

彭俊, 卢艳松, 夏凯. 一种基于 SOM 的多维农业时空数据可视化分析方法[J]. 江苏农业科学, 2017, 45(20): 244-249.
doi:10.15889/j.issn.1002-1302.2017.20.061

一种基于 SOM 的多维农业时空数据可视化分析方法

彭俊¹, 卢艳松¹, 夏凯^{1,2}

(1. 浙江农林大学信息工程学院, 浙江临安 311300; 2. 浙江省林业智能监测与信息技术研究重点实验室, 浙江临安 311300)

摘要:随着“3S”、物联网等技术逐渐被应用于农业生产过程中, 多维农业时空数据大量产生。如何挖掘和分析多维时空数据在时空上的内在价值, 对推动现代农业发展有重要意义。针对多维农业时空数据存在维数高, 数据关系复杂, 难以直接挖掘分析的问题, 提出一种基于 SOM 的多维农业时空数据可视化分析方法。该方法利用 SOM 算法对多维数据降维, 结合数据表、平行坐标系、时空立方体等可视化组件, 实现数据的多角度显示和分析, 达到深层次挖掘的目的。基于该方法, 设计完成可视化分析平台, 并以美国农产品数据为例, 讨论该方法的可行性。

关键词:多维数据; 时空数据; SOM 神经网络; 可视化分析

中图分类号: S126 **文献标志码:** A **文章编号:** 1002-1302(2017)20-0244-05

伴随全球定位系统、遥感技术等技术的不断突破及在农业领域的广泛应用, 产生了大量多维时空数据。这些数据有效记录并展示了事物各阶段的发展状况, 是一种结构复杂、多层嵌套、具有空间和时态特征的高维数据^[1]。由于数据在时间和空间上具有很强的相关性, 蕴含巨大的挖掘潜力。如何挖掘和分析这些数据对于精细农业的发展、农业生产的进步及现代社会的发展有着极大的研究意义。

作为探索数据内在隐藏信息的一种手段, 时空数据可视化方法通过对多维时空数据进行图形化表示, 能够全面展示数据基本特征, 实现数据之间的交互^[2]。当前, 多维时空数据的可视化方法除了传统的 2D、3D 显示技术, 还可分为基于几何的技术、基于层次的技术、基于图标的技术、面向像素的技术和降维映射技术等^[3-5]。除了降维映射技术, 其他多维时空数据可视化方法多是将数据各个维度直接映射于某一实体, 降维映射技术能够将多维或高维数据投影到二维或三维空间, 以数据散点图显示数据集的聚类结构和数据分布, 并显示聚类结果类标信息(不同的类可以用数字或颜色区分)^[6]。因此被广泛应用于维度较高、数据量较大的信息可视化。

降维可视化方法关键在于降维或映射技术的选择^[7]。当前降维可视化方法主要有主成分分析法、投影寻踪法、自组织映射法(SOM)和多维标度法等。雷君虎等通过主成分分析法实现对烟用香料图谱数据进行数据降维和香料分类^[8]。闫绪娴运用投影寻踪聚类方法对中西部地区 2000—2011 年自然灾害社会易损性时空数据进行 3 个层面的空间特征和区

域差异分析, 从而对地区自然灾害社会易损性进行综合评价^[9]。芮小平等利用自组织映射降维技术实现了对 2003 年中国“非典”的多维信息进行数据降维和可视化分析^[10]。李权利利用 Sammon 投影(一种典型的多维标度法)将多维数据映射到二维平面, 从而实现对海量微博数据进行可视化分析^[11]。

本研究采用 SOM 算法实现对高维数据的降维。原因在于 SOM 神经网络实现对数据降维的同时, 能保留其拓扑结构不变, 从而避免“维数灾难”的发生。同时, SOM 将输入数据转换成数据量较小的矢量数据, 数据在被其他投影技术分析和表达的同时, 在较少计算量的情况下, 能产生良好的投影结果。

考虑到多维数据多包含大量时空信息, 单纯通过 SOM 降维可视化技术进行数据可视化, 并不能很好地反映当前信息的时空特征。同时, 针对某一问题的研究, 由于采集数据的属性维度较多且存在差异性, 即使采用多属性的时空数据可视化表达也无法保证其能够达到良好的可视化效果。有学者考虑使用多种可视化分析方法集成的模式对时空数据进行可视化分析表达, 例如: Andrienko 等利用 SOM 降维数据对美国 41 年间各州犯罪率时空数据进行降维分类, 并在地图上进行各州犯罪情况的可视化展示, 取得了良好的数据挖掘效果, 提升了数据分析效率^[12]。然而, 这种多可视化方法集成的模式更多地还是集中于对原始数据的可视化表达, 相互关联性并不强, 且运算量较大, 并不能很好地满足当前对时空数据全面、高效、动态的深入分析和挖掘要求。

针对单一的可视化方法无法满足时空数据多角度表示和分析的要求, 集成的可视化方法更多也仅是对时空数据可视化工具的组合, 本质上还是独立地对时空数据进行表达说明。本研究提出一种基于 SOM 的多视图协同可视化分析方法, 该方法从对降维数据的可视化角度出发, 集成多种可视化分析工具, 针对降维后数据进行可视化表达, 既解决了传统可视化工具无法对高维多属性时空数据进行可视化的问题, 又做到各表达工具之间的联动, 实现对数据的实时多角度可视化表达和分析, 增强分析者对隐藏信息的挖掘能力。本研究

收稿日期: 2016-07-27

基金项目: 浙江省自然科学基金(编号: Y5110145); 浙江农林大学科研发展基金(编号: 2014FR086); 浙江农林大学智慧农林业研究中心项目(编号: 2013ZHNL01); 浙江省林业智能监测与信息技术研究重点实验室项目。

作者简介: 彭俊(1990—), 男, 浙江宁海人, 硕士研究生, 主要从事地理信息系统、数据挖掘与数据分析研究。E-mail: pengjun691918986@hotmail.com。

通信作者: 夏凯, 副教授, 硕士生导师, 主要从事地理信息系统、数据分析和数据可视化研究。E-mail: laxiakai@live.cn。

以美国农业时空数据为例,搭建基于该方法的动态可视化分析平台,进行时空数据可视化分析,验证该方法的可行性,旨在为推进我国精细农业的发展提供依据。

1 基于 SOM 的降维可视化

1.1 基于 SOM 的聚类方法

1982 年芬兰学者 Kohonen 根据人脑的生物学、生理学、心理学的研究成果,提出自组织映射网络 (self-organizing feature map, SOM), 别称 kohonen 网络,它是无监督学习的神经网络^[13]。SOM 网络主要面向高维数据,通过降维将数据从完全混乱转换为整体有序,从而挖掘感兴趣的特征。其学习过程是一个竞争过程,即调整一部分输入数据的权值使得权向量更接近或更偏离输入矢量,最终使得所有权向量都在输入空间内相分离,形成各自所代表的一类输入空间模式。具体算法过程如下:

设有 m 个输入样本,每个输入样本有 n 个属性,即输入为 n 维向量 $X = (x_1, x_2, \dots, x_n)$,从而构成一个二维的神经元输入层。输出层由 P 个神经元节点构成,每个结点与一个权值相联系。其连接权值 w_{ij} 表示第 i 个输入神经元节点与第 j 个输出神经元节点之间的连接权值。

(1) 初始化:为网络的权向量在 $[0, 1]$ 区间随机赋予 1 个值,要求互不相同,从而确定邻域的初始值。邻域是指以确定的获胜神经元为中心包含若干神经元的区域范围。

(2) 接收输入并寻找获胜节点:对输入数据 X 进行归一化处理,计算权向量与 X 之间的欧氏距离,找出最小距离,确定获胜神经元 i 。

$$d_j = \sum_{i=1}^n [x_i(t) - w_{ij}(t)]^2;$$

$$i(x) = \min d_j.$$

式中: $x_i(t)$ 表示 t 时刻输入样本数据值, $i(x)$ 即为最匹配神经元。

(3) 定义优胜邻域:以获胜神经元为中心,确定权值调整值。

$$h_{j,i(x)}(t) = \exp\left(-\frac{\|r_{i(x)} - r_j\|}{2\sigma^2(t)}\right).$$

式中: $r_{i(x)}$ 、 r_j 分别为输入节点 $i(x)$ 、 j 的位置, σ 表示邻域范围, $h_{j,i(x)}(t)$ 为值域函数,表示获胜神经元周围的邻域函数,随时间单调递减。

(4) 调整权值:对邻域内所有神经元与输入层神经元之间的权向量进行修正。

$$w_{ij}(t+1) = w_{ij}(t) + \eta(t) h_{j,i(x)}(t) [X_i - w_{ij}(t)].$$

式中: $\eta(t)$ 表示学习速率 $[0 < \eta(t) < 1]$,随时间单调递减,保证学习过程收敛性。

(5) 结束判定:更新学习速率及邻域,重复上述学习过程,直到对所有样本进行学习。

最后通过训练,连接权向量不再进行调整,所得网络拓能够近似描述输入向量分布情况,实现模式分类的目的。

1.2 基于 SOM 的降维可视化

基于 SOM 的高维数据可视化过程如图 1 所示,此前首先需要对输入的数据集进行预处理,规范其数据结构,保证数据为矩阵形式数据。对数据进行规范处理后,通过 SOM 神经网络

算法进行数据降维。这个过程中,高维数据将通过降维映射到二维平面的神经元上,自组织映射网络中的神经元代表所需分析的高维数据,其所在位置能够反映数据的聚类信息,并保证其拓扑关系不变。然而 SOM 算法本身并不能将其所分析获得数据的聚类结构通过图形的方式展现,必须借助其他可视化方法进行显示。因此,SOM 算法可视化过程的最后一步就是通过自组织映射网络可视化算法对自组织映射网络上的神经元进行数学运算,获得结果再由图形图像的形式进行可视化表达。当前,自组织映射网络的可视化方法使用最多的是 U-Matrix 算法可视化方法。

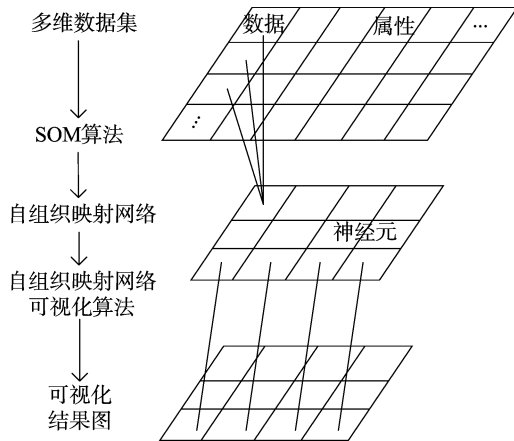


图1 SOM神经网络可视化示意

U-Matrix 算法(unified distance matrix) 又称为统一距离矩阵,由 Ultsch 等于 1990 年为了实现对 Kohonen 提出的 SOM 算法进行可视化表示而提出。该算法解决了 SOM 算法无法对聚类结果进行抓取的缺点,在 SOM 算法保证数据拓扑关系的情况下实现对聚类结构的提取和现实。基于 U-Matrix 算法的自组织映射网络可视化方法为一种基于距离度量的方法,其基本结构如图 2 所示,为一种二维网络形式的结构。SOM 自组织映射网络中每个神经元分别对应 U 矩阵算法中 4 个网络节点,节点用于存储每一神经元与相邻神经元距离参数值。如节点 1 存储 a 与水平相邻神经元 b 的权向量之间距离,节点 2 存储 a 与纵向相邻神经元 c 的权向量之间距离,节点 3 存储 a 与对角线相邻神经元 d 的权向量之间距离,节点 4 存储这 3 个距离平均值。

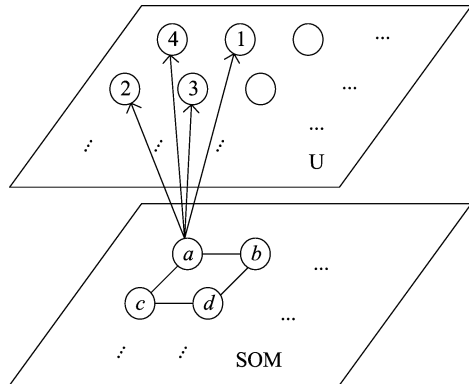


图2 U-Matrix算法结构示意图

考虑到数据进行 SOM 神经网络降维聚类后,不同的聚类数据之间相似性较小。通过距离度量表示就是距离越大其差

异性越大,数据类型越不同,反之则为同一聚类。这就是 U-Matrix 可视化能够表达 SOM 神经网络聚类结果的主要依据。U-Matrix 可视化通过对自组织映射网络中各神经元的权向量与其相邻神经元权向量之间进行距离运算获得运算结果,将之存储于 U-Matrix 网络中,并用不同颜色对数据值进行表示,从而实现自组织映射网络中的聚类结构进行分离和显示^[14]。通常,U 矩阵二维网格多由六边形来表示,当 U 矩阵数值较大,则表示该位置为聚类的边缘部分,反之则为聚类本身。

2 基于 SOM 的协同多视图可视化

协同多视图可视化方法相较于传统的基于 SOM 的降维数据可视化方法,在利用 U-Matrix 算法可视化实现对其聚类结果拓扑关系显示的同时,引入平行坐标和时空立方体等可视化工具。前者通过折线为聚类后特征数据的属性关系进行直观二维平面可视化表达,补充说明数据各属性之间的关系。后者利用平面对聚类后特征数据的时空关系进行直观三维平面可视化展示,补充说明数据在时间和空间维度上特征关系。通过多种工具协同可视化,利用计算机组件实现数据的实时交互,对颜色等特征表示方法进行统一说明,从而增强各可视化工具之间的联系,达到时空数据多窗口协同效果,能够实现对高维数据的全方面可视化表达,增强可视化效果。

2.1 平行坐标可视化

作为高维数据在二维平面上较为经典的可视化表达方法,平行坐标(parallel coordinates)由 Alfred 在 1985 年提出,平行坐标是基于几何可视化技术的代表。其基本思想是将 n 维数据属性空间通过 n 条等距离的平行轴映射到二维平台上,每一属性维由一条轴线表示,轴线上的取值范围由对应属性的最小值到最大值均匀分布^[8]。

平行坐标的优点在于能够通过射影几何解释和对偶特征直观地表达数据之间的关系,而不必使用矢量或其他可视图标,易于理解。缺点在于随着数据量的增大,折线密度增大导致大量交叠线,难于辨识^[15]。研究对原始数据进行降维,再进行平行坐标可视化,能够很好地规避其缺点,并充分利用其优势对多维数据关系进一步补充说明。图 3 为基于 SOM 降维的平行坐标可视化结果,针对 SOM 降维所得聚类结果,获得每类中的数据典型作为输入变量,参考数据抽象、坐标轴交换、维数控制等基于平行坐标的可视化技术,按照数据属性维度设置平行轴,实现时空数据的可视化表达,并为各分类分配对应颜色,方便观察分析。

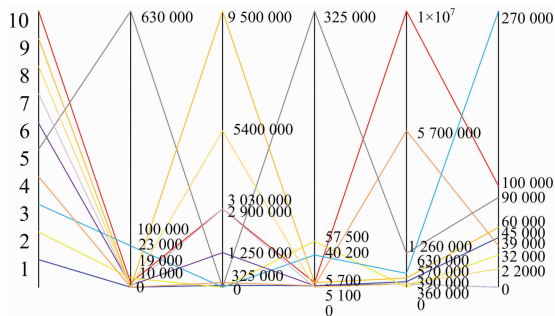


图3 平行坐标图

2.2 时空立方体可视化

时空立方体的概念最初于 1970 年由 Hagerstraand 提出,主要用于实现对时空路径的表达^[16-17]。如图 4 所示,每个平面代表某地理现象在某时间段的变化情况,为一时间平面,由多个时间平面构成时空立方体,以此来表达时空数据的变化^[18-19]。其主要优点在于能够将时空数据在三维的立方体中进行表达,突出表现地理现象随时间的变化情况。与平行坐标相似,当数据量较大且属性维度较多时,将造成平面交叠、路径混乱、多属性难以表示等问题。通过 SOM 降维技术对数据进行聚类降维后,数据属性维度减少,可以利用颜色表示其分类情况,借助地图等地理显示工具,完成时空数据表达,展示 SOM 降维后数据的时空属性关系,从而实现对高维时空数据的可视化。

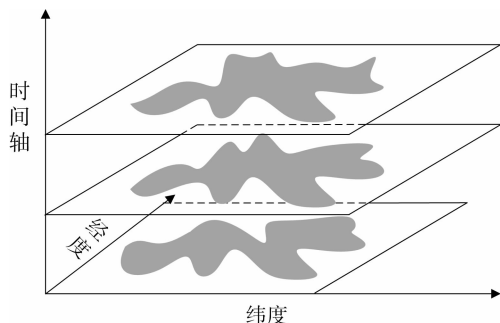


图4 时空立方体概念图

3 平台搭建及试验效果

3.1 平台搭建

为验证时空数据可视化分析方法的实际应用效果,以美国农产品时空数据为例,通过 Visio studio 2010 平台,利用 C# 语言,建立基于 SOM 的美国农产品时空数据可视化分析平台。研究平台主要分为 3 个模块,分别为数据层、挖掘层、可视化界面层(图 5)。

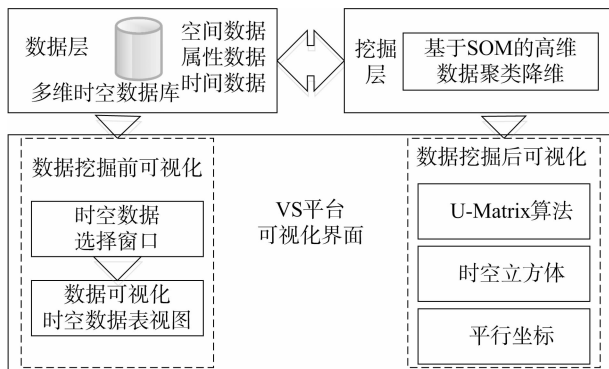


图5 基于 SOM 的农业多维时空数据集成可视化平台模块设计

数据层用于存储时空数据,支持各模块进行数据调用。挖掘层基于 SOM 的高维数据降维和挖掘进行数据分类,为可视化提供数据基础。可视化界面层为平台可视化展示层,用于对数据进行可视化表达。可视化主要包括数据挖掘前后的可视化表示方式 2 种,前者主要用于对数据进行检索,通过数据检索获取感兴趣的数据源,后者主要是对前者选取的数据源进行基于 SOM 降维聚类后的数据可视化表达。主要表达

工具有 U - Marrix 算法、平行坐标、时空立方体等。

3.2 研究数据源

研究数据源为美国 NASS 网站的农业调查数据,其原始数据按照年份与地域尺度(州、县)进行划分,主要分为农作物、动物及产品、农业经济、农业环境 4 部分数据,其中农业经济数据包括农业资产和农业租借费用两方面。鉴于原始数据的缺陷,首先对数据进行筛选处理,最终用于本研究的数据源为其农作物数据源。数据主要包含美国 50 个州 2001—2010

年 19 种农产品各自的总面积、总收成及总产量。

3.3 试验效果

对数据进行归一化处理,输入 SOM 神经网络进行数据降维,并对结果进行可视化展示。将可视化流程和平台设计最终所得可视化结果通过农业时空数据可视化平台的可视化界面层进行展示,达到数据的多角度显示和分析的作用,结果如图 6 所示。

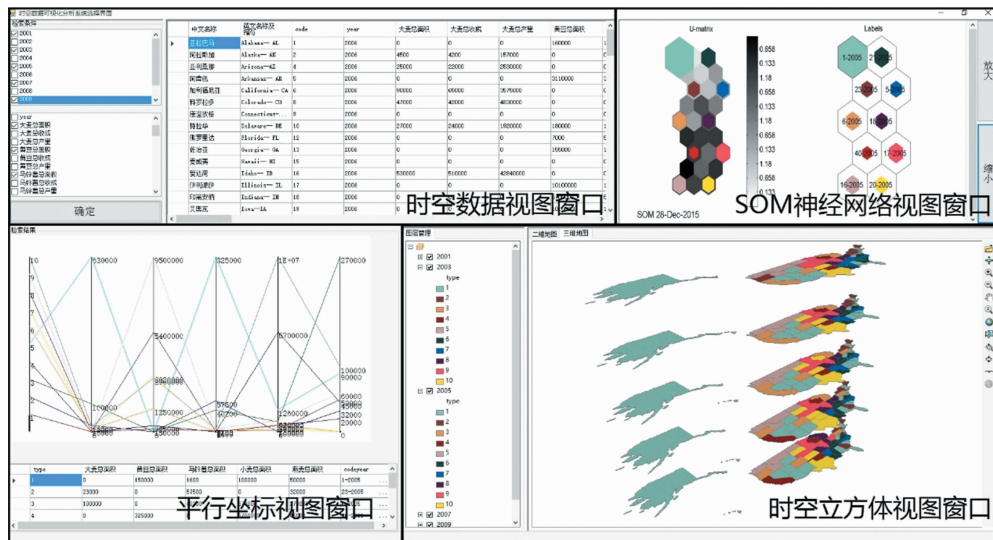


图6 基于 SOM 的多维农业时空数据集成可视化分析平台界面

3.3.1 时空数据视图窗口 时空数据视图窗口由时空数据选择窗口和时空数据表视图窗口两部分组成。通过时空数据选择窗口对最初的多维时空数据进行选择,并将筛选出的数据以二维表的形式在时空数据表视图窗口显示。其目的在于实现用户与系统之间交互,通过对海量数据的筛选,达到任意时空、属性维度数据挖掘和分析的目的。本研究按照年份和属性选择 2001、2003、2005、2007、2009 年的大麦、大豆、马铃薯、小麦、燕麦的总面积等 5 个属性数据,探索其间关系。

3.3.2 SOM 神经网络视图窗口 SOM 神经网络视图窗口显示数据进行 SOM 神经网络聚类降维后,进行 U - Marrix 算法可视化的结果。窗口左图为 U 矩阵图,右图为 labels 图,按照 U 矩阵图,灰度深浅表示相邻区域距离的远近,颜色大小是权重的多少,labels 图上显示最为代表性的区域及其时间。2005 年亚拉巴马州的数据(1—2005)和 2005 年俄克拉何马州的数据(40—2005)分别最为接近此时权重的代表值。研究基于 SOM 的神经网络可视化主要通过 Matlab 软件实现,采用混合编程的模式,利用 SOM Toolbox 作为 SOM 神经网络的软件工具,利用 Matlab 软件中 Deploytool 工具将 m 函数编译为 .NET 程序集,再由 C# 软件调用。

3.3.3 平行坐标视图窗口 平行坐标视图窗口用于对 SOM 神经网络视图窗口中显示的代表性区域的某年数据信息的补充说明。如在 SOM 神经网络视图显示的权重最小的 2005 年俄克拉何马州数据,其大麦、大豆、马铃薯、小麦、燕麦的总面积的参数分别为 0、325 000、0、5 700 000、45 000。同时,通过轴线形象化表示,达到最直观的比较数据参数的目的。如在代表性数据中,大豆、马铃薯总面积最大的是 2005 年爱达荷

州(16—2005),燕麦总面积最大的是加利福尼亚州(6—2005)。平台平行坐标可视化功能直接在 VS 上由 C# 中 GUI + 界面设计完成。

3.3.4 时空立方体视图窗口 时空立方体窗口的每一平面代表某一年份以显示地理空间属性数据,其区域颜色与 SOM 神经网络算法颜色匹配。研究主要通过 VS 平台上应用 ArcEngine 组件库,将 ArcScene 的 3D 建模移植到平台窗口,进行时空立方体的构建。通过观察颜色的变化情况,可直观看出随年份的变化各地区颜色变化情况。如俄勒冈州,颜色从 2001、2003、2005、2007 年的橙色到 2009 年的绿色;华盛顿州、蒙他拿州、爱达荷州、北达科州等没有发生颜色的变化。

3.3.5 平台协同多视图实现 通过在时空数据视图窗口对研究数据进行筛选,获取所需研究数据进行 SOM 降维,在对降维数据进行可视化表达后,获得各个窗口的显示结果,通过对结果进行观察可获得数据内在联系,挖掘有用信息。例如,对时空立方体视图窗口进行观察,由图 7 可知,南达科他州的颜色变化从 2001、2003、2005、2007 年的深粉色变成 2009 年的黄色,同样的还有南卡罗来纳州,其颜色变化从 2001、2003、2005、2007 年的浅蓝色变成 2009 年的灰蓝色,说明 2 个州在前 4 年作物种植面积相似,而 2009 年有一定变化。针对这一现象,对时空数据视图窗口显示的相关数据进行观察(图 8),发现南达科他州的作物种植面积的确在 2009 年较前几年有明显改变,表现在大麦、燕麦总面积有明显减少。南卡罗来纳州也存在相似情况,2009 年大豆总面积有明显提升。同样的,对 SOM 神经网络视图窗口所得结果观察可知,浅蓝色在时空数据显示上最具代表性,数据向 2005 年亚拉巴马(1—

2005)数据聚拢。这一结论亦可由时空立方体视图窗口显示结果观察获得:从时空立方体平面图展示结果来看,2001、2003、2005、2007、2009 年作物生产面积上大部分显示浅蓝

色。当需要研究这一聚类数据维度关系具体情况时,可通过平行坐标窗口观察获得。

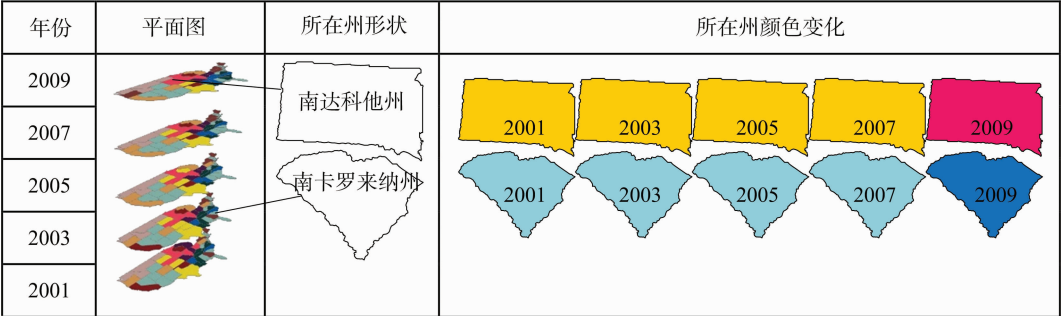


图7 时空立方体检索结果

	中文名称	英文名称及缩写	type	codeyear	大豆总面积	大豆总面积	马铃薯总面积	小麦总面积	燕麦总面积
	南达科他	Southdako...	10	46-2005 ...	65000	3900000	0	3315000	380000
	南达科他	Southdako...	10	46-2007 ...	56000	3250000	0	3508000	330000
	南达科他	Southdako...	10	46-2003 ...	75000	4250000	1000	3078000	420000
	南达科他	Southdako...	10	46-2001 ...	90000	4500000	2800	3020000	350000
	南达科他	Southdako...	9	46-2009 ...	48000	4250000	0	3209000	200000
	南卡罗来纳	Southcaro...	1	45-2005 ...	0	430000	0	170000	35000
	南卡罗来纳	Southcaro...	1	45-2007 ...	0	460000	0	160000	33000
	南卡罗来纳	Southcaro...	1	45-2003 ...	0	430000	0	200000	40000
	南卡罗来纳	Southcaro...	1	45-2001 ...	0	440000	0	200000	50000
	南卡罗来纳	Southcaro...	6	45-2009 ...	0	590000	0	165000	30000

图8 时空数据检索结果

总体而言,利用 SOM 降维技术对高维数据进行聚类降维后,通过 U - Marrix 算法、平行坐标和时空立方体等可视化工具对时空数据进行协同可视化表达,能够解决传统的 SOM 降维可视化表达对高维时空数据时空关系说明存在的缺陷,清晰展现多维农业时空数据在时间维度上的地理空间变化,说明数据的聚集和分布情况,方便对数据的潜在关系进行挖掘和分析。

4 结论

通过聚类降维算法和多种其他可视化方法的结合使用,实现多视窗协同可视化,克服了单一的可视化方法存在的维度、样本量限制等问题,较大地提高挖掘效率,为多维时空数据的分析挖掘提供新思路,在对海量农业数据挖掘和分析中,能够提供良好的技术支持,有利于精细农业的发展和推广,具有一定的经济效益和社会效益。然而,不同聚类方法对数据集的拓扑特性拟合能力不同,导致聚类准确性存在差异,同时不同可视化方法数据可视化表达侧重点各有不同。对此,针对不同农业数据集,如何寻找合适的聚类降维算法,如何确定其拓扑分布,判断其聚类准确性,选用合适可视化工具展示其时空关系,值得进一步研究。

参考文献:

[1] 钟运琴,方金云,赵晓芳. 大规模时空数据分布式存储方法研究[J]. 高技术通讯,2013,23(12):1219-1229.
[2] 齐志. 基于 SOM 神经网络的聚类可视化方法研究[D]. 长春:东北师范大学,2009.

[3] 李杨,郝志峰,谢光强,等. 质量度量指标驱动的数据聚合与多维数据可视化[J]. 智能系统学报,2013,8(4):299-304.
[4] 杨彦波,刘滨,祁明月. 信息可视化研究综述[J]. 河北科技大学学报,2014,35(1):91-102.
[5] Thakur S, Hanson A J. A 3D visualization of multiple time series on maps[C]. Proceedings of the Proceedings of the International Conference Information Visualisation,2010:336-343.
[6] 刘芳,Tian K,Zhou Z G,et al. 基于 SOM 聚类的可视化方法及应用研究[J]. 计算机应用研究,2012,29(4):1300-1303,1306.
[7] 王开军. 多维数据的聚类结果可视化技术综述[J]. 福建师范大学学报(自然科学版),2012,28(4):115-124.
[8] 雷君虎,杨家红,钟坚成,等. 基于 PCA 和平行坐标的高维数据可视化[J]. 计算机工程,2011,37(1):48-50.
[9] 闫绪娴. 中西部地区自然灾害社会易损性空间特征分析[J]. 经济地理,2014,34(5):34-40.
[10] 芮小平,张立强. 基于 SOM 的多维信息可视化研究[J]. 应用基础与工程科学学报,2011,19(3):379-388.
[11] 李权. 面向多维数据及微博社交网络的可视分析技术的研究[D]. 北京:清华大学,2012.
[12] Andrienko G, Andrienko N, Bremm S, et al. Space - in - time and time - in - space self - organizing maps for exploring spatiotemporal patterns[J]. Computer Graphics Forum,2010,29(3):913-922.
[13] Obu - Cann K, Fujimura K, Tokutaka H. Data mining of power transformer database using self - organising maps[J]. Proceedings of ICU International Conferences on Info - tech and Info - net,2001,18(4):44-49.
[14] Vesanto, Alhoniemi E. Clustering of the self - organizing map[J]. IEEE Transactions on Neural Networks,2000,11(3):586-600.

刘家福,王鑫全,韩程,等. 基于 GIS 与 CA - Markov 的吉林西部土地覆被演变及预测[J]. 江苏农业科学,2017,45(20):249-253.
doi:10.15889/j.issn.1002-1302.2017.20.062

基于 GIS 与 CA - Markov 的吉林西部 土地覆被演变及预测

刘家福¹,王鑫全¹,韩程²,李林峰¹

(1. 吉林师范大学旅游与地理科学学院,吉林四平 136000; 2. 吉林师范大学博达学院,吉林四平 136000)

摘要:为探究吉林西部地区土地利用类型演变规律,选取 1987 年、2000 年、2013 年 3 期专题制图仪(TM)影像为数据源,分析 2000—2013 年吉林西部地区土地利用类型覆被演变;运用 CA - Markov 模型模拟 2013 年预测数据,与解译数据进行模型验证,整体 kappa 系数为 76.3%,模拟精度较高;并基于 2000 年、2013 年解译数据,模拟预测了 2026 年和 2039 年研究区土地利用格局,对各时段土地动态进行对比分析。结果表明,吉林西部地区土地利用以旱田为主,占研究区总面积 50% 以上;在 2000—2013 年年间,水田、居民工矿用地和交通用地面积增加明显,林地和沼泽地面积减幅较大,其中面积增加最多的是水田;2013—2039 年土地类型大体保持 2000—2013 年的演变趋势,即随着人类活动加强,居民及工矿用地、水田面积持续快速增加,林地和沼泽地面积不断减少,吉林西部土地利用结构正在发生改变,生态保护形势不容乐观。

关键词:CA - Markov 模型;吉林西部;土地利用;模拟预测

中图分类号:F323.211;S127 **文献标志码:**A **文章编号:**1002-1302(2017)20-0249-05

近年来,随着社会经济的快速发展和全球气候变化的不稳定,人地关系矛盾日益突出,正在加速的土地利用类型变化或土地覆盖类型变化会对区域内自然生态环境和社会经济环境产生一系列影响,因此土地利用类型变化的研究愈发重要^[1-3]。土地利用类型变化的研究通过对土地利用格局的动态模拟,重建了过去某一时段内土地覆被所发生的系列变化,并对未来土地利用类型可能发生的变化趋势作出科学合理的预测,可为生态环境保护和区域可持续发展提供科学的决策依据,成为国际研究最热门的领域之一^[4-6]。

目前,Markov 模型在土地利用模拟预测中被广泛应用,CA 模型则多被应用于森林防护、水土保持、灾害预测等领域,一般在应用时多采用与其他方法相结合的方式^[7-10]。将 CA 与 Markov 模型相结合应用于土地利用类型分析与预测可保留 Markov 模型长期预测的优势,同时结合了 CA 模拟空间

复杂状态的功能,可有效模拟土地利用变化。吉林西部地区位于农牧业交错地带,生态环境脆弱,土地利用类型复杂多样,人类活动对生态环境产生了重要影响,在分析与预测吉林西部土地覆被结构与变化特点时引入 CA - Markov 模型,可有效揭示未来土地利用类型的变化特点与趋势,为地区土地利用可持续发展提供重要参考依据。

1 研究区概况

吉林西部地区总面积 4.689 万 km²,占全省土地面积约 27.2%,地理坐标为 121°38'~126°11'E,43°15'~46°18'N,行政区划包括白城市和松原市。该区位于松嫩平原南部,地处农牧交错地带,海拔为 150~200 m,地势平坦低平,大部分地区为平原,该地区属于温带半干旱半湿润季风气候区,四季分明,全年温度变化较大,春秋多风少雨,冬季严寒少雪,雨热同期,年平均降水量 350~500 mm,河流稀少。吉林西部地区是省内重要的农、牧业生产基地和能源生成基地,近年来伴随着经济发展与区域开发,水域面积不断减少,湿地退化与草场退化十分严重,对当地的发展造成严重的挑战。

2 研究方法

2.1 数据来源及预处理

选取吉林西部地区 1987 年、2000 年和 2013 年 LANDSAT

Conference on IEEE,2004:145-152.

[18] Hadlak S, Tominski C, Schulz H J, et al. Visualization of attributed hierarchical structures in a spatiotemporal context[J]. International Journal of Geographical Information Science, 2010, 24(10): 1497-1513.

[19] 吴正升, 胡艳, 何志新. 时空数据模型研究进展及其发展方向[J]. 测绘与空间地理信息, 2009, 32(6): 15-17, 21, 24.

收稿日期:2016-11-15

基金项目:中国科学院重点部署项目(编号:KZZD-EW-08-02);

吉林省科技发展计划(编号:20150204047SF);吉林省高校科技与“十二五”科研规划(编号:吉教科合字[2015]第 222 号)。

作者简介:刘家福(1975—),男,吉林敦化人,博士,副教授,从事遥感与 GIS 在资源环境、自然灾害等领域的应用研究。Tel:(0434) 3296005;E-mail:liujiafu750506@126.com。

[15] 孙扬,唐九阳,汤大权,等. 改进的多变元数据可视化方法[J]. 软件学报,2010,21(6):1462-1472.

[16] Qi F, D Fei. Tracking and visualization of space-time activities for a micro-scale flu transmission study[J]. International Journal of Health Geographics, 2013, 12(1): 6.

[17] Gatalaky P, Andrienko N, Andrienko G. Interactive analysis of event data using space-time cube[C]. Proceedings Eighth International