

基于PCA-SOM的深圳产业空间结构

章 文^{1,2}, 王佳璆³

(1. 中山大学地理科学与规划学院, 广州 510275; 2. 广东省城市化与地理环境空间模拟重点实验室, 广州 510275; 3. 伦敦大学学院高级空间分析中心, 伦敦 W1T 4TJ)

摘要: 利用深圳市企业空间分布数据,以街道为基本空间单元,运用主成分分析提取因子达到产业变量降维和抽象目的,在此基础上构建SOM神经网络进行聚类分析,通过PCA-SOM耦合模型实现了对城市内部产业空间分类和结构描述。研究表明:街道的产业功能可以通过提取各产业企业分布数据的主成分因子来表征;PCA-SOM耦合模型将深圳产业空间划分为六种类型区,分类结果与实际吻合;深圳城市产业结构存在空间分异,有别于传统城市的同心环形模式,深圳城市中心区位于地理空间的底部,以集聚型现代服务业为中心、整体呈扇型辐射,并具有明显的路径依赖特征;深圳东部分区产业功能现状仍不明朗,需要对该区域做进一步的空间结构规划和政策支持,以形成多中心的产业辐射模式。

关键词: 城市结构;产业功能;主成分分析;SOM聚类;深圳

DOI: 10.11821/dlyj201409014

1 引言

城市空间结构是地理学研究关注重点之一,在进行职能分类^[1]、竞争力评价^[2,3]、产业结构^[4]、城市群间联系^[5,6]、空间功能分异^[7]等涉及区域功能结构的研究时,研究对象较为集中在城市或城市群构成的宏观空间。随着统计数据和计量研究方法的完善,对城市内部微观地域结构研究有了持续发展^[8]。近年来,李捷定量综合评估城市发展状况指标明确北京市各区县城市发展因素和优势^[9],叶昌东等在对广州的企业规模、产出强度、工业区位商等原始数据变量的选取和处理基础上研究了广州工业空间分异^[10],卢明华等以“乡镇/街道”为基本空间单元研究了北京都市区城市功能格局的特征与模式^[11],陈春林等根据长春市就业空间结构特征按照就业类型将城市内部划分为六种类型区^[12],官卫华等采用定量和定性相结合的方式分析了南京市现代服务业空间布局特征^[13]。而现有对深圳城市内部结构的研究还较少。

城市空间结构研究方法主要有主成份分析(PCA)等方法^[9-12,14,15]和层次聚类方法(hierarchical methods, 又称系统聚类)^[16-19]以及上述两者的耦合模型。层次聚类在小样本情况下,有时仍需辅以定性分析来调整聚类结果^[16,17],或先将离群点挑出、聚类后再人工归入合适的分类^[11],以便反映空间分布的实际情况,这样的分析处理方式需要对研究区域和研究对象有较高的认知程度,在地理单元观测样本较多时也可能使聚类结果难以解释。因此SOM神经网络逐渐走入研究学者的视线。李适宇等在区域发展空间差异定量分

收稿日期: 2013-11-11; 修订日期: 2014-05-23

基金项目: 国家自然科学基金项目(41001385); 国家科技支撑计划项目(2012BAJ15B02)

作者简介: 章文(1978-), 男, 安徽安庆人, 博士, 主要从事城市产业规划与模拟研究。

E-mail: zhangwen@ustc.edu

济行业分类》(GB/T 4754-2002)的字母门类代码对企业进行分类,其中S公共管理和社会组织和T国际组织不属于企业的范畴;A农、林、牧、渔业,B采矿业,D电力燃气及水的生产和供应业因深圳全部的企业数较少(<1000家),在分析中不予考虑;P教育业(收集到的企业主要类型为驾校)和Q卫生、社会保障和社会福利业(收集的企业主要类型为宠物医院),不能真实反应教育卫生产业的本质,也不参加分析。有效分类后共得到13类产业465780家企业的信息,并空间化到街道单元中。

3 主成分分析过程

在研究城市产业空间结构时,经常会搜集尽可能多的产业数据信息,同时又希望对产业空间结构有一个较全面、综合的认识,对众多产业变量全部加以考虑就很难抽象出空间单元上产业功能的实质,因此需要对初始产业变量进行降维处理。主成分分析是一种常用的降低变量维度的因子提取方法,将众多产业变量转化为少数几个彼此不相关的因子指标,以此把街道产业功能高度概括出来。

获取的空间数据包括13类产业465780家企业分布信息,运用主成分分析方法对13类产业变量做降维处理。首先利用SPSS 20对地理样本进行取样适合性的KMO(Kaiser-Meyer-Olkin)度量和Bartlett球形度检验,KMO检验结果为0.828>0.8,Bartlett's球形度检验 χ^2 值为1008.411,显著度为0.000,表明原样本变量矩阵适合进行因子降维。

主成分分析提取结果(Extraction Sums of Squared Loadings)提取了初始特征值大于1的产业主成分3个(表1中显示大于0.1的特征变量),解释了原变量方差的84.717,接近85%,解释效果较好。未旋转时全部信息量的64.561%集中在第一主成分上,主成分结构还不够清晰。在方差最大正交旋转(Varimax)后提取因子结果(Rotation Sums of Squared Loadings)中,其累计解释方差百分比与未旋转的主成分分析提取结果相同,但经旋转后得到的因子重新分配了原始变量的解释方差,使得因子之间载荷更为接近,便于问题分析。

表2显示的是旋转后的主因子载荷矩阵,为清楚描述各主因子对应的载荷结构,对载荷绝对值小于0.5的变量不予显示。从中可以看出,第三产业中的G信息传输、计算机服务和软件业,I住宿和餐饮业,J金融业,K房地产业,L租赁和商务服务业,N水利、环境和公共设施管理业,O居民服务和其他服务业,R文化、体育和娱乐业在主因子 f_1 上

表1 主成分分析的总方差解释表
Tab. 1 Total variance explained

成分	初始特征值			提取因子的载荷平方和			旋转后提取因子的载荷平方和		
	合计	方差百分比 (%)	累计方差 (%)	合计	方差百分比 (%)	累计方差 (%)	合计	方差百分比 (%)	累计方差 (%)
1	8.393	64.561	64.561	8.393	64.561	64.561	6.329	48.687	48.687
2	1.583	12.176	76.737	1.583	12.176	76.737	3.542	27.247	75.934
3	1.037	7.980	84.717	1.037	7.980	84.717	1.142	8.783	84.717
4	0.638	4.911	89.628						
5	0.455	3.503	93.131						
6	0.278	2.139	95.270						
7	0.264	2.032	97.302						
8	0.133	1.024	98.327						

的载荷较高;第二产业的C制造业以及与之配套的第三产业H批发和零售业和M科学研究、技术服务和地质勘查业在主因子 f_2 上的载荷较高;第三产业中的F交通运输、仓储和邮政业单独在主因子 f_3 上有着较高载荷。而第二产业中的E建筑业在主因子 f_1 和 f_2 上均有较高载荷,这可能与建筑业和第三产业中房地产业有着紧密联系有关(两者Pearson相关系数为0.84)。

根据旋转后主成分提取的因子,主因子 f_1 代表集聚型现代服务业和生活性服务业。集聚型现代服务业包括有信息传输和计算机软件业、金融业、房地产业、租赁和商务服务业、文化体育和娱乐业等产业,这些产业的全局Moran's I 统计量最小值为0.16,最大值为0.40,呈现不等的显著空间正相关性,并具有产业集聚特性,根据城市区位理论,该类型现代服务业高值聚集的区位是城市的中心区,或称中心商务区;生活性服务业(或称公共与消费服务业)主要包括住宿和餐饮业、公共设施管理业以及居民服务和其他服务业等直接向居民提供物质生活消费产品及服务,此两类服务业的方差贡献率为48.69%。主因子 f_2 代表了制造业以及与制造业配套、连接产业链上下游的服务业(包括批发零售业、科学研究技术服务和地质勘查业), f_2 的方差贡献率为27.25%。主因子 f_3 代表的是物流业(深圳交通运输、仓储和邮政业Moran's I 值为0.05,基本无空间自相关性),其方差贡献率为8.78%。用上述3个主因子 f_1 、 f_2 、 f_3 替代原始的13个产业变量来描述街道的产业功能,达到了降维的目的,抽象的结果使得待研究问题简化明了。

根据回归算法(Regression)计算出得分系数矩阵,可以得到产业因子得分函数:

$$f_1=-0.229C+0.018E-0.059F+0.137G-0.040H+0.212I+0.206J+0.153K+0.209L-0.094M+0.108N+0.137O+0.093R \tag{1}$$

$$f_2=0.437C+0.209E+0.004F-0.038G+0.248H-0.182I-0.136J-0.020K-0.174L+0.301M+0.096N+0.009O+0.074R \tag{2}$$

$$f_3=0.099C-0.177E+0.802F-0.008G+0.053H+0.174I+0.004J+-0.028K+0.187L+0.080M-0.420N-0.063O-0.109R \tag{3}$$

由此可计算出深圳各街道在3个主因子上的得分,按照自然间断点分级法(Jenks)分为5个等级,图2为各因子得分空间分布图。主因子 f_1 得分处于第一等级是福田区的福田街道和罗湖区的南湖街道,主因子 f_2 得分处于第一等级的是宝安区的西乡街道,主因子 f_3 得分处于第一等级的是罗湖区的南湖街道、盐田区的盐田街道和宝安区的福永街道。

4 SOM聚类分析

自组织映射(Self-Organizing Map, SOM)网络由荷兰学者Kohonen提出基于模型聚类的神经网络,又称为Kohonen网络^[25],该方法具有对噪声不敏感和聚类质量较高的

表2 旋转后的因子载荷矩阵*
Tab. 2 Rotated component matrix*

	主因子		
	f_1	f_2	f_3
C		0.895	
E	0.626	0.708	
F			0.875
G	0.746		
H		0.783	
I	0.911		
J	0.893		
K	0.888		
L	0.921		
M		0.816	
N	0.679		
O	0.849		
R	0.672		

注:提取方法:主成分分析;旋转法:具有Kaiser标准化的正交旋转法;*旋转在4次迭代后收敛。

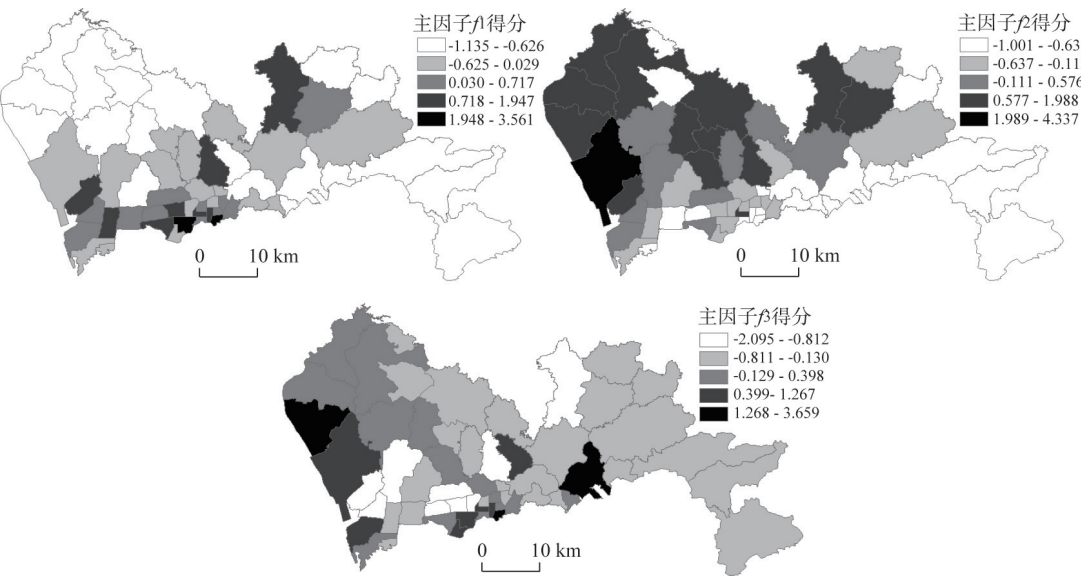


图2 各主因子得分空间分布图
Fig. 2 Distribution of scores of principal components on Shenzhen subdistricts

特点。深圳街道主因子得分存在着若干离群值，其他大部分的得分差别并不明显，适合采用SOM网络来进行聚类分析。产业得分向量不像一些经典聚类测试数据（如UCI数据库中 iris flower^[26]），在映射到竞争层神经元后可用裸眼根据SOM的邻接权重距离（Neight Weight Distances）进行再分类，且非大样本数据要求对SOM输出权值再利用K-means等聚类算法进行两步聚类^[27]，因此可采用2-D SOM网络竞争层神经元输出直接分类的方式。

2-D SOM网络算法步骤如下：

- （1）网络初始化，将网络连接权 $\{w_{ij}\}$ 赋 $[0,1]$ 区间随机值，迭代计数器清0，确定学习率 $\eta(t)$ 初始值 $\eta(0)(0<\eta(t)<1)$ ，确定邻域的初始值，赋值迭代总次数 T ；
- （2）选择新的输入样本 $X^k:X^k=\{x_1^k,x_2^k,\dots,x_n^k\}$ ， $k=1,2,\dots,q$ ，输入到网络上，提供给网络输入神经元， n 为输入向量维度，等于输入层神经元数， q 为样本总数；
- （3）计算模式 X_k 和所有输出单元的距离 d_j^k ，并选择和 X_k 距离最小的神经元为获胜神经元，记为 j^* ， $d_j^*=\min_j\{d_j^k\}=\min_j\{\|X^k-W_j^k\|\}$ ， $j=1,2,\dots,s$ ， s 为映射层神经元数；
- （4）权值学习，更新神经元 j^* 及其邻域结点的连接权值为 $\Delta w_{ij}=\eta(t)(x_i(t)-w_{ij}(t))$ ，式中学习率 $\eta(t)$ 随着时间逐渐减小；
- （5）选取另一学习样本作为SOM网络的输入数据，转至步骤（3），直至 q 个学习样本全部提供给神经网络；
- （6）令 $t=t+1$ ，返回步骤（2），直至 $t=T$ 结束。

在自组织映射模型的学习过程中，邻域随 t 增加而收缩，学习率 $\eta(t)$ 渐趋向零，保证了神经网络学习过程的收敛性。

考虑到深圳街道样本总数，过粗或过细的分类均不适宜，故采用 2×3 的二维SOM竞争层神经元作为聚类的容器。竞争层神经元拓扑采用六边形连接，通过Matlab 2012a编程实现，输入层向量与竞争层输出神经元的对应矩阵使用vec2ind函数转换以确定样本与聚类类别的映射，迭代次数为1000次，聚类结果如表3。分类输出也可以选择一维SOM

竞争层神经元^[28], 此时同样需要根据研究对象的实际情况谨慎选择输出神经元的个数。

对比层次聚类方法, 当采用基于个案的聚类(即Q聚类)对因子得分进行树状图分析, 度量标准选择欧氏平方距离(Squared Euclidean distance), 以离差平方和法(Ward's method)计算类与类之间距离, 会发现聚类的类别要么过少、要么过多(距离为10时分为4类、距离为5时分为9类), 这种以距离为依据的不相似测度方法因离群值的存在可能还会造成一些类别只有一个样本, 而另一些类别下样本数过多, 但从因子结构上判断这些样本并不适合作为一种类型来分析。即便对输入数据进行归一([0,1]或[-1,1])标准化处理后, 层次聚类的结果仍难以解释。在地理空间样本数较多(本文为58个)、且存在若干离群值时, SOM神经网络聚类较层次聚类方法更为适用。

SOM聚类根据主因子得分将深圳城市内部产业空间结构划分为6类产业类型区(图3): ① 集聚型现代和生活性服务业及物流业混合区; ② 集聚型现代和生活性服务业及制造业混合区; ③ 集聚型现代和生活性服务业类型区; ④ 制造业类型区; ⑤ 物流业类型区; ⑥ 混合功能散布区。

集聚型现代和生活性服务业及物流业混合区为主因子 f_1 和 f_2 得分较高的街道, 涉及福田区的福田街道和罗湖区的南湖街道(均拥有陆路口岸); 集聚型现代和生活性服务业及制造业混合区为主因子 f_1 和 f_2 得分较高的街道, 主要分布在传统关内(福田、罗湖和南山区)和关外(宝安区、龙华新区和龙岗区)“二线关”结合部以及龙岗区的腹地; 集聚型现代和生活性服务业类型区为主因子 f_1 得分高的街道, 主要集中在福田、罗湖和南山

表3 SOM网络聚类结果

Tab. 3 Results of SOM clustering

神经元编号	聚类结果
1	福田, 南湖
2	布吉, 莲花, 龙城, 龙岗, 梅林, 南头, 新安
3	翠竹, 东门, 桂园, 华强北, 黄贝, 南山, 南园, 沙头, 香蜜湖, 园岭, 粤海
4	坂田, 大浪, 公明, 观澜, 龙华, 民治, 沙井, 石岩, 松岗, 西乡
5	大鹏, 东湖, 东晓, 福保, 光明, 海山, 横岗, 华富, 坑梓, 葵涌, 莲塘, 梅沙, 南澳, 南湾, 平湖, 坪地, 坪山, 清水河, 沙河, 沙头角, 蛇口, 笋岗, 桃园, 西丽, 招商
6	福永, 盐田

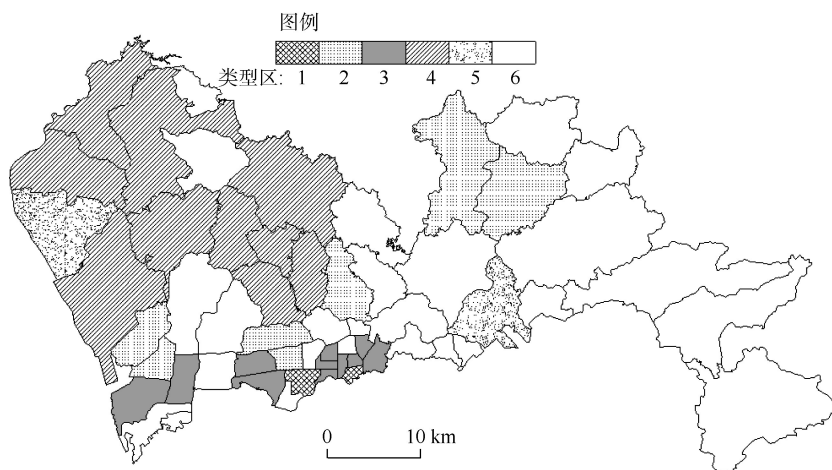


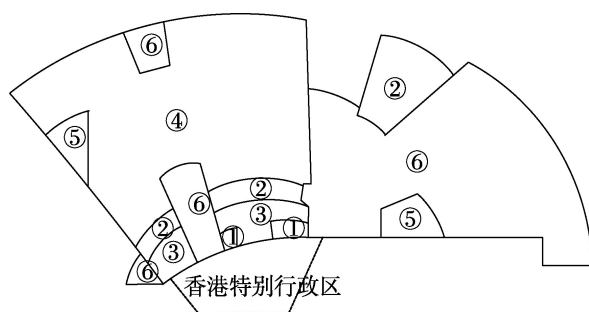
图3 深圳产业类型区街道分布格局

Fig. 3 Classification of industrial zones on Shenzhen subdistricts

区；制造业类型区为主因子 f_2 得分高的街道，主要集中在宝安区和龙华新区；物流业类型区为主因子 f_3 得分高的街道，包括盐田区的盐田街道（盐田港所在街道、其远洋集装箱班轮密度国内第一）和宝安区的福永街道（深圳宝安国际机场所在街道、通过高速公路邻接东莞和西部通道连接香港）；混合功能散布区在3个主因子上的得分均不突出，在传统关内地区表现为产业功能退化区（如深圳特区产业起步的策源地蛇口街道等）或居民集中居住区（如沙河街道等），在关外主要为山地湖泊、生态休闲旅游区（如大鹏新区，包括葵涌、大鹏和南澳街道）或从企业分布上来看产业功能尚不明朗的区域（如坪山新区，包括坪山和坑梓街道）。

5 产业功能空间结构和特征

深圳产业功能结构空间分异明显，通过主成分分析提取因子方法刻画出各街道的产业特征，经SOM聚类形成的产业类型区专题图上可看出产业功能空间分异的存在，反映出不同产业对地租的承受能力，这种产业空间分异也是城市专业化分工深化的空间表现。有别于国内GDP排名靠前的其他3个城市：北京^[9,11]、上海^[29]和广州^[10,30]，以及其他传统大中城市如长春^[12]、南京^[13]等，这些城市的产业空间结构更多表现为同心环模式，深圳的产业中心区并不在空间质心附近，而是在靠近香港的区域，整体产业功能结构呈扇形模式（图4）。由内至外产业功能类型分别是：集聚型现代和生活性服务业类型区（包括集聚型现代和生活性服务业及物流业混合区）、集聚型现代和生活性服务业及制造业混合区、制造业类型区、混合功能散布区，自南向北单向扩展。集聚型现代和生活性服务业处在城市的中心区，集聚型现代和生活性服务业及制造业混合区可视为由集聚型现代和生活性服务业类型区向制造业类型区转变的过渡区，物流业类型区集中在口岸、机场和港口等交通枢纽所在街道。扇形模式向东部分区的渗透力较弱，原因是由于东莞和深圳间的经济吸引力系数较高^[31]，引起深圳产业布局的西倾。



注：①集聚型现代和生活性服务业及物流业混合区；②集聚型现代和生活性服务业及制造业混合区；③集聚型现代和生活性服务业类型区；④制造业类型区；⑤物流业类型区；⑥混合功能散布区

图4 深圳产业扇形模式空间结构图

Fig. 4 Fan-shaped model of Shenzhen's industrial space structure

经济发达地区的产业扇形模式，一方面确保了深圳从一个边陲小镇快速发展为大都市的可能性：作为内地唯一与香港接壤的城市，使得深圳自发展初期开始，就能便利地通过香港引进获取资金、技术、设备和管理方法，从早期的“三来一补”到现代服务业如金融业，港资都占据了较大比重。另一方面，扇形模式相比同心环模式而言，产业结构中心向外辐射面偏窄、涓滴效应相对较弱。

这种扇形模式也反映了深圳产业发展与深港之间的联系存在着密切的路径依赖。扇形底部是集聚型现代服务业和生活性服务业类型区，深圳的6个一类陆路口岸主要集中在该区域内，分别是罗湖、文锦渡、皇岗、沙头角、深圳湾和福田口岸，与香港直接接驳。基于CEPA框架下^[32]的深港双城经济一体化，又进一步强化了这种路径趋向。深圳市现代服务业发展“十二五”规划所提出着力建设的三大综合功能区，即加快建设前海深

中心区在城市边缘、贴近历史上

港现代服务业合作区、加快完善深圳福田中央商务区和加快打造罗湖国际消费中心^[33], 也无一例外地分布在深圳扇形的底部。

现代服务业主导着国际大都市中心区的发展, 决定着城市经济的繁荣及其国际竞争力的高低^[34]。深圳金融业等现代服务业聚集在更邻近香港的城市底部地区, 形成的城市经济中心以地缘优势确保了城市的可持续性发展和城市竞争力。同时也应看到, 深圳东部分区产业发展仍不明朗, 龙岗区的龙城和龙岗街道的集聚型现代和生活性服务业及制造业虽已成型, 但仍属于集聚型现代服务业和制造业的混合区, 而周边的坪山新区的产业功能目前还无法判定。产业功能空间结构在路径形成后会被锁定, 由于循环累积因素效应引发的自组织演化, 加上现有的产业发展政策, 会形成强势的路径依赖。只有通过政策性支持引导的干预, 才能实现路径解锁, 创造出新的发展路径^[13,35]。因此深圳有必要在现有产业格局的基础上, 重视产业政策向城市东北部的倾斜力度。这样既能缓解“空间有限难以为继”的困境, 发挥城市产业发展的潜能, 也可以以龙岗区作为城市副中心, 弥补扇形模式产业辐射面较窄的缺点, 带动坪山新区乃至东北部惠州邻近地区的产业演进, 兼起到深圳传统产业向粤东转移的桥梁作用。

6 结论与讨论

(1) 明确地理单元的产业功能是界定城市产业空间结构的前提, 以各产业企业分布信息为基础, 利用主成分分析方法经旋转后提取深圳街道的产业主因子构成及计算相应得分, 可以刻画出街道代表性产业的特征, 进而以因子得分作为输入, 基于SOM神经网络进行城市内部产业功能区域的划分。

(2) 根据PCA-SOM模型评价分类结果, 深圳街道的产业功能呈现空间分异, 产业中心区极大偏离地理质心而位于城市空间的底部, 最大限度与香港保持地理上接近, 产业发展路径依赖明显, 产业空间结构以金融业等集聚型现代服务业为核心整体呈扇形由南向城市北部辐射。

(3) 深圳东部分区产业功能尚不明晰, 龙岗区的龙城、龙岗街道集聚型现代服务业和制造业虽已具规模, 但仍为交错式的产业过渡带, 产业扩散效应不够显著, 坪山等新区还属混合功能散布区, 产业发展潜力有待挖掘, 因此该区域的产业定位和规划需要进一步加强, 以形成多中心的城市产业空间结构, 平衡单一扇形模式辐射面较窄的不足。

致谢: 在2013年8月13日深圳大学举办的第一届“空间信息智能感知与服务暑期研讨会”上进行了摘要报告, 感谢香港大学建筑学院李卫锋副教授等给予的建议和帮助。

参考文献(References)

- [1] 周一星, 孙则昕. 再论中国城市的职能分类. 地理研究, 1997, 16(1): 11-22. [Zhou Yixing, Sun Zexing. Redisussion on China's urban function classification. Geographical Research, 1997, 16(1): 11-22.]
- [2] 许学强, 程玉鸿. 珠江三角洲城市群的城市竞争力时空演变. 地理科学, 2006, 26(3): 257-265. [Xu Xueqiang, Cheng Yuhong. Spatial-temporal changes of urban competitiveness in urban cluster of Pearl River Delta. Scientia Geographica Sinica, 2006, 26(3): 257-265.]
- [3] 陈思源. 区域旅游目的地竞争优势的空间聚类分析: 以浙江省为例. 地域研究与开发, 2012, 31(1): 85-88. [Chen Siyuan. The spatial clustering on the competitive advantage of regional tourism destinations: A case study of Zhejiang Province. Areal Research and Development, 2012, 31(1): 85-88.]
- [4] 任崇强, 宗跃光, 王燕军. 京津冀地区产业结构和竞争力空间分异研究. 地域研究与开发, 2012, 31(3): 1-5. [Ren

- Chongqiang, Zong Yueguang, Wang Yanjun. Study on the Spatial differentiation of industry in Beijing-Tianjin-Hebei region. *Areal Research and Development*, 2012, 31(3): 1-5.]
- [5] 陈园园, 李宁, 丁四保. 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. *地理科学*, 2011, 31(12): 1461-1467. [Chen Yuanyuan, Li Ning, Ding Sibao. Spatial combination capacity and classification based on SOM network of urban agglomerations: A case study of central and southern Liaoning urban agglomerations. *Scientia Geographica Sinica*, 2011, 31(12): 1461-1467.]
- [6] 方创琳, 宋吉涛, 张茜, 等. 中国城市群结构体系的组成与空间分异格局. *地理学报*, 2005, 60(5): 827-840. [Fang Chuanglin, Song Jitao, Zhang Qiang, et al. The formation, development and spatial heterogeneity patterns for the structures system of urban agglomerations in China. *Acta Geographica Sinica*, 2005, 60(5): 827-840.]
- [7] 杜瑜, 樊杰. 基于产业-人口集聚分析的都市经济区空间功能分异: 以我国三大都市经济区为例. *北京大学学报: 自然科学版*, 2008, 44(3): 467-474. [Du Yu, Fan Jie. Research on spatial function diversity of metropolitan economic regions based on "industries & population" agglomeration analysis: Illustrations from Chinese "three large metropolitan economic regions". *Acta Scientiarum Naturalium Universitatis Pekinensis*, 2008, 44(3): 467-474.]
- [8] 冯健, 周一星. 中国城市内部空间结构研究进展与展望. *地理科学进展*, 2003, 22(3): 204-215. [Feng Jian, Zhou Yixing. A review and prospect on urban internal spatial structure research in China. *Progress in Geography*, 2003, 22(3): 204-215.]
- [9] 李健. 基于因子分析的北京城市功能空间布局研究. *城市发展研究*, 2005, 12(4): 57-62. [Li Jie. A research on Beijing's urban functions distribution based on the methods of factor analysis. *Urban Studies*, 2005, 12(4): 57-62.]
- [10] 叶昌东, 周春山, 刘艳艳. 近10年来广州工业空间分异及其演进机制研究. *经济地理*, 2010, 30(10): 1664-1669. [Ye Changdong, Zhou Chunshan, Liu Yanyan. The industrial spatial differentiation and its evolution mechanism of Guangzhou in recent ten years. *Economic Geography*, 2010, 30(10): 1664-1669.]
- [11] 卢明华, 李国平, 孙铁山. 北京都市区城市功能格局及其变化: 基于经济普查数据的分析. *地理研究*, 2011, 30(11): 1970-1982. [Lu Minghua, Li Guoping, Sun Tieshan. Study on the functional pattern of the Beijing metropolitan region and its changes: Based the analysis of data from the economic unit census. *Geographical Research*, 2011, 30(11): 1970-1982.]
- [12] 陈春林, 梅林, 刘继生, 等. 转型期长春市就业结构的时空分异. *地理科学*, 2011, 31(5): 581-588. [Chen Chunlin, Mei Lin, Liu Jisheng, et al. Spatial pattern of employment structure of Changchun city in transformation era. *Scientia Geographica Sinica*, 2011, 31(5): 581-588.]
- [13] 官卫华, 陈雯. 大都市现代服务业空间组织机理研究: 以南京为例. *地理科学进展*, 2013, 31(3): 341-353. [Guan Weihua, Chen Wen. Mechanisms of spatial organization of modern service industry in metropolises: Taking Nanjing as an example. *Progress in Geography*, 2013, 31(3): 341-353.]
- [14] 于正松, 李同昇, 李献波, 等. 西北地区农业技术扩散环境的时空分异: 以陕、甘、宁县域为例. *地理科学进展*, 2013, 32(4): 618-626. [Yu Zhengsong, Li Tongsheng, Li Xianbo, et al. Spatial differences of agri-technique diffusion environment in northwestern China: A case study of counties in Shaanxi-Gansu-Ningxia region. *Progress in Geography*, 2013, 32(4): 618-626.]
- [15] 丛明珠, 欧向军, 赵清, 等. 基于主成分分析法的江苏省土地利用综合分区研究. *地理研究*, 2008, 27(3): 574-582. [Cong Mingzhu, Ou Xiangjun, Zhao Qing, et al. Division of land use degree in Jiangsu province based on principal component analysis. *Geographical Research*, 2008, 27(3): 574-582.]
- [16] 李健, 宁越敏. 1990年代以来上海人口空间变动与城市空间结构重构. *城市规划学刊*, 2007, 168(2): 20-24. [Li Jian, Ning Yuemin. Population spatial change and urban spatial restructuring in Shanghai since the 1990s. *Urban Planning Forum*, 2007, 168(2): 20-24.]
- [17] 蒲欣冬, 刘兰涛, 陈怀录, 等. 甘肃省广河县土地利用分区研究. *干旱区资源与环境*, 2013, 27(2): 64-68. [Pu Xindong, Liu Lantao, Chen Huailu, et al. The zoning of land utilization division in Guanghe, Gansu. *Journal of Arid Land Resources and Environment*, 2013, 27(2): 64-68.]
- [18] 胡玮, 王心源. 安徽省城镇居民收入水平时空分异特征. *安徽师范大学学报: 自然科学版*, 2003, 26(1): 75-80. [Hu Wei, Wang Xinyuan. The spatio-temporal difference of urban residents' income level in Anhui province. *Journal of Anhui Normal University: Natural Science*, 2003, 26(1): 75-80.]
- [19] 李军军, 周利梅. 基于聚类分析视角的我国省域经济综合竞争力研究. *福建师范大学学报: 哲学社会科学版*, 2008, 148(1): 85-90. [Li Junjun, Zhou Limei. Evaluation for overall competitiveness of China's provincial economy on cluster analysis. *Journal of Fujian Normal University: Philosophy and Social Sciences Edition*, 2008, 148(1): 85-90.]
- [20] 李适宇, 厉红梅. 广东省区域发展空间差异SOM网络判定. *科技通报*, 2007, 23(4): 609-616. [Li Shiyu, Li Hongmei.

- Analysis on space differentiations of regional development by SOM networks. *Bulletin of Science and Technology*, 2007, 23(4): 609-616.]
- [21] Park Y, Tison J, Lek S, et al. Application of a self-organizing map to select representative species in multivariate analysis: A case study determining diatom distribution patterns across France. *Ecological Informatics*, 2006, 1(3): 247-257.
- [22] 张学儒, 张懿铨, 刘林山, 等. 基于SOFM神经网络模型的土地类型分区尝试: 以青藏高原东部样带为例. *地理研究*, 2013, 32(5): 839-847. [Zhang Xueru, Zhang Yili, Liu Linshan, et al. Zoning by land types based on SOFM network: A case study on transect of eastern Tibetan Plateau. *Geographical Research*, 2013, 32(5): 839-847.]
- [23] Frenkel A, Bendit E, Kaplan S. The linkage between the lifestyle of knowledge-workers and their intra-metropolitan residential choice: A clustering approach based on self-organizing maps. *Computers, Environment and Urban Systems*, 2013, 39(5): 151-161.
- [24] 罗海平. 是什么成就了深圳. *改革与开放*, 2011, (7): 13-14. [Luo Haiping. What brought about the achievement of Shenzhen. *Reform & Opening*, 2011, (7): 13-14.]
- [25] Kohonen T. The self-organizing map. *Proceedings of the IEEE*, 1990, 78(9): 1464-1480.
- [26] Bache K, Lichman M. Uci machine learning repository. Irvine, Ca: University of California, School of Information and Computer Science. <http://archive.ics.uci.edu/ml>, 2013-01-01/2013-06-08.
- [27] Vesanto J, Alhoniemi E. Clustering of the self-organizing map. *IEEE Transactions on Neural Networks*, 2000, 11(3): 586-600.
- [28] 于翊. 基于一维SOM神经网络的聚类及数据分析方法研究. 天津: 天津大学博士学位论文, 2009. [Yu Yan. A study of clustering and data analysis methods based on one-dimensional SOM. Tianjin: Doctoral Dissertation of Tianjin University, 2009.]
- [29] 王桂新, 魏星. 上海从业劳动力空间分布变动分析. *地理学报*, 2007, 62(2): 200-210. [Wang Guixin, Wei Xing. Change of labor force spatial structure in Shanghai from 1996 to 2001. *Acta Geographica Sinica*, 2007, 62(2): 200-210.]
- [30] 闫小培, 姚一民. 广州第三产业发展变化及空间分布特征分析. *经济地理*, 1997, 17(2): 41-48. [Yan Xiaopei, Yao Yimin. Analysis on characteristics of development and spatial distribution of the tertiary sector in Guangzhou. *Economic Geography*, 1997, 17(2): 41-48.]
- [31] 李红锦, 李胜会. 基于引力模型的城市群经济空间联系研究: 珠三角城市群的实证研究. *华南理工大学学报: 社会科学版*, 2011, 13(1): 19-24. [Li Hongjin, Li Shenghui. Gravity model based on spatial connection of urban agglomeration economies: An empirical study of the Pearl River Delta city group. *Journal of South China University of Technology: Social Science Edition*, 2011, 13(1): 19-24.]
- [32] 钟韵, 闫小培. 改革开放以来香港生产性服务业对广州同行业的影响. *地理研究*, 2006, 25(1): 151-160. [Zhong Yun, Yan Xiaopei. Impact of the producer services of Hong Kong on the same industry of Guangzhou since the reform. *Geographical Research*, 2006, 25(1): 151-160.]
- [33] 深圳市发展和改革委员会. 深圳市现代服务业发展“十二五”规划. <http://www.sz.gov.cn/>, 2012-1-16/2013-05-05. [Development and Reform Commission of Shenzhen Municipality. The twelfth five-year plan of Shenzhen modern service industry. <http://www.sz.gov.cn/>, 2012-01-16/2013-05-05.]
- [34] 杨亚琴, 王丹. 国际大都市现代服务业集群发展的比较研究: 以纽约、伦敦、东京为例的分析. *世界经济研究*, 2005, (1): 61-66. [Yang Yaqin, Wang Dan. Comparative research on the modern service industry cluster development of the international metropolis: Case study of New York, London and Tokyo. *World Economy Study*, 2005, (1): 61-66.]
- [35] 尹贻梅, 刘志高, 刘卫东. 路径依赖理论及其地方经济发展隐喻. *地理研究*, 2012, 31(5): 782-791. [Yin Yimei, Liu Zhigao, Liu Weidong. Path-dependence and its implication for regional development. *Geographical Research*, 2012, 31(5): 782-791.]

A research on Shenzhen's industrial spatial structure based on PCA-SOM

ZHANG Wen^{1,2}, WANG Jiaqiu³

(1. Geography and Planning School of Sun Yat-sen University, Guangzhou 510275, China; 2. Key Laboratory of Urbanization and Geographical Environment Simulation of Guangdong Province, Guangzhou 510275, China; 3. Centre for Advanced Spatial Analysis of University College London, London W1T 4TJ, Britain)

Abstract: By using spatial distribution data of Shenzhen enterprises, regarding subdistricts as the basic spatial unit, this paper adopts the principal component analysis (PCA) method to extract factors to achieve the objective of dimension reduction and properties abstraction on industry variables. Based on the above, self-organizing map (SOM) neural network clustering model is built with three industrial characters of the subdistricts resulted from PCA and factors extraction as input data. Through PCA-SOM coupling model, this paper brings about the spatial classification and description of urban internal industrial function structure. Then the following conclusions are drawn. (1) The industrial functions of subdistricts can be represented through extracting the factors of enterprises spatial distribution data by principal component analysis. (2) PCA-SOM coupling model divides Shenzhen's industrial spatial function into 6 type-zones. In this way the classification results accord with the actual situation and avoid subjective and arbitrary assessment. (3) Shenzhen's industrial structure presents spatial differentiation in inner city. Unlike concentric ring model of traditional city, Shenzhen's city center is located at the bottom of geographic space. Moreover the industrial spatial structure of Shenzhen takes clustered modern service industry as the core, spreads as fan-shaped radiation from south to north area on the whole, and has obvious path dependence characteristics. (4) The industrial spatial function of eastern division in Shenzhen is still indistinct, which means further industrial spatial structure planning and more policy support should be implemented in this region in order to develop polycentric industrial radiation mode.

Key words: urban structure; industrial function; principal component analysis; SOM clustering; Shenzhen