

引用格式: Gong Chen, Li Xinwu, Wu Wenjin. Study on the Model of Ecological Vulnerable Human-land System based on Big Data Analysis Framework[J]. Remote Sensing Technology and Application, 2020, 35(5): 1187-1196. [宫晨, 李新武, 吴文瑾. 基于大数据分析框架的生态脆弱型人地系统模式研究[J]. 遥感技术与应用, 2020, 35(5): 1187-1196.]
doi:10.11873/j.issn.1004-0323.2020.5.1187

基于大数据分析框架的生态脆弱型人地系统模式研究

宫晨^{1,2}, 李新武¹, 吴文瑾¹

(1. 中国科学院遥感与数字地球研究所, 北京 100094;

2. 中国科学院大学, 北京 100049)

摘要:针对生态脆弱型人地系统模式研究中存在的数据处理繁杂、模式识别偏主观、内在机理复杂等问题,提出了基于云平台及大数据方法的模式分析框架。通过遥感及社会经济云平台实现数据的云上收集及处理,利用自组织映射神经网络聚类(Self-Organizing Map, SOM)方法实现无先验知识的模式识别;同时利用知觉图从社会经济发展与生态友好性两个角度分析变化轨迹,利用关联规则方法筛选社会经济与生态环境之间的潜在规律。以“一带一路”65国进行实验分析,实验结果将“一带一路”65个国家有效划分为10类模式,并分析了10类模式的变化轨迹及关系规律。结果表明:该分析框架能够快速实现数据获取及处理、人地系统多模式识别、变化轨迹可视化和规律探测等功能,有效弥补了人地系统多模式研究中的不足。

关键词:生态脆弱型人地系统; SOM; 知觉图; 关联规则

中图分类号: X22; P20 **文献标志码:** A **文章编号:** 1004-0323(2020)05-1187-10

1 引言

人地系统是地理学的研究主题与研究核心^[1],且生态脆弱型人地系统是人地系统研究领域的重要环节。生态脆弱型人地系统因其内部自然要素和社会经济要素的复杂作用关系,和面临的生态环境脆弱与经济社会活动不合理的双向瓶颈制约,对典型性人地系统进行模式及规律研究是该领域的重点问题^[2-7]。因此,识别生态脆弱型人地系统的典型发展模式,挖掘模式内要素相互关系及规律,不仅可促进区域人文—自然复合系统研究进一步深化,还可为人地关系理论在区域可持续发展研究中提供指导作用^[8]。

导致生态环境脆弱的原因,不仅与自然因素有

关,还与人类后期不合理的开发活动所产生的负面影响有关,因此是自然因素与人为因素共同影响和彼此叠加导致的结果^[9]。现有的一些研究已对区域生态脆弱型人地系统模式进行了探索分析。Iriarte等^[10]以智利南部具有典型意义的Patagonia为例,提出基于生态系统服务的区域开发管理战略。Shukla等^[11]分析了气候变化背景下印度的生态脆弱现状以及脆弱地区的自适应管理行动来应对全球变化。刘晓琼和张军涛等^[12-13]分别运用定量方法研究了陕西省榆林市和东北农牧交错生态脆弱区的发展模式,根据影响区域发展的关键因子,提出可持续发展建设的对策建议。冷疏影等^[14]分析了中国生态脆弱区的人口—资源—环境与发展矛盾,提出了典型生态脆弱区可持续发展指标体系。董锁成等^[15]

收稿日期:2018-12-26;修订日期:2020-08-22

基金项目:海南省重点研发计划(ZDYF2019005),中国科学院国际合作局对外合作重点项目(131C11KYSB20160061)资助。

作者简介:宫晨(1993—),男,甘肃兰州人,硕士研究生,主要从事资源环境遥感研究。E-mail: gongchen@radi.ac.cn

通讯作者:李新武(1973—),男,四川内江人,博士,研究员,主要从事全球环境变化遥感研究。E-mail: lixw@radi.ac.cn

研究了切合黄土高原生态脆弱典型区——陇西县实际的循环经济发展模式,提出了发展循环经济的对策建议。但现有生态脆弱型人地系统模式研究还存在如下一些问题:

(1)全球或大区域尺度、长时间序列的社会经济与生态环境研究所需数据量大、数据处理繁杂,需要投入大量人力物力。

(2)在研究生态脆弱型人地系统模式过程中,往往需要通过人工划分区域、预设类别等方式引入专家先验知识及主观因素,因此模式识别的主观性较强。

(3)由于影响因子众多、关系复杂等原因,目前生态脆弱型人地系统模式的内在机理及规律研究,只能在局部开展,较少扩展到多模式系统研究。

为解决上述问题,研究设计了基于大数据方法的生态脆弱型人地系统模式分析框架,该框架基于遥感及社会经济云平台,利用大数据分析方法识别人地系统模式,并对模式的变化轨迹进行可视化,绕开了复杂的内在机理,转而探究模式外在表征规律,为模式规律研究提供了一种新的思路。

2 研究方法

针对现有生态脆弱型人地系统模式识别存在的问题,设计了基于云平台及大数据方法的系统模式分析框架(图1),该框架能够实现数据获取及处理、多模式识别、变化轨迹可视化和规律探测等功能。其中数据的收集和处理通过遥感等云平台云操作;获取生态环境与社会经济数据后,利用SOM聚类方法实现系统模式识别;同时,该框架分别利用知觉图从社会经济发展与生态友好性两个方面分析不同模式变化轨迹,利用关联规则方法得到每类模式的社会经济与生态环境之间的规律,从而实现了人地系统模式研究中从数据处理到模式规律筛选等一整套流程。

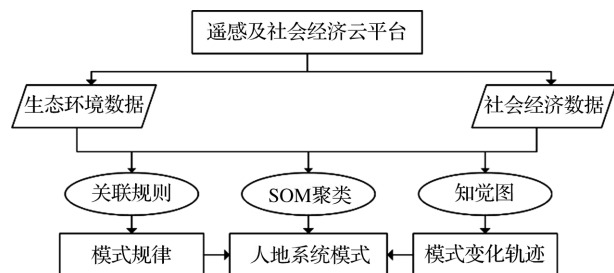


图1 基于大数据方法的生态脆弱型人地系统模式分析框架

Fig.1 A framework for ecological vulnerable human-land system model analysis based on big data method

2.1 基于云平台的数据收集与处理

在传统生态脆弱型人地系统研究方法中,由于涉及到的社会经济数据和生态环境数据往往所需数据量大、处理步骤繁琐,因此整个研究过程需投入大量人力物力,耗费时间较长。而遥感及社会经济云平台的出现,彻底改变了数据的获取和计算方式,云平台的超强存储和运算能力使得数据的获取和处理变得轻松方便。

很多关于自然环境的指标可通过遥感手段获得,如表征气候条件的温度、降雨数据,表征生态环境的植被参数数据,表征生态压力的土地利用及人口数据等。以上大部分数据均可以通过遥感云平台获取和处理,从而解决了遥感数据量大、处理繁杂的问题。由谷歌公司开发的 Google Earth Engine^[16]和 Climate Engine^[17]云平台是专门处理卫星图像和其他地球观测数据的云端运算平台,其拥有超过 40 a 的近地卫星数据存档,包括全球 DEM、土地覆盖、温度、降雨和人口密度等数据,且所有的数据均经过预处理和地理校正,直接可用。这些云平台不仅可以在线展示和下载这些数据,还可以直接云上操作,输出最终处理结果。这一处理模式不仅大大缩减了大数据量遥感影像的处理时间,还省去了传统的下载、拼接、波段叠加等基础操作,彻底改变了传统的遥感影像处理模式。

除遥感数据可在云平台进行收集外,很多社会经济数据也可以从云平台收集获取。如由哈佛大学开发的 World Map 平台收集了世界各国的社会经济数据,并将其进行了地图可视化,方便进行查询和检索。世界银行开发的 World Bank 云平台中包括了世界发展指标数据库、人口统计数据库、教育统计数据库等多种类型社会经济数据,可以进行图表统计和地图化展示。

2.2 基于 SOM 神经网络的模式识别

在生态脆弱型人地系统模式识别探究中,由于涉及大量社会经济与生态环境等的复杂数据,利用大数据分析中的聚类能够较好地对这些数据进行分类及预测^[18],从而从数据驱动的角度确定发展模式。利用 SOM 聚类方法识别不同人地系统的发展模式,避免了通过人工划分区域、预设类别等方式引入先验知识及主观因素导致的不确定性,使模式识别结果能够较为客观。

SOM 神经网络是由芬兰学者 Teuvo Kohonen 于 20 世纪 80 年代提出的^[19],是一种无监督的具有

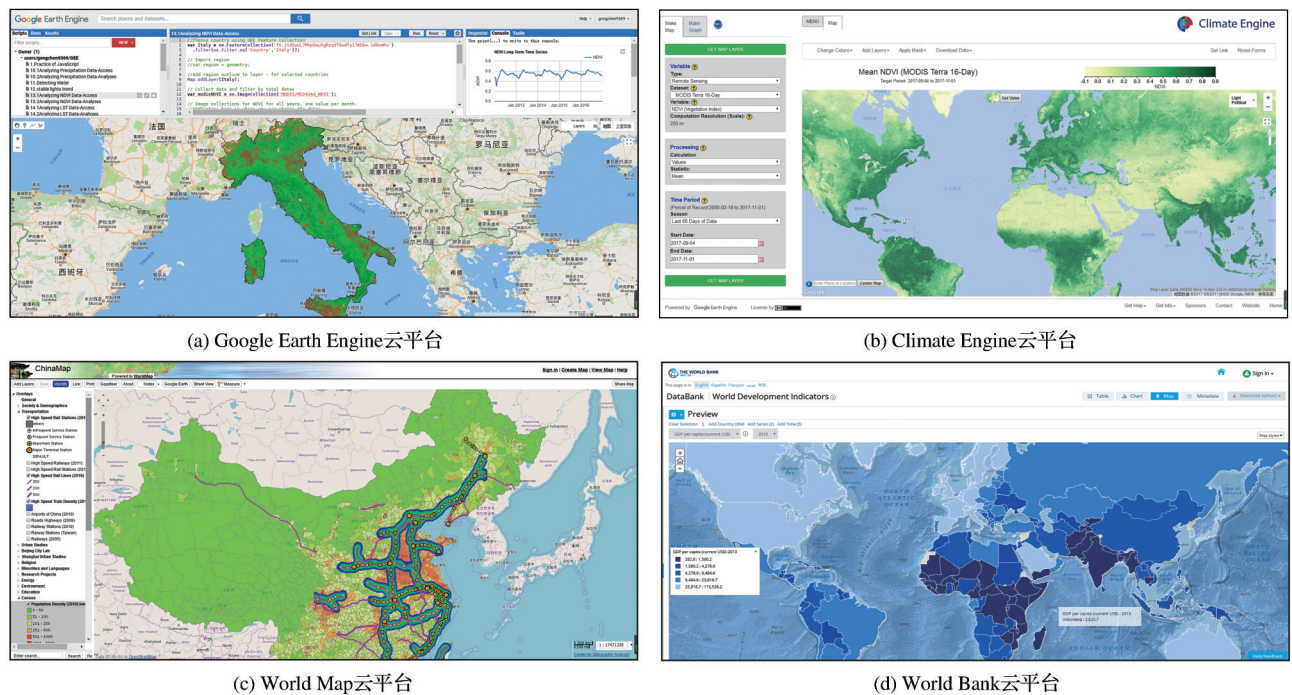


图 2 数据获取及处理云平台(a:Google Earth Engine 云平台;b:Climate Engine 云平台;c:World Map 云平台;d:World Bank 云平台)

Fig.2 Cloud platform of data acquisition and processing

自我学习能力的人工神经网络,具有保序映射、数据压缩、特征抽取等特性^[20],其拓扑结构一般分为两层:输入层和输出层(也称为竞争层)(图 3)。具体实现原理如下:首先初始化从输入神经元到输出神经元之间的权值,接着向神经网络中输入向量,计算该向量与所有输出神经元之间的距离,以内积最大者作为获胜神经元。获胜神经元及其领域范围内的神经元获得调整权值的机会,获胜神经元对周围神经元的影响是随距离远近逐渐变化的,距离近的表现兴奋,距离远的表现为抑制。通过以获胜神经元为中心范围划定“优胜邻域”的方式实现,“优胜邻域”内的神经元按各自距离获胜神经元的距离调整其自身的权值,以强化其对于输出的响应。完成权值调整后,继续输入其他向量对神经网络进行训练,在训练过程中“优胜邻域”的半径逐渐缩小,最终达到收敛状态,使神经元趋于聚类中心。

与传统模式识别方法不同的是,SOM方法可以反映影响因子之间的本质差异,减少人为要素在一致性准则中造成的影响。同时算法自身可以通过无监督学习提取数据集中的规律和特征,能够将原数据集中的度量关系和拓扑结构有效保留在变换后的数据之中^[21-22]。SOM已成为是神经网络最为热门的研究领域之一,其中的 WTA(Winner Takes

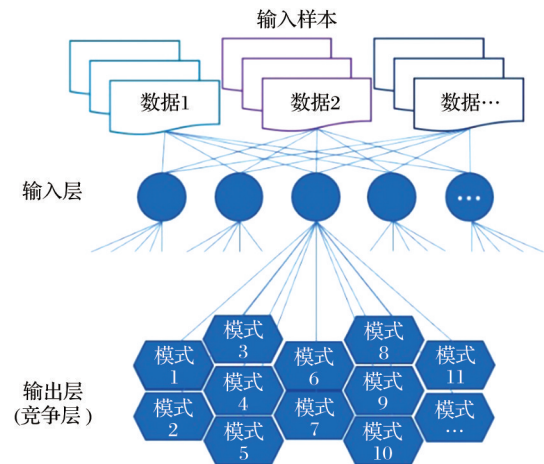


图 3 基于 SOM 聚类的发展模式识别

Fig.3 Development model recognition based on SOM clustering

All)竞争机制反映了自组织学习最根本的特征,目前已被广泛地应用于不同领域各种复杂数据集的有效信息提取。

2.3 基于知觉图的模式变化轨迹可视化

基于 SOM 聚类方法得到生态脆弱型人地系统发展模式后,为了探究发展模式在人类活动与生态环境两方面的变化规律,本研究引入知觉图的概念。将模式内的众多影响因子实现降维,从社会经济发展与生态友好性两个维度进行变化规律可视化展示,能够清晰认识每类发展模式的在社会经济

与生态环境两方面的特点及其发展规律。

知觉图是指受多要素影响的事务在两个或多个维度上进行认知,以达到思维的整体利用,可以分析、比较各个维度的特点。知觉图可以将多影响因子实现有效降维,通过少数几个维度展示出事务的规律^[23]。知觉图一般用平面坐标或三维坐标进行可视化展示,其中坐标轴代表事务的特征因子,图中的散点位置代表该事务在各特征因子上的表现评价。采用因子分析的方法进行知觉图的绘制。因子分析与主成分分析的最大不同点在于,主成分分析强调新变量贡献的比例方差,不关心是否有明确的实际意义;而因子分析着重要求新变量具有实际的意义,能解释原始变量间的内在结构。具体实现过程包括:①KMO 检验与 Bartlett's 检验:首先检验变量间的相关性,一般认为 KMO 值大于 6, Bartlett's 检验的显著性 P 值 < 0.001 。说明变量之间存在相关性,适合进行因子分析^[24];②提取主成分和公因子:得到成分分析的结果后,根据对应的“特征值”,即因子的贡献率,判断占比较大的主成分,一般来说,选择“特征值”大于 1 的成分作为主成分;③因子旋转:成分矩阵中所提取的公因子往往是大而全的“综合因子”,不存在实际意义,利用最大方差法实现因子的旋转,从而获得更好的解释;④知觉图绘制:根据旋转后的成分矩阵,得到最终两个因子得分,在平面坐标中绘制知觉图。

正是利用知觉图清晰易懂的优点,结合发展模式及其影响因子,设计了社会经济与生态环境知觉图。通过社会经济发展与生态友好性两个维度的描述,展示出人地系统中各类发展模式在人类活动与生态环境之间的潜在关系及变化情况。

2.4 基于多维关联规则的模式规律挖掘

由于生态脆弱型人地系统中影响因子众多、内在机理复杂,在涉及多模式研究中做到精确研究每一类模式的机理往往需要投入大量时间精力。因此本分析框架绕开了多模式研究中复杂的内在机理,转而探究模式的外在表征规律,从而为模式规律研究提供了一种新的思路。因为人类活动与生态环境之间存在较大的差异,而大数据分析方法中的关联规则可以作为此类问题的解决手段。关联规则方法可以挖掘出不同模式间的典型规律,这些规则可代表不同模式的时空异质性,可以作为模式表征规律判断的依据。

关联规则挖掘是大数据分析方法中的一种重

要方法,也是一种不需要先验知识的无指导学习方法。关联规则的基本形式是 $X \rightarrow Y$,其中 X, Y 互斥,即 $X \cap Y = \emptyset$ 。它表示当事件 X 发生时,事件 Y 也很可能发生,此处为 X 规则前提, Y 为规则结果^[25]。关联规则由项、项集、支持度和置信度等概念组成。

(1)项与项集:数据库中不可分割的最小单位信息,称为项目,用符号 i 表示。项的集合称为项集。设集合 $I = \{i_1, i_2, \dots, i_k\}$ 是项集, I 中项目的个数为 k -项集。

(2)事务:设 $I = \{i_1, i_2, \dots, i_k\}$ 是由数据库中所有项目构成的集合,一次处理所含项目的集合用 T 表示, $T = \{t_1, t_2, \dots, t_n\}$ 。每一个 t_i 包含的项集都是 I 子集。

(3)项集的频数(支持度计数):包括项集的事务数称为项集的频数(支持度计数)。

(4)关联规则:关联规则是形如 $X \rightarrow Y$ 的蕴含式,其中 X, Y 分别是 I 的真子集,并且 $X \cap Y = \emptyset$ 。 X 称为规则的前提, Y 称为规则的结果。关联规则反映 X 中的项目出现时, Y 中的项目也跟着出现的规律。

(5)关联规则的支持度(Support):关联规则的支持度指事务数据库中既包含 X 又包含 Y 的事务与数据库中总事务数的比值,记为 $\text{support}(X \rightarrow Y)$,即:

$$\text{support}(X \rightarrow Y) = \text{support}(X \cup Y) = P(XY) \quad (1)$$

支持度反映了 X 和 Y 中所含的项在事务集中同时出现的频率。

(6)关联规则的置信度(Confidence):关联规则的置信度指事务数据库中包含 X 的事务同时也包含 Y 的比值,记为 $\text{confidence}(X \rightarrow Y)$,即:

$$\text{confidence}(X \rightarrow Y) = \frac{\text{support}(X \cup Y)}{\text{support}(X)} \quad (2)$$

置信度反映了包含 X 的事务中,出现 Y 的条件概率。

(7)最小支持度与最小置信度:通常用户为了达到一定的要求,需要指定规则必须满足的支持度和置信度阈值,当 $\text{support}(X \rightarrow Y)$ 、 $\text{confidence}(X \rightarrow Y)$ 分别大于等于各自的阈值时,认为 $X \rightarrow Y$ 是有趣的,此两个值称为最小支持度阈值(min_sup)和最小置信度阈值(min_conf)。其中, min_sup 描述了关联规则的最低重要程度, min_conf 规定了关联规则必须满足的最低可靠性。

(8)频繁项集:设 $U = \{u_1, u_2, \dots, u_n\}$ 为项目的集

合,且 $U \subseteq I$, $U \neq \emptyset$, 对于给定的最小支持度 \min_sup , 如果项集 U 的支持度 $\text{support}(U) \geq \min_sup$, 则称 U 为频繁项集, 否则, U 为非频繁项集。

(9) 强关联规则: $\text{support}(X \rightarrow Y) \geq \min_sup$ 且 $\text{confidence}(X \rightarrow Y) \geq \min_conf$, 称关联规则 $X \rightarrow Y$ 为强关联规则, 否则称 $X \rightarrow Y$ 为弱关联规则。

采用经典的 Apriori 算法来生成关联规则。Apriori 算法的基本思想^[26-27]是: 首先找出所有频繁项集, 这项集出现的频繁性至少和预定义的最小支持度一样。然后由频集产生强关联规则, 这些规则必须满足最小支持度和最小置信度。接着使用第 1 步找到的频集产生期望规则, 产生只包含集合项的所有规则, 其中每一条规则的右部只有一项。一旦这些规则被生成, 那么只有那些大于用户给定的最小置信度的规则才被留下来。为了生成所有频集, 使用了递推的方法。图 4 为 Apriori 算法流程图。

针对挖掘模式规律的目标, 提出了一个适合多维关联规则挖掘的框架, 框架主要包含 4 步: ① 指标数据预处理: 针对多维原始数据, 利用聚类方法进

行离散化处理, 形成适合关联规则挖掘的事务数据库; ② 频繁项集生成: 扫描事务数据库, 找出所有满足最小支持度阈值的项集, 即频繁项集; ③ 根据最小置信度, 筛选关联规则; ④ 关联规则评价: 从挖掘出的强关联规则中筛选出值得重点关注的规则, 能够代表不同模式。

3 实验案例

3.1 实验案例简介

实验选择“一带一路”沿线经过的 65 个国家作为案例分析。该区域为欧亚大陆中纬度地区, 具有复杂的地理环境和多样性生态系统, 存在草地退化、森林减少、湖泊萎缩等多种生态环境问题。同时, 该区域经济发展不均衡, 呈现出“两头高, 中间低”态势, 东亚、西欧等地区经济发展速度较快, 中亚、西亚等区域经济发展缓慢。正是由于该区域经济发展模式不一, 生态系统多样, 两者之间关系复杂, 因此确定该区域为生态脆弱型人地系统研究案例。

3.2 数据获取及处理

为了开展“一带一路”区域生态脆弱型人地系统模式研究, 实验需要用到生态数据和社会经济数据。对于生态数据, 实验选择了能够表征气候变化和社会经济活动变化的植被参数数据, 主要用到了 Landsat5 卫星的 NDVI 产品, NOAA 气象卫星上来自 AVHRR 传感器的 NPP 产品, MODIS 卫星的 NPP 产品和 NDVI 产品(表 1)。为了能快速有效地处理海量遥感数据, 本研究利用谷歌公司开发的 Google Earth Engine(GEE)遥感云平台, 计算得到 NPP 与 NDVI 的不同国家年平均数据。

社会经济数据主要通过 World Bank 云平台上的世界发展指标数据库收集得到。为能充分表征社会经济发展对生态环境变化带来的影响, 从人口、经济、产业三方面进行收集。其中人口指标主要指城市人口占总人口比例, 经济指标包括国家总 GDP 和人均 GDP 两项, 产业指标包括第一产业比例、第二产业比例和第三产业比例。表 2 为社会经济数据集的详细信息。

3.3 生态脆弱型人地系统模式识别

综合上述生态遥感数据与社会经济数据, 利用 SOM 聚类方法挖掘不同国家生态脆弱型人地系统典型模式。SOM 聚类结果如表 3 所示, 将 65 个国家共分为 10 类发展模式。

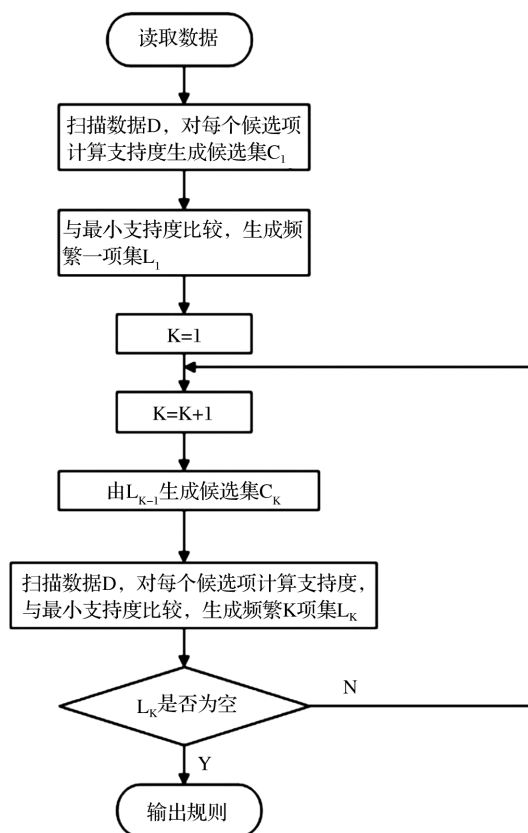


图 4 Apriori 算法流程图

Fig.4 Apriori algorithm flow chart

表 1 生态遥感数据集

Table 1 Ecological remote sensing data sets

序号	数据集	GEE ImageCollection ID	时间范围	空间分辨率	时间分辨率
1	AVHRR NPP	users/ gongchen9369/ 1992- 2000_npp_latlon	1992~2000	8 km	1 a
2	MODIS NPP	MODIS/ 055/ MOD17A3	2001~2014	1 km	1 a
3	Landsat 5 NDVI	LANDSAT/ LT05/ C01/ T1_8DAY_NDVI	1992~2000	30 m	8 d
4	MODIS NDVI	MODIS/ MCD43A4_NDVI	2001~2014	250 m	16 d

表 2 社会经济数据集

Table 2 Socio-economic data sets

序号	数据集	数据来源	时间范围	空间分辨率	时间分辨率
1	城市人口占总人口比例	United Nations World Urbanization 2014 Prospects	1992~2014	国别	1 a
2	总 GDP	World Bank national accounts and OECD data files	1992~2014	国别	1 a
3	人均 GDP	World Bank national accounts and OECD data files	1992~2014	国别	1 a
4	第一产业比例	World Bank national accounts and OECD data files	1992~2014	国别	1 a
5	第二产业比例	World Bank national accounts and OECD data files	1992~2014	国别	1 a
6	第三产业比例	World Bank national accounts and OECD data files	1992~2014	国别	1 a

表 3 SOM 聚类分析结果

Table 3 SOM clustering analysis results

发展模式	国家
SOM 1	巴林、伊朗、伊拉克、科威特、阿曼、卡塔尔、沙特阿拉伯、阿联酋
SOM 2	哈萨克斯坦、俄罗斯、土耳其
SOM 3	阿塞拜疆、中国、埃及、蒙古、叙利亚、土库曼斯坦、也门
SOM 4	亚美尼亚
SOM 5	阿富汗、不丹、柬埔寨、印度、吉尔吉斯斯坦、缅甸、尼泊尔、巴基斯坦、塔吉克斯坦、乌兹别克斯坦
SOM 6	以色列、约旦、新加坡
SOM 7	文莱、保加利亚、克罗地亚、塞浦路斯、捷克、爱沙尼亚、希腊、匈牙利、拉脱维亚、黎巴嫩、立陶宛、波兰、斯洛伐克、斯洛文尼亚
SOM 8	白俄罗斯、乌克兰
SOM 9	波黑、格鲁吉亚、印度尼西亚、马其顿、马来西亚、马尔代夫、黑山、菲律宾、罗马尼亚、塞尔维亚、泰国
SOM 10	阿尔巴尼亚、孟加拉、老挝、摩尔多瓦、斯里兰卡、越南

3.4 模式变化轨迹可视化

在得到“一带一路”区域的典型发展模式之后，为研究每类模式在近 23 年间的变化情况，利用因子分析的方法进行知觉图绘制，从社会经济发展与生态友好性两个维度进行变化规律可视化展示。

表 4 为原始数据的 KMO 检验和 Bartlett’s 检验，其中 KMO 值为 0.650，大于 6。Bartlett’s 检验的显著性 P 值 <0.001。综合两个指标，说明变量之间存在相关性，说明实验所采用的数据适合进行因子分析。

表 4 KMO 和 Bartlett 检验

Table 4 KMO and Bartlett test

Kaiser-Meyer-Olkin测量取样适当性	0.65	
Bartlett's球形度检验	近似卡方	5 361.56
	自由度	28
	显著性	0.000

正交旋转后的因子矩阵表如表 5 所示，因子 1 包括：城市人口占总人口比例、人均 GDP 和第三产业比例，这些特点代表经济发展程度高，因此将因子 1 命名为社会经济发展。因子 2 主要包括 NPP、NDVI 和第三产业比例，代表生态环境得到保护，因此将因子 2 命名为生态友好性。

表 5 正交旋转后的因子矩阵

Table 5 Orthogonal rotation of the factor matrix

关键因子	成分 1	成分 2
城市人口占总人口比例	0.84	-0.24
总 GDP	0.04	-0.17
人均 GDP	0.69	-0.30
第一产业比例	-0.89	0.15
第二产业比例	0.07	-0.77
第三产业比例	0.67	0.58
NPP	-0.23	0.67
NDVI	-0.11	0.70

在利用因子分析提取出两个因子,并得到了社会经济与生态环境的两个因子得分后,以5a为一个阶段划分出1992~2014年间的4个时间范围,绘制10类发展模式在各阶段的知觉图,其中X轴代表社会经济发展属性,Y轴代表生态友好属性,从两方面分析了10类发展模式的变化规律。图5为10类发展模式的细部图,详细展示了各类模式在近23年的变化轨迹。

总的来说,10类发展模式在近23年间经济状况始终呈正增长趋势,除SOM1模式在1997~2001年间的倒退外,从未出现经济衰退。而生态环境本身

差异较大,SOM7、SOM9和SOM10发展模式生态环境较为优秀,而SOM1~SOM4关系模式的生态环境较为恶劣,且10类模式中大多数模式在1992~2011年间生态环境普遍出现衰退,但在2012年之后变得平稳。每类模式之间还存在异质性,如SOM10模式的变化轨迹与SOM9模式相似,均是除在1997~2001年间出现生态环境衰退外,其余时间均较为平稳。但不同的是SOM10模式经济发展程度并不如SOM9模式高(更加靠近X正半轴),且SOM10模式经济发展虽然一直是正增长趋势,但经济增速是逐渐变缓的。

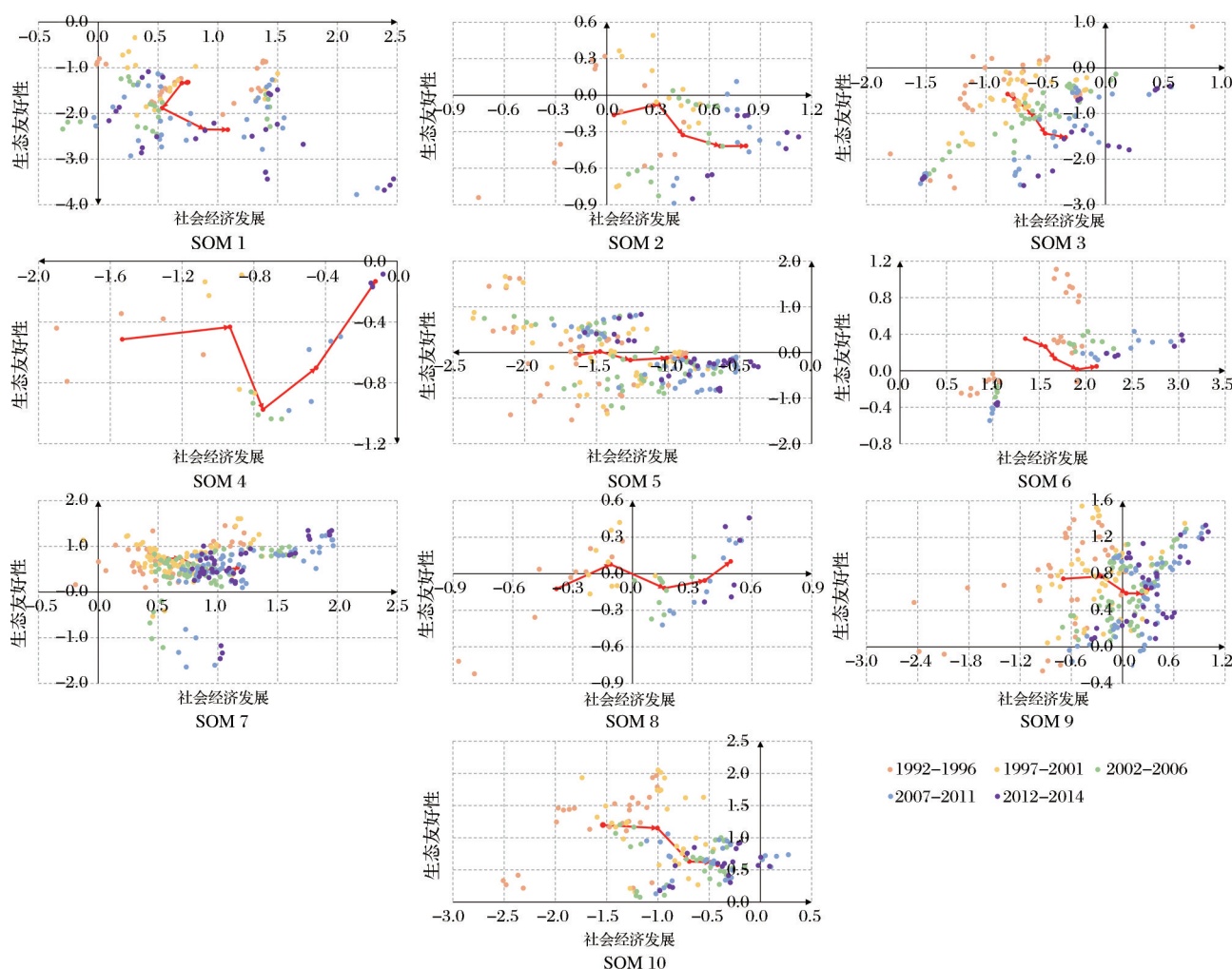


图5 10类发展模式的知觉图变化轨迹

Fig.5 Perceptual map trajectories of 10 development models

3.5 模式规律挖掘

为确定每一类模式的特殊规律,利用关联规则方法进行社会经济与生态环境关系研究。首先,基于K-means聚类算法,对以上的指标因子进行离散化预处理。一般来说越细致越准确的规则越有效,但过多地分割会大大增加算法的复杂度,降低算法

的效率,同时也会造成满足支持度阈值的频繁项过少从而无法完成所需规则的挖掘。经过多次试验,本实验将社会经济数据的分割数定为4,生态遥感数据的分割数定为3,经过聚类后得到聚类中心的结果如表6所示。

依据最邻近原则对原始数据进行离散化后,得

到规范化的适合关联规则挖掘的社会经济与生态环境事务表,如表 7 所示。其中 I_{ij} 代表社会经济因子属性项(i 分别代表“城市人口占总人口比例”,“总 GDP”,“人均 GDP”,“第一产业比例”,“第二产业比例”,“第三产业比例”, j 分别代表类别“分类 1”,“分类 2”,“分类 3”,“分类 4”), E_{ij} 代表生态环境属性项(i 分别代表“NPP”,“NDVI”, j 分别代表类别“分类 1”,“分类 2”,“分类 3”)。

表 6 关键因子离散化				
Table 6 Discretization of key factors				
评价指标	中心 1	中心 2	中心 3	中心 4
城市人口占总人口比例/%	28.45	51.12	69.17	88.88
总 GDP(US \$)	1.01E+10	4.85E+10	9.84E+10	1.61E+11
人均 GDP(US \$)	747.50	2390.73	4731.78	7744.12
第一产业比例/%	4.51	15.12	26.98	44.08
第二产业比例/%	20.56	30.44	42.94	63.51
第三产业比例/%	26.86	41.02	53.36	66.96
NPP/gCm ⁻²)	202.63	720.25	1616.67	
NDVI	0.15	0.46	0.64	

表 7 规范化社会经济与生态环境事务表								
Table 7 Standardization table of social-economic and environmental affairs								
ID	Un	GDP	GDP per	Ag	In	Se	NPP	NDVI
BHR1992	I14	I21	I34	I41	I53	I64	E11	E21
BHR1993	I14	I21	I34	I41	I53	I64	E11	E21
BHR1994	I14	I21	I34	I41	I53	I64	E11	E21
BHR1995	I14	I21	I34	I41	I53	I63	E11	E21
BHR2007	I14	I21	I34	I41	I53	I63	E11	E21
...

注:Un:城市人口占总人口比例;GDP:总 GDP;GDP per:人均 GDP;Ag:第一产业比例;In:第二产业比例;Se:第三产业比例;NPP:净初级生产量;NDVI:归一化植被指数。

为了充分挖掘可能有重要意义的关联规则,同时考虑规则的可信度,本实验设定最小支持度为 60%,最小置信度为 70%。根据多维关联规则挖掘方法对规范化社会经济与生态环境时空事务表进行挖掘,得到部分关联规则如表 8 所示。

基于设定的最小支持度和最小置信度,将 1992~2014 年 10 类发展模式进行挖掘后总共得到 346 条规则,其中 SOM1 有 46 条,SOM2 有 37 条,SOM4 有 168 条,SOM6 有 9 条,SOM7 有 4 条,SOM8 有 82 条,而 SOM3、5、9 和 10 在此支持度和置信度下不存在关联规则,说明关系模式较为复

表 8 关联规则部分挖掘结果			
Table 8 Part results of association rules			
ID	Association Rules	ST_sup	ST_conf
1	E11 → I41	88.98%	88.98%
2	I41 → E21	88.98%	100%
3	E11 → E21, I41	88.98%	88.98%
4	I34 → E11	77.17%	100%
5	E21 → I34	77.17%	77.17%
6	I34 → E11, E21	77.17%	100%
...

杂,此情况下社会经济与生态环境间不存在明显关系。这里以 SOM1 发展模式中有意义的 2 条规则进行举例说明:

SOM1 模式国家主要是包括巴林、伊朗、伊拉克、科威特、阿曼、卡塔尔、沙特阿拉伯和阿联酋在内的 8 个西亚国家。自然环境方面,SOM1 模式国家地形以高原为主,大部分地区水资源短缺,气候干旱,草原和沙漠广布,植被稀疏,种类较少,且有显著耐干旱特征。经济方面,SOM1 模式的 8 个国家均属于石油输出国,石油业在国民生产总值、国民收入和出口值中的比重都占据绝对优势。且建筑业、运输业、加工业和商业都是以石油生产为其发展基础,战后经济发展非常迅速,人均国民生产总值居世界前列。受制于环境限制,西亚国家的农产品自给率低,成为世界农牧产品主要进出口区之一。

典型规则 1: E21→I41 (sup=88.98%; conf=88.98%);

此条规则说明当植被指数处于 0.07~0.31 较低值范围段时,第一产业比例也处于 0.03%~7.30% 的较低值范围。此条规则的置信度高达 88.98%,说明这条规律的可信度极高,这也与我们的日常认知比较一致。西亚的地形以高原为主,除地处两河流域的伊朗外,西亚大部分地区降水稀少,水资源短缺。西亚自然植被主要是热带和亚热带荒漠及半荒漠草原,植被稀疏,因此植被指数较低。在这种干旱条件下,西亚的第一产业主要以畜牧业和灌溉农业为主,耕地较少,农产品自给率低,是世界农牧产品主要进出口区之一。由此可见,不适宜植被生长的环境,造成植被指数偏低,同时第一产业比例也属于较低值范围。

典型规则 2: E21, I54→I34 (sup=76.38%; conf=76.68%);

此条规则说明当植被指数处于 0.07~0.31 较低值范围段,第二产业比例处于 53.60%~96.62% 较

高值范围段时,人均GDP处于世界领先6 254.95~88 564.82美元的范围。尽管SOM1模式的植被覆盖率低,但西亚地区石油资源丰富,包括沙特、阿联酋、巴林在内的8个国家均属于石油输出国家,石油作为国家经济命脉,在国民生产总值和出口值中的比重都居绝对优势。因此此类关系模式国家的第二产业比例在“一带一路”范围内属于最高水平,带来的经济效益也使得国民收入处于“一带一路”前列。

4 结 语

本研究创新性地设计了一个基于云平台及大数据方法的系统模式分析框架,并以“一带一路”65国作为分析案例,利用生态植被、人口经济等8项参数,将其划分为10类发展模式,研究发现10类模式总体变化趋势是从经济落后、生态环境尚可向经济迅速发展、生态环境遭到破坏的状态变化,各模式受气候、产业结构等因素影响,植被环境状况变化不一。该分析框架能够快速实现数据获取及处理、人地系统多模式识别、变化轨迹可视化和规律探测等功能,有效弥补了人地系统多模式研究中的不足。与传统方法最大的不同点在于,研究不仅局限于模型的创新或指标的选择,整个分析框架涵盖了从数据处理到模式规律筛选等一整套流程,具有现实的可操作性。

本研究目前仅从社会经济与生态环境两方面进行了人地系统模式分析,暂未考虑其他影响因素,下一步可以扩展加入气候要素、地形要素、政策要素等等,使人地系统模式分析更加具有综合性、动态性、可调控性。同时可以关键技术框架为核心,融合面向模式分析与决策服务的时空多源监测数据组织与管理策略,对各个模块进行进一步封装、集成,形成一体化决策支持服务系统。框架分析结果将有利于我们认清区域生态环境的脆弱程度、形成原因和变化趋势,是有针对性地采取技术措施进行生态环境治理的基本前提,也可为区域发展提供理论基础和科学依据。

参考文献(References):

- [1] Pattison, William D. The Four Traditions of Geography[J]. *Journal of Geography*, 1990, 89(5):202-206.
- [2] Newbold T, Bentley L F, Hill S L, *et al.* Global Effects of Land Use on Biodiversity Differ Among Functional Groups[J]. *Functional Ecology*, 2020, 34(3):684-693. doi: 10.1111/1365-2435.13500.
- [3] Probert J R, Parr C L, Holdo R M, *et al.* Anthropogenic Modifications to Fire Regimes in the Wider Serengeti-Mara Ecosystem[J]. *Global Change Biology*, 2019, 25(10):3406-3423. doi: 10.1111/gcb.14711.
- [4] Fu B J, Tian T, Liu Y X, *et al.* New Developments and Perspectives in Physical Geography in China[J]. *Chinese Geographical Science*, 2019, 29(3):363-371.
- [5] Wu Z H, Lei S G, He B J, *et al.* Assessment of Landscape Ecological Health: A Case Study of a Mining City in a Semi-Arid Steppe[J]. *International Journal of Environmental Research and Public Health*, 2019, 16(5):752. doi: 10.3390/ijerph16050752.
- [6] Gon S M, Tom S L, Woodside U. 'Aina Momona, Honua Au Loli-Productive Lands, Changing World: Using the Hawaiian Footprint to Inform Biocultural Restoration and Future Sustainability in Hawai'i[J]. *Sustainability*, 2018, 10(10):3420. doi:10.3390/SU10103420.
- [7] Liu Yanxu, Fu Bojie, Wang Shuai, *et al.* Research Progress of Human-earth System Dynamics based on Spatial Resilience Theory[J]. *Acta Geographica Sinica*, 2020, 75(5):891-903. [刘焱序, 傅伯杰, 王帅, 等. 空间恢复力理论支持下的人地系统动态研究进展[J]. *地理学报*, 2020, 75(5):891-903.]
- [8] Lu Dadao. Theoretical Studies of Man-land System as the Core of Geographical Science[J]. *Geographical Research*, 2002, 21(2):135-145. [陆大道. 关于地理学的“人—地系统”理论研究[J]. *地理研究*, 2002, 21(2):135-145.]
- [9] Ma Hailong. The study on the Coordination and Adjustment Mechanism of Human-land Relationship in the Industry-Development-Restricted Region of Western China[D]. Beijing: Institute of Geographical Sciences and Natural Resources Research, CAS, 2008. [马海龙. 西部限制开发区人地关系调控的机理与途径[D]. 北京: 中国科学院地理科学与资源研究所, 2008.]
- [10] Iriarte J L, González H E, Nahuelhual L. Patagonian Fjord Ecosystems in Southern Chile as a Highly Vulnerable Region: Problems and Needs[J]. *AMBIO-A Journal of The Human Environment*, 2010, 39(7):463-466.
- [11] Shukla R, Chakraborty A, Joshi P K. Vulnerability of Agro-ecological Zones in India Under the Earth System Climate Model Scenarios[J]. *Mitigation and Adaptation Strategies for Global Change*, 2017, 22(3):399-425.
- [12] Liu Xiaoqiong, Liu Yansui. Research on the Sustainable Development of Vulnerable Ecological Region based on AHP—A Case Study of Yulin City[J]. *Journal of Arid Land Resources and Environment*, 2009, 23(5):19-23. [刘晓琼, 刘彦随. 基于AHP的生态脆弱区可持续发展评价研究——以陕西省榆林市为例[J]. *干旱区资源与环境*, 2009, 23(5):19-23.]
- [13] Zhang Juntao, Fu Xiaofeng. Research on the Sustainable Development of the Vulnerable Eco-Regions[J]. *China Population, Resources and Environment*, 2005, 15(5):58-62. [张军涛, 傅小锋. 东北农牧交错生态脆弱区可持续发展研究[J]. *中国人口·资源与环境*, 2005, 15(5):58-62.]
- [14] Leng Shuying, Liu Yanhua. A Conceptual Framework of Sustainable Developmental Indicators of Fragile Environment of China[J]. *China Population, Resources and Environment*, 1999, 9(2):3-5. [冷疏影, 刘燕华. 中国脆弱生态区可持续发展指标体系框架设计[J]. *中国人口·资源与环境*, 1999, 9

- (2):3-5.]
- [15] Dong Suocheng, Li Dai, Li Rongsheng, *et al.* A Circular Economic Pattern in the Eco-fragile Region of the Loess Plateau: A Case Study in Longxi County of Gansu Province. *Resources Science*, 2005, 27(4):82-88. [董锁成, 李岱, 李荣生等. 黄土高原生态脆弱区循环经济发展模式研究[J]. 资源科学, 2005, 27(4):82-88.]
- [16] Gorelick N, Hancher M, Dixon M, *et al.* Google Earth Engine: Planetary-scale Geospatial Analysis for Everyone [J]. *Remote Sensing of Environment*, 2017, 202: 18-27.
- [17] Hegewisch K, Daudert B, Morton C, *et al.* Climate Engine - Monitoring Drought with Google Earth Engine [C]// American Geophysical Union, Fall Meeting, 2016.
- [18] Liu Y, Weisberg R H. A Review of Self-organizing Map Application in Meteorology and Oceanography [M]. Rijeka: IntechOpen, 2011.
- [19] Kohonen T. Self-Organized Formation of Topologically Correct Feature Maps [J]. *Biological Cybernetics*, 1982, 43(1): 59-69.
- [20] Kalteh A M, Hjorth P, Berndtsson R. Review of the Self-Organizing Map (SOM) Approach in Water Resource: Analysis, Modeling and Application [J]. *Environmental Modeling & Software*, 2008, 23(7): 835-845.
- [21] Tsakovski S, Joanna Żukowska, Peter Bode, *et al.* Self-organizing Maps Classification of Epidemiological Data and Toenail Selenium Content Monitored on Cancer and Healthy Patients from Poland [J]. *Environmental Letters*, 2010, 45(3):7. doi:10.1080/10934520903467840.
- [22] Gholami V, Khaleghi M R, Salimi, E T. Groundwater Quality Modeling Using self-organizing Map (SOM) and Geographic Information System (GIS) on the Caspian Southern Coasts [J]. *Journal of Mountain Science*, 2020, 17(7):1724-1734. doi: 10.1007/s11629-019-5483-y.
- [23] Zhang Youyin, Gu Jing. Analysis of British Tourists' Hotel Property Perception based on Perceptual Mapping [J]. *Tourism Forum*, 2015(3):8-13. [张佑印, 顾静. 基于知觉图的旅华英国游客酒店属性感知分析[J]. 旅游论坛, 2015(3): 8-13.]
- [24] Xie Kun, Zhang Junfang. Short-term Wind Speed Forecasting Using PCA-WNN based on KMO-Bartlett Typical Wind Speed Selection [J]. *Power Equipment*, 2017, 31(2):86-91. [解坤, 张俊芳. 基于KMO—Bartlett典型风速选取的PCA-WNN短期风速预测[J]. 发电设备, 2017, 31(2):86-91.]
- [25] Chen M S, Han J, Yu P S. Data Mining: An Overview from a Database Perspective [J]. *IEEE Transactions on Knowledge and Data Engineering*, 1996, 8(6):866-883.
- [26] Wu X, Kumar V, Quinlan J R, *et al.* Top 10 Algorithms in Data Mining [J]. *Knowledge and Information Systems*, 2008, 14(1):1-37.
- [27] Dayal U, Whang K Y, Lomet D, *et al.* Proceedings of the 32nd International Conference on Very Large Data bases [C]// International Conference on Very Large Data Bases, 2006.

Study on the Model of Ecological Vulnerable Human-land System based on Big Data Analysis Framework

Gong Chen^{1,2}, Li Xinwu¹, Wu Wenjin¹

(1. *Institute of Remote Sensing and Digital Earth Chinese Academy of Sciences, Beijing 100049, China;*

2. *University of Chinese Academy of Sciences, Beijing 100049, China)*

Abstract: To solve the problems of complex data processing, subjective model recognition and complex internal mechanism in the study of ecological vulnerable human-land system, a model analysis framework based on cloud platform and big data methods was proposed. Remote sensing and socio-economic cloud platform are used to collect and process data. Self-organizing mapping neural network clustering (SOM) method is used to recognize model without prior knowledge. The trajectories was analyzed from the perspective of social-economic development and ecological friendliness by using perceptual map, and the laws between social economy and ecological environment was selected by using association rules. The experimental analysis was carried out in 65 Belt and Road countries. The experimental results effectively divided 65 countries into 10 models, and analyzed the trajectories and relationship rules of 10 models. The results show that the framework can perform the functions of data acquisition and processing, multi-model recognition of human-land system, trajectories visualization and rules detection. It effectively makes up for the deficiencies in the multi-model study of human-land system.

Key words: Ecological vulnerable human-land system; SOM; Perceptual map; Association rules