comtrade(<http://comtrade.un.org>)为众多不同领域研究做出了重要贡献.

1) UN Comtrade 为丰富贸易理论,加强贸易系统规律及其驱动因素认知提供数据支撑. 例如: Fadeyi

等^[1]利用引力模型的泊松伪极大似然规范,评价自由贸易协定(FTA)对南非农业贸易的影响; Veninga 等^[2]基于 UN Comtrade 数据研究埃及国内政治不稳定对小麦贸易的影响. 此外,部分研究分析了 FTA 对欧洲农产品贸易特征^[3],拉丁美洲出口多样性^[4]的影响.

2) UN Comtrade 为国际贸易特征计量方法的发展提供了应用指导. 复杂网络分析方法与复杂性理论为揭示贸易关系的内在结构与演变,以及贸易伙伴之间相互依赖关系提供研究范式. 大部分相关研究是以国家(或地区)为节点,以贸易流为边,以贸易特征(例如贸易额、显示性比较优势指数)为权重,构建网络以表达特定类别产品在不同国家之间的贸易关系及其驱动因素^[5-8]. 例如: Dong 等^[5]构建小麦贸易竞争网络,分析气候变化对全球小麦贸易流动的影响,进而提出一个政策框架以促进小麦贸易环境的健康稳定; Ermann 等^[6-7]提出,贸易网络具有生态网络的典型特征,即高度嵌套性,并构建国家间多产品世界贸易的 Google 矩阵以分析贸易流特征. 针对网络分析中产品种类数受限问题,部分研究提出通过构建多层网络增加贸易产品数量^[8-10]. 另一方案是通过二分网络模型表达众多类别产品之间的关系^[11-12]. 其中代表性研究是 Hidalgo 等^[13]在 2007 年提出了“产品空

^{*} 第二次青藏高原综合科学考察研究资助项目(2019QZKK0608); 国家自然科学基金资助项目(42171250, 41901316)

[†] 通信作者: 宋长青(1961—), 博士, 教授. 研究方向: 地理区域综合研究、全球化与地缘关系等. E-mail: songcq@bnu.edu.cn

收稿日期: 2021-11-19

间”的概念,设置可以串联生产的产品(即“生产过程需要近似的生产部门、基础设施、物理因素、技术工艺等”的产品)在二分网络中的近邻分布.在此基础上,将国家和产品之间的联系纳入“产品空间”网络,评价国家生产结构的复杂性^[14-15].其他被广泛应用的计量方法还包括引力模型^[3, 16-17]、聚类模型^[18]等.

3) UN Comtrade 为描述全球贸易格局与变化过程提供重要支持.众多研究提出众多指数来评估不同国家间特定商品或产业链的贸易关系(例如比较优势、互补性、相似性、技术复杂性等).例如:Zheng 等^[19]计算并比较了 30 个国家新能源产品技术复杂度指数及其区域异质性;Cao 等^[20]评估了中国出口环境类产品技术复杂性的演变及其在国际产业价值链中的地位;Hao 等^[21]分析了铁矿石进口竞争格局、进口竞争区域和主要进口国的总体特征;Ye 等^[22]应用显性比较优势指数分析了中美商品贸易的竞争格局.

4) UN Comtrade 以其全球化视角被越来越广泛地应用于地理学研究,为全球生态保护^[21-23]、水-能源-食物-土地系统认知^[23-26]、污染防治^[27-28]、能源管理^[29-31]及国家安全^[32]等其他主题研究提供重要数据支撑.Moran 等^[33]计算了不同国家在各自贸易伙伴边界内的生态足迹;类似的方法也被用于分析 19 世纪美国和英国之间主要农产品贸易生态足迹^[25].为丰富全球电子垃圾的概念和政策,Lepawsky^[27]从 1996—2012 年超过 9400 份贸易报告中,量化了 206 个国家之间的电子垃圾贸易的规模 and 方向.Dalin 等^[26]将农产品贸易量与省级农产品虚拟水含量估算相结合,构建了中国国内外虚拟水贸易网络,并分析虚拟水流动格局及其节水效益.Meyfroidt 等^[24]提出评估国家植树造林工程及其对碳减排的贡献应综合考虑隐藏在农林产品贸易背后的森林砍伐与耕地开垦转移,并计算净木材贸易对国家植树造林碳减排效益的抵消比率.Chen 等^[29]利用 UN Comtrade 能源贸易数据,从相关性、多样性、国际关系影响等方面评估哈萨克斯坦和土库曼斯坦(出口国)、吉尔吉斯斯坦(进口国)的能源安全.

在数据共享方面,UN Comtrade 数据库集成开放了基于应用程序界面互操作和基于 JSON/XML 应用程序接口等不同形式的共享方案,但在实际共享应用中存在限制:首先,全球贸易研究往往涉及多国家间大量不同类型商品长时间序列进出口额数据记录,基于 web service 的数据共享方式随着查询结果增加而趋于不稳定;其次,共享方案中预先设定产品类型、年份、reporter、partner、商品/服务编码、贸易流种类等若干种检索因素,简化共享互操作的同时也限制了

检索条件的扩展与优化(例如用户难以在检索过程中集成分组、排序及聚合函数),并使得贸易特征计量模型难以与共享方案集成应用.这些不足在涉及大量地理要素时空格局与过程研究中表现得尤为突出.

针对上述问题,本文在设计框架体系结构与数据表结构基础上,基于 ORACLE 数据库管理系统构建 UN Comtrade 数据共享平台,以期对相关领域研究提供数据与工具支撑.平台基于 Python 3.6(集成 Scrapy 1.5.2)框架研发贸易数据自动化聚合组件,并通过分表策略与开放 ODBC 接口提升数据检索指令执行效率与可扩展性.实际应用表明,平台具有动态稳定更新、高并发检索效率、高可扩展性等优点,可以为贸易特征计算与分析研究提供更为便捷快速、形式多样的数据共享服务.需要说明的是,平台仅聚合年度双边商品贸易记录,不包含 UN Comtrade 数据库中服务贸易或月度贸易记录等其他数据集.

1 平台框架体系结构

考虑到平台低耦合、高灵活性与易维护性的要求,UN Comtrade 数据共享平台应用分层体系结构^[34-36],包括基础设施层、数据层与应用层,各层在下一层提供的功能与服务基础上构建,解决问题的一部分并为上一层提供功能与服务接口,如图 1 所示.由于双边贸易数据的结构化特征,基础设施层服务器端侧重纵向扩展,并应用万兆网络保障物理传输性能;数据层部署于 Linux 操作系统上,基于 Oracle 数据库管理系统构建混合存储模式,包括应用关系型数据表存储管理双边商品贸易数据集,并结合文件系统存储原始数据、备份文件与共享数据.此外通过开放不同形式的访问接口,SQL 功能可以被直接应用于设计检索指令;应用层由一个数据更新节点与若干应用客户端组成,数据更新节点基于 Python 3.6 & Scrapy 1.5.2 执行数据爬取—解析—核查—加载过程并根据数据层反馈指令动态调整计算任务,应用客户端兼容 JAVA、.NET 等不同计算环境,向数据层发送检索指令并接收相应查询结果数据.

2 数据表结构设计

平台数据表逻辑结构如图 2 所示,其中:

1) MTRADE_TRADEDATA 为核心数据表,每条记录表示在特定年份(YR),单一 reporter(RTCODE)与 partner(PTCODE)间对于某种商品(CMDCODE)基于特定贸易关系(RGCODE)的贸易额(TRADEVALUE)、贸易数量(QUANTITY)与商品净重(NETWEIGHT)等.例如,中国(RTCODE)与美国(PTCODE)间对于

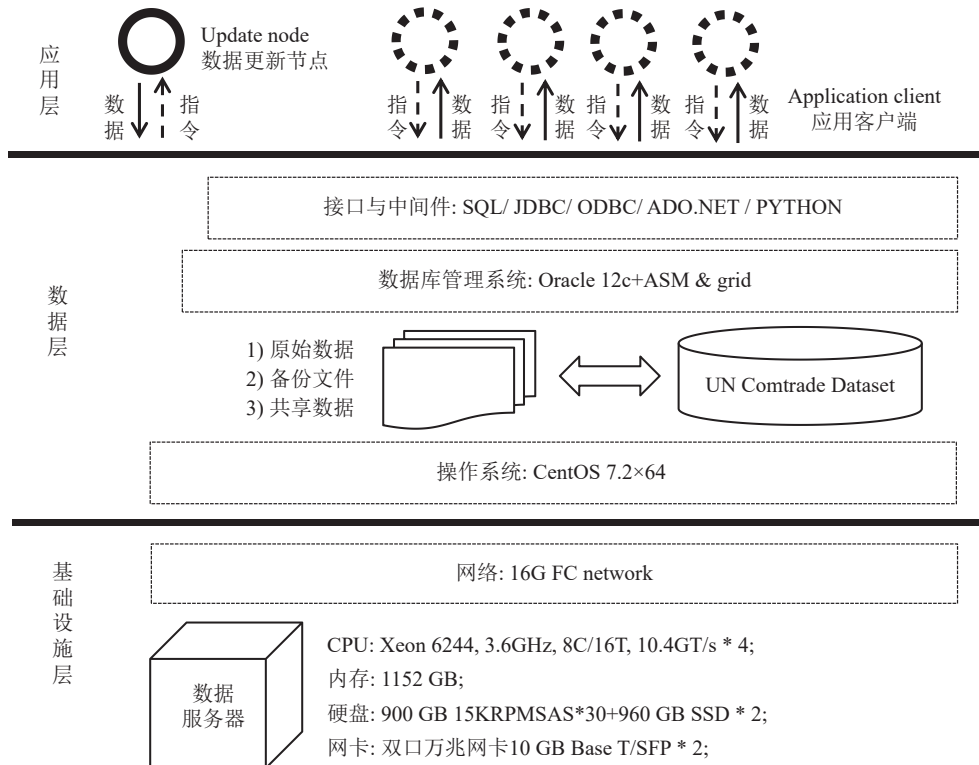


图 1 UN Comtrade 数据共享平台框架体系结构

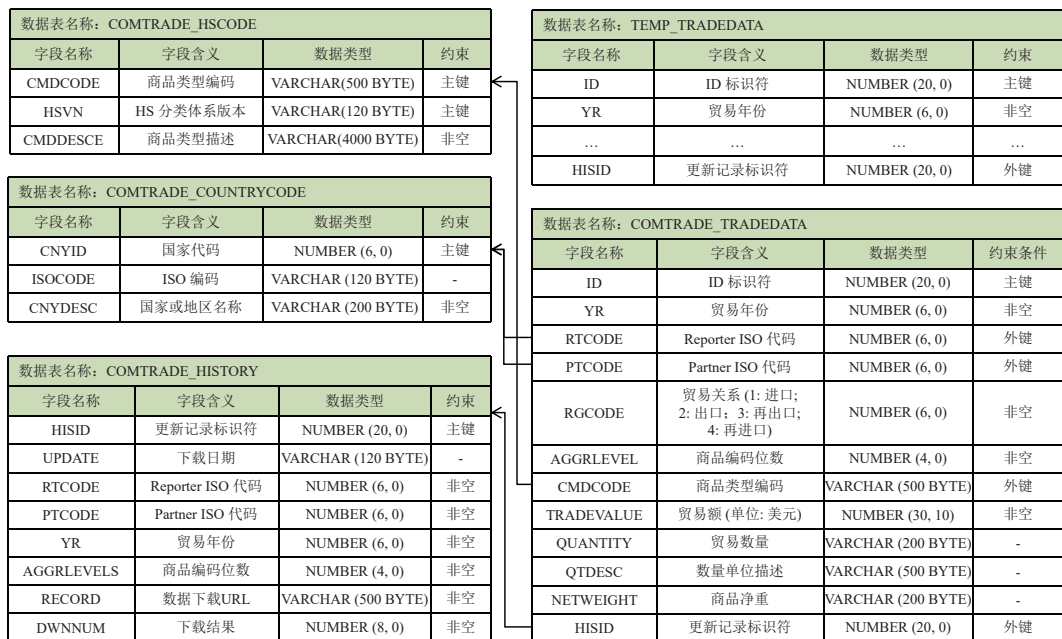


图 2 平台数据表逻辑结构设计

大豆(HS 4 位编码: 1201, CMDCODE)进口(RGCODE)的贸易额。

2)COMTRADE_COUNTRYCODE 与 COMTRADE_HSCODE 为外键表, 分别用于存储不同版本的 HS 商品种类编码信息与国家信息。

3)COMTRADE_HISTORY 用于存储贸易数据下载信息, 每条记录对应一次下载任务, URL 下载指令与下载结果数量分别存储于字段 RECORD 与 DWNNUM。

4)临时数据表TEMP_TRADEDATA 与COMTRADE_TRADEDATA 表结构一致, 下载记录则被解析、加载到临时数据表中, 经过检测无误后转移到核心数据表; 如存在错误则回滚下载过程, 自动修正下载指令并重新执行。

3 平台关键技术

3.1 数据聚合策略 针对 UN Comtrade 数据难以稳

定爬取的难题,数据共享平台通过综合集成数据爬取模块与数据加载模块实现海量商品贸易数据的动态稳定聚合,2 模块通过触发器松散耦合以降低聚合过程中突发错误的影响.数据爬取模块基于 Python Scrapy 爬虫框架研发.该框架具有高效、高鲁棒性等特点,可以快速抓取 Web 站点并从网页提取结构化数据,符合从 UN Comtrade 大批量获取数据的需求.实际爬取过程中,模块每小时自动触发 1 次,按照年份、reporter、partner、贸易流种类、商品分类尺度的顺序动态穷举检索条件,进而构造相应页面 URL(样例: <https://comtrade.un.org/api/get?max=50000&type=c&freq=a&px=hs&ps=1996&r=124&p=804&rg=all&cc=ag2,ag4,ag6>) 执行爬取任务;爬取成功后模块自动提取、解析网页数据记录,导入临时数据表并触发数据加载模块.数据加载执行数据完整性与乱码检测,对于通过检测的数据记录直接导入数据表,同时记录页面 URL 与处理日期;对于爬取失败或检测失败的任务则标识错误记录,触发修正程序,回滚指令并跳过当前任务.循环上述过程直至聚合全部数据.考虑到自动化聚合具有不稳定性,引起爬取失败、检测失败的原因众多(包括网络故障、数据编码错误、网络服务超时、检索结果数据量超标或无返回数据等),模块修正程序嵌套采用多种方式修正爬取错误,包括:1)重复执行爬取指令;2)拆分检索条件以减少反馈数据量;3)标识乱码并写入日志;4)将爬取指令写入缓存表,延迟重复执行;5)通过延迟发送爬取指令,保证每分钟最多执行一个爬取任务;6)对于修正失败的爬取任务则转为人工修正.此外,在上述过程中针对不同 reporter 通过随机抽样检验来测试自动化数据爬取的可靠性,测试表明,数据爬取与加载模块运行稳定且具有符合要求的自适应检错能力.在数据聚合效率检测方面,我们随机执行 5 组持续 10 min 的数据聚合试验,结果表明模块平均每 10 min 完成近 7 个爬取任务,聚合数据记录 10 700 条.截至 2021 年 7 月,该数据表汇聚记录 571 485 528 条.

3.2 数据高效检索策略 适用于多用户并发访问环境的高效检索机制是 UN Comtrade 数据共享平台的另一个挑战.由于双边商品贸易数据具有典型的结构化特征,数据共享平台未采用 NOSQL 存储技术,而是选择更适用于结构化数据存储管理的关系型数据表,并应用基于 ASM 镜像组织的 15 000 转机械硬盘构建适用于数据库文件高效读写的物理环境.针对数据表文件太大降低检索效率问题,平台针对核心数据表 COMTRADE_TRADEDATA 采用范围分区技术,以字段 YR 作为分区条件,将不同年份数据记录存储到不同物理子表中,并通过将子表文件分发存储到不

同的磁盘以均衡系统 IO 负载.各子表在逻辑上是一个整体,不影响前端应用复杂度.此外,由于数据检索需求主要以按年查询为主,并涉及 reporter 和 partner 这 2 类检索条件,平台以 YR、RTCODE、PTCODE 等为索引列构建本地分区复合索引,以期通过减少磁盘扫描提高检索效率.

在此基础上,我们在应用客户端基于 JMeter 5.2.1 JDBC 组件部署测试环境,分别模拟不同数量用户并发访问数据共享平台,发送数据检索指令并获取检索结果的全周期耗时与检索指令执行耗时(即仅执行检索任务,不包括数据提取、传输与读写过程),如表 1 所示.5 项数据检索指令分别面向:①单一分区表大量数据检索,②特定国家间或③特定产品长时间序列数据,以及④、⑤将贸易特征计算过程融入检索任务的集成指令.试验表明,虽然数据获取效率会因并发访问数量增加而降低,本文数据共享平台能够稳定反馈 80 用户并发检索指令,具有实用意义.对比指令①~③可见,查询条件对检索效率的影响显著高于检索结果记录数,指令①仅遍历 2016 年分区表,因此检索效率最高;指令②遍历记录数与指令③近似,但因查询条件为分区复合索引列而表现出显著较小的检索耗时,特别是其执行耗时在不同并发条件下相对稳定;由各指令执行过程中全周期耗时与执行耗时的差异特征可见,数据读写(包括提取、传输、写入)耗时是影响实际数据获取效率的重要因素,并且读写耗时会随着数据离散化分布程度增加而增加;此外,本文单节点测试环境在应用多线程模拟用户并发访问时,检索效率会受到磁盘读写速度、CPU 性能、网络传输速度等物理环境制约,数据共享平台在实际应用中可以通过横向扩展数据处理节点实现更高的并发指令处理性能.指令④检索过程覆盖了全部 HS4 位编码产品,相关遍历数据量约是指令③的 145 倍,并在检索过程中集成平均值计算,因此其执行耗时显著长于指令③,但由于计算过程减少了数据吞吐量,指令④全周期耗时小于指令③,这表明将计算过程融入检索任务,可以更加有效利用服务器端资源并节省数据传输与读写耗时,具有效率更高、简化数据处理过程等优点.指令⑤比较优势指数 $I'_{x,RCA}$ 计算式为

$$I'_{x,RCA} = \frac{P_{x,EX}^i / P_{all,EX}^i}{P_{x,EX}^{world} / P_{all,EX}^{world}}, \quad (1)$$

式中: $P_{x,EX}^i$ 表示一定时期内 i 国商品 x 的出口额; $P_{all,EX}^i$ 表示 i 国全部商品的总出口额; $P_{x,EX}^{world}$ 与 $P_{all,EX}^{world}$ 分别表示全球不同国家商品 x 的出口总额及全部商品的总出口额.

由式 1 可见,指令⑤集成 4 项查询任务及其简单

表1 数据共享平台并发检索效率对比

指令编号	指令描述	不同数量用户并发全周期耗时/执行耗时/s ¹⁾						检索结果 记录数/条
		1用户	5并发	10并发	25并发	50并发	80并发	
①	查询2016年中国全部HS6位编码商贸记录	33.5 / 20.2	49.3 / 19.5	44.0 / 19.3	64.4 / 19.6	107.6 / 20.1	131.7 / 22.7	521543
②	查询中-美1992—2018年HS4位编码商贸记录	87.6 / 24.8	310.3 / 32.8	419.5 / 39.7	1038.0 / 51.8	1906.0 / 80.5	2786.6 / 105.8	57352
③	查询全球不同国家1992—2018年大豆记录	94.7 / 51.9	299.7 / 116.6	534.4 / 200.9	1242.9 / 368.9	2643.0 / 583.1	3751.6 / 721.2	53296
④	查询计算全球不同国家1992—2018年HS4位编码商品进出口额平均值	101.7 / 93.4	239.3 / 228.4	429.2 / 415.8	987.7 / 971.1	1682.6 / 1653.7	2124.8 / 2113.6	520882
⑤	计算2017年中国HS6位商品比较优势指数	36.7 / 36.6	35.7 / 35.7	36.4 / 36.4	37.2 / 37.2	38.2 / 38.2	41.6 / 41.6	5002

¹⁾ JMeter Ramp-up period参数设为1 s.

计算, 由于不跨分区表查询并将计算任务集中在服务器端, 检索过程高效稳定.

4 平台应用

通过开放数据访问接口, UN Comtrade 数据共享平台可以为贸易特征计算与分析研究提供便捷快速、形式多样的数据共享服务. 在此基础上, 本文在应用层基于.NET研发并部署基于B/S架构的数据查询共享客户端V1.0, 由用户管理模块、数据共享模块、扩展模型库与系统设置模块组成, 如图3所示.

其中: 用户管理模块承担用户注册、权限管理、登录验证与用户信息管理等功能; 数据共享模块支持用户通过自定义SQL语句查询并下载数据; 扩展模型库中集成贸易特征指数计算表达模型, 并预先制备贸易显性比较优势指数(RCAI)计算参考模板. 计算表达流程如图3所示: 步骤1, 用户在参考模板基础上构建查询计算特定贸易特征指数的SQL指令; 步骤2, 用户向平台服务器端发送指令, 接收服务器端反馈的检索结果; 步骤3, 将贸易特征计算结果映射到贸易特征格网图实现空间化表达, 并以.shp格式输出.

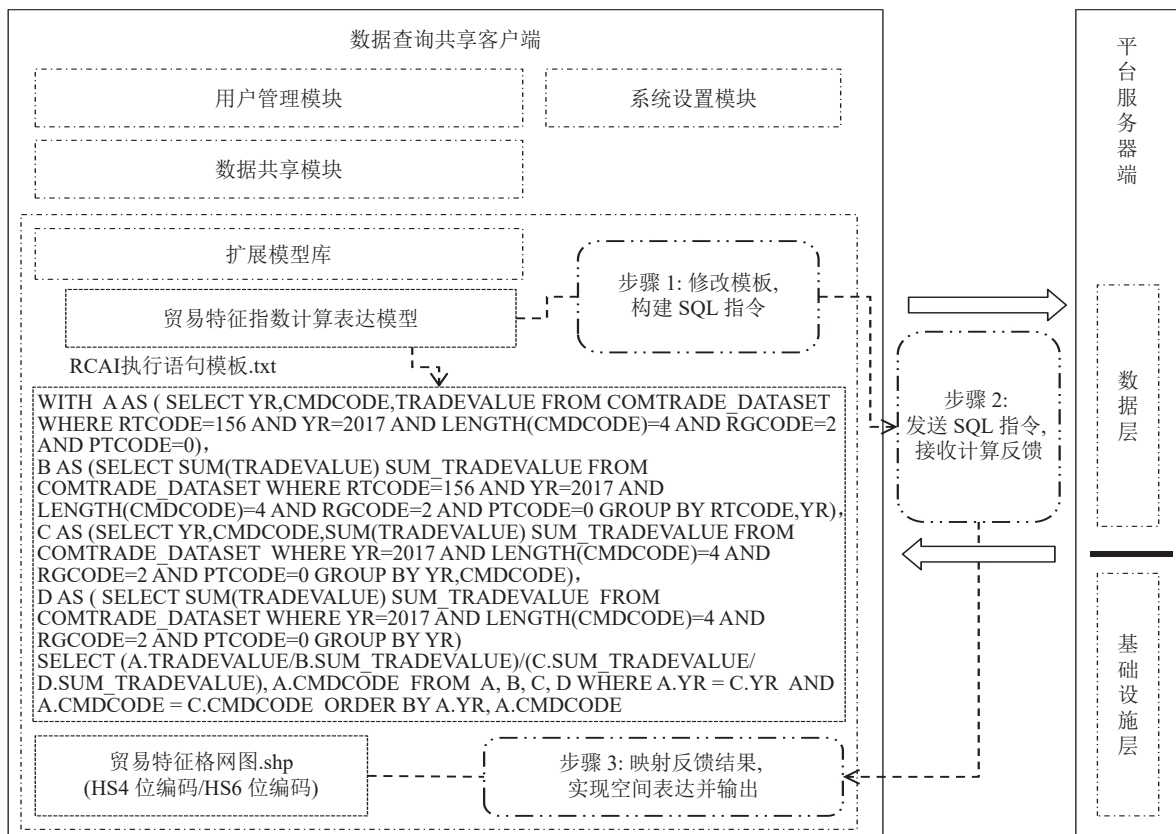


图3 数据查询共享客户端模块结构及其交互流程示例

本文应用客户端模型库贸易特征指数计算表达模型,分别查询计算 2017 年中国与美国 HS4 位编码商品 RCAI,并输出贸易特征格网图^[37],商品竞争力依据 RCAI 分为 4 级,如图 4 所示.贸易特征格网图通过逆向生成 Hilbert 曲线将一维商品类型编码序列映射到二维格网空间,建立商品类型与格网一一对应关系,以实现上千种商品贸易特征的空间化表达.在此基础上,相似种类的商品分布在邻近的格网,贸易特征分析可以综合考虑特定格网及其邻近格网的聚集性分布特征,空间分析模型可以被引入以丰富贸易特征集成分析的方法体系.对比分析图 4-a 与图 4-b,2017 年中国具有极强国际竞争力(4 级)的 HS4 位编码商品数量高于美国,且表现出更强聚集特征;中国 3、4 级商品主要聚集在纺织原料及纺织制品(HS2 位编码:50~63),鞋、帽、伞、杖等生活制品(64~67),石料、陶瓷、玻璃制品(68~70),钢铁制品(73),贱金属工具、器具制品(82、83),机械器具、电机电器设备及其零件(84、85),乐器、家具、玩具等生活娱乐用品(92、94~96),以劳动密集型制成品为主,处于产业链末端(获利低位);而美国 3、4 级商品则在化学工业及其相关产品(28~38),塑料及其制品(39),镍及其制品(75),光学、照相、电影、计量、检验、医疗或外科用仪器设备及零件(90),武器(93)以及艺术收藏品(97),多处于产业链上游,以原材料与技术密集型零部件为主,具备资源壁垒或技术壁垒.

此外,本文叠加中-美商品 RCAI 分级信息,将分

级叠加信息划分为“弱-弱”“弱-强”“强-弱”“强-强”4 种组合,以直观表达两国间商品出口优势的相互作用特征,如图 5 所示:中-美商品出口优势表现出显著互补特征,“强-弱”与“弱-强”组合分布广泛;竞争激烈(“强-强”关系)的商品数量较少,主要集中在有机或无机化学品,对于其他商品则多呈离散分布;竞争激烈的商品分布于“强-弱”组合商品邻近格网.此外,数据共享平台服务并不局限于数据查询共享客户端,用户还可以通过 Oracle 客户端直接访问数据表,也可以在 .NET、Java、Python、Matlab 等不同研发环境下定制数据检索 SQL 语句或研发数据检索-计算-表达应用程序.基于 UN Comtrade 数据共享平台的其他应用案例还包括双边贸易统计失衡评估^[38]、中美及其周边地区贸易特征分析^[37、39]、“一带一路”国家轨道交通产品的贸易格局及时空变化分析^[40]等.

5 结论与展望

5.1 结论 UN Comtrade 数据库是全球最大且应用最为广泛的国际贸易数据库,为众多不同领域研究做出重要贡献.针对 UN Comtrade 数据库在数据共享方式、数据检索接口方面的不足,在框架体系结构设计与数据表结构设计基础上,基于 Oracle 数据库管理系统构建 UN Comtrade 数据共享平台,以期为地理学研究提供数据与工具支撑.

1)平台基于 Python 3.6(集成 Scrapy 1.5.2)框架研发贸易数据自动化聚合组件,通过综合集成数据爬取

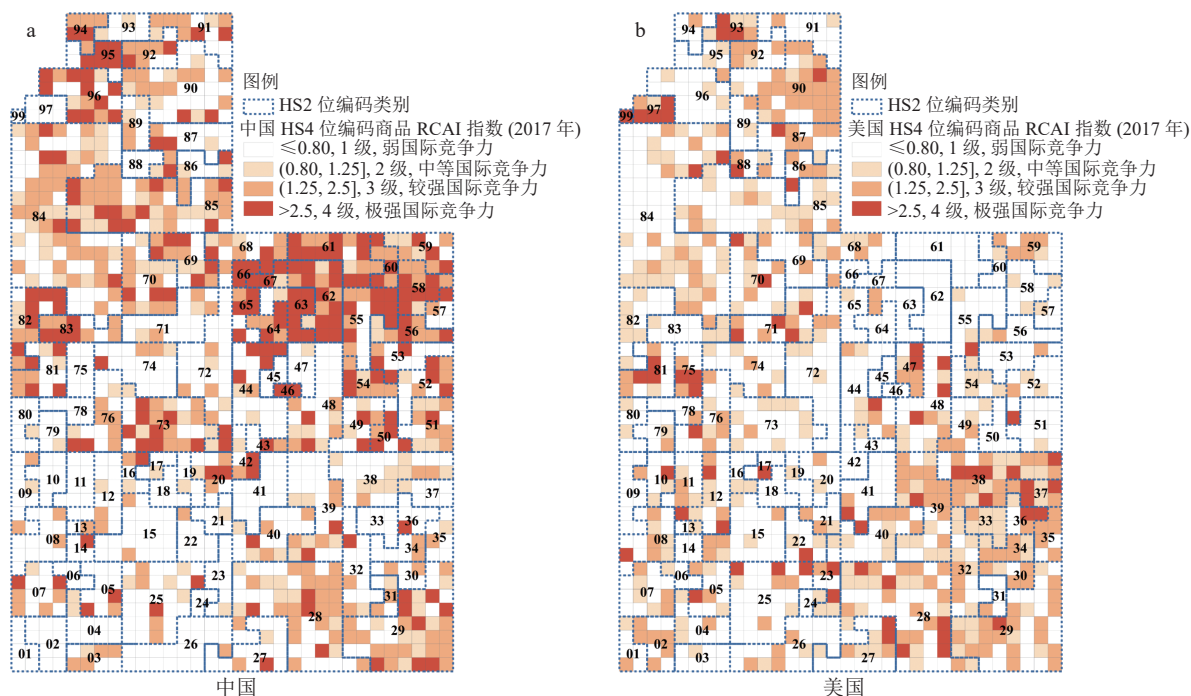
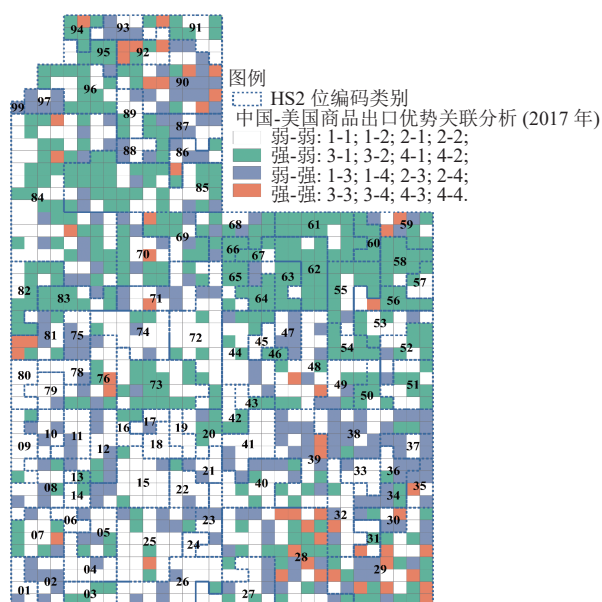


图 4 中国-美国 2017 年 HS4 位编码商品显性比较优势指数格网图



注: 以“1-1”为例, 前1个数字对应商品美国显性比较优势级别, 后1个数字对应商品中国显性比较优势级别, 显性比较优势分级与图4一致。

图5 中国-美国 2017 年 HS4 位编码商品显性比较优势关联分析

模块与数据加载模块并嵌套采用多种错误修正方法, 实现海量商品贸易数据的动态、稳定、高容错聚合;

2) 平台针对数据结构化特征, 通过范围分区、分区复合索引及开放 ODBC 接口提升数据检索指令执行效率与可扩展性。

在此基础上, 本文研发数据查询共享客户端, 并将其应用于商品显性比较优势特征检索-计算-表达, 试验表明平台具有高效、稳定并发检索效率, 高可扩展性等优点, 可以为贸易特征计算与分析研究提供更为便捷快速、形式多样的数据共享服务。

5.2 展望

1) 根据平台用户数量增加引起的检索效率降低问题, 横向扩展数据库节点并研究支撑高效数据检索-计算-传输的负载均衡模式;

2) 进一步完善数据共享平台内容, 构建双边服务贸易数据集、GDELT 数据集聚合与管理体系统;

3) 研究贸易特征分析方法, 扩展数据查询共享客户端, 将空间分析模型、复杂网络模型、聚类模型及深度学习模型等嵌入模型库并开放服务接口。

致谢: 感谢王艳东在平台建设中的帮助, 感谢北京师范大学地理科学学部地理数据与应用分析中心地学高性能计算平台支持 [https://gda.bnu.edu.cn/].

6 参考文献

[1] FADEYI O, BAHTA T, OGUNDEJI A, et al. Impacts of the

SADC Free Trade Agreement on South African agricultural trade[J]. The International Journal of Entrepreneurship and Innovation, 2014, 43(1): 53

[2] VENINGA W, IHLE R. Import vulnerability in the Middle East: effects of the Arab spring on Egyptian wheat trade[J]. Food Security, 2018, 10(1): 183

[3] MATKOVSKI B, RADOVANOV B, ZEKIC S. The effects of foreign agri-food trade liberalization in South East Europe[J]. Ekonomicky Casopis, 2018, 66(9): 945

[4] DINGEMANS A, ROSS C. Free Trade Agreements in Latin America since 1990: an evaluation of export diversification[J]. CEPAL Review, 2012, 108(108): 27

[5] DONG C, YIN Q, LANE K, et al. Competition and transmission evolution of global food trade: a case study of wheat[J]. Physica A: Statistical Mechanics and its Applications, 2018, 509: 998

[6] ERMANN L, SHEPELYANSKY D. Ecological analysis of world trade[J]. Physics Letters Section A: General Atomic & Solid State Physics, 2013, 377(3/4): 250

[7] ERMANN L, SHEPELYANSKY D. Google matrix analysis of the multiproduct world trade network[J]. The European Physical Journal B, 2015, 88(4): 84

[8] 陈小强, 袁丽华, 沈石, 等. 中国及其周边国家间地缘关系解析[J]. 地理学报, 2019, 74(8): 1534

[9] KIVEL M, ARENAS A, BARTHELEMY M, et al. Multilayer networks[J]. Journal of Complex Networks, 2014, 2(3): 203

[10] LIU X M, STANLEY H E, GAO J X. Breakdown of interdependent directed networks[J]. PNAS, 2016, 113(5): 1138

[11] LATAPY M, MAGNIEN C, VECCHIO N. Basic notions for the analysis of large two-mode networks[J]. Social Networks, 2008, 30(1): 31

[12] VIDMER A, ZENG A, MEDO M, et al. Prediction in complex systems: the case of the international trade network[J]. Physica A, 2015, 436: 188

[13] HIDALGO C, KLINGER B, BARABSI A, et al. The product space conditions the development of nations[J]. Science, 2007, 317(5837): 482

[14] HIDALGO C A, HAUSMANN R. The building blocks of economic complexity[J]. PNAS, 2009, 106(26): 10570

[15] MEALY P, FARMER J D, TEYTELBOYM A. Interpreting economic complexity[J]. Science Advances, 2019, 5(1): eaau1705

[16] 杜彬, 高培超, 宋长青, 等. 青海省出口贸易格局演化及其影响因素解析[J]. 北京师范大学学报(自然科学版), 2022, 58(1): 161

[17] MUPELA E, SZIRMAI A. Communication costs and trade in Sub Saharan Africa: a gravity approach[C]//

- BISSYANDÉ T, VAN STAM G. E-Infrastructure and e-services for developing countries. AFRICOMM 2013. Lecture notes of the institute for computer sciences, social informatics and telecommunications engineering, Springer: Cham, 2014: 135
- [18] DE F, DE S, VIEIRA J. Evaluation of logistic performance indexes of Brazil in the international trade[J]. RAM. Revista de Administração Mackenzie, 2015, 16: 213
- [19] ZHENG H, WANG Z. Measurement and comparison of export sophistication of the new energy industry in 30 countries during 2000–2015[J]. Renewable and Sustainable Energy Reviews, 2019, 108: 140
- [20] CAO X, HANSON-RASMUSSEN N. Dynamic change in the export technology structure of China's environmental goods and its international comparison[J]. Sustainability, 2018, 10: 3508
- [21] HAO X, AN H, SUN X, et al. The import competition relationship and intensity in the international iron ore trade: from network perspective[J]. Resources Policy, 2018, 57(1): 45
- [22] YE S, CHENG C, SONG C, et al. Visualizing bivariate local spatial autocorrelation between commodity revealed comparative advantage index of China and USA from a new space perspective[J]. Environ Plan A Econ Sp, 2021, 53(2): 223
- [23] PETERS G, MINX J, WEBER C, et al. Growth in emission transfers via international trade from 1990 to 2008[J]. Proceedings of the National Academy of Sciences, 2011, 108: 8903
- [24] MEYFROIDT P, RUDEL T K, LAMBIN E F. Forest transitions, trade, and the global displacement of land use[J]. PNAS, 2010, 107(49): 20917
- [25] HORNBORG A. Footprints in the cotton fields: the Industrial Revolution as time–space appropriation and environmental load displacement[J]. Ecological Economics, 2006, 59(1): 74
- [26] DALIN C, HANASAKI N, QIU H, et al. Water resources transfers through Chinese interprovincial and foreign food trade[J]. Proceedings of the National Academy of Sciences, 2014, 111(27): 9774
- [27] LEPAWSKY J. The changing geography of global trade in electronic discards: time to rethink the e-waste problem[J]. The Geographical Journal, 2015, 181(2): 147
- [28] BROOKS A, WANG S, JAMBECK J. The Chinese import ban and its impact on global plastic waste trade[J]. Science Advances, 2018, 4(6): 131
- [29] CHEN X, CHENG S, WU L. Quantitative analysis of central Asian countries' energy security and its political influence factors[J]. Journal of Resources and Ecology, 2018, 9(4): 434
- [30] 洪菊花, 骆华松, 梁茂林. 主体间性视角下的“一带一路”能源安全共同体研究[J]. 世界地理研究, 2017, 26(2): 11
- [31] 姜巍, 高卫东. 中国能源消费增长特征及影响因素分析[J]. 世界地理研究, 2013, 22(3): 160
- [32] JACKSON M O, NEI S. Networks of military alliances, wars, and international trade[J]. PNAS, 2015, 112(50): 15277
- [33] MORAN D D, WACKERNAGEL M C, KITZES J A, et al. Trading spaces: calculating embodied Ecological Footprints in international trade using a Product Land Use Matrix (PLUM)[J]. Ecological Economics, 2009, 68(7): 1938
- [34] 叶思菁, 朱德海, 姚晓闯, 等. 基于移动GIS的作物种植环境数据采集技术[J]. 农业机械学报, 2015, 46(9): 325
- [35] 叶思菁, 张超, 王媛, 等. GF-1遥感大数据自动化正射校正系统设计与实现[J]. 农业工程学报, 2017, 33(S1): 266
- [36] 叶思菁. 大数据环境下遥感图谱应用方法研究: 以作物干旱监测为例[J]. 测绘学报, 2018, 47(6): 892
- [37] YE S, SONG C, CHENG C, et al. digital trade feature map: a new method for visualization and analysis of spatial patterns in bilateral trade[J]. ISPRS Int J Geo Inf, 2020, 9(6): 363
- [38] HU L, SONG C, YE S, et al. Spatiotemporal statistical imbalance: a long-term neglected defect in UN Comtrade Dataset [J]. Sustainability, 2022, 14(3): 1431
- [39] 陈小强, 袁丽华, 宋长青, 等. 中美在中国周边地区的商品贸易发展及影响力对比[J]. 地理研究, 2022, 41(3): 663
- [40] WANG Y, SONG C, SIGLEY G, et al. Using social networks to analyze the spatiotemporal patterns of the rolling stock manufacturing industry for countries in the Belt and Road Initiative[J]. ISPRS Int J Geo-Inf, 2020, 9, 431

Design and implementation of UN Comtrade data sharing platform

YE Sijing¹⁾ SONG Changqing^{1, 2)} CHENG Changxiu^{1, 2)} GAO Jianbo^{1, 2)}
GAO Peichao¹⁾ HU Luoming²⁾

(1) State Key Laboratory of Earth Surface Processes and Resource Ecology, Beijing Normal University, 100875, Beijing, China;

2) Faculty of Geographical Science, Beijing Normal University, 100875, Beijing, China)

Abstract United Nations international trade statistics database (UN Comtrade) has provided important data support and application guidance on strengthening the recognition of trade system rules and their driving factors, developing international trade measurement methods, depicting global trade pattern and its changing process. Furthermore, it has been more and more widely applied to global ecological conservation, water-energy-food-land systems, pollution control, energy management, national security and other topics of geographical researches. In this paper, we establish the UN Comtrade data sharing platform with Oracle DBMS based on design of framework and data table structure. This platform is designed to make up for the database's deficiencies in data sharing methods and retrieval interfaces as well as provide data and tool support for geographic research. We develop an automated data archiving module with Python 3.6 and its Scrapy framework 1.5.2, which achieves a dynamic, stable, and highly fault-tolerant processing of more than 500 million trade records by comprehensively integrating data crawling module, data loading module and nesting multiple error correction methods. In addition, the platform improves the efficiency and scalability of data retrieval instruction execution through range partitioning, partitioned composite indexing, and open ODBC/JDBC interfaces based on the structured characteristics of data. Experiments show that the platform can stably execute different types of retrieval instructions in a concurrent mode of 80 users. By invoking the ODBC/JDBC interface to integrate the calculation process into the retrieval task, the system can more effectively use server-side resources and save time for data transmission, reading and writing with higher efficiency and simplified data processing. Based on the platform, we develop a data query and sharing client and apply it to the retrieval, calculation, grid representation and comparative analysis of the explicit comparative advantage characteristics of Chinese and American products in 2017. The application shows that the platform has the advantages of high efficiency, stable concurrent retrieval efficiency and high scalability. It can provide more convenient, fast, and diverse data sharing services for the calculation and analysis of trade characteristics research.

Keywords international trade; big data; grid; revealed comparative advantage index; GIS

【责任编辑: 刘先勤】