

# 互联网金融空间聚集分析及系统性风险防范

## ——基于 t-SNE 机器学习模型

米传民, 徐润捷, 陶 静

(南京航空航天大学经济与管理学院, 江苏 南京 210016)

**摘 要:** 互联网金融同传统金融具有不同的空间聚集特征。互联网金融在带来金融开放、门槛降低、效率提升、成本下降的同时, 也给互联网金融体系、乃至整个金融系统带来风险新问题。大量研究表明互联网金融在宏观经济冲击、内部脆弱性等影响下, 往往具有与以往不同的系统性金融风险特征。本文利用北京大学的31个省和335个地市区域的互联网金融发展指数有关数据, 运用 t-SNE 机器学习模型进行我国互联网金融发展的降维和聚类分析, 得到我国互联网金融空间聚集和不同业务模式发展的分布特征, 发现在区域发展程度上存在尖峰厚尾, 在业务模式上存在不均衡现象。基于此, 提出了考虑互联网金融发展区域差异造成的三方面系统性风险, 并为防范互联网金融系统性风险提出建议。

**关键词:** 互联网金融; 系统性风险; 降维聚类; t-SNE 算法

**中图分类号:** F832

**文献标识码:** A

**文章编号:** 1004-4892(2019)08-0053-10

## 一、引 言

互联网给商业、金融、工业等众多行业带来了机遇与挑战, 其中互联网金融充分地利用了互联网的优势, 在一定程度上提高了金融资源配置效率, 促进了金融的普惠发展<sup>[1][2]</sup>。但随着网络支付、P2P 信贷、众筹等互联网金融新模式的普及, P2P 平台违约、跑路等互联网金融风险问题频发。互联网金融的发展以及由此带来的系统性风险成为学界和业界关注的重要问题。

2007 年金融危机爆发以来, 系统性金融风险的防范已经成为各国金融监管部门的重要工作。多元正态分布是刻画系统性风险的重要工具, 但难以拟合金融时间序列的尖峰厚尾和非对称相依结构等特征<sup>[3]</sup>。另一方面, 金融系统性风险的影响因素具有动态变化特征, 例如互联网金融新商业模式层出不穷、监管滞后、金融市场流动性问题等等。如果不考虑这些现实因素的影响, 则造成金融系统风险分析失真。随着金融监管理论、大数据、以及互联网金融发展, 一些新的研究方法进入研究的视野, 如通过降维来估计系统性风险的 CoVaR<sup>[4]</sup>、MES<sup>[5]</sup>等方法。机器学习算法领域的 Laurens van der Maaten 和 Geoffrey Hinton 提出了 t-SNE 降维聚类算法<sup>[6]</sup>。本文运用 t-SNE 算法, 研究我国互联网金融发展的空间聚类, 通过对互联网金融多种业务发展的指数数据降维来提取局部空间

收稿日期: 2019-01-23

基金项目: 国家社会科学基金资助项目(17BGL055)

作者简介: 米传民(1976-), 男, 山东聊城人, 南京航空航天大学经济与管理学院副教授, 博士; 徐润捷(1999-), 男, 安徽六安人, 南京航空航天大学经济与管理学院学生; 陶静(1983-), 女, 安徽寿县人, 南京航空航天大学经济与管理学院博士生。

结构,进而观察我国互联网金融发展聚类情况,并在此基础上提出三类系统风险问题。

## 二、文献综述

互联网金融的虚拟线上空间、跨区域发展等带来了互联网金融发展区域和空间聚集的不均衡。一些学者对互联网金融区域发展、空间聚集进行了系统研究。北京大学互联网金融研究中心课题组编制了互联网金融发展指数,并对互联网金融空间聚集效应进行了分析<sup>[7][8]</sup>。郭海凤和陈霄(2015)深入考察了中国网贷平台发展的地区差异<sup>[9]</sup>。廖理等(2014)系统分析了网贷平台借贷中的地域歧视<sup>[10]</sup>。王赛芳(2016)通过构建指标体系,基于因子分析和聚类分析方法实证研究了我国 31 省市的互联网金融发展水平及空间分异<sup>[11]</sup>。李清磊和王旺(2018)基于北京大学数字普惠金融指数数据,定量分析了安徽省金融发展的地域差距<sup>[12]</sup>。但总体上,受制于互联网金融数据缺乏,目前的研究多采用统计分析的方法进行空间聚焦和差异分析。

巴塞尔资本协议 III 提出了宏观审慎监管的思路,我国提出了“守住不发生系统性金融风险”的金融监管底线<sup>[13]</sup>。但业界和学界关于金融系统性风险的定义尚没有完全统一。从金融稳定视角,有的学者将系统性风险的根本原因归结为金融系统本身的内在不稳定性<sup>[14]</sup>。Allen(2000)从经济平衡的角度解释系统性风险,认为金融危机是由于系统不均衡引起的<sup>[15]</sup>。关于金融系统性风险的量化研究,一种思路是利用商业银行或金融市场的数据来建立系统性金融风险度量模型。例如 Iling 和 Liu(2006)构建金融压力指数(FSI)研究金融系统风险问题<sup>[16]</sup>。陈守东和王妍(2014)将极值理论引入到系统性金融风险度量中,从而证明金融机构的市场价值的非正态分布特征<sup>[17]</sup>。王培辉和袁薇(2017)基于 CCA 和动态因子 Copula 模型对我国金融机构系统风险进行了评估<sup>[3]</sup>。另一种思路将金融系统视为网络,利用网络科学的理论方法建模研究整体风险的静态和动态特征。巴曙松(2013)<sup>[18]</sup>、Acemoglu 等(2015)<sup>[19]</sup>研究了金融网络及传染对金融稳定的影响, Axel(2016)利用贝叶斯网络进行系统性金融风险评价<sup>[20]</sup>, Hamed(2016)研究了不均匀金融网络的风险传染问题<sup>[21]</sup>, 贾彦东(2011)<sup>[22]</sup>、隋聪等(2016)<sup>[23]</sup>从网络视角进行了金融结构系统重要性分析和银行业系统性风险度量,邓超(2014)<sup>[24]</sup>、欧阳红兵(2015)<sup>[25]</sup>利用复杂网络研究系统性风险传染问题。

互联网金融带来金融系统结构的内在变化,复杂性和关联性增强,金融系统性风险呈现区别于传统金融的不同特征。Onay 和 Ozsoz(2014)分析了互联网金融与商业银行的贷款审批业务,发现互联网金融可以享受比较低的贷款利率<sup>[26]</sup>。吴成颂(2019)等运用沪深股市上市商业银行数据进行了互联网金融对商业银行金融系统性风险影响的量化实证<sup>[27]</sup>。顾海峰和杨立翔(2018)基于中国银行业数据研究了互联网金融与银行的风险共担问题<sup>[28]</sup>。Klaft(2008)对 P2P 行业进行分析,得出互联网金融的交易风险源于对贷款业务经验的缺乏<sup>[29]</sup>。张李义和涂奔(2018)从信息优势角度研究了互联网金融对同业市场利率的影响<sup>[30]</sup>。吴珂和谢晋雯(2018)基于 Z 值评分模型识别互联网金融风险诱因和严重程度<sup>[31]</sup>,陈耀辉和杨宁(2018)建立 La-VaR 模型研究了互联网金融流动性风险<sup>[32]</sup>。兰翔(2017)选择中证 800 金融指数,运用 VaR 分析与 Copula 的方法对互联网金融进行了风险测量,结果表明互联网金融市场的 VaR 与 ES 均高于传统市场,互联网金融市场本身可能具有更高的系统性风险<sup>[33]</sup>。王立勇和石颖(2016)采用二层次 CRITIC - 灰色关联模型构建互联网金融风险评价体系,运用 VaR 方法度量互联网金融风险<sup>[34]</sup>。另外,针对互联网金融风险传染问题,贾楠(2018)基于层次分析法分析了互联网金融在技术风险、操作风险、法律风险、信用风险和业务风险方面的影响,得出信用与技术是互联网金融中重要的风险传染因素<sup>[35]</sup>。朱宸和华桂宏(2018)根据互联网金融和银行业之间的相互作用,研究了互联网金融影响下的银行业系统性风险引发和传染机制<sup>[36]</sup>。张婷和米传民(2016)构建了超网络模型研究互联网金融均衡问题<sup>[37]</sup>,并考虑互联网金融中社交网

络关系研究了系统性金融风险传染问题<sup>[38]</sup>。

综上可以看出,互联网金融快速发展,带来金融开放、效率提升、成本降低的同时,也给互联网金融体系乃至整个金融系统带来系统性风险新问题。大量研究表明互联网金融无论从宏观金融体系层面,还是从微观业务层面,都给传统金融带来了影响,相应的金融风险问题也值得深入研究。目前对互联网金融下系统性金融风险的研究还处于初步阶段,虽然取得了一些成果,但相对于互联网金融的快速发展,有必要深入、系统研究其内在机理。随着互联网金融发展带来的数据和资料急剧增加,以及机器学习和人工智能的普及发展和深度应用,运用大数据、人工智能的思路和方法,研究互联网金融发展区域差异、空间聚集,从宏观、中观层面找到系统性金融风险点并进行监管,是一个新视角,也符合巴塞尔资本协议 III 的监管发展思路。本文基于北京大学互联网金融研究课题组收集的“互联网金融发展指数”数据,运用 t-SNE 算法构建模型,进行互联网金融区域发展的降维和聚类分析,解决互联网金融数据复杂性高和变量多而导致的维度过多问题,得到我国互联网金融空间聚集和不同业务模式(支付、货币基金、保险、投资)发展的分布特征,找出中观层面的系统性风险因素,进而提出互联网金融发展区域差异造成的三方面系统性风险,并提出防范互联网金融系统性风险的建议。

### 三、数据来源和研究方法

#### (一)数据来源

“互联网金融发展指数”由北京大学联合国内主要互联网金融企业,根据 2014 年 1 月至 2015 年 12 月的互联网金融数据发展编制而成<sup>[7]</sup>。本文选择了 31 个省(不含港澳台,下同)与 335 个地级市,共计 35136 个数据用于实证研究。该数据囊括了四种业务模式指标,并通过时间戳体现在每一个城市上,变量多且维数高。传统的多元统计方法在处理这类实际数据时会遇到数据不符合正态分布的情况,也难以直接观察空间结构。本文采用 t-SNE 机器学习降维算法对众多数据进行训练、可视化和聚类,来研究互联网金融的空间聚焦与业务发展的关系。

该指数将国内的互联网金融分为互联网支付、互联网基金、互联网信贷、互联网保险四种业务,是目前国内为数不多的互联网金融研究数据。在数据集中包括分地区互联网金融指数。分地区指数在计算过程中考虑了各地区指数点位之间的横向可比性。不同地区在某时期下相对交易渗透率、人均交易金额、人均交易笔数三个指标的计算公式如下:

$$A_{h,i,j,t} = \frac{X_{h,i,j,t}}{X_{i,j,t}} \quad (1)$$

其中,  $A_{h,i,j,t}$  表示  $t$  时刻下,  $h$  地区中第  $i$  类业务的第  $j$  个指标相对于全国总指数的相对值,  $X_{i,j,t}$  表示  $t$  时刻全国总指数的同业务同指标,  $X_{h,i,j,t}$  表示  $t$  时刻  $h$  地区的同业务同指标。

不同地区在某时期的互联网支付、基金、信贷、保险业务相对于全国系数计算公式如下:

$$B_{h,i,t} = \sum_{j=1}^3 m_j A_{h,i,j,t} \quad (2)$$

其中,  $B_{h,i,t}$  表示在  $t$  时期下  $h$  地区中第  $i$  类业务相对于全国总指数的相对系数,  $m_1$ 、 $m_2$ 、 $m_3$  分别表示上述的交易渗透率、人均交易额、人均交易笔数应占的权重,该指数中  $m_1 = 50\%$ ,  $m_2 = 25\%$ ,  $m_3 = 25\%$ 。

#### (二)t-SNE 机器学习算法

t 分布随机邻嵌入算法(t-distributed stochastic neighbor embedding, t-SNE), 由 Laurens van der Maaten 和 Geoffrey Hinton(2008)提出<sup>[6]</sup>, 后期经过 Maaten 的改进, 分别在 2015 年和 2016 年提出了

LINE 与 LargeVis 算法,很大程度上降低了训练复杂度<sup>[39][40]</sup>。该算法目前在降维、聚类、可视化的应用上取得了良好效果, Gordon Berman(2014)将 t-SNE 算法运用到对果蝇地面自由运动(即除了飞行)录像上进行降维分析<sup>[41]</sup>。刘丰等(2017)基于 t-SNE 算法实现了鸟类音频情报辨识的数据可视化<sup>[42]</sup>。詹威威和王彬等(2018)在高维脑网络状态观测矩阵中使用 t-SNE 算法,有效的解决了分散、交叉和散点<sup>[43]</sup>。本文选择的省级和地级市数据存在数据维度高且复杂的性质,包含四个业务模式、三百余个度量区域、二十四四个时间节点。经典数据处理方法,如回归分析、主成分分析、相关性分析等针对稀疏和多变量复杂数据集往往只能体现总体的关联,忽视了局部之间的联系,而 t-SNE 算法使用高低维二者的联合概率,能够有效解决优化困难与维度拥挤的问题,使得降维后的数据较好保持原有流形结构。本文运用 t-SNE 算法可将互联网金融指数有效映射为二维图像,有利于对整个互联网金融的发展进行观察,进而研究其系统性风险情况。

在高维空间中, t-SNE 算法采用高斯分布,对于高维数据点  $x$  的低维对应点  $y$  而言,可以计算条件概率  $p_{ij}$  来表示  $i$  点与  $j$  点之间的关系:

$$p_{ij} = \frac{(1 + \|y_i - y_j\|^2)^{-1}}{\sum_{k \neq i} (1 + \|y_k - y_i\|^2)^{-1}}, \forall i, j \text{ 且 } i \neq j \quad (3)$$

而在低维空间,采用自由度为 1 的  $t$  分布,以  $q_{ij}$  表示。

$$q_{ij} = \frac{q_{ij} + q_{ji}}{2n}, \forall i, j \text{ 且 } i \neq j \quad (4)$$

复杂数据降维过程中,最理想的状态是高维空间样本点之间的相似度与低维空间样本点之间的相似度相同。t-SNE 利用 Kullback-Leibler 散度作为目标函数来判断差异,从而实现最佳参数。

$$C = KL(P \| Q) = \sum_i \sum_j p_{ji} \log \frac{p_{ij}}{q_{ij}} \quad (5)$$

利用梯度下降法实现的最小化迭代公式如下:

$$\frac{\delta C}{\delta y_i} = 4 \sum_j (p_{ij} - q_{ij})(y_i - y_j)(1 + \|y_i - y_j\|^2)^{-1} \quad (6)$$

在实验过程中有 4 个因素影响降维效果,即概率模型的困惑度(perplexity)、前期放大系数(early exaggeration factor)、学习率(learning rate)、最大迭代次数(maximum number of iterations)。困惑度是对后期结果影响最为关键的因素,它的作用是用来获得高斯分布的方差。条件概率矩阵  $P$  的任意行困惑度可以定义为:

$$Perp(P_i) = 2^{H(P_i)} \quad (7)$$

其中,  $H(P_i)$  为  $P_i$  的香农熵。如果高维空间概率分布的熵越大,则数据集的变量不确定性越大,从而造成处理后的数据在分布上更加平坦。反之,高维空间概率分布的熵越小,则数据的关联性越高,处理后得到的数据越具有分布规律。香农熵与高维空间的概率分布关系为:

$$H(P_i) = - \sum_j p_{ij} \log_2 p_{ij} \quad (8)$$

#### 四、省级区域的互联网金融发展及系统性风险分析

为了检验 t-SNE 算法的有效性,通过与经典 PCA(principal component analysis)的比较研究,对 31 个省级区域(不含港澳台)的数值进行降维、聚类、成像和分析。

##### (一) 基于 PCA 的省级互联网金融发展及系统性风险分析

PCA 作为经典的数据挖掘算法被应用于许多数据挖掘场景。PCA 通过线性投影简化数据,将

高维数据映射到低维空间,尽可能保留原数据的内在信息。本文使用 PCA 将每个省份互联网金融发展状况降到 2 维,并在坐标轴上可视化,所得的成像效果见图 1(a),横坐标表示各省的发展体量指数大小,数值越大则发展情况越领先;纵坐标表示互联网的支付、保险、基金、信贷四大业务之间发展的均衡动荡程度,数值越大表明四大业务发展的体量越不均衡。

从 PCA 分析结果来看,31 个省级数据总体呈现出:个别省份发展迅猛、多数省份聚集的尖峰厚尾特征。但从互联网金融业务发展均衡程度上来观察,互联网金融整体发展优异的区域在业务间发展的差异性也低。这得益于发达省份互联网金融基础设施完善、互联网普及率较高、经济活力与消费水平旺盛。根据图来看,互联网金融总体发展程度较低的区域也表现出了不同业务之间的发展均衡现象,此现象可能是由于这些区域本身互联网经济发展动力不足导致的。相较于前两者,发展中游的城市出现了业务发展不均衡现象,说明了互联网金融在发展中城市的巨大潜力,也令尾部依赖这一特征在系统性风险积累过程中的显得尤为重要。

依照 PCA 处理后的数值结果,将其依照大小排序,见表 1。发展最快的几个城市,例如北京市,相比于全国平均值高出 8 倍之多;沿海东部省份互联网金融发展速度远远超过内陆城市,西部省份大部分处于互联网金融发展落后阶段。该排名同时反映出在互联网金融的系统性风险考量中,需要考虑城市在互联网金融发展中的重要程度。

表 1 基于 PCA 的 31 省级区域互联网金融排名

排序	1	2	3	4	5	6	7	8	9	10	11
区域	北京	上海	浙江	广东	江苏	福建	天津	湖北	山东	重庆	辽宁
数值(x)	816.8	806.5	524.9	248.5	218.3	142.7	106.3	-19.8	-27.2	-49.1	-50.8
排序	12	13	14	15	16	17	18	19	20	21	22
区域	陕西	海南	安徽	四川	山西	河北	江西	黑龙江	河南	吉林	新疆
数值(x)	-68.1	-79.7	-86.3	-89.4	-97.3	-109.2	-116.9	-122.1	-128.0	-135.2	-136.3
排序	23	24	25	26	27	28	29	30	31		
区域	宁夏	湖南	西藏	广西	内蒙古	云南	青海	贵州	甘肃		
数值(x)	-136.6	-139.1	-151.2	-172.1	-173.3	-179.0	-186.7	-200.8	-207.4		

## (二) 基于 t-SNE 的省级互联网金融发展系统性风险分析

本文使用 python 进行 t-SNE 算法计算,对 31 个省级区域的互联网金融系统风险分析。由于 t-SNE 是从非线性降维出发,将高维空间样本投影到多个二维空间映射图上,使用高低维双向概率分布,得出的结果具有随机性特点。本文结合了流行正则化技术来控制映射图,从而使得投影到可视化空间的样本点不但可以保持高维数据的整体结构,也保持了局部近邻点的关系。在 t-SNE 的数据处理中包括两个阶段,分别为前期放大阶段与后期性能选择阶段。实验中前期放大系数设置为 12.0,空间的联合概率通过乘以前期放大系数的方式逐步增加。重点调整困惑度与学习率,使得在单点近似区域实现迅速而准确的训练效果。不同参数下的分布形态、聚类效果见表 2。

表 2 不同 t-SNE 参数下的分布形式

困惑度	Per = 30.0		Per = 20.0		Per = 10.0		Per = 5.0	
训练次数	L = 200	L = 500	L = 200	L = 500	L = 200	L = 500	L = 200	L = 500
分布形式	均匀分布	均匀分布	均匀分布	均匀分布	聚类分布	离散分布	聚类分布	聚类分布

31 个省份的聚类分布在  $P_{er} = 5.0$  时更具备稳定的聚类特征,聚类效果如图 1(b)、(c) 所示,存在 31 个省级区域的聚类划分,大致划分为:

I 级:北京、上海、浙江、广东。

II 级:江苏、福建、天津、湖北、山东、重庆、辽宁。

Ⅲ级：陕西、海南、安徽、四川、山西、河北、江西、黑龙江、河南。

Ⅳ级：吉林、新疆、宁夏、湖南、西藏、广西、内蒙古、云南、青海、贵州、甘肃。

比较图 1(b)和(c)，基于 t-SNE 的结果囊括了发展强、较强、中游、以及落后的省份，聚类效果体现了我国国内互联网金融发展特征为：个别省份极端发展、多数省份聚集的尖峰厚尾现象。t-SNE 聚类结果更能体现互联网金融发展的区域性，有助于进行互联网金融的系统性风险分析。

从图 1 来看，PCA 算法够清晰反映各省互联网金融发展水平，解释不同区域在诱发互联网金融系统性风险中的重要性。但 PCA 成像在分布上出现了较为无序的子集杂糅，不能很好体现统计特性。一部分城市之间相互吸引，而另一部城市之间却远距离分散，难以直观观察到不同区域在互联网金融发展程度上的关系特点。

t-SNE 降维聚类结果呈现清晰的聚类团，不同互联网金融区域发展程度的界定更为明显。作为非线性算法，t-SNE 通过恢复数据低维度状态下的流行结构起到降噪作用，并体现数据内在关系，从而更好地反映降维前的系统特征<sup>[44]</sup>。使用 t-SNE 对 31 个省份的聚类，反应出我国互联网金融发展过程中区域间相互关联的特性，这为监管部门对互联网金融系统性风险进行分级分区域监管提供了数据驱动的决策参考。

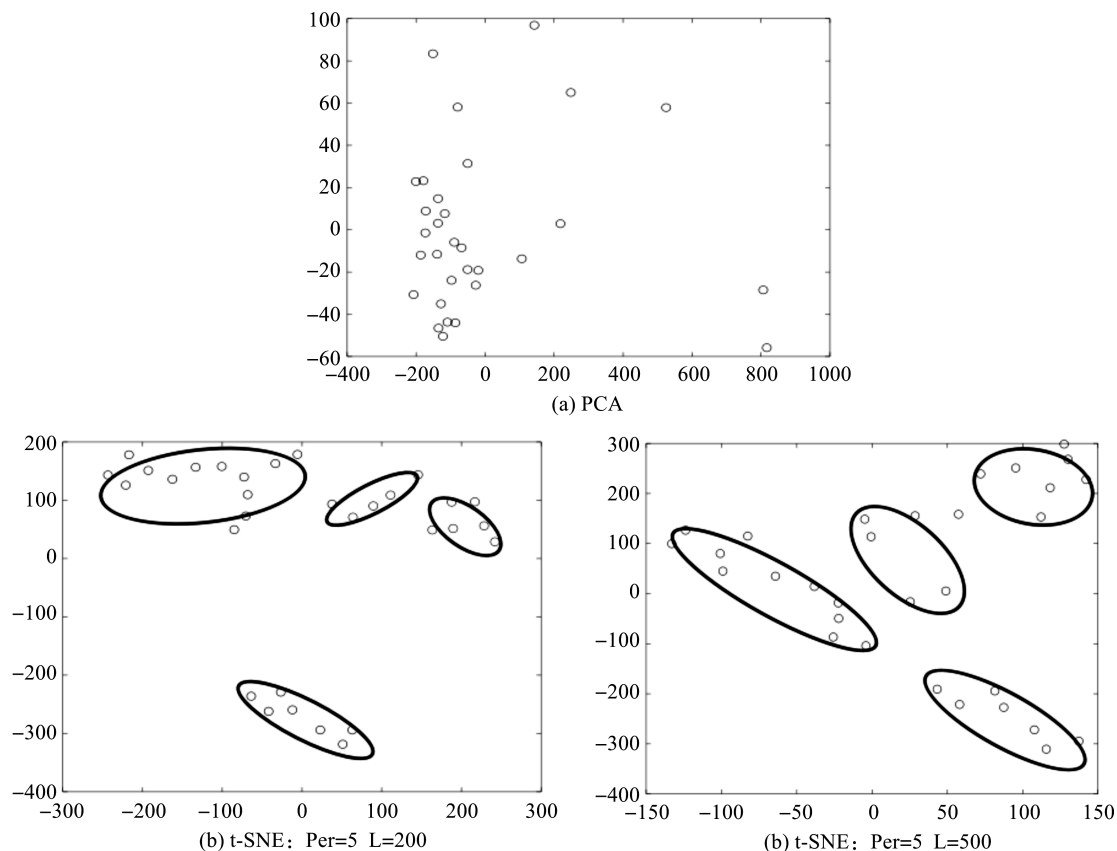


图 1 PCA 与 t-SNE 的省级互联网金融发展类效果对比

## 五、地级市的互联网金融区域发展及系统性风险分析

在第四部分，主要从省级层面进行了数据分析。但省级数据的横切面只有 31 类，无法体现更细节的互联网金融系统性风险特征。本节将使用全国 335 个地级市（包括自治州、盟、地区）的多

业务发展数据进行实验与分析。第四部分的 PCA 分析,在处理数据的过程中难以避免维度拥挤和数据杂糅的问题,无法产生良好的聚类效果,因此本节主要探讨基于 t-SNE 机器学习算法的地级市层面的互联网金融发展系统风险聚类效果。

#### (一)地级市互联网金融发展聚类特征分析

在地级市 t-SNE 机器学习聚类过程中,参数设置如下:前期放大系数为 12.0,困惑度为 30.0,训练次数为 5000,然后使用 K-means 算法对降维后的数据进行聚类,得到的聚类结果如图 2 所示,地级市的互联网金融发展情况汇聚成 7 大类,另外有一地级市的点分散在 7 大主要类的附近。对于较大相似度的地级市的点,t 分布在低维空间中的距离稍小一点,即同一簇内的点聚合的更紧密;而对于低相似度的地级市的点,t 分布在低维空间中的距离需要更远,即不同簇之间的点更加疏远。

#### (二)地级市互联网金融发展关联性特征分析

在利用 t-SNE 机器学习聚类过程中,通过参数调整,我们发现:当概率分布困惑度  $Per = 10.0$  时,数据分布产生了整体杂糅性,如下图 3 所示。这说明,第一,互联网金融发展在地区之间的聚集分布不是完全性的,虽然主要是依从一个整体,但是存在空间聚集差异。第二,在图 3 中,地级市互联网金融发展区域分布以横纵坐标原点为中心向外扩散。横坐标的大小代表相对于平均水平的各城市互联网金融发展情况,纵坐标表示互联网金融的支付、基金、信贷、保险四大分业务发展的均衡动荡程度,这与上文 PCA 算法得到的效果类似,也与互联网金融发展指数局部 Moran's 散点图得到的效果相似<sup>[7]</sup>。游离在聚焦原点外的数据大部分是一、三、四象限分布,少数在第二象限。第二象的坐标具体含义是:互联网金融发展指数较低、分业务发展的差异性很高,但实验结果是第二象限呈现极少的散点分布。结合我国互联网金融发展实际情况,以及上述 PCA 聚类结果,可以看出,中国西部城市在互联网金融整体水平较低的情况下,分业务发展的均衡性却与发达城市相同的现象。这也印证了本文第四部分得出的中国互联网金融发展情况是个别省份极端发展、多数省份相互聚集的尖峰厚尾现象。

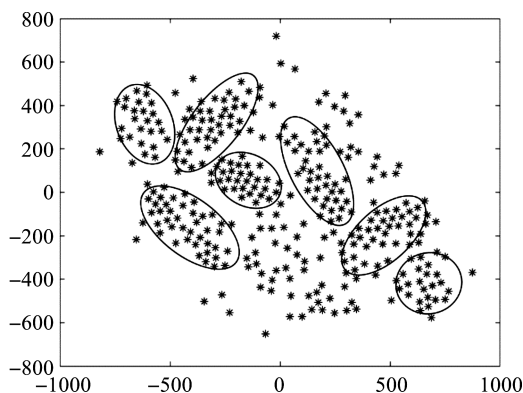


图 2 t-SNE 下的地级市互联网金融发展聚类图

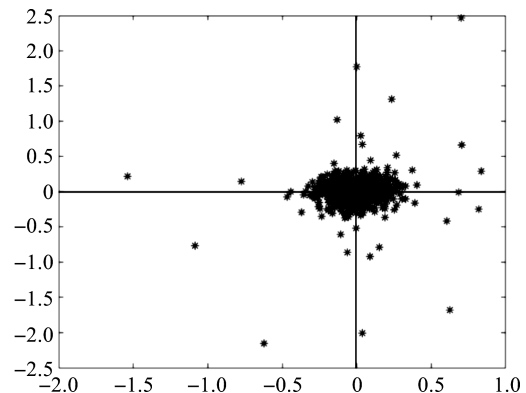


图 3 t-SNE 下的地级市互联网金融发展关联图

## 六、互联网金融空间聚集对系统性风险监管的管理启示

结合李建平(2010)等风险相关性与集成的研究成果<sup>[45]</sup>,从管理启示角度,本文认为我国互联网金融系统性风险在区域上的防控可从以下三个方面关注。

### (一) 地理空间聚集传染风险

我国互联网金融系统性风险需要注意区域性防范问题,避免区域性传染风险。从上述聚类结果看,中国互联网金融发展主要集聚在沿海东部地区,呈现区域特征。这一方面增大了互联网金融系统性风险的传染性,另一方面也反映出对聚集区域进行针对性监管的必要性。具体来说,东部发达地区体现出互联网发展速度极端化特点,该类区域经济金融发展水平高、互联网基础设施建设完善、互联网金融发展的需求与供给充足。这极大地促进了互联网金融发展在空间区域选择上的倾斜;经济发展良好的地区在互联网金融的发展速度上增速很高,且伴随有互联网金融不同业务之间的发展差异;经济欠发达的西部地区虽然受到互联网跨时空障碍、方便快捷的优势影响,但互联网金融发展的速度与动力要落后于东部地区。这种互联网金融发展的区域性差异,可能带来互联网金融的传染性系统风险,在宏观监管过程中需要重点关注。

### (二) 业务发展差异性风险

业务发展的差异性,可在中观层面带来系统性风险,需要关注。从分级业务的发展情况来看,互联网金融发展领先的城市在互联网支付、基金、信贷、保险四大业务上发展均衡,而发展中等水平的城市呈现了业务间发展高低不平衡的特征。这同 Allen 的研究一致,即从经济平衡的角度解释系统性风险,考虑到经济发展不仅体现在总体量上的差距,实际上也需要参考各级市场的发展情况<sup>[15]</sup>。我们的研究认为,从系统内在联系的角度,互联网金融发展处于中游的城市,其不均衡的业务发展意味着互联网金融发展存在市场差异,这也可能带来系统风险,需要在监管过程中关注业务发展的差异性。

### (三) 互联网金融发展系统性重要城市风险

从网络科学角度,网络中的节点间存在联系,不同节点的重要性不同,对网络的影响也不同。如果将全国互联网金融看成一个网络,根据上述聚类实证研究发现,北京、上海、深圳、杭州等城市是我国互联网金融发展突出的城市,对周边地区辐射效应也最为强烈。北京、上海基于传统金融中心基础,深圳依托微信相关互联网金融基础,以及杭州依托蚂蚁金融为主的互联网金融发展,为这些城市互联网金融发展提供了良好的机会。朱晓谦等(2018)的研究发现,单个金融机构的危机可能导致整个金融系统陷入危机,用概率可以度量系统性风险<sup>[46]</sup>。巴塞尔资本协议 III 针对 2007 年金融危机提出了系统性重要金融机构的概念,加强对系统性重要金融机构的监管。基于第五部分研究的结果,我们认为,考虑到互联网跨时空、集聚效应更强和风险传播更快的特点,除了关注系统性重要金融机构,有必要从中观层面,关注系统性重要城市可能给金融体系带来的系统性风险。

## 七、结 论

互联网金融给经济金融带来深刻影响,其风险问题也不容小觑。区别于传统的系统性风险度量研究方法,本文从风险传染视角,对高维数据进行降维和聚类处理,研究互联网金融在空间地理上的可视化展现与业务分布上的结构差异,得到了互联网金融发展的系统性风险区域特征。从实证结果来看,t-SNE 算法能够较好地捕获系统性风险发生的特征与薄弱环节,并得出互联网金融市场存在的尖峰、厚尾等特征,在此基础上本文提出三方面互联网金融系统性风险防控建议。

下一步的研究,一方面,可以利用互联网金融机构内部数据或仿真数据,进行更细颗粒度地建模,或许发现更有价值的结论。另一方面,本文提出的三方面互联网金融系统性风险有必要从金融风险传染、金融监管等角度进行建模研究,如通过研究区域聚集系数从而得到较为精确的区域风险比重;结合更具体数据,对互联网金融系统性城市、以及重要金融机构判断方法进行深入研究;运用科学的理论方法找到不同互联网金融业务发展的合适比例。

## 参考文献:

- [1] 谢平, 邹传伟. 互联网金融模式研究 [J]. 金融研究, 2012, (12): 11-22.
- [2] 刘柳, 屈小娥. 互联网金融改善了社会融资结构吗?——基于企业融资选择模型 [J]. 财经论丛, 2017, (3): 38-48.
- [3] 王培辉, 袁薇. 我国金融机构系统性风险动态监测——基于 CCA 和动态因子 copula 模型的研究 [J]. 财经论丛, 2017, (12): 43-53.
- [4] Adrian, Tobias, Markus K., Brunnermeier. CoVaR [R]. National Bureau of Economic Research, 2011.
- [5] Acharya V., Engle R. F. Richardson M. Capital Shortfall: A New Approach to Ranking and Regulating Systemic Risks [J]. American Economic Review, 2012, 102(3): 59-64.
- [6] L. J. P. van der Maaten, Hinton G. E. Visualizing High-dimensional Data Using t-SNE [J]. Journal of Machine Learning Research, 2008, (9): 2579-2605.
- [7] 北京大学互联网金融研究中心课题组. 互联网金融发展指数的编制与分析 [J]. 新金融评论, 2016, (1): 101-129.
- [8] 郭峰, 孔涛, 王靖一. 互联网金融空间聚集效应分析——来自互联网金融发展指数的证据 [J]. 国际金融研究, 2017, (8): 75-85.
- [9] 郭海凤, 陈霄. P2P 网贷平台综合竞争力评价研究 [J]. 金融论坛, 2015, (2): 12-23.
- [10] 廖理, 李梦然, 王正位. 聪明的投资者: 非完全市场化利率与风险识别——来自 P2P 网络借贷的证据 [J]. 经济研究, 2014, (7): 125-137.
- [11] 王赛芳. 我国互联网金融发展水平及空间分异研究 [J]. 农村金融研究, 2016, (11): 33-37.
- [12] 李清磊, 王旺. 安徽省数字普惠金融的聚类分析 [J]. 市场周刊(理论研究), 2018, (1): 132-133.
- [13] 巴曙松, 朱元倩, 金玲玲. 巴塞尔协议 III 与金融监管大变革 [M]. 北京: 中国金融出版社, 2015.
- [14] 王妍, 陈守东. 系统性金融风险研究——基于金融不稳定的视角 [J]. 数量经济研究, 2014, (2): 12-21.
- [15] Allen F. and Gale, D. Financial Contagion [J]. Journal of Political Economy, 2000, 108(1): 1-33.
- [16] Illing M., Liu Y. Measuring Financial Stress in a Developed Country: An Application to Canada [J]. Journal of Financial Stability, 2006, 2(3): 243-265.
- [17] 陈守东, 王妍. 我国金融机构的系统性金融风险度量——基于极端分位数回归技术的风险度量 [J]. 中国管理科学, 2014, (7): 10-17.
- [18] 巴曙松, 左伟, 朱元倩. 金融网络及传染对金融稳定的影响 [J]. 财经问题研究, 2013, (2): 3-11.
- [19] Acemoglu D., Asuman T. Systemic Risk and Stability in Financial Network [J]. American Economic Review, 2015, 105(2): 564-608.
- [20] Axel G., Luitgard A. A Bayesian Methodology for Systemic Risk Assessment in Financial Network [J]. Management Science, 2016, 63(12): 4428-4446.
- [21] Hamed A., Andreea M. Inhomogeneous Financial Networks and Contagious Links [J]. Operations Research, 2016, 64(5): 1109-1120.
- [22] 贾彦东. 金融机构的系统重要性分析——金融网络中的系统风险衡量与成本分担 [J]. 金融研究, 2011, (10): 17-33.
- [23] 隋聪, 迟国泰, 王宗尧. 基于网络视角的银行业系统性风险度量方法 [J]. 中国管理科学, 2016, (5): 54-64.
- [24] 邓超, 陈学军. 基于复杂网络的金融传染风险模型研究 [J]. 中国管理科学, 2014, (11): 11-18.
- [25] 欧阳红兵, 刘晓东. 中国金融机构的系统重要性及系统性风险传染机制分析——基于复杂网络视角 [J]. 中国管理科学, 2015, (10): 30-37.
- [26] Onay C., Ozsoz E. The Impact of Internet-banking on Brick and Mortar Branches: The Case of Turkey [J]. Journal of Financial Services Research, 2014, 44(2): 187-204.
- [27] 吴成颂, 王超, 倪清. 互联网金融对商业银行系统性风险的影响——基于沪深股市上市商业银行的证据 [J]. 当代经济管理, 2019, (2): 90-97.
- [28] 顾海峰, 杨立翔. 互联网金融与银行风险承担: 基于中国银行业的证据 [J]. 世界经济, 2018, (10): 75-100.
- [29] Klaffit, Michael. Peer to Peer Lending: Auctioning Microcredits over the Internet [C]. Proceedings of the International Conference on Information Systems, Technology and Management, Dubai, 2008.
- [30] 张李义, 涂奔. 互联网金融信息优势对同业市场利率影响的实证研究——基于商业银行经营决策分析 [J]. 财经论丛, 2018, (2): 47-57.
- [31] 吴珂, 谢晋雯. 互联网金融风险诱因和严重程度的识别——基于 Z 值评分模型 [J]. 商业经济研究, 2018, (22): 154-156.
- [32] 陈耀辉, 杨宁. 基于 La-VaR 模型互联网金融流动性风险的研究 [J]. 时代金融, 2018, (32): 341-342.
- [33] 兰翔. 基于 VaR 分析与 Copula 方法的互联网金融风险度量 [D]. 济南: 山东大学硕士学位论文, 2017.

- [34] 王立勇, 石颖. 互联网金融的风险机理与风险度量研究——以 P2P 网贷为例 [J]. 东南大学学报(哲学社会科学版), 2016, (2): 103-112, 148.
- [35] 贾楠. 中国互联网金融对银行业风险影响及其系统性风险度量研究 [J]. 经济问题探索, 2018, (4): 145-157.
- [36] 朱辰, 华桂宏. 互联网金融对中国银行业系统性风险的影响——基于 SCCA 模型及逐步回归法的实证研究 [J]. 金融经济学研究, 2018, (2): 50-59.
- [37] 张婷, 米传民. 基于超网络的互联网金融均衡问题研究 [J]. 复杂系统与复杂科学, 2016, (2): 36-43.
- [38] 米传民, 李丹丹, 张婷, 等. 考虑社交关系水平的互联网金融超网络模型及均衡研究 [J]. 中国管理科学, 2018, (12): 56-65.
- [39] Tang J., Qu M., Wang M. et al. Line: Large-scale Information Network Embedding [C]. Proceedings of the 24th International Conference on World Wide Web, 2015: 1067-1077.
- [40] Tang J., Liu J., Zhang M. et al. Visualizing Large-scale and High-dimensional Data [C]. Proceedings of the 25th International Conference on World Wide Web, 2016: 287-297.
- [41] Gordon Berman, Daniel M. Choi William Bialek, Joshua W. Shaevitz. Mapping the Stereotyped Behaviour of Freely Moving Fruit Flies [J]. Journal of the Royal Society Interface, 2014, 11(99): 1-12.
- [42] 刘丰, 赵疏涛, 廖晨宇, 等. 基于 t-SNE 声音情报识别的可视化系统在 E-learning 平台应用——以鸟类音频情报识别为例 [J]. 企业技术开发, 2017, (4): 9-12.
- [43] 詹威威, 王彬, 薛洁, 等. 自适应加权 t-SNE 算法及其在脑网络状态观测矩阵降维中的应用研究 [J]. 计算机应用研究, 2018, (7): 2055-2058, 2070.
- [44] 姚雪曼. 基于流形学习的降维技术的研究与实现 [D]. 北京: 北京邮电大学硕士学位论文, 2017.
- [45] 李建平, 丰吉闯, 宋浩, 等. 风险相关性下的信用风险、市场风险和操作风险集成度量 [J]. 中国管理科学, 2010, (1): 18-25.
- [46] 朱晓谦, 李靖宇, 李建平, 等. 基于危机条件概率的系统性风险度量研究 [J]. 中国管理科学, 2018, (6): 1-7.

## Analysis of Internet Financial Spatial Aggregation and Systematic Risk Prevention ——Based on t-SNE Machine Learning Model

MI Chuanmin, XU Runjie, TAO Jing

(College of Economics and Management, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

**Abstract:** The development of Internet finance has different spatial focusing characteristics from traditional finance. Internet finance gives rise to the new problem of systemic risk in the Internet financial system, and even in the financial system as a whole, while bringing about advantages of financial opening, lowering the threshold, improving efficiency and reducing the cost. A large number of studies have shown that Internet finance often has different characteristics of systemic financial risk compared with the traditional finance because of its characteristics such as macroeconomic shock or internal vulnerability. Using the data of Internet financial development in 31 provinces and 335 metropolitan areas, this paper uses t-SNE machine learning algorithm to analyze the health of Internet finance, reveals the distribution characteristics of Internet financial space polymerization and the development of different business models in China, and finds that there are spikes and thick tails in the degree of regional development. There is also an imbalance in the business models. Based on the above analysis, the paper not only suggests that the three aspects of the systemic risk caused by regional differences in Internet financial development should be taken into consideration, but also puts forward some proposals to prevent the systemic risks in Internet finance.

**Key words:** Internet Finance; Systemic Risk; Dimensionless Clustering; t-SNE Algorithm

(责任编辑: 原 蕴)