

# 基于空间聚类方法的京津冀城市群多层次空间结构研究

张 珣<sup>1,3</sup>, 陈健璋<sup>1</sup>, 黄金川<sup>2,3,4\*</sup>, 于重重<sup>1</sup>, 陈秀新<sup>1</sup>

(1. 北京工商大学计算机与信息工程学院 食品安全大数据技术北京市重点实验室, 北京 100048;  
2. 中国科学院 区域可持续发展分析与模拟重点实验室, 北京 100101; 3. 中国科学院地理科学与资源研究所,  
北京 100101; 4. 中国科学院大学资源与环境学院, 北京 100049)

**摘要:**在京津冀协同发展战略背景下,以建设京津冀世界级城市群为引领,遵循城市发展规律,优化城市空间布局,明确京津冀城市群等级结构及其空间特征具有重要意义。本文以京津冀城市群156个区县为研究对象,从经济中心性、交通中心性、信息中心性、人口中心性4个角度,利用4种空间聚类方法进行5个等级的聚类分析,并基于克氏中心地理论对京津冀城市群等级划分结果进行空间结构分析。结果显示,自组织特征映射神经网络算法(SOM)较适合京津冀城市群的等级划分;京津冀城市群正从以北京城区为单核心的圈层空间结构向3条带型空间结构转变,其中京津都市发展走廊发育成熟,沿海都市发展带也初具规模,而包括雄安新区在内的京石都市发展带正在孕育。

**关键词:**空间聚类;空间结构;京津冀城市群

## 1 引言

随着全球化进程日益加快,区域竞争的主体已经从区域范围内城市间的竞争转向城市群经济实体在更广阔范围内的全球竞争。城市群逐渐成为中国经济增长的核心,是中国新型城镇化的重要载体及国家区域政策的重要空间单元(黄金川等, 2014)。由于城市一区域系统的空间异质性、空间复杂性和空间开放性特点,城市间存在较大差异并且表现出发展不协调(牛方曲等, 2015)。城市群空间从模块化向网络化发展的趋势,迫切需要对城市群网络层级结构进行研究。城市的中心地理论广泛用于划分城市等级层次,描述城市体系空间交互

作用的结构,还可对城市体系进行横向、纵向比较提供指导(Preston, 1970; Marshall, 1989; 周一星等, 2001)。

近年来,利用空间聚类方法探索城市区域系统逐渐成为研究热点。李新延等(2005)利用DB-SCAN(Density-Based Spatial Clustering of Applications with Noise)算法,基于空间相似性对某城市商业网点和中小学等公共设施进行了空间聚类分析,该研究尚未考虑属性数据影响;Gong(2010)应用ARMA-GRNN神经网络模型对1991-2008年北京—上海的空中客流进行了分析预测;陈园园等(2011)运用SOM神经网络分级模型,基于经济和交通指标,评价辽中南城市群10个节点城市的空间联

收稿日期:2017-07;修订日期:2017-10。

**基金项目:**教育部人文社会科学研究青年基金项目(15YJCZH224);北京市自然科学基金青年项目(9164025);首都流通业研究基地内设课题一般项目(JD-YB-2017-024);北京市教委科研计划面上项目(KM201510011010)[**Foundation:** Research Project of Humanities and Social Science on Youth Fund of the Ministry of Education, No.15YJCZH224; Beijing Natural Science Foundation, No.9164025; Opening Fund of Capital Circulation Industry Research Base, No.JD-YB-2017-024; Scientific Research Project of Beijing Educational Committee, No.KM201510011010]。

**作者简介:**张珣(1986-),男,吉林辽源人,副教授,硕导,从事商业地理分析、GIS软件技术研究,E-mail: zhangxun@btbu.edu.cn。

**通讯作者:**黄金川(1973-),男,河南开封人,副研究员,主要从事城市地理与区域规划研究,E-mail: huangjc@igsnr.ac.cn。

**引用格式:**张珣,陈健璋,黄金川,等. 2017. 基于空间聚类方法的京津冀城市群多层次空间结构研究[J]. 地理科学进展, 36(11): 1359-1367. [Zhang X, Chen J Z, Huang J C, et al. 2017. Multi-level spatial structure analysis of urban agglomeration in the Beijing-Tianjin-Hebei region based on spatial clustering algorithms[J]. Progress in Geography, 36(11): 1359-1367.]. DOI: 10.18306/dlkxjz.2017.11.005

系能力并进行层次划分;杨志民等(2015)构建基于K-means方法的金融中心登记识别模型,对长三角城市群的各级城市金融中心级别进行了识别。顾朝林等(2015)采用因子分析和聚类分析方法对绍兴城市群基于功能区的行政区调整进行了研究。方碧琪等(1991)利用基于距离划分的聚类方法研究北京市各区县经济分类;马林靖等(2010)应用聚类方法对进行研究,分析天津12个涉农区县的均衡发展情况;袁媛等(2014)应用SOFM网络对河北省县域贫困度进行分类。

可以看出,空间聚类算法对于区域空间层级分析具有重要价值,但是相关研究绝大部分只应用了基于欧氏距离的聚类方法,虽然这种算法易于理解和实现,但是对分布规律性不明显的现象聚类结果往往不太理想,层级划分的准确度较低。而且,目前针对京津冀城市群区域空间结构研究的分级指标考虑较单一,划分结果说服力不足。因此,本文着重从政治地位、经济实力、城市规模、区域辐射力等多个角度构建中心性指标体系,试验4种聚类方法在城市群空间结构划分中的适用性并进一步基于中心地理论深入分析京津冀城市群网络层级的空间结构特征。

## 2 研究区域与研究方法

### 2.1 京津冀城市群

京津冀城市群为中国三大城市群之一。按照国家发展和改革委员会界定,京津冀城市群范围涵盖北京市、天津市和河北省的石家庄、唐山、保定、秦皇岛、廊坊、沧州、承德、张家口8个地市及其所属的通州新城、顺义新城、滨海新区和唐山曹妃甸工业新区,面积为18.84万 $\text{km}^2$ (刘卫东等, 2011)。2016年,京津冀城市群生产总值68857.15亿元,总人口1.1亿(李国平等, 2017)。目前,随着国家京津冀协同战略的实施,京津冀城市群存在核心城市带动作用不明显、区域发展不协调等问题(张旺等, 2012)。本文以京津冀156个区县为研究对象进行等级划分和空间结构分析,其中,北京、天津2个直辖市和河北省8个地级市所辖城区如表1。

### 2.2 空间聚类算法

空间聚类算法是典型的无监督机器学习算法,能在描述数据是如何组织或聚类的时候,从总体中给出的样本信息对总体做出某些推断(闫友彪等,

2004)。经过聚类后的数据,可根据其数据特征生成以簇为单位的数据对象集合。同一个簇中的对象彼此相似,不同簇中的对象彼此相异(金建国, 2014)。聚类算法旨在生成对象彼此相似的簇,且实现簇内对象相似度最大,簇间对象相似度最小(Ferrari et al, 2015)。将聚类算法应用于城市群内网络层级划分,是将城市群内区县为研究对象,基于多维指标数据,对所属区县进行更为综合和系统的划分。划分的结果应体现出簇内各区县特征相对一致,簇间各区县数据特征相差最大,即同一发展特征的区县聚为一簇。空间聚类算法的分析重点在于对基于空间距离的空间相似性和基于属性的非空间相似性的融合。从算法的角度,需要讨论算法是否有较好的可伸缩性、效率,算法能否处理不同类型属性的数据,是否能够发现任意形状的聚类,以及是否具有较好的处理噪声数据的能力(Grubestic et al, 2014)。主要的聚类算法包括以下4种(Duda et al, 2001):基于划分的聚类算法(Hagen et al, 2006)、基于层次的聚类算法(Gómez et al, 2015)、基于密度的聚类算法(Zhou et al, 2011)、基于神经网络的聚类算法(Lin et al, 2007)。本文对4种聚类算法进行比较试验研究。

#### (1) K-means

K-means是基于划分的聚类方法。其聚类过程为:首先,需要确定 $K$ 个质心,即希望聚成的簇的数量;其次,鉴于每个指标代表空间内一个维度的坐

表1 直辖市、地级市城区统计表

Tab.1 Urban districts of provincial-level municipalities and prefectural-level cities

城区名称	包含区
北京城区	东城区、西城区、朝阳区、丰台区、海淀区、石景山区
天津城区	和平区、河东区、河西区、南开区、河北区、红桥区
保定城区	莲池区、竞秀区(含高新区)、徐水区、清苑区、满城区
沧州城区	新华区、运河区
承德城区	双桥区、双滦区、高新区、营子区
邢台城区	桥东区、桥西区
张家口城区	桥东区、桥西区、宣化区、下花园区、万全区、崇礼区
唐山城区	路北区、路南区、古冶区、开平区、丰润区、丰南区、曹妃甸区
石家庄城区	桥西区、新华区、长安区、裕华区、井陉矿区、藁城区、鹿泉区、栾城区
廊坊城区	安次区、广阳区
邯郸城区	丛台区、复兴区、邯山区、峰峰矿区
秦皇岛城区	海港区、山海关区、北戴河区、抚宁区

标,计算各指标的欧氏距离,并集成计算对象间的空间距离;第三,多次运行,使每次每组随机选择质心,直到选取具有最小误差平方和的簇集;第四,计算所有像素点到质点的距离,假使像素点 $P$ 距离质点 $S$ 最近,那么 $P$ 属于 $S$ 簇;最后,迭代算法直至聚类结果稳定。

### (2) Chameleon

Chameleon 聚类算法属于自下而上的层次聚类算法。但与一般的凝聚聚类算法不同,Chameleon 算法以树的形式分为2个主要阶段(Gómez et al, 2015)。第一个阶段将分散的对象点,采用 $k$ 最近邻算法连接起来,形成若干小簇。簇的每一个节点表示数据项,每一条边表示数据项的相似度,各指标集成计算出的空间欧氏距离即为每一条边的长度;第二个阶段计算任意2个簇的相对互连性 $RI$ 和相对互连性 $RC$ 和度量函数 $Metric$ ,合并2个子簇。合并过程为:首先给定度量函数距离阈值 $MinMetric$ 。访问每个簇,计算它与邻近的每个簇的相对互连性 $RI$ 和相对互连性 $RC$ ,通过度量函数公式计算出临时距离 $TempMetric$ 。找到最大的临时距离,如果最大的临时距离超过距离阈值,则将簇与此值对应的簇合并。如果找到的最大的临时距离没有超过距离阈值,则表明此聚簇已合并完成,移除聚簇列表,加入到结果聚簇中。递归此步骤,直到待合并聚簇列表最终大小为空(Gómez et al, 2015)。

$$RI(C_i, C_j) = \frac{2 \times |EC(C_i, C_j)|}{|C_j|EC(C_i) + |C_i|EC(C_j)} \quad (1)$$

$$RC(C_i, C_j) = \frac{(|C_i| + |C_j|)EC(C_i, C_j)}{|C_j|EC(C_i) + |C_i|EC(C_j)} \quad (2)$$

$$Metric = RI(C_i, C_j) \times RC(C_i, C_j)^\alpha \quad (3)$$

式中: $|C_i|$ 为簇 $i$ 内数据点的个数; $EC(C_i)$ 为簇 $i$ 内所有边的加权和; $EC(C_i, C_j)$ 为跨越2个簇的所有边的加权和。

### (3) DBSCAN

DBSCAN 是基于密度的聚类算法,在计算中需先确定2个参数:一是簇扫描半径 $Eps$ ,它表示以给定点 $P$ 为中心, $Eps$ 为半径的邻域范围;二是以 $P$ 为中心的邻域内最小包含点数阈值。DBSCAN 聚类的过程为:首先,扫描整个数据集,将各指标通过计算欧氏距离的方法集成计算对象点间的空间距离,每个指标代表空间内一个维度的坐标;其次,确定

一个核心点 $P$ ,寻找从该核心点出发的所有密度相连的数据点对点 $P$ 所在簇进行扩充(Baseli et al, 2015)。遍历该核心点 $P$ 为中心, $Eps$ 为半径的邻域内的所有核心点,寻找与这些数据点密度相连的点,直到没有可以扩充的数据点为止,使聚类成簇的边界节点都是非核心数据点;第三,重新扫描剩下的数据集,寻找没有被聚类的核心点,再重复上面的步骤,对该核心点进行扩充直到数据集中没有新的核心点为止。

### (4) SOM

SOM 聚类由 Tenno Kohonen 提出(Kohonen, 1990),是一种无监督学习的聚类算法(Bação et al, 2005)。它由 $m$ 个神经元组成的输入层和由 $a \times b$ 个神经元组成的二维平面阵列竞争层(或映射层)组成,两层之间神经元全部连接。为每个输入在竞争层中寻找最佳匹配点,即胜出点,并用梯度下降法更新节点参数。SOM 聚类过程为:首先,对网络初始化,即对输入层和竞争层间的权值进行随机初始化,对 $m$ 个输入神经元到输出神经元连接权值赋予较小权值,每个节点的参数个数与输入维度相同;其次,把输入向量 $X$ 输入给输入层,找到与其最匹配的节点,计算映射层的权值向量和输入向量的欧氏距离。

$$d_j = \|X - W_j\| = \sqrt{\sum_{i=1}^m (x_i(t) - w_{ij}(t))^2} \quad (4)$$

式中: $w_{ij}$ 为输入层的 $i$ 神经元和竞争层的 $j$ 神经元之间的权值。

通过计算,得到具有最近距离的胜出神经元即为 $j^*$ ,并给出临近神经元集合。根据式(5)修正 $j^*$ 及其邻近神经元的权值。

$$\Delta w_{ij} = w_{ij}(t+1) - w_{ij}(t) = \eta(t)(x_i(t) - w_{ij}(t)) \quad (5)$$

式中: $\eta$ 为在(0,1)区间内的常数,随时间变化逐渐下降到0。

因为不同数据集的数据特征不同,所以并不存在一种聚类算法适用于所有的数据集。4种聚类算法的优缺点见表2。

可以看出,K-means 和 Chameleon 算法只能对点数据和基于距离的空间相似性进行分析聚类,无法做到对不同属性的数据进行相似性分析。DBSCAN 可以同时考虑空间和属性数据的相似性,SOM 作为辅助性聚类方法也同时考虑了空间和属性特征。

表 2 4 种聚类算法效果比较表  
Tab.2 Comparison of the characteristics of four clustering algorithms

算法	算法效率	适合的数据类型	发现的聚类类型	对噪声数据或异常数据的敏感性	对输入顺序的敏感性	其他
K-means	高	数值	凸形或球形	不敏感	一般	易于理解和实现,时间复杂度低
Chameleon	一般	任意	任意形状	不敏感	不敏感	算法灵活但执行时间长
DBSCAN	一般	数值	任意形状	敏感	敏感	处理高纬度数据效果差
SOM	高	任意	任意形状	不敏感	敏感	结果有特征映射,能可视化

### 3 研究数据

京津冀城市群各区县的等级应由该区县的中心性来衡量,中心性越大,为周围地区提供的货物越多,则所要求的门槛人口越多,吸引和影响的范围就越大,中心地的等级越高;反之,中心地等级越低(Bell et al, 2015)。不同研究对中心性指标的选取不尽相同(马林靖等, 2010; 韩玉刚等, 2016)。为全面地衡量各区县的中心性,本文从经济中心性、信息中心性、交通中心性和人口中心性4个角度,综合评价各区县的等级。其中,以GDP表征经济中心性;以社交网络(新浪微博)签到数表征信息中心性;以铁路吸引子和公路路网密度表征交通中心性(马林靖等, 2010);以人口总数表征人口中心性。鉴于新浪微博为代表的新型网络社区已成为当代人们交流的新型媒介(王波等, 2016),以社交网络签到数据表征区域的信息中心性。通过互联网爬虫技术推算的客流量刻画交通中心性。网络爬取12306铁路列车班次、始发站、终点站、余票量等信息,并依此测算各区县铁路交通吸引力。选定2016年7月25-29日余票总数作为淡季余票的样本,2016年7月30-31日作为旺季余票的样本;同时选定2016年10月6日的余票作为“十一黄金周”特殊时段的余票样本。

本文假设:已获取的工作日、休息日余票数据样本作为实验样本,代表铁路客运正常运营水平。特殊节假日的数据,由“十一黄金周”客流情况作为代替。代入铁路日均客流量计算,铁路承载的客流量如下:

$$p = (d_1 \times (n_{11} - n_{12}) + d_2 \times (n_{21} - n_{22}) + d_3 \times (n_{31} - n_{32})) / 365 \quad (6)$$

式中:  $p$  为铁道交通日均人流量、 $d_i$  为旺季、淡季和节假日天数;  $n_{1i}$  为旺季放票日、发车日当天票数;  $n_{2i}$  为淡季放票日、发车日当天票数;  $n_{3i}$  为特殊节假日放票日、发车日当天票数。

为量化各个区县客流吸引能力,使之能够作为区县分级的重要指标之一,定义吸引因子 $\varepsilon$ 如下:

$$\varepsilon = p_i / p \quad (7)$$

式中:  $p_i$  为第  $i$  个区县日均到达人数;  $p$  为铁道交通日均人流量。

### 4 实验过程及结果分析

#### 4.1 聚类过程

在K-means算法实验中,通过计算不同K值聚类结果的各簇内对象点的距离均值和类间距离均值的比值,当K为5时,类内距离均值和类间距离均值比值曲线趋于平滑,且距离均值和类间距离均值比值标准差较小,所以选择聚类簇数为5。同理,SOM算法的输入神经元个数也为5。Chameleon和DBSCAN算法的各参数由实验时对照不同实验结果的参数进行微调得到(表3)。可以看出,K-means算法和SOM算法的迭代次数均是使聚类结果稳定的相对较小迭代次数。

#### 4.2 聚类结果

各聚类算法最终计算结果如表4和图1所示。从聚类结果看,K-means能成功地将京津冀城市群区域分为合理的5级,但未能识别出城市群核心区,且大多数区县被划在最低级别,这种基于简单距离划分的聚类方法,未能发现大小差别很大的簇。Chameleon算法能适应线性分布的数据几何形状,但是划分出的6类数据缺乏逻辑关系,即数值明显

表 3 聚类算法参数设置表

Tab.3 Parameters of the four clustering algorithms

聚类算法	参数设置
K-means	质心个数 $k=5$ ; 迭代次数 $n=15$
Chameleon	$k$ 最邻近值设定为 $k=3$ ; 度量函数距离阈值 $MinMetric=0.015$
DBSCAN	中心点 $P$ 的扫描邻域半径 $Eps=3$ ; 以 $P$ 为中心的邻域内最小包含点数阈值 $min=3$
SOM	输入层输入神经元 $k=5$ ; 迭代次数 $n=300$

表 4 4 种聚类算法结果  
Tab.4 Results of the four cluster algorithms

	第一类	第二类	第三类	第四类	第五类	第六类
K-means	北京市城区等 3 个区县	保定市城区等 8 个区县	北辰区等 9 个区县	霸州市等 40 个区县	安国县等 96 个区县	
Chameleon	成安县等 10 个区县	霸州市等 19 个区县	安平县等 25 个区县	北京市城区等 22 个区县	安新县等 35 个区县	安国市等 40 个区县
DBSCAN	固安县等 6 个区县	隆化县等 6 个区县	广宗县等 7 个区县	滦平县等 7 个区县	大名县等 15 个区县	噪声数据(115 个区县)
SOM	北京市城区	石家庄、天津市城区	保定市城区等 21 个区县	霸州市等 42 个区县	安国市等 90 个区县	

较小的区县被划在了较高簇群,而数值高的区县则没有被安放在较高层。DBSCAN 算法将大部分数据归于噪声,也就是说当数据集的密度不均、类间距差相差很大时,基于密度的 DBSCAN 算法聚类效果不好。

SOM 算法通过输入模式反复学习,获取各个输入数据特征,从空间数据相似性出发,基于多维数据的非空间属性特征进行自组织,基于神经元之间的距离,得到 5 个神经元内分别含有 90,42,21,2,1 个数据节点,整体呈现正三角形的空间稳定结构。从结果来看,SOM 算法将京津冀城市群各区县分为 5 级,且竞争层节点间存在逻辑和拓扑关系(图 2-3)。最上方小正六边形代表 SOM 聚类结果的第五类(安国市等 90 个区县),第二个小正六边形代表第四类(霸州市等 42 个区县),第三个小六边形代表第三类

(保定市城区等 21 各区县),第四个小六边形代表第二类(石家庄和天津市城区),最下方小六边形代表第一类(北京市城区)。由此可见,SOM 算法能准确识别区县间的数据差别并设定类间距离,并在结果中显示,比较适合城市群区域空间网络层级划分。

### 5 基于中心地理论的城市群多层次空间结构分析

根据克氏中心地理论,德国南部地区在市场、交通、行政三原则共同作用下,一个地区城镇等级体系应当由 1 个一级城市,2 个二级城市,6~12 个三级城市,42~54 个四级城市,118 个五级城市组成(Bell et al, 2015)。下面对基于 SOM 聚类算法得到的京津冀城市群 5 级结构进行空间分析。

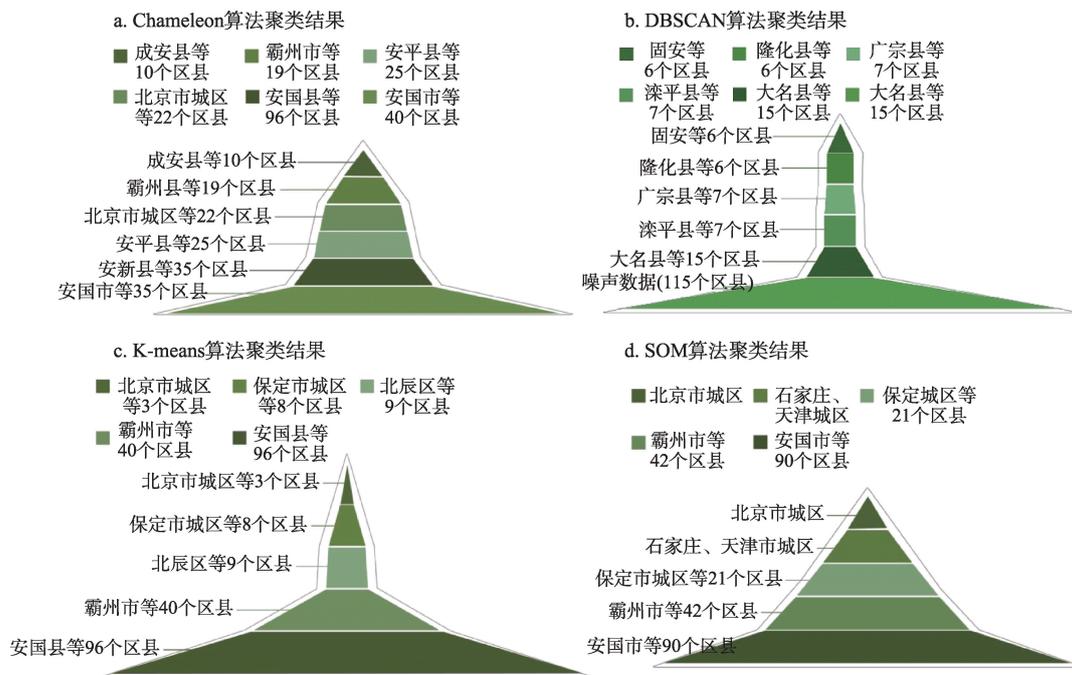


图 1 聚类结果网络层次结构特性漏斗图

Fig.1 Funnel plot of network hierarchical characteristic of the clustering result

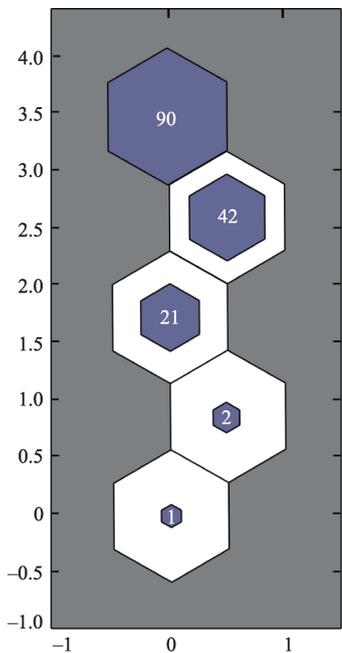


图2 神经元分布情况拓扑结构  
Fig.2 Topology of neuron distribution

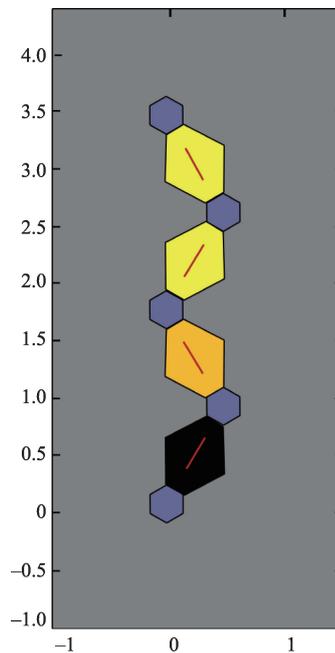


图3 临近神经元距离情况  
Fig.3 Nearby neuron distance

京津冀区域标准化后的各区县数据如表5。可以看出,北京作为全国的政治中心、文化中心、国际交流中心、科技创新中心,具有明显优势,是无可争议的一级区域辐射核心。其中,北京在人口中心性、经济中心性和信息中心性等方面优势尤为显著。从京津冀协同发展的角度,北京城区应加快推动与首都功能不符合的功能转移,辐射带动周边区县快速发展,缓解核心城区的交通、人口、就业等压力,并力争在生态、住房、教育、养老、医疗、环境、科技等领域取得引领性突破。二级区县包括石家庄城区和天津城区。其中,石家庄城区作为京津冀南部地区唯一核心,承担着联合南部区县加强承接京津两市的功能疏解和协同任务。第三级别的区县中,位于京津冀城市群东北部的唐山城区和秦皇岛城区以及位于北部的承德城区,对京津冀城市群区域协同均有重要作用。环京津的欠发达区县多处

于第四、五级,大部分位于京津冀西部和北部地区。就吸引因子而言,第二级区县路网密度平均值高于第一级,表明第二级区县的交通基础设施较第一级区县更为完善;京津冀城市群区域各区县交通中心性差距较小,说明区域发展到一定程度后,交通设施已经接近区域均等化,成为京津冀协同发展的基本条件。

从京津冀各级区县的空间结构和发展态势来看,京津冀城市群正从以北京为单核心的圈层结构向带型空间结构转变(图4)。其中,京津两市之间的区域是先行发展成熟的都市走廊,对周边区县辐射效果明显,总体发展在城市群之中处于领先水平。包括秦皇岛、唐山、天津、沧州等区县的沿海都市发展带也已初具规模,基本连片,成为京津冀沿海战略推进的重要空间。其中,天津应集中提升滨海新区的发展规模和层级,打造为北方地区的物流枢纽中心,形成最便利的海关通关,经济承接北京城区的功能疏解和科技成果转化任务。包括雄安新区、石家庄等在内的沿京广交通走廊都市发展带正在孕育,将成为京津冀城市群未来增长潜力最大的带型区域。尤其是雄安新区的开发建设,将是京津冀城市群京石都市发展带的重大引爆点,推动京津冀城市群区域结构的加速优化。作为北京、天津功能疏解的主要承接地,河北省各区县应继续加强交通网络基础设施建设,提高交通可达性。尤其是石家

表5 各级别区县数据标准化后平均值

Tab.5 Average values of every level of districts and counties after normalization

	人口	GDP	路网密度	吸引因子	微博签到数
第一级	1.000000	1.000000	0.566022	0.925568	1.000000
第二级	0.187632	0.323590	0.712439	0.824197	0.089826
第三级	0.079778	0.066928	0.263885	0.202920	0.016356
第四级	0.015132	0.036220	0.096532	0.052529	0.003622
第五级	0.008050	0.025172	0.048087	0.013577	0.000620

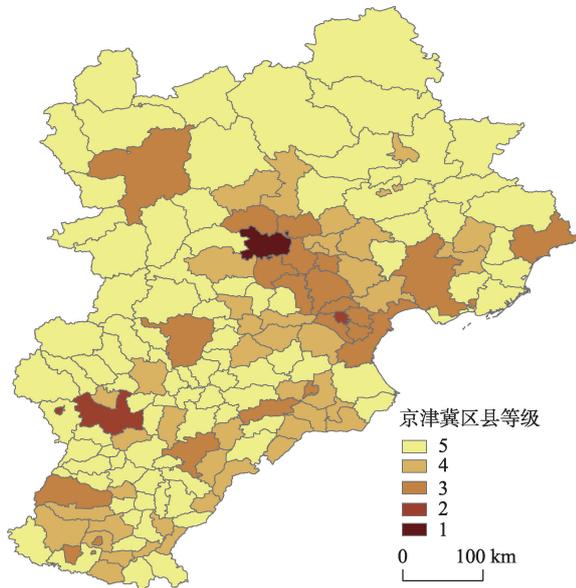


图 4 聚类各级别结果空间分布图

Fig.4 Spatial distribution of every level of clusters

庄、唐山、秦皇岛 3 城区应积极发挥次区域辐射核心的带动作用,加快城市群南部地区和东北部地区各区县的协同发展。承德和张家口作为西部地区的三级城区,应分别承担起带动京津冀城市群西北部和北部协同发展的任务,通过加强与京津两市科技、文化交流和产业对接,促进西北部和北部山区的绿色产业发展。

## 6 结论与探讨

通过 4 种聚类算法对比试验发现,非监督学习的自组织特征映射(SOM)算法比较适合于城市群区域的等级划分:一是基于神经网络的自组织特征映射(SOM)算法结果的区县簇等级结构较为合理,由低级别至高级别呈明显的三角型稳定结构;二是簇间区别明显,能够准确甄别出区域核心城区及其他级别区县。

基于 SOM 算法对京津冀城市群的聚类结果显示:①京津冀城市群正从以北京为单核心的圈层结构向 3 条带型空间结构转变;其中京津城区优势明显,在多数维度上与其他区县差距较大,应该凸显其作为区域辐射核心的带动作用;②京津冀城市群的五级数量结构符合克里斯塔勒的中心地理论,城市群空间结构较稳定;③京津冀城市群正在形成 3 条空间发展带,推动京津冀协同发展,其中,京津都市发展走廊已经发育成熟,沿海都市发展带也已初

具规模,包括雄安新区的京石都市发展带正在加速孕育。

下一步研究可集中在以下两方面:一是进一步挖掘京津冀城市群的等级数量结构和空间结构的内涵,形成更完善的城市群空间网络体系;二是将基于神经网络的自组织特征映射(SOM)聚类算法扩大到全国城市群区域的等级结构分析,根据不同城市群地区的数据特点,调整参数,完善方法,最终形成一种可适用于广泛区域的城市群等级划分通解。

## 参考文献(References)

- 陈园园, 李宁, 丁四保. 2011. 城市群空间联系能力与 SOM 神经网络分级研究: 以辽中南城市群为例[J]. 地理科学, 31(12): 1461-1467. [Chen Y Y, Li N, Ding S B. 2011. Spatial combination capacity and classification based on SOM network of urban agglomerations: A case study of central and southern Liaoning urban agglomerations[J]. Scientia Geographica Sinica, 31(12): 1461-1467.]
- 方碧琪, 汪永新. 1991. 北京市各区县经济分析的聚类方法[J]. 数理统计与管理, (1): 25-30. [Fang B Q, Wang Y X. 1991. Clustering analysis on the economy of Beijing regions[J]. Application of Statistics and Management, (1): 25-30.]
- 顾朝林, 王颖, 邵园, 等. 2015. 基于功能区的行政区划调整研究: 以绍兴城市群为例[J]. 地理学报, 70(8): 1187-1201. [Gu C L, Wang Y, Shao Y, et al. 2015. Research on administrative divisions based on functional areas analysis: A case of Shaoxing metropolitan area[J]. Acta Geographica Sinica, 70(8): 1187-1201.]
- 韩玉刚, 叶雷. 2016. 中国欠发达省际边缘区核心城市的选择与区域带动效应: 以豫皖省际边缘区为例[J]. 地理研究, 35(6): 1127-1140. [Han Y G, Ye L. 2016. Core city selection in Chinese undeveloped provincial border-region and its promoting effects on regional development: A case study of Anhui-Henan Provincial Border-region[J]. Geographical Research, 35(6): 1127-1140.]
- 黄金川, 刘倩倩, 陈明. 2014. 基于 GIS 的中国城市群发育格局识别研究[J]. 城市规划学刊, (3): 37-44. [Huang J C, Liu Q Q, Chen M. 2014. The identification of urban agglomeration distribution in China based on GIS analysis [J]. Urban Planning Forum, (3): 37-44.]
- 金建国. 2014. 聚类方法综述[J]. 计算机科学, 41(Z2): 288-293. [Jin J G. 2014. Review of clustering method[J]. Computer Science, 41(Z2): 288-293.]
- 李国平, 罗心然. 2017. 京津冀地区人口与经济协调发展关

- 系研究[J]. 地理科学进展, 36(1): 25-33. [Li G P, Luo X R. 2017. Coordinated development between population and economy in the Beijing-Tianjin-Hebei region[J]. Progress in Geography, 36(1): 25-33.]
- 李新延, 李德仁. 2005. DBSCAN空间聚类算法及其在城市规划中的应用[J]. 测绘科学, 30(3): 51-53. [Li X Y, Li D R. 2005. DBSCAN spatial clustering algorithm and its application in urban planning[J]. Science of Surveying and Mapping, 30(3): 51-53.]
- 刘卫东, 金凤君, 张文忠, 等. 2011. 中国经济地理学研究进展与展望[J]. 地理科学进展, 30(12): 1479-1487. [Liu W D, Jin F J, Zhang W Z, et al. 2011. Progress in economic geography (2006-2011)[J]. Progress in Geography, 30(12): 1479-1487.]
- 马林靖, 周立群. 2010. 基于聚类分析法的县域均衡发展指数构建与测度应用研究: 以天津12个涉农区县为例[J]. 经济经纬, (2): 57-60. [Ma L J, Zhou L Q. 2010. A research into the construction and measurement of county balanced development indexes and its application based on cluster analysis method[J]. Economic Survey, (2): 57-60.]
- 牛方曲, 刘卫东, 宋涛, 等. 2015. 城市群多层次空间结构分析算法及其应用: 以京津冀城市群为例[J]. 地理研究, 34(8): 1447-1460. [Niu F Q, Liu W D, Song T, et al. 2015. A multi-level spatial structure analysis algorithm for urban agglomeration study in China[J]. Geographical Research, 34(8): 1447-1460.]
- 王波, 甄峰. 2016. 网络社区交流中距离的作用: 以新浪微博为例[J]. 地理科学进展, 35(8): 983-989. [Wang B, Zhen F. 2016. The role of distance in online social networks: A case study of Sina micro-blog[J]. Progress in Geography, 35(8): 983-989.]
- 闫友彪, 陈元琰. 2004. 机器学习的主要策略综述[J]. 计算机应用研究, 21(7): 4-10, 13. [Yan Y B, Chen Y Y. 2004. A survey on machine learning and its main strategy[J]. Application Research of Computers, 21(7): 4-10, 13.]
- 杨志民, 化祥雨, 叶娅芬, 等. 2015. 金融空间联系及K-means聚类中心等级识别研究: 以长三角为例[J]. 地理科学, 35(2): 144-150. [Yang Z M, Hua X Y, Ye Y F, et al. 2015. Spatial combination of finance and center level identify based on K-means clustering: A case study of the Changjiang river delta[J]. Scientia Geographica Sinica, 35(2): 144-150.]
- 袁媛, 王仰麟, 马晶, 等. 2014. 河北省县域贫困度多维评估[J]. 地理科学进展, 33(1): 124-133. [Yuan Y, Wang Y L, Ma J, et al. 2014. Multidimensional evaluation of county poverty degree in Hebei Province[J]. Progress in Geography, 33(1): 124-133.]
- 张旺, 申玉铭. 2012. 京津冀都市圈生产性服务业空间集聚特征[J]. 地理科学进展, 31(6): 742-749. [Zhang W, Shen Y M. 2012. The spatial characteristics of producer service agglomeration in Beijing-Tianjin-Hebei metropolitan region[J]. Progress in Geography, 31(6): 742-749.]
- 周一星, 张莉, 武悦. 2001. 城市中心性与我国城市中心性的等级体系[J]. 地域研究与开发, 20(4): 1-5. [Zhou Y X, Zhang L, Wu Y. 2001. Study of China's urban centrality hierarchy[J]. Areal Research and Development, 20(4): 1-5.]
- Baço F, Lobo V, Painho M. 2005. Self-organizing maps as substitutes for K-means clustering[C]//Proceedings of the 5th international conference on computational science. Berlin, Heidelberg, Germany: Springer-Verlag: 476-483.
- Baselice F, Coppolino L, D'Antonio S, et al. 2015. A DBSCAN based approach for jointly segment and classify brain MR images[C]// Proceedings of the 37th International Conference of the IEEE Engineering in Medicine and Biology Society. Milan, Italy.
- Bell T L, Lieber S R, Rushton G. 2015. Clustering of services in central places[J]. Annals of the Association of American Geographers, 64(2): 214-225.
- Duda R O, Hart P E, Stork D G. 2001. Pattern classification [M]. 2nd ed. New York, NY: A Wiley-Interscience Publication, 2001.
- Ferrari D C, de Castro L N. 2015. Clustering algorithm selection by meta-learning systems: A new distance-based problem characterization and ranking combination methods[J]. Information Sciences, 301: 181-194.
- Gong W W. 2010. ARMA-GRNN for passenger demand forecasting[C]//2010 Sixth international conference on natural computation (ICNC). Yantai, China: IEEE.
- Gómez D, Zarrazolab E, Yáñez J, et al. 2015. A Divide-and-Link algorithm for hierarchical clustering in networks[J]. Information Science, 316: 308-328.
- Grubestic T H, Wei R, Murray A T. 2014. Spatial clustering overview and comparison: Accuracy, sensitivity, and computational expense[J]. Annals of the Association of American Geographers, 104(6): 1134-1156.
- Hagen L, Kahng A B. 2006. New spectral methods for ratio cut partitioning and clustering[J]. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 11(9): 1074-1085.
- Kohonen T. 1990. The self-organizing map[J]. Neurocomputing, 21(1-3): 1-6.
- Lin G F, Wu M C. 2007. A SOM-based approach to estimating design hyetographs of ungauged sites[J]. Journal of Hy-

- drology, 339(3-4): 216-226.
- Marshall J U. 1989. The structure of urban systems[M]. Toronto, Canada: University of Toronto Press.
- Preston R E. 1970. Two centrality models[J]. Yearbook of the Association of Pacific Coast Geographers, 32: 59-78.
- Zhou Y H, Guo H, Dong Y S, et al. 2011. Cluster analysis for drilling-quality based on the modified algorithm of InDB-SCAN[J]. Key Engineering Materials, 480-481: 877-882.

## Multi-level spatial structure analysis of urban agglomeration in the Beijing-Tianjin-Hebei region based on spatial clustering algorithms

ZHANG Xun<sup>1,3</sup>, CHEN Jianzhang<sup>1</sup>, HUANG Jinchuan<sup>2,3,4\*</sup>, YU Chongchong<sup>1</sup>, CHEN Xiuxin<sup>1</sup>

(1. Beijing Key Laboratory of Big Data Technology for Food Safety, Beijing Technology and Business University, Beijing 100048, China; 2. Key Laboratory of Regional Sustainable Development Modeling, CAS, Beijing 100101, China; 3. Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China; 4. College of Resources and Environmental Sciences, University of Chinese Academy of Sciences, Beijing 100049, China)

**Abstract:** With the main focus of regional competition shifting from cities to urban agglomerations, it is important to analyze the spatial structure and direction of coordinated development in urban agglomerations. This is especially significant in the Beijing-Tianjin-Hebei urban agglomeration, which is a core urban agglomeration in China. Machine learning algorithms are relatively new methods for addressing geographical problems. Clustering method, as unsupervised learning, is useful for classifying geographical units without the need for priori knowledge. Using data from 156 counties in the Beijing-Tianjin-Hebei urban agglomeration, this study applied four clustering algorithms: the K-means, density-based spatial clustering of applications with noise (DBSCAN), Chameleon, and self-organizing map (SOM) methods, for classifying counties and districts in the Beijing-Tianjin-Hebei urban agglomeration from the perspectives of economic centrality, traffic centrality, information centrality, and population centrality. GDP of the counties in 2014 was used to represent economic centrality; density of road networks in counties and attraction factor, calculated by the unsold train tickets in different time periods of the year, represent traffic centrality; Sina Weibo check-in data were used to represent information centrality; and county/district population represents population centrality. The result classifies the urban agglomeration into several levels. Respectively, K-means algorithm classifies counties into five levels; DBSCAN algorithm classifies counties into six levels; Chameleon algorithm classifies counties into six levels; and SOM algorithm classifies counties into five levels. SOM is the most applicable algorithm for the division of the urban agglomeration because the structure of counties is stable. This study further analyzed the spatial structure of the urban agglomeration with the central place theory, which points out that an agglomeration should contain certain number of counties in every level. The result of the SOM algorithm matches the central place theory. This research shows that there were remarkable gaps between different levels of the urban agglomeration. The central area of Beijing, as the core of the region, has strong radiation effect on the surrounding areas, but its functions are shared by the nearby counties. Moreover, the second and third level central cities distribute evenly and play an important role in regional development.

**Key words:** spatial clustering; spatial structure; Beijing-Tianjin-Hebei Region Agglomeration