
基于文本分析的武汉市社区治理趋势与需求探究

谭星¹ 卢思远² 郭泉³¹

(1. 武汉理工大学 管理学院, 湖北 武汉 430070;

2. 武汉理工大学 信息学院, 湖北 武汉 430070;

3. 武汉理工大学 计算机学院, 湖北 武汉 430070)

【摘要】: 武汉市凭借着丰富的经验有效地控制住疫情, 对于社区治理方面的研究热点久居不下, 在该背景下, 把握文本信息并对当前武汉市社区治理的趋势和需求分析显得尤为重要。本文采用改进的 TF-IDF 算法和语义网络从政策信息和居民动态两个方面的文本信息进行了趋势和需求分析, 最后从基层治理、技术创新和人才引进三个角度给出建议。

【关键词】: 社区治理 改进的 TF-IDF 语义网络 文本分析

【中图分类号】:F2 **【文献标识码】:**A

1 研究设计

待分析的有两方面内容: 一是社区治理的趋势; 二是社区治理的需求。对于趋势的分析, 可采用关键词提取算法针对工作报告、意见等政策类文本进行关键词挖掘; 对于需求的探究, 可采用语义网络对社交平台上武汉市社区居民发布的动态进行网络分析。

1.1 数据来源与处理

本文所分析的全部为文本型数据。其中, 趋势分析的数据来源于武汉市民政局官网(mzj.wuhan.gov.cn), 在该官网中采集的数据具体为: 武汉市 2012-2021 年政府工作报告、政府公开的政策与意见和关注度较高的新闻。需求分析的数据来源于各个社区用户微博、知乎等社交平台的动态正文(不包括评论与回复)。文本信息的检索词设定为社区治理、基层治理, 数据采集时间为 2022 年 2 月 10 日。

通过 Python 的 Scrapy 框架, 给定不同的初始网址, 可以较快速的抓取所需数据, 之后进行预处理, 包括机械压缩、无效内容去除、构建停用词典等, 最终获取的有效信息情况所示。

作者简介: 谭星(2001-), 男, 汉族, 湖北武汉人, 武汉理工大学管理学院本科在读, 研究方向: 自然语言处理。

基金项目: 武汉理工大学创新创业训练项目(S202110497056)

1.2 分析方法

1.2.1 改进的 TF-IDF 算法

TF-IDF 是一种较为有效的关键词提取算法，它是根据词频与逆文本频率的大小来综合判断某个词的重要程度，在政治论述类文章里词频的大小往往反映重要性程度，因而相较于 TextRank、LDA 主题算法，更能体现出它的优越性。算法分为两部分：一部分为 TF 算法；另一部分为 IDF 算法，在实际使用中，TF 的计算式如下：

$$tf_i = \frac{n_{ij}}{\sum_k n_{kj}}$$

其中 n_{ij} 表示词 i 在文档 j 中的出现频次，对词频进行归一化后，分母就是统计文档中每个词出现次数的总和。

$$idf_i = \log \left(\frac{|D|}{1 + |D_i|} \right)$$

IDF 的计算方式为：

其中， $|D|$ 为文档集中的总文档数， $|D_i|$ 为文档集中出现词 i 的文档数量，分母加 1 采用了拉普拉斯平滑，避免有部分新的词段没有在语料库中出现而导致分母为零的情况出现，增强算法的健壮性。

那么对于词 i 的 TF-IDF 值，有：

$$tfidf_i = tf_i \times idf_i$$

但传统的 TF-IDF 忽略了类间集中度和类内分散度造成的问题，因此需要引入修正因子 η 来进行改进。 η 表示为：

$$\eta = \frac{m_i}{m_i + k_i} \times \frac{m_i}{m_i + q_i}$$

其中， m 和 k 分别表示为某文档集和其他文档中包含词 i 的数目， q 为不包含词 i 的数目，因而修正系数可以表示为集中度和分散度的乘积。最终改进的 TF-IDF 计算公式如下：

$$tfidf'_i = tf_i \times idf_i \times \eta$$

1.2.2 语义网络分析

对于居民用户的动态数据，适合采用语义网络分析，整体分为两步：第一步还是关键词的提取；第二步为共现分析，根据关键词的重要程度得到共现矩阵，再根据共现矩阵来进一步理解各词之间的关系与亲疏程度。

1.3 研究工具

本研究采用的工具为 Python、ROSTCM6 以及 Gephi0.9.2。用 Python 进行数据爬取与预处理并利用改进的 TF-IDF 算法进行关键词提取，接着利用 ROSTCM6 进行语义网络分析，最后利用 Gephi0.9.2 将语义网络分析得到的共此矩阵可视化。

2 结果分析

2.1 从政策信息看武汉市社区治理的趋势

将所收集到的政府工作报告、公开政策与意见和相关热点新闻整理至 csv 文件中并代入编写好的改进 TF-IDF 算法框架程序中运行，得到的各关键词的 TF-IDF 值降序排列。

针对 TF-IDF 的值对其进行词云可视化，可视化结果如图 1 所示。

可以较为清晰地看出，网格化、智慧化和信息化是当前武汉市社区治理较为明显的趋势特征，“基层”在文本中的重要程度反映了当前社区治理的中心有所下沉。“数据”“信息”“平台”和“精准”等词体现了将先进技术、智能设施以及大数据相结合的重要性。



图 1 政策文本词云图

2.2 从居民动态看武汉市社区治理的需求

需求部分的分析利用的数据为居民发布在平台上的动态信息，先使用 Python 的 jieba 库进行分词并储存为 csv 格式文件，然后将该文件导入 ROSTCM6 进行语义网络分析，再将语义网络分析得到的共词矩阵导入 Gephi0.9.2 进行可视化，结果如图 2 所示。

语义网络的结点大小和边的粗细分别对应着该社区治理网络中词的重要程度和共现度，而某词中心度的大小可由与之相连接的词的数目来衡量。从上图可以看出，“智能”的中心度最高且连接着“问题”，说明当前武汉市的社区智慧化程度还有待提高，对于各个治理模块，智能化都是关键要素。再与“智能”相连接的词中，“数字”“服务”“治理”和“反馈”的重要程度和中心度紧随其后。此外，从“基层”和“反馈”两词及其相连接的词可以看出，基层治理仍需要进行加强和创新，对于治理意见的反馈平台存在着一定的问题，需要进一步提高资源的共享性。

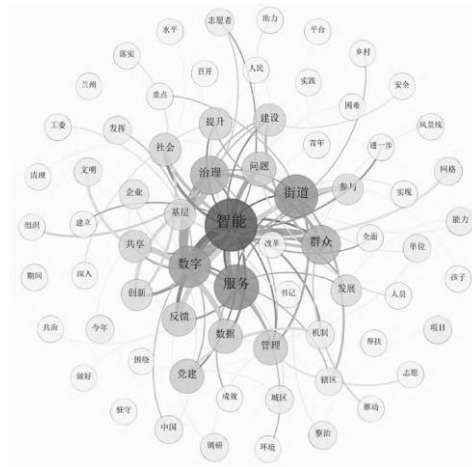


图 2 语义网络分析结果图

3 结论与建议

在社区地位不断提高的背景下，武汉市在社区治理这方面有扎实的基础与丰富的经验。网格化、智慧化、信息化和重心下沉化是从疫情爆发以来不变的趋势，信息化平台或网站的缺乏给居民反馈和相关人员的管理造成一定的难度，党建引领、智慧赋能、信息共享是当前武汉市居民的主要需求，加强社区治理的创新并打造社区治理的高效模式具有较大意义。现从基层治理、技术创新和人才引进三个角度给出建议：

- (1)落实基层治理，将党员干部下沉社区。

不断强化居民需求导向，更有效地完善基层治理的协同合作模式。做到“两个加强、两个引入”，加强党的影响力与基层行政能力，引入新的社区服务与社区治理机构。在此基础上，强化多元共治，推进“三治”结合。

- (2) 加快先进技术的引入与创新，提高智慧化程度。

运用数字技术推动社区服务与管理的精细化,利用数字技术赋权打造共治共享的新格局,消除治理主体之间的交流阻碍,尽快实现智能设施在武汉市社区的普及,对现有基础设施进行数字化改造升级,跨越“数字鸿沟”,提高社区数字化治理效能。

- (3) 坚持人才引领发展的战略地位。

政府应出台一系列政策加强专业人才的引进，鼓励技术人员从实践中创新，提高对居民意见、反馈等平台文本信息的重视程度并不定期结合 NLP 前沿领域的算法进行文本分析，为数字化治理效能的提高夯实基础。

参考文献:

- [1] 韦仁忠, 张作程. 新时代城市社区治理的现实境遇和实践向度——基于协同共治的视角[J]. 领导科学论坛, 2022, (01):31-37.
- [2] Hanguang Hu. Research on the Construction of Grass Roots Governance System of “Three Governance

Combination”[J].International Journal of Higher Education Teaching Theory,2020,1(4).

[3]周莹莹. 社会主体多元化治理中的社会参与问题浅析[J]. 经贸实践, 2018, (06):7-8.

[4]Yang Yutang.Knowledge Map of Community Governance Research in the Last 10 Years—Visual Analysis Based on CiteSpace[J]. 社会科学前沿, 2021,10(12).

[5]马晗. 中国城市社区治理的研究热点与发展趋势——基于 CiteSpace 的可视化分析[J]. 社会福利(理论版),2021, (04): 37-45.

[6]Yili Yao,Xianghong Zhou.Impacts of the internet on perceptions of governance at the community level:the case of Jiangqiao Township in Shanghai,China[J].Public Administration and Policy,2021,24(2).

[7]滕方炜. 基层社会治理现代化: 时代逻辑与路径选择——基于社区治理的文本分析[J]. 传播力研究, 2018,2(18):1-3.

[8]胡媛敏, 张寿明. 基于信息论的 TF-IDF 算法在文本分类中的应用研究[J]. 中国水运(下半月),2021,21(12):31-33.

[9]Xinming Zhang,Zhang Xinming,Shi Yuanbo,Wei Haiping.Research on TFIDF Algorithm Based on Weighting of Distribution Factors[J].Journal of Physics:Conference Series,2020,1621(1).

[10]Christensen Alexander P,Kenett Yoed N.Semantic network analysis (SemNA):A tutorial on preprocessing, estimating,and analyzing semantic networks[J].Psychological methods,2021.