doi:10.3772/j.issn.2095-915x.2024.03.004

开放科学（资源服务）标识码（OSID）

基于 LDA-ARIMA 的我国智能手机关键技术主题识别与演化分析

庞庆华　姚玉康　张丽娜

河海大学商学院　常州 213022

摘要：[目的 / 意义] 专利是技术能力的表现形式，包含关键技术主题信息。以智能手机专利数据为基础，提出关键技术主题的识别和演化分析方法，帮助企业获取行业内的技术信息，调整专利研究的成本与精力投入。[方法 / 过程] 首先，选择专利数据库和高级检索内容，下载和导出专利标题和摘要数据，并对数据进行去停用词和 jieba 分词等处理；其次，构建困惑度求解模型，确定最优主题数，再将已经分好词的文本导入 LDA 模型进行主题挖掘，得到每个技术主题下关键词语的分布；再次，将主题热度转化为时间序列，进行平稳性检测和白噪声检验，确定 ARIMA 模型参数后应用预测；最后，依据每个技术主题下提取的特征词确定关键技术主题并进行解读，通过时间序列预测结果对关键技术主题进行演化分析。[结果 / 结论] 以智能手机为研究对象，成功识别出屏幕、电池与充电、生物识别系统等 15 个关键技术主题，挖掘出各主题不同的演化发展特征，根据演化趋势分析提出建议，验证了本文主题识别与演化分析方法的可行性与有效性。

关键词：关键技术；主题识别；LDA 模型；主题演化；ARIMA 模型；智能手机

中图分类号：G35; TP391

Identification and Evolution Analysis of Key Technologies Topics of Smartphones in China Based on LDA-ARIMA

PANG Qinghua　YAO Yukang　ZHANG Lina

School of Business, Hohai University, Changzhou 213022, China

Abstract: [Objective/Significance] Patent is a manifestation of technological capability and contains information on key technological topics. Based on smartphone patent data, the identification and evolution analysis method of key technological topics is proposed to help enterprises obtain technological information in the industry and adjust the cost and effort invested in patent research. [Methods/Processes] Firstly, the patent database and advanced search content are selected, and the patent title and abstract data are downloaded and exported. The data is then processed, including deactivating words and jieba-splitting.

基金项目　河海大学 2022 年度“优秀专业硕士学位论文选拔培育”（422003524）。

作者简介　庞庆华（1977-），博士，教授，主要研究方向为竞争情报；姚玉康（1999-），硕士研究生，主要研究方向为数据分析与挖掘；E-mail：yyk13382892903@163.com；张丽娜（1986-），博士，副教授，主要研究方向为资源规划与决策。

引用格式　庞庆华 , 姚玉康 , 张丽娜 . 基于 LDA-ARIMA 的我国智能手机关键技术主题识别与演化分析 [J]. 情报工程 , 2024, 10(3): 49-62.

doi:10.3772/j.issn.2095-915x.2024.03.004

Subsequently, a perplexity solution model is built to determine the optimal number of topics. The preprocessed text is then imported into the LDA model for topic mining, resulting in the distribution of key words under each technical topic. The topic's heat degree is transformed into a time series, followed by smoothness and white noise tests. The parameters of the ARIMA model are determined and applied for prediction. Finally, the feature words extracted under each technical topic are used to identify and interpret the key technical topics. The evolution of these topics is analyzed through the results of the time series prediction. [Results/Conclusions] By focusing on smartphones as the research subject, this study successfully identifies 13 key technological topics such as screen, battery and charging, and biometric system. The different developmental characteristics of each theme are analyzed. Moreover, suggestions are provided based on the analysis of evolutionary trends, thus confirming the feasibility and effectiveness of the theme identification and evolutionary analysis method proposed in this paper.
Keywords: Key Technology; Topic Recognition; LDA Model; Topic Evolution; ARIMA Model; Smart Phone

引言

智能手机作为现代社会不可或缺的一部分，经历了令人瞩目的发展历程。从 2012—2022 年，智能手机行业经历了巨大的变革和创新，全球市场的规模也得到了显著扩大。互联网数据中心（Internet Data Center，简称 IDC）公布 2017 年智能手机出货量达到 15.5 亿台，触及顶峰，而后开始有所下滑。据 IDC 显示，2023 年全球智能手机出货量规模约为 11.7 亿台，同比下降 3.2%。在智能手机市场上，以苹果、三星为代表的美韩企业一直处于领先地位。苹果首次超越三星成为第一位，出货量 2.35 亿台；三星位于第二，出货量 2.27 亿台。值得注意的是，中国的智能手机企业在过去几年中取得了令人瞩目的成绩，在全球市场上崭露头角。2023 年小米、OPPO 分别以 1.46、1.03 亿台的成绩处于第三、第四名，vivo 也在 2023 年第四季度实现同比正增长。根据数据显示，2023 年中国企业的市场份额约占据了全球智能手机市场的三分之一以上，反映了中国企业在技术创新和市场推广方面取得的巨大进步。

关键技术是在产业内起到引导与扩张作用的重要技术，对行业和领域的发展有关键决定作用 [1]。随着智能手机技术的发展迅速，产业变革急速推进，智能手机关键技术的研发和创新是国内各大主流智能手机生产厂商的研究重心。由于芯片、操作系统等受制于国外的技术封锁，我国的智能手机企业存在着短时间难以突破的技术壁垒。在这样的背景下，我国智能手机企业更要加强对关键技术的攻关，合理配置科技资源。

专利是技术能力的表现形式，包含关键技术主题信息。为了更好获取智能手机行业内的重要信息，本文基于专利数据库中的智能手机专利摘要数据，构建了主题识别模型对智能手机关键技术主题进行识别与解读，并将主题热度转化为时间序列进行演化分析。本文旨在帮助我国智能手机企业随着技术市场的转变积极调整技术战略方向，抢夺技术先机，从而在智能手机国际市场的同行业竞争中取得优势。

1 文献综述

1.1 关键技术主题识别相关研究

关键技术主题指对于关键技术的相关特征与内容的总结，关键技术主题识别的过程主要

有两步，分别是：①文本预处理，主要采用 jieba 分词方法 [2] 以及 TF-IDF 算法 [3] 等。其中，传统的分词方法存在识别误差，已有学者对此进行改进，如吕鲲等 [4] 将分词方法相组合，将 jieba 分词（词典分词）和 TF-IDF 算法（统计分词方法）组合运用，提高分词效率，同时也提升了关键词提取的正确率。②文本挖掘，主要使用了 LDA 模型 [5] 以及 SCAN 算法 [6] 等。在文本挖掘方面，伊惠芳等 [7] 改进了增加词汇语境的 Content-LDA 模型，从综述学术论文结束语中挖掘领域技术问题，分析领域主体在生产和创新活动中遇到的各类技术需求，辅助确定关键技术主题。Xu 等 [8] 构建 LDA 主题模型时，在给予响应变量的基础上加入类别标签，提升了中文与外语语料库分类的精细度。王秀红等 [9] 将 BERT 与 LDA 相结合，充分考虑文本的语义信息和上下文信息，明显提高关键技术识别时主题的连贯性及细粒度划分的精准度。

1.2 主题演化分析相关研究

国内外对于技术主题演化分析的方法主要有 ARIMA 模型 [10] 和 SAO 结构法 [11] 等。Bian 等 [12] 组合运用了 ARIMA 模型与 GM 模型、广义神经回归网络模型，研究了趋势预测的组合分析模型。Lv 等 [13] 将 LDA 与 NMF 模型结合，构建双层识别模型，并构建相似性和差异性评价指标，通过关键技术主题演化分析对比主题的发展性和关键性。岳丽欣等 [14] 将 ARIMA 模型的趋势预测功能与人为定性判别方法相结合，利用时间序列模型提高了趋势预测的准确性和科学性，让学科情报分析更加严谨。Hu 等 [15] 从重要性和受欢迎程度两个方面分析了技术话题的演变趋势，并通过构建 ARIMA 模型，预测 2021—2025 年各技术主题在新能源汽车中的流行程度和发展趋势。李勇等 [16] 以生命周期为划分阶段，统计专利摘要词频分析技术热点演变，构建语义网络识别技术关联及技术主题的应用侧重点，更加精准地识别数字文旅关键技术及趋势。

1.3 智能手机专利分析相关研究

智能手机作为互联网时代的产物，在近年来备受关注，对于智能手机技术的研究也有较高的热度。随着专利分析的发展，部分学者尝试将技术识别的方法运用到智能手机技术领域的相关研究中。Sun 等 [17] 提取智能手机技术专利引文数据，根据相关专利参数获取了智能手机核心专利，规划了智能手机相关专利的扩散路径，依据 BCRW 算法实现了专利申请人之中的核心企业的识别。Wang 等 [18] 认为智能手机相关企业需要完善硬件与软件的相关专利布局，同时将智能手机与其他科技融合领域的专利布局进行拓展，将智能手机作为核心，实现了全面的智能科技生态格局。苏敬勤等 [19] 基于德温特数据库中的智能手机专利数据，将专利引用量和被引用量作为指标构建关于时间的变化曲线，分析了颠覆性技术的演化趋势。闯家梁 [20] 基于专利引用量、专利数量等指标构建了智能手机相关技术发展前景的评价体系，分析了我国智能手机专利的技术创新能力，为我国的智能手机相关产业的发展提供了技术指导。彭英等 [21] 制作了专利组合图像来研究智能手机企业的技术竞争力，以专利数量作为指标，分析智

doi:10.3772/j.issn.2095-915x.2024.03.004

能手机企业的技术能力，定义了智能手机专利的相对技术优势进而计算分析了技术强度。

综上所述：一方面，大部分学者将研究重心放在以文本聚类为基础的关键技术主题识别上，关键技术主题识别和演化分析的综合性分析需要进一步深入；另一方面，学者的研究主要集中在研究智能手机的软件层面，而对于硬件技术方面的研究有待加强，另外在智能手机技术领域中关键技术主题识别和主题演化分析方法也需要进一步补充。

针对现有研究所存在的不足，本文将 LDA 主题识别模型与 ARIMA 时间序列分析模型相结合，提出一种基于 LDA-ARIMA 的关键技术识别与演化分析方法。本文以专利标题与摘要作为研究数据，构建智能手机专利数据的主题分析模型，挖掘专利数据中的关键技术主题分布情况；将专利研究热度作为时间序列导入趋势预测模型，对智能手机关键技术主题进行演化分析，能够获取智能手机各主题的专利研究热度，帮助企业抓住新的发展机遇。

2 研究设计

2.1 研究思路

本文以智能手机为研究对象，采用 LDA 与 ARIMA 组合方法进行关键技术主题识别与演化分析。本文的研究框架流程分为如下阶段：①数据获取与预处理：选择专利数据库和高级检索内容，下载和导出专利标题和摘要数据，并对数据进行去停用词和 jieba 分词等处理；②基于 LDA 模型的技术主题识别：构建困惑度求解模型，确定最优主题数，再将已经分好词的文本导入 LDA 模型进行主题挖掘，得到每个技术主题下关键词语的分布，并对结果进行聚类以及可视化分析；③基于 ARIMA 模型的技术趋势预测：将主题热度转化为时间序列，进行平稳性检测和白噪声检验，确定模型参数，最后应用预测；④关键技术主题解读与演化分析：依据每个技术主题下提取的特征词确定关键技术主题并进行解读，通过时间序列模型预测结果对关键技术主题进行演化分析，具体研究框架如图 1 所示。

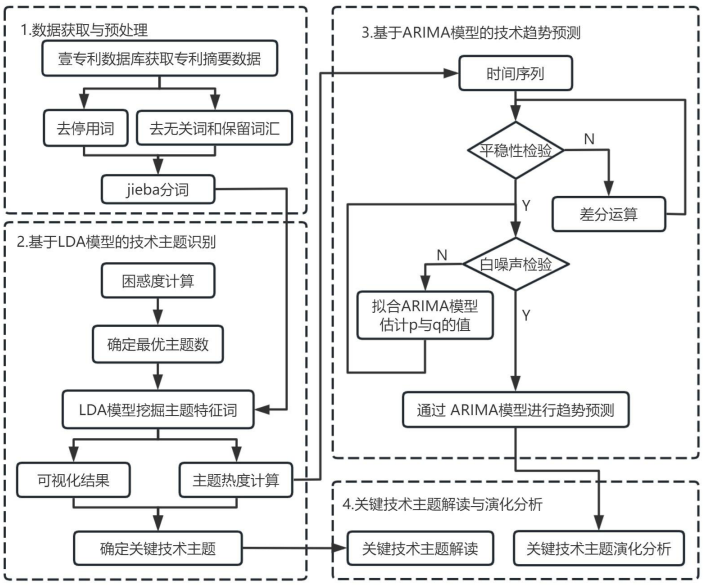

图 1 研究思路

2.2 研究方法

2.2.1 LDA模型

本文利用 LDA 主题模型求得智能手机专利集中每篇专利摘要的主题概率分布，再根据主题分布进行主题聚类和文本分类。其中，一篇专利摘要可以包含多个主题，根据计算产生主题分布概率，可以认为专利摘要中的每个词都由其中的一个主题生成。本文 LDA 模型结构原理如图 2 所示：

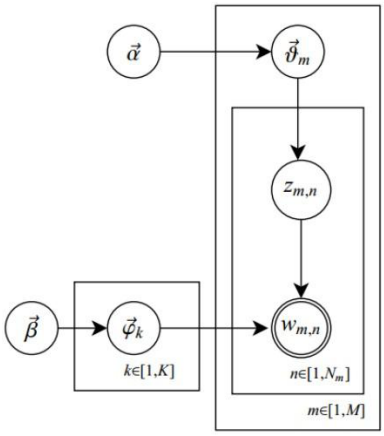

图 2 LDA 模型结构

主题数量通常视不同场景进行调整，本文采用计算困惑度（Perplexity）[22] 的方法来衡量选取智能手机的不同主题数量的优劣：

$$Perplexity(D_{test}) = \exp\{\frac{-\sum_{d=1}^{M}\ln(p(w_d))}{\sum_{d=1}^{M}N_d}\} \quad (1)$$

其中，M 是测试语料的大小（文档的数量），N_d 是第 d 篇文本大小。随着 K（主题个数）值的增大，困惑度逐渐减小。根据手肘法 [23]，困惑度图像存在比较显著的拐点时，选取最佳的 K 值。

本文使用 LDA 模型进行主题识别，利用困惑度确定 LDA 模型的最优主题数，将已经分好词的专利数据导入 LDA 模型，制定好模型参数，选取出现特征词语的频次数，根据困惑度曲线选择主题数量，再确定每个主题提取的特征词数量，将专利标题和摘要以及分词结果存入 EXCEL 结果文件，并输出每条专利数据对应概率最大的主题标号和每条数据对应各个主题的概率，选取各个主题中具有代表性的词作为特征词。

2.2.2 ARIMA模型

ARIMA（自回归积分移动平均）是一种广泛使用的时间序列分析方法。它是一种利用以前的观察结果预测未来趋势的统计模型。ARIMA 模型由三个部分组成：AR（自回归）、I（积分）和 MA（移动平均）。在智能手机专利领域，研究热度趋势具有时间序列特征，ARIMA 模型可以反映智能手机专利研究热度的趋势、季节性特征和随机变化，提供良好的未来趋势预测。

本研究使用的 ARIMA 模型建立步骤如下：首先需要对主题热度时间序列进行平稳性检验，若检验结果不合格，便对时间序列差分运算直至数据达到平稳；其次对时间序列进行白噪声检验；如果主题热度是平稳非白噪声时间序列，就计算 ACF（自相关系数）、PACF（偏自相关系数），确定 ARIMA 模型中参数 p、q 的值，通过拟合精度和参数确定模型的综合最优化配置。

3 研究过程

3.1 数据获取与预处理

3.1.1 数据获取

壹专利数据库覆盖了全球 159 个国家的 1.65 亿多条专利数据，其中囊括专利名称、摘要、

doi:10.3772/j.issn.2095-915x.2024.03.004

专利号、专利价值度、专利状态等权威内容，是国内相对权威的专利数据库，能够成为精准、高效的专利检索和分析工具，因此本文选择壹专利数据库中的智能手机专利信息作为数据来源。

然后构建专利检索式：由于 2011 年为中国智能手机元年，国内智能手机企业如中兴、华为、魅族等开始生产基于 Android 系统的智能手机，于是将检索时间段确定为 2011 年 1 月 1 日—2023 年 7 月 1 日，检索时间为 2023 年 7 月 1 日。由于本文的研究对象主要是国内智能手机企业，故数据范围选择“国内”，申请人类型（中国）选择“企业”且专利状态选择“有效”。本次检索获得 124282 条专利，专利公开的年度趋势如图 3 所示。专利公开量随年份增长急剧增加，其主要原因是智能手机开始普及，国内各大主流智能手机生产厂商加大了专利研发力度。专利数量统计的申请人排名情况如图 4 所示，其中 OPPO、华为、联想、小米、维沃、努比亚是我国知名的通讯设备制造厂商，可见其对专利申请、技术创新的重视程度。

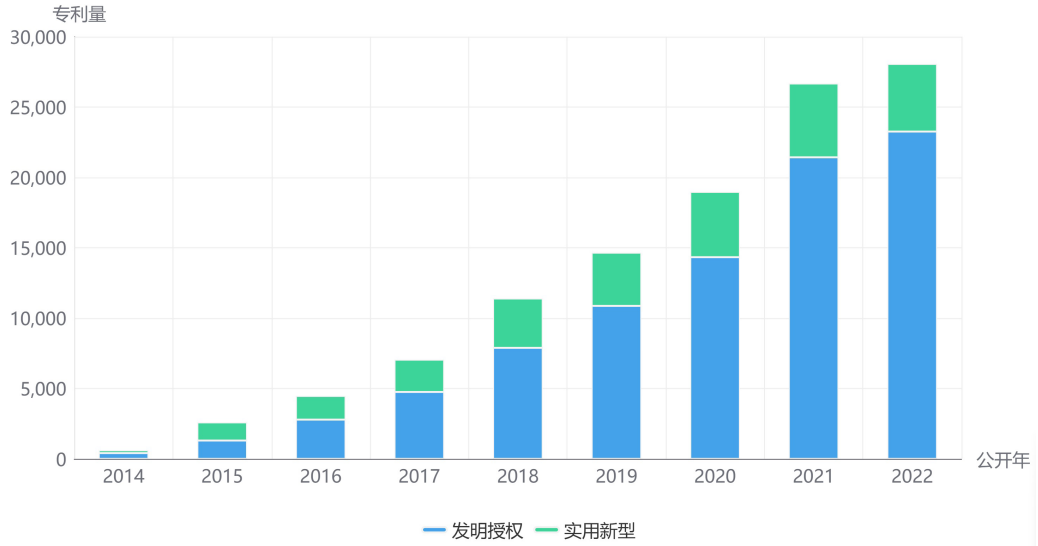

图 3　专利公开的年度趋势

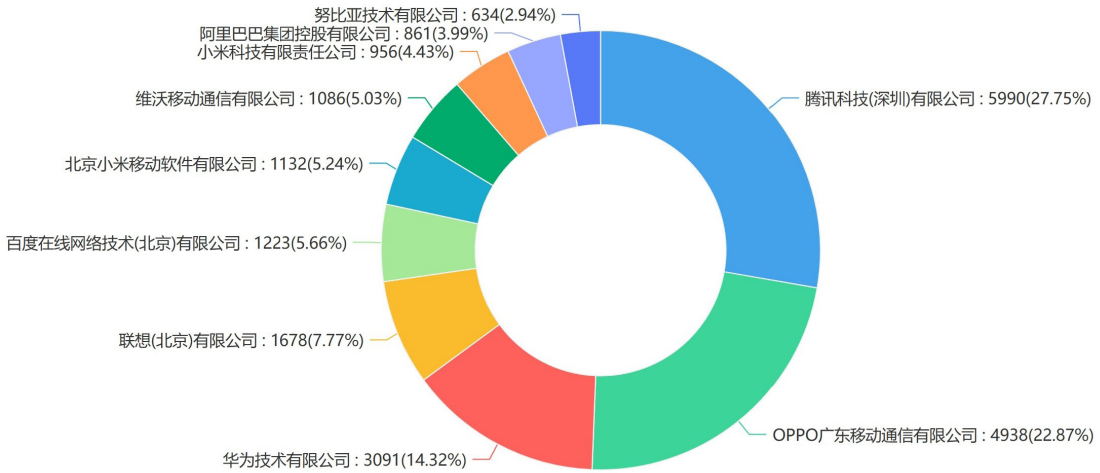

图 4　专利数量统计的申请人排名情况

专利检索数量庞大，其中包含许多干扰性专利，故需要对专利进行筛选。壹专利平台从高专利质量、高技术性、高经济性、发展前景四个维度对专利价值进行评估，专利价值度指标的数值范围为 0~100，数值越高，则专利价值越高。为提高数据的可靠性，模型需采用高质量、高价值的发明专利数据，故本文选取专利价值度大于 50 的专利（共 51057 条）进行导出分析。最终检索式为 TACD=(智能手机) AND 日期 =(2011.01.01:2023.07.01) and selectitem ((申请人类型 =(“企业”) and ipc=(“G” or “H”) and 专利状态 =(“有效”) and 主 ipc=(“G” or “H”) and 专利类型 =(“发明授权” or “实用新型”) and 价值度区间 =(“50~100”)))

3.1.2 数据预处理

本文选择中文分词组件 jieba 对智能手机专利摘要数据进行预处理。先建立停用词列表用于删除停用词，其次创建无关词列表用于去除专利摘要中的无关词，例如“本发明”“方法”等。本文在基本的中文分词基础上做改进：创建自定义保留词典用于保留智能手机新兴短语，例如“安卓”“蓝牙”等。准备工作完成后，对数据进行中文分词处理，分词结果如表 1 所示。

表 1　专利标题与摘要数据 jieba 分词结果

专利编号	分词结果									
1	平行	驱动	线路	平行	线路	传感器	平行	驱动	线路	...
2	恒温器	中央	电子	显示屏	方式	用户	温度	设置	图形	...
3	模组	模塑	感光	组件	方法	模塑	感光	组件	电路板	...
4	光固化	树脂	树脂层	图像	部件	表面	透光性	部件	制作	...
5	信号	方法	电子	设备	保护套	电子	设备 .	保护套	供电	...
6	电极	活性	材料	电极	电池	方法	电极	活性	材料	...
7	液晶显示	装置	显示装置	基板	传感	布线	基板	传感	布线	...
8	基板	聚四氟乙烯	介电	常数	填料	介电	常数	填料	介电	...
9	家用电器	设备	用户	界面	面板	区域	表面	平面	玻璃	...
...	...									

3.2 基于LDA模型的技术主题识别

3.2.1 最优主题数确定

在进行主题识别前，本文通过评估不同主题数模型的困惑度来选择最优的模型主题数。如图 5 困惑度图所示，在主题数增加后，困惑度也随之变化。根据手肘法，当 K 约等于 15 时，困惑度图像存在比较显著的拐点：当 K 小于 15 时，曲线单调递减且斜率较大；当 K 大于 15 时，曲线基本趋于平稳；所以 15 为 K 的最佳值，因此在本文所选定的语料集中，LDA 主题的生成数量选定为 15。

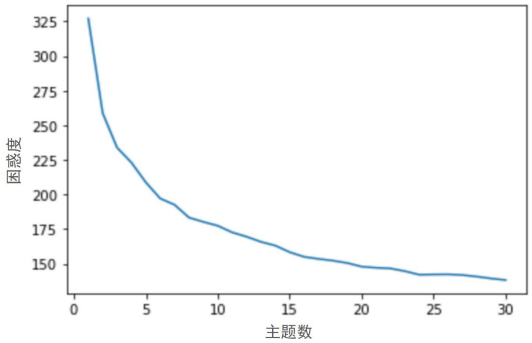

图 5　不同主题数下的困惑度

doi:10.3772/j.issn.2095-915x.2024.03.004

3.2.2 主题识别

本文使用 python 中的 sklearn 工具包进行主题识别工作。将已经分好词的专利数据导入 LDA 模型，提取出现频次排名前 1000 的特征词语，根据困惑度曲线将主题数量定为 15 个，由于数据量较大，可能会出现主题分类效果一般的高频特征词（如“智能”“组件”等），为保证识别效果，每个主题暂提取 25 个特征词，再根据主题特征进行筛选，可视化结果如图 6 所示，其中圆圈大小代表每个专利包含专利特征词数量的多少，圆圈越大则包含特征词越多，反之则越少。专利标题和摘要以及分词结果存入 EXCEL 结果文件，并输出每条专利数据对应概率最大的主题标号和每条数据对应各个主题的概率。各个主题中列取较为典型且具有代表性的特征词如表 2 所示。

热点主题可视化结果

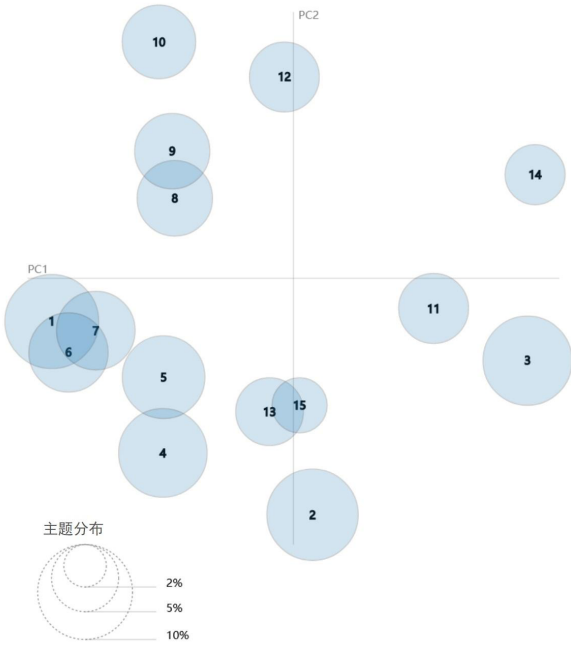

图 6 主题数为 15 时的热点主题聚类可视化结果

表 2 智能手机关键技术主题特征词

主题编号	具有代表性的特征词
主题 1	排列 亮度 分辨率 刷新 柔性 材质 视觉 密度 峰值 功耗 像素点 疏油层 光源 采样 表面
主题 2	协议 电池 电压 装置 状态 温度 电源 调节 散热 电流 供电 处理单元 电力 电量 指令
主题 3	图像 检测 目标 人脸 指纹 摄像头 场景 亮度 画面 短焦 特征 调整 对焦 投影 调节
主题 4	透镜 光学 测试 镜头 表面 折力 光轴 物侧 凸面 材质 传感器 焦度 焦距 镜片 非球面
主题 5	解调 电路 频率 解码 基带 射频 兼容 电波 数字 制式 网络 频段 传输 逻辑 节点
主题 6	频率 处理器 制程 核心 运算 指令 单元 架构 精度 能力 渲染 性能 集成 图形 控制
主题 7	像素 动态 光圈 场景 曝光 环境 画质 防抖 范围 面积 景深 计算 分辨率 插值 画面
主题 8	网络 无线 传输 网页 基站 信道 数据包 系统 链路 对象 地址 接入点 分配 协议 无线网络
主题 9	用户 车辆 屏幕 检测 运动 轨迹 距离 触摸屏 按键 车机 传感器 方向 环境 报警 车载
主题 10	存取 融合 程序 存储器 介质 分区 数据 半导体 系统 电容 组件 读取 稳态 刷新 地址
主题 11	参数 音频 页面 元素 耳机 装置 布局 扬声器 混合 效果 单位 音箱 系数 录音 音频文件
主题 12	终端 用户 指令 账号 密码 通话 体验 联系人 时间 号码 功能 系统 次数 短信 方式
主题 13	互联 管理 家居 服务 语音 用户 控制 指令 监控 感应 平台 智慧 社区 电动 物联网
主题 14	应用 文件 界面 图片 用户 指令 列表 目标 图标 功能 程序 代码 系统 软件 内存
主题 15	材料 体验 质量 玻璃 装置 方向 实体 外观 曲率 属性 按键 材质 盖板 测量 塑料

在 LDA 模型的主题识别结果中，将每篇专利所对应的公开日以时间顺序排列。通过计算各个主题近两年每个月的专利公开数量，可以评估各个主题专利研究的热度。ARIMA 模型结

果的准确性高度依赖于数据的可靠性和差异性，由于专利的申请和授权需要一定的时间，且专利研究的热度在近些年变化幅度明显，故本文选取近两年（2021 年 1 月至检索时间 2023 年 7 月）的专利研究的热度作为原始数据进行模型的拟合和预测，以月作为记录周期，刻画专利研究的热度随时间的变化趋势。

3.3 基于ARIMA模型的技术趋势预测

本文使用 python 中的 statsmodels 包，依次将各个技术主题的研究热度数据导入到 ARIMA（p,d,q）模型。

以主题 1 为例，首先判断数据是否具有平稳性以及是否为白噪声：原数据有单调递增趋势，并不具有平稳性，于是对数据进行差分处理，一阶差分的结果相较于原数据更稳定，使用 python 计算一阶差分的 ADF 检验结果 p 为 $0.0225 < 0.05$（拒绝原假设），可知一阶差分序列不是非平稳序列，具有平稳性。一阶差分序列的白噪声检验结果 p 值为 $0.0007 < 0.05$（拒绝原假设），可知一阶差分序列不是纯随机序列，是白噪声。故将模型参数中 d 的值定为 1。

然后确定 ARIMA 模型中参数 p、q 的值，本文利用自相关系数 ACF 和偏自相关系数 PACF 分别确定 p、q 的值，由图 7 可知自相关系数 ACF 在 1 阶后截尾，故将 p 值定为 1。偏自相关系数 PACF 拖尾，故将 q 的值定为 0。

接着检验 ARIMA 模型的残差是否是平均值为 0 且方差为常数的正态分布。利用 QQ 图检验，由图 7 可见结果为线性即为正态分布，故残差符合正态分布检验。使用 Durbin-Watson test（D-W）检验残差的自相关性，DW 值为 1.97，接近于 2 时，则不存在（一阶）自相关性。使用 Ljung-Box test（L-B）检验残差是否为白噪声，LB 值为 0.0039，则残差是白噪声。故残差符合检验，数据适用于 ARIMA 模型。最后对未来 6 个月的专利研究热度进行预测。

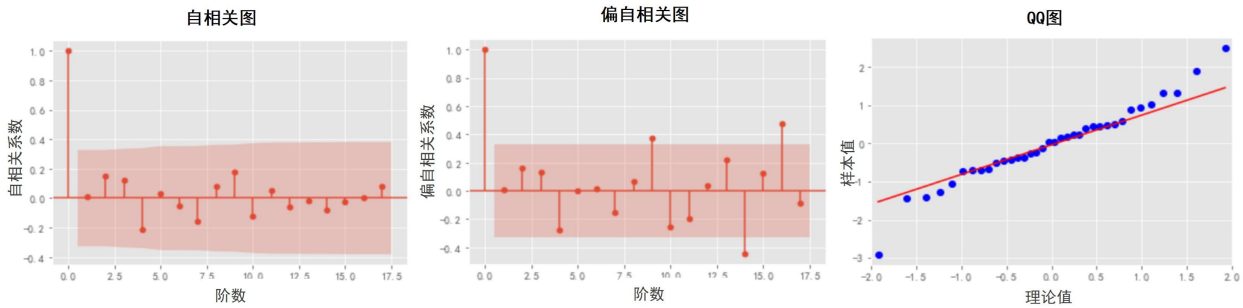

图 7　参数确定与检验结果

3.4 关键技术主题解读

根据表 2 可知主题 1 中的特征词“亮度”“视觉”等关注智能手机的屏幕显示；“分辨率”“像素点”“排列”对应智能手机屏幕相关参数，可以得知主题 1 主要对应的关键技术主题为“屏幕”。

主题 2 中的特征词“电压”“电流”“协议”等对应智能手机充电方面技术特征词；“散热”“功耗”“续航”等关注智能手机电池续航，可以得知主题 2 主要对应的关键技术主题为“电

doi:10.3772/j.issn.2095-915x.2024.03.004

池与充电”。

主题 3 中的特征词“指纹”“短焦”“对焦”则对应智能手机的屏下指纹解锁功能；“人脸”“摄像头”“特征”等与智能手机的前置摄像头相关功能有关，综合分析判断其对应的并非智能手机摄像技术，而是人脸识别解锁技术。可以得知主题 3 主要对应的关键技术主题为“生物识别系统”。

主题 4 中的特征词“透镜”“光学”“镜头”“传感器”“焦距”“镜片”等可以很明显发现该主题与摄影摄像相关，但由于在主题 3 的判断中也曾出现摄影摄像相关技术主题词，而摄影摄像覆盖的技术面较为广泛（包括拍照算法，摄像头传感器硬件等），故将主题 4 暂定为摄影摄像相关技术主题。

主题 5 中的特征词“解调”“基带”“解码”等对应智能手机基带技术特征词；“电波”“网络”“频段”“射频”等对应智能手机射频相关技术。两方面都为智能手机主板中通信功能的关键组成部分，综合考虑可以得知主题 5 主要对应的关键技术主题为“通信”。

主题 6 中的特征词“制程”“核心”“运算”等对应智能手机 CPU；“渲染”“集成”“图形”“性能”等对应智能手机 GPU。可以得知主题 6 主要对应的关键技术主题为“CPU 和 GPU”。

主题 7 中的特征词“光圈”“曝光”“环境”“画质”“防抖”“景深”“计算”等同样与摄影摄像相关。与主题 4 不同的是，主题 7 的特征词更多聚焦于摄影摄像的软件算法；主题 4 的特征词更多关注摄影摄像的硬件参数。故得知主题 7 主要对应的关键技术主题为“摄影摄像算法”；主题 4 主要对应的关键技术主题为“摄影摄像硬件”。

主题 8 中的特征词“传输”“基站”“信道”“链路”“地址”等与主题 5 类似，但频繁出现“网络”“无线”“无线网络”等特征词。可以得知主题 8 主要对应的关键技术主题为“网络”。

主题 9 中的特征词“车辆”“车机”“车载”与汽车关联紧密；同时根据特征词“传感器”“检测”“方向”“环境”等可以得知主题 9 主要对应的关键技术主题为“手机与车机互联”。

主题 10 中的特征词“存取”“融合”“存储器”“电容”“介质”“半导体”等多次提到存储，智能手机中的存储器与计算机类似，分为随机存储器（RAM）和只读存储器（ROM），故主题 10 主要对应的关键技术主题为“RAM 和 ROM”。

主题 11 中的特征词“参数”“扬声器”“耳机”“音箱”“录音”“音频文件”等与智能手机各项音频相关，故主题 11 主要对应的关键技术主题为“音频”。

主题 12 中的特征词“用户”“指令”“体验”等对应智能手机用户操作；“账号”“密码”“系统”等对应智能手机账户功能，可以得知主题 12 主要对应的关键技术主题为“用户与账户”。

主题 13 中的特征词“用户”“控制”“管理”等同样对应智能手机用户操作；而“家居”“服务”“互联”等对应的是家居服务，综合推断可以得知主题 13 主要对应的关键技术主题为“智能家居”。

主题 14 中的特征词“应用程序”“界面”“图片”“图标”等对应智能手机 UI 样式；“程序”“代码”“软件”“系统”等对应智能手机操作系统软件代码，可以得知主题 14 主要对应的关键

技术主题为“操作系统”。

主题 15 中的特征词“材料”“玻璃”“实体”“外观”“曲率”“按键”“材质”“盖板”“塑料”等对应智能手机材质属性以及外观，可以得知主题 15 主要对应的关键技术主题为“外观设计”。

各个主题编号与对应关键技术主题名称如表 3 所示。

3.5 关键技术主题演化分析

由 ARIMA 模型可以得到未来半年内的专利研究热度预测结果如图 8 所示，本文对包含特征词数量前三的主题（“屏幕”“电池与充电”“生物识别系统”）进行演化分析。

表 3 主题编号对应关键技术主题名称

主题编号	关键技术主题名称	主题编号	关键技术主题名称
主题 1	屏幕	主题 9	手机与车机互联
主题 2	电池与充电	主题 10	RAM 和 ROM
主题 3	生物识别系统	主题 11	音频
主题 4	摄影摄像硬件	主题 12	用户与账户
主题 5	通信	主题 13	智能家居
主题 6	CPU 和 GPU	主题 14	操作系统
主题 7	摄影摄像算法	主题 15	外观设计
主题 8	网络		

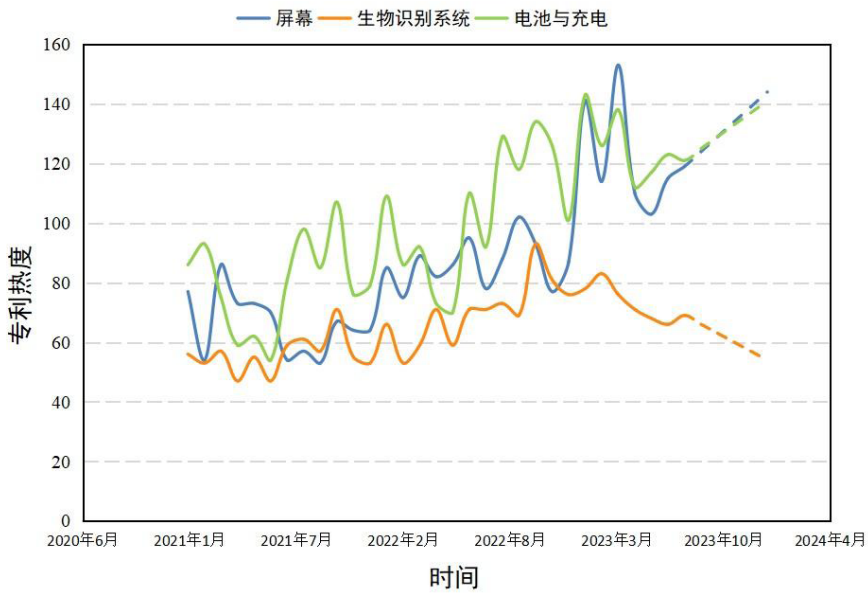

图 8 智能手机专利研究热度预测

3.5.1 智能手机屏幕技术

根据主题“屏幕”专利研究热度进行的预测结果，可以发现智能手机屏幕相关专利在 2021 年至 2022 年上半年间保持平稳，在 2022 年下半年持续增长，从 2023 年开始热度陡然上升，专利数量几乎较 2021 年初翻倍，年底下滑后再次回暖。ARIMA 预测其趋势为逐步上升，虽然上升幅度不如 2023 年大，但是仍保持较高的增长率，分析其原因如下：

智能手机屏幕材质以 OLED（Orangic Light Emitting Diode）[24] 居多，中文名称为有机发光二极管。以往的各家手机厂商的高端旗舰机屏幕供应商基本均为三星，三星屏幕凭借其峰值亮度高、分辨率高、画面清晰、色彩艳丽饱和

doi:10.3772/j.issn.2095-915x.2024.03.004

等优点获得了国内手机厂商的青睐。

国产智能手机屏幕供应商例如京东方、华星光电、天马、维信诺等近年来进步巨大，分别研发了蓝钻、珍珠、风车等屏幕类钻石排列，达到了较高的等效 ppi，在清晰度上与三星 Pentile 排列屏幕接近。在材料方面，我国智能手机屏幕辅助材料国内市场供应占 12%，发光层材料占 5%，关键材料依赖进口。今年来，柯达、UDC 等海外屏幕材料供应商的专利陆续到期，国内对于磷光金属配合物和关键配体等环节正值专利突破的时机。在屏幕制造设备方面，关键设备蒸镀机被日本 Tokki 垄断，高端蒸镀机几乎被三星买断，近年来我国合肥欣奕华首台硅基 OLED 蒸镀机导入客户量生产线，攻克了蒸镀机技术，我国蒸镀工艺蓬勃发展。

综上所述，国产屏幕在材料与设备面临专利发展的机遇，应加大研发攻克难题；在屏幕素质方面接近三星高端屏幕，其他方面例如峰值亮度、功耗等与三星屏幕存在差距，有较大进步空间，另外国内供应商研发了高频 PWM 技术，在护眼程度方面实现了超越，前景光明。

3.5.2 智能手机生物识别系统技术

根据主题“生物识别系统”专利研究热度进行的预测结果，可以发现智能手机生物识别系统相关专利研究热度自 2021 年起保持平稳，并且有缓慢增长的趋势，直至 2022 年底以较快速度提升，从 2023 年开始下滑。ARIMA 预测其趋势为逐步下降，下降速度与 2023 年初保持一致，分析其原因如下：

生物识别技术 [25] 是指通过计算机与光学、声学、生物传感器和生物统计学原理等高科技手段结合，利用人体固有的生理特征（指纹、人脸、虹膜等）和行为特征（笔迹、声音等）进行个人身份鉴定，其中指纹与人脸识别技术常用在智能手机领域。早在 2013 年，苹果公司就在 Iphone5s 上配备了实体指纹识别 TouchID 技术，并且在 2017 全面取消指纹识别，改用更为先进的 FaceID 人脸识别。

近年来部分国内厂商在指纹识别上实现弯道超车：vivo 研发的广域超声波指纹识别技术能实现在湿手或有污渍的情况下，解锁场景十分广泛：利用超声波穿透屏幕面板扫描捕捉手指纹理，由手指不同纹理材质返回超声波的不同时间来计算距离，从而建立对应指纹的 3D 图像来实现指纹识别。在人脸识别方面，华为于 2019 年开始自研 3D 人脸硬件、结构光模组、人脸识别算法，在需要解锁时，红外相机、散斑投射器和红外补光灯捕捉人脸数据，与机主人脸信息对比进行解锁，并且应用于此后发布的旗舰机中。

综上所述，国内手机厂商仅有 vivo 与华为分别在人脸与指纹识别技术上实现了突破，但其他厂商却与其有较大差距。生物识别系统的好坏在消费者的日常感知中十分明显，应该得到各企业的充分重视。ARIMA 模型预测其专利研究热度有降低的趋势，其他企业应该警惕，需要紧跟技术发展趋势，加大研发力度，努力引进 vivo、华为等先行者的相关技术，并进行创新发展。

3.5.3 智能手机电池与充电技术

根据主题“电池与充电”专利研究热度进行的预测结果，可以发现智能手机电池与充电技术相关专利研究热度在 2021 至今起伏较大，但总体趋势是持续上升的，其中 2022 年间增长

率较高，到 2023 年后增长变缓。ARIMA 预测其趋势为逐步上升，仍然维持较高增长率，分析其原因如下：

国内智能手机的充电和电池技术相较于国际大厂（例如三星、苹果）已是遥遥领先。三星在 2016 年发布的智能手机 Samsung Galaxy Note 7 就曾出现过严重的安全隐患，该机型发布一个多月，就已在全球范围内出现了近四十起因电池爆炸而发生的安全事故；苹果手机则因为其低功率充电、电池容量小而被贴上了“充电慢、续航差”的标签。

近年来国内厂商将充电与电池技术作为研发重点，并取得了成效：充电方面，Realme 开发 240W 充电架构，通过定制的三路电荷泵，将高功率充电所带来的高热量进行分散，并引进了 4 ：2 大功率充电架构，将大电流分解成三路小电流进行充电，降低了充电通路的阻抗，提升了整体的转化效率，同时也减少了热量的堆积；Redmi 全链路自研了 300W 快充架构，定制了 6 ：2 电荷泵芯片，芯片最高转换效率高达 98%，多颗电荷泵并联后，直接对电池充电。电池方面，Redmi 研发 15C 高倍率电池，使用三明治双电芯堆叠，引入了新型硬碳材料，相比以前的石墨要更加疏松无序，提升充电速度的同时还能兼顾高能量密度；荣耀研发了硅碳负极电池技术，在相同电池体积的情况下提供了更大的电池容量，实现了旗舰手机电池容量的突破。

综上所述，高功率充电技术方面国内的研发已趋于饱和，相关企业可以从温度控制、安全性方面着手突破；在电池方面，锂电池几乎达到化学极限，能量密度提升比较困难，各企业可以朝着新的电池形态（例如固态电池）寻找突破口。未来充电与电池技术将成为国内智能手机的卖点，各厂商应努力维护目前国内充电与电池技术的领先地位。

4 结论

本文通过实证研究确定智能手机关键技术主题分别为屏幕、电池与充电、生物识别系统、摄影摄像硬件、通信、CPU 和 GPU、摄影摄像算法、网络、手机与车机互联、RAM 和 ROM、音频、用户与账户、智能家居、操作系统、外观设计，并根据演化趋势分析提出建议。有助于企业把握技术发展趋势，及时调整战略。

然而，本文仅使用了国内智能手机专利摘要数据进行分析，对于其余重要数据例如论文、舆情数据等并未进行分析；另外对于专利研究热度仅使用各月份的专利公开量作为参考也不够严谨。在未来的研究中将扩大数据源，对国内专利、论文、舆情数据综合分析，并结合国外各项数据分析结果进行对比参考，构建专利研究热度指标进行计算，提高 ARIMA 模型预测的准确性，对智能手机行业起到更好的导向和促进发展的作用。

参 考 文 献

[1] 滕飞，张奇，曲建升，等. 基于专利竞争力指数和 Doc-LDA 主题模型的关键核心技术识别研究 - 以新能源汽车为例 [J/OL]. 数据分析与知识发现，1-19[2024-04-26]. http://kns.cnki.net/kcms/detail/10.1478.g2.20231118.1750.002.html.

[2] YU C C, CAO L, YIN W B, et al. Automatic word segmentation on Lizu spoken annotation corpus[J]. Application Research of Computers, 2017, 34(5): 1325-1328.

doi:10.3772/j.issn.2095-915x.2024.03.004

[3] 陈稳，陈伟．科学与技术对比视角下的前沿主题识别与演化分析 [J]. 情报杂志，2022, 41(1): 67-73, 163.

[4] 吕鲲，陈箫羽，靖继鹏．基于组合分词方法和 LDA 模型的区块链金融产业关键技术识别研究 [J]. 图书情报工作，2022, 66(19): 110-121.

[5] LIU Z Y, WANG Y, WANG W P, et al. Bayesian Parameter Estimation in LDA:[C]//International Conference on Computer Information Systems and Industrial Applications. Bangkok, Thailand, 2015: 837-840.

[6] 郑荣，魏明珠，高志豪，等．基于 SCAN-CPM 的产业新兴技术识别与演化路径分析：以新能源汽车产业为例 [J]. 图书情报工作，2022, 66(11): 100-109.

[7] 伊惠芳，刘细文，龙艺璇．基于多文本分析的领域关键技术问题挖掘方法研究 [J]. 情报理论与实践，2023, 46(1): 187-196.

[8] XU T T, HUANG H J. An Improved Supervised-LDA Text Model and Its Application[J]. Computer Engineering, 2018, 44(1): 69-73, 78.

[9] 王秀红，高敏．基于 BERT-LDA 的关键技术识别方法及其实证研究——以农业机器人为例 [J]. 图书情报工作，2021, 65(22): 114-125.

[10] DE ARAÚJO MORAIS L R, DA SILVA GOMES G S. Forecasting daily Covid-19 cases in the world with a hybrid ARIMA and neural network model[J]. Applied Soft Computing, 2022, 126: 109315.

[11] 马铭，王超，周勇，等．基于语义信息的核心技术主题识别与演化趋势分析方法研究 [J]. 情报理论与实践，2021, 44(9): 106-113.

[12] BIAN Z L, TANG J Q, NI C H, et al. Analysis on prevalence of pneumoconiosis in Jiangsu Province using ARIMA-GRNN combined model[J]. Chinese Journal of Environmental & Occupational Medicine, 2019, 36(8): 755-760.

[13] LV L C, ZHOU J, WANG X Z, et al. Technology Evolution Analysis Framework Based on Two-Layer Topic Model and Application[J]. Data Analysis and Knowledge Discovery, 2022, 6(2/3): 18-32.

[14] 岳丽欣，周晓英，陈旖旎．基于 ARIMA 模型的信息构建研究主题趋势预测研究 [J]. 图书情报知识，2019(5): 54-63, 72.

[15] HU R, MA W, LIN W, et al. Technology Topic Identification and Trend Prediction of New Energy Vehicle Using LDA Modeling[J]. COMPLEXITY, 2022, 2022: 1-20.

[16] 李勇，欧志梅，黄格，等．基于专利挖掘的数字文旅关键技术识别和趋势分析 [J]. 图书馆论坛，2023, 43(1): 121-130.

[17] SUN B, XU X F, SU X. Identification of the Main Path of Technology Diffusion and Core Enterprises: A Case Study of Patent Citation Network in Mobile Phone Chips[J]. Journal of the China Society for Scientific and Technical Information, 2019, 38(2): 201-208.

[18] WANG X W, SONG Y Y. Artificial Intelligence Scientific Management: Artificial Intelligence Patents Layout in the Smartphone Field[J]. Science & Technology for Development, 2018, 14(3): 142-148.

[19] 苏敬勤，刘建华，王智琦，等．颠覆性技术的演化轨迹及早期识别——以智能手机等技术为例 [J]. 科研管理，2016, 37(3): 13-20.

[20] 闯家梁．基于专利分析的中国智能手机企业技术创新能力研究 [J]. 中国集体经济，2020(14): 153-154.

[21] 彭英，闯家梁，杨映雪．专利视阈下石墨烯智能手机技术创新研究 [J]. 电子商务，2019(10): 42-43, 65.

[22] HASAN M, RAHMAN A, KARIM MD R, et al. Normalized approach to find optimal number of topics in latent dirichlet allocation (lda)[C]//2nd International Conference on Trends in Computational and Cognitive. Savar, Bangladesh, 2021: 341-354.

[23] HUANG L, MA J, CHEN C, et al. Topic Detection from Microblogs Using T-LDA and Perplexity[C]//24th Asia-Pacific Software Engineering Conference. Nanjing, PEOPLES R CHINA. 2017: 71-77.

[24] LIU D X, ZHONG J Y, TANG B, et al. Research progress of flexible and printed OLED[J]. Chinese Journal of Liquid Crystals and Displays, 2021, 36(2): 217-228.

[25] AMIRTHALINGAM G, GOVINDARAJU R. Comparison on Multi-modal Biometric Recognition Method[C]//1st International Conference on Computational Intelligence and Information Technology. Pune, INDIA, 2011: 524-526.