

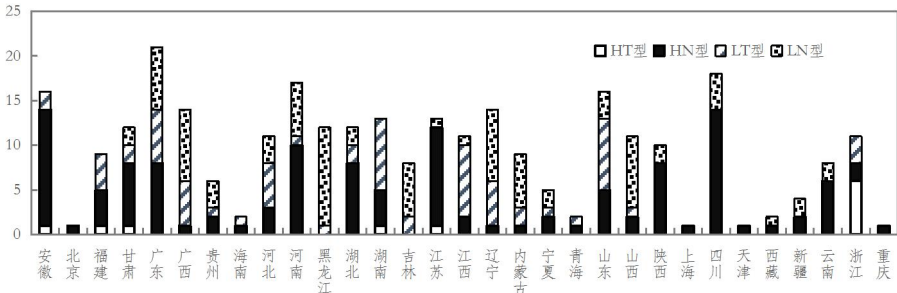
附表 1 基于七大地理分区的谱聚类变量均值 (%)			
地区	创新活跃度	技术创新偏向度	非技术创新偏向度
全国 2014	40.0668	68.3843	85.5727
全国 2020	48.1325	75.1976	76.7898
东北 2014	26.1610	62.5132	87.2849
东北 2020	37.9201	67.7920	80.0880
华北 2014	36.1527	61.5347	85.9109
华北 2020	41.7651	73.1787	74.8350
华东 2014	46.0255	74.1268	80.6774
华东 2020	58.9050	84.9547	71.2709
华南 2014	37.3692	68.5602	85.6520
华南 2020	42.6266	72.8498	77.8985
华中 2014	41.4292	72.0702	84.6553
华中 2020	49.7919	75.3328	76.9219
西北 2014	44.7123	69.2996	89.7510
西北 2020	44.0819	67.4842	81.9900
西南 2014	44.5130	65.4781	90.6869
西南 2020	48.0055	72.4218	81.3365

附表 2 聚类变量最高、最低、变化最大的城市汇总			
	创新活跃度	非技术创新偏向度	技术创新偏向度
2014 最高	镇江、乐山、泰州、巴中、无锡	嘉峪关、张家界、日喀则、平凉、吐鲁番	陇南、岳阳、张掖、庆城、海口
2014 最低	松原、巴彦淖尔、通辽、抚顺、乌兰察布	台州、营口、岳阳、绍兴、佛山	临汾、忻州、朔州、吕梁、通辽
2020 最高	嘉兴、深圳、丽水、南京、衢州	海东、三亚、拉萨、白山、吐鲁番	萍乡、嘉兴、舟山、丽水、马鞍山
2020 最低	贵港、松原、黑河、玉林、吐鲁番	郴州、舟山、益阳、南通、台州	三亚、海东、商丘、延安、吐鲁番
增幅最大	德州、淄博、怀化、张掖、衡阳	随州、三亚、岳阳、海东、营口	中卫、淮安、泰州、秦皇岛、连云港
降幅最大	庆阳、三亚、克拉玛依、乌鲁木齐、乐山	郴州、运城、聊城、湘潭、中卫	海东、三亚、陇南、周口、商丘

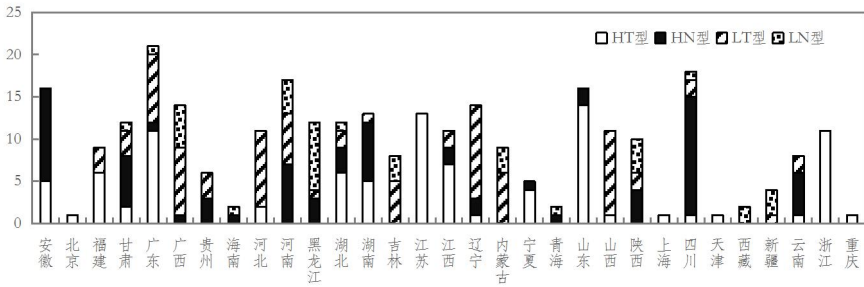
附表 3 聚类参数选择结果							
评估标准	聚类方法	聚类个数	相似矩阵构建方式	gamma	degree	coef0	得分
Davies-Bouldin 指数	谱聚类	4	sigmoid	1.00	—	4	0.7062
		4	rbf	0.12	—	—	0.7636
		4	rbf	0.18	—	—	0.7637
		4	rbf	0.16	—	—	0.7758
		4	rbf	0.14	—	—	0.7761
	k-means 聚类	6	—				1.0964
		5					1.1498
		4					1.1803
	轮廓系数	5	poly	1.00	3	4	0.2851
		4	rbf	0.14	—	—	0.2802
		5	poly	1.00	3	3	0.2774
		4	rbf	0.10	—	—	0.2743

		4	rbf	0.18	—	—	0.2737
k-means 聚类		4	—				0.28245
		5					0.2718
		6					0.2706

注：参数调试方案为：k 近邻法需指定 k，调试范围为 4~15；多项式核函数，高斯核函数与 sigmoid 核函数的参数 gamma 调试范围为 0.10~4.00；多项式核函数的参数 degree 调试范围为 3~6；多项式核函数与 sigmoid 核函数的参数 ceof0 调试范围为 1~4。由于本文基于创新结构和创新活跃度两个维度构建城市企业创新发展模式，因此目标聚类个数至少为 4，最终聚类个数尝试了 4~6。附表 3 展示了各种参数选择方案中轮廓系数最大与 Davies-Bouldin 指数最小的 5 个结果，并将 k-means 聚类的结果也列入比较。



附图 1 2014 年 31 个省级行政单位各模式城市数



附图 2 2020 年 31 个省级行政单位各模式城市数

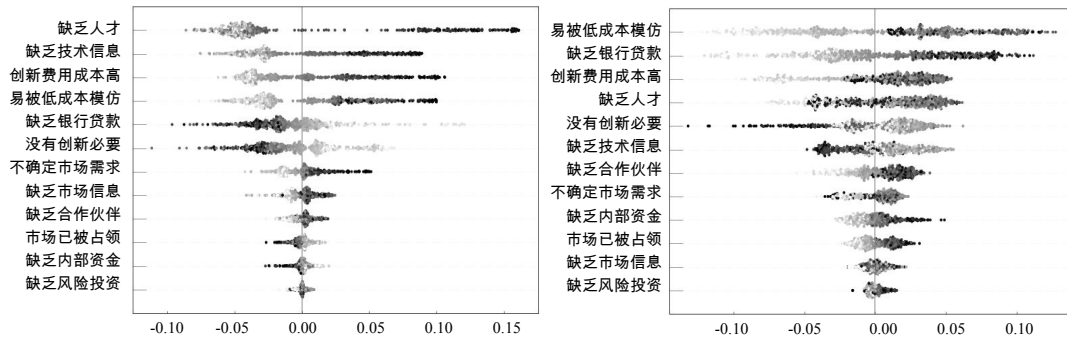
附表 4	2014 年与 2020 年部分城市群各模式的城市个数（个）							
	2014				2020			
	HT 型	HN 型	LT 型	LN 型	HT 型	HN 型	LT 型	LN 型
京津冀城市群	—	5	5	3	4	—	9	—
长江三角洲城市群	5	19	2	—	21	5	—	—
粤港澳大湾区	—	5	4	—	9	—	—	—
长江中游城市群	1	11	13	3	14	9	5	—
山东半岛城市群	—	5	8	2	14	1	—	—
粤闽浙沿海城市群	2	7	8	3	13	—	7	—
成渝城市群	—	12	—	4	2	12	2	—
中原城市群	—	15	4	10	3	12	10	4
关中平原城市群	1	6	—	4	1	4	4	2
北部湾城市群	—	1	5	4	—	2	5	3
哈长城市群	—	—	3	7	—	2	5	3

注：“—”表示不含该类型城市。

附表 5

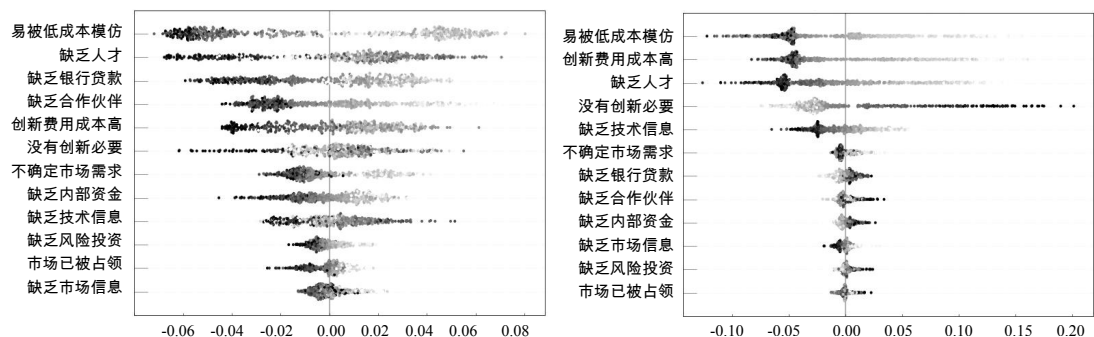
随机森林特征重要性排序

特征	特征重要性
缺乏人才	0.1668
创新成果易被低成本模仿	0.1506
创新费用成本高	0.1372
没有创新必要	0.1110
缺乏银行贷款	0.0931
缺乏技术信息	0.0921
缺乏合作伙伴	0.0565
不确定市场需求	0.0495
缺乏内部资金	0.0421
缺乏市场信息	0.0403
市场已被占领	0.0332
缺乏风险投资	0.0275



附图 3 基于随机森林的 SHAP 解释图（左：活跃度高-技术创新主导，右：活跃度高-非技术创新主导）

注：横坐标为 SHAP 值，每行代表一个特征变量即阻碍因素（本文包含 12 个阻碍因素即 12 行），每个点代表一个样本（在本文中即代表一个城市），每一行上的样本点数量是相同的，特征取值用样本点颜色深浅反映，样本点颜色越深代表其所在行对应的特征取值越大。以附图 3 左图（HT 型）第一行（缺乏人才）为例，深色样本点多分布于右侧，浅色样本点多分布于左侧，即“缺乏人才”这一特征取值越大，SHAP 值越大，也就是说，如果一个城市受“缺乏人才”这一因素阻碍的程度较大，那么模型更倾向于将该城市识别为 HT 型。即在不同模式的城市中，企业的发展瓶颈应是呈现右边深色、左边浅色的特征。附图 4 同。



附图 4 基于随机森林的 SHAP 解释图（左：活跃度低-技术创新主导，右：活跃度低-非技术创新主导）