

基于LDA主题模型对AIGC的影响力分析

翟其玲, 张佳怡, 刘宝瑞, 黄龙彬, 陈立伟, 金秀玲*

闽江学院数学与数据科学学院, 福建 福州

收稿日期: 2023年9月9日; 录用日期: 2023年10月9日; 发布日期: 2023年10月17日

摘要

ChatGPT的出现, 引领生成式AI迅速渗透进人们生产生活的方方面面, 故了解人工智能内容生成的发展趋势和未来方向非常重要。基于抓取微博平台收录的关于人工智能内容生成的文本内容, 先利用jieba分词进行文本预处理, 通过LDA主题模型的方法得到七个人工智能内容生成领域的热点主题, 包括: 人工智能概念股、智能数字内容创作、数字经济下的科技股投资趋势、自然语言技术的突破等。研究揭示了当前AIGC的舆情热点, 有助于相关人员掌握生成式AI领域的发展状态, 对人工智能未来的应用提供参考。

关键词

ChatGPT, LDA模型, 人工智能内容生成, 热点分析

Influence Analysis of AIGC Based on LDA Topic Model

Qiling Zhai, Jiayi Zhang, Baorui Liu, Longbin Huang, Liwei Chen, Xiuling Jin*

College of Mathematics and Data Science, Minjiang University, Fuzhou Fujian

Received: Sep. 9th, 2023; accepted: Oct. 9th, 2023; published: Oct. 17th, 2023

Abstract

The emergence of ChatGPT has led the rapid penetration of generative AI into all aspects of people's production and life, so it is very important to understand the development trend and future direction of artificial intelligence content generation. Based on grabbing the text content of artificial intelligence content generation collected on Weibo platform, jieba segmentation is first used for text pre-processing, and seven hot topics in the field of artificial intelligence content generation are ob-

*通讯作者。

tained by LDA topic model, including: Artificial intelligence concept stocks, intelligent digital content creation, technology stock investment trends under the digital economy, breakthroughs in natural language technology, etc. The research reveals the current public opinion hot spots of AIGC, which is helpful for relevant personnel to grasp the development status of the field of generative AI and provide reference for the future application of artificial intelligence.

Keywords

ChatGPT, LDA Model, Artificial Intelligence Generation Content, Hot Spot Analysis

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

2022年11月,人工智能公司 OpenAI 推出了生成式人工智能,并命名为 ChatGPT。发布后短短两个月吸引了超1亿用户。ChatGPT 强大的内容生成能力,引起了全世界的广泛关注, AIGC (是指利用人工智能技术生成内容)概念由此走上风口并渗透进各行各业[1]。在文学领域, AIGC 可以生成文章、诗歌和故事,扩展了创作者的想象力和创作能力[2] [3]。在设计领域, AIGC 可以辅助自动生成艺术作品、建筑设计和虚拟场景等,为创意行业带来新的可能性。在娱乐领域, AIGC 可以生成虚拟角色、游戏关卡和剧情,提供个性化和互动性的游戏体验[4]。

然而,随着 AIGC 技术的快速发展,也带来了一些热点问题。了解和分析这些热点问题,对于深入理解 AIGC 技术的发展方向、潜在挑战和应用前景至关重要。本文通过翻阅和研究大量参考文献,旨在基于 LDA (潜在狄利克雷分配)主题模型对 AIGC 技术发展的热点进行分析。利用爬虫技术从微博上爬取文本并利用 LDA 主题模型从中确定研究主题,并深入探讨 AIGC 技术发展中的热点问题和前沿动向,这将有助于揭示 AIGC 技术的现状、趋势以及未来发展的方向,为相关人员和从业者提供有益的指导和启示。

2. 研究思路和方法

2.1. 研究思路

本文将获取到 AIGC 相关的微博文本内容通过文本预处理切割得到原始语料,通过困惑度和一致性大小指标确定最优主题数,然后进行 LDA 主题模型建模,从热点话题中确定研究主题,识别 AIGC 技术发展的热点问题,并进行深入探讨。

本文研究框架如图 1 所示。具体流程为:

- 1) 收集关键词为“AIGC”的文本数据;
- 2) 文本预处理;
- 3) 特征提取
- 4) 通过困惑度和一致性来确定最优主题数目;
- 5) 通过模型实验得到主题 - 词分布;
- 6) 确定主题类别。

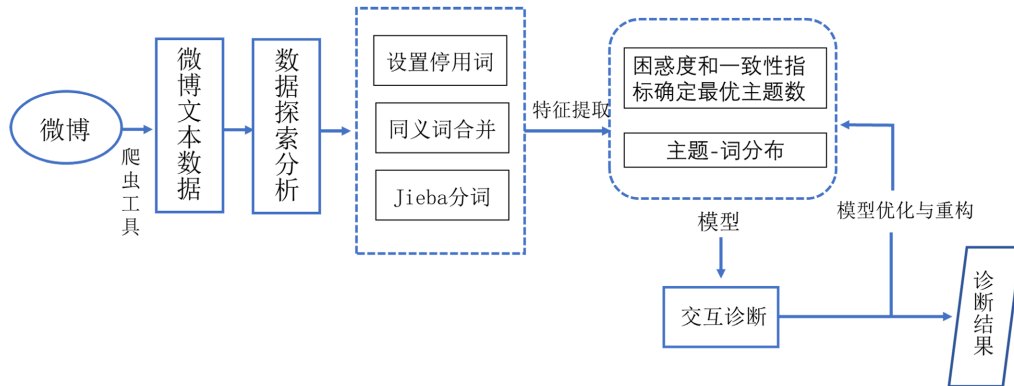


Figure 1. Analysis process
图 1. 分析流程

2.2. 研究方法

2.2.1. LD 主题模型

本文运用 LDA 主题模型识别 AIGC 技术发展的热点问题。LDA (Latent Dirichlet Allocation) 是一种主题模型，用于无监督学习和文本分析。它将文档集表示为词的集合，假设每个文档由多个主题构成，其中每个词都由一个主题生成。LDA 可以通过概率分布给出每篇文档的主题，从而实现对文章的主题归纳。与线性判别分析(Linear Discriminant Analysis)不同，LDA 主要应用于模式识别领域，如人脸识别和图像识别。LDA 的训练过程不需要手工标注的训练集，只需要文档集和指定的主题数量。另一个优点是 LDA 可以为每个主题选择描述性词语。选择主题数量是手动设置的参数，然后每篇文章都会被分配一个主题的概率，并给出每个主题下词语的概率。

LDA 假设文档的生成过程如下，流程图如图 2 所示：

- 1) 对每个主题 $k \in [1, K]$ ，生成“主题 - 词项”分布 $\varphi_k \sim Dir(\beta)$ ；
- 2) 生成文档 m 的“文档 - 主题”分布 $\theta_m \sim Dir(\alpha)$ ；
- 3) 生成文档 m 的长度 $N_m \sim Poiss(\xi)$ ；
- 4) 对文档 m 中的每个词 $n \in [1, N_m]$ ，生成当前位置的所属主题 $Z_{m,n} \sim Cat(\theta_m)$ ；
- 5) 根据之前生成的主题分布，生成当前位置的词的相应词项 $w_{m,n} \sim Cat(\varphi_{z_{m,n}})$ 。

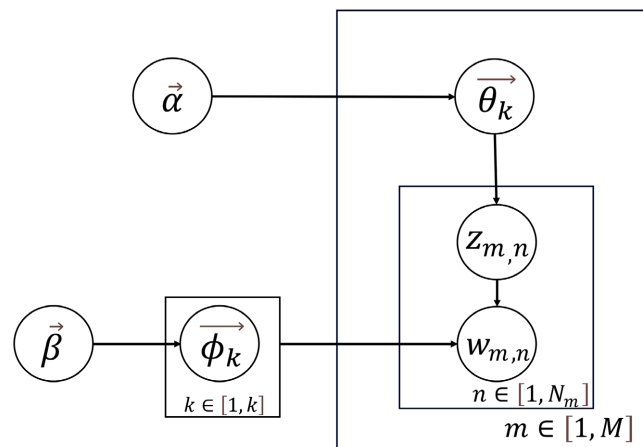


Figure 2. LDA assumes document generation
图 2. LDA 假设文档的生成

2.2.2. 文本特征提取

计算机本身不能读懂文字含义，在文本挖掘过程中需通过计算文字权重进而转换为可以量化的特征词方便进一步数据挖掘。故文本特征提取是 LDA 建模过程中必不可少的环节。LDA 主题模型构建有两种方式，分别是基于 TF-IDF 的构建方法和基于词袋(Corpus)的构建方法。词袋构建法是将每一篇文档视为一个词频向量，从而将文本信息转化为了易于建模的数字信息。TF-IDF 技术是一种信息检索与数据挖掘的加权技术，TF-IDF 中 TF 代表词条在文本中出现的频率[5]。其中， $n_{i,j}$ 表示词条 t_i 在文档 d_j 中出现的次数， $TF_{i,j}$ 就是表示词条在文档 d_j 中出现的频率。

$$TF_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}}$$

IDF 代表关键词的普遍程度，它是将特征词在文档中出现的次数除以包含该特征词的文档数作为该词的权重，常用来做文本分类。

$$IDF_i = \log \frac{|D|}{1 + |j:t_i \in d_j|}$$

其中 $|D|$ 表示所有文档的数量， $|j:t_i \in d_j|$ 表示包含词条 t_i 的文档数量，通过 TF-IDF 将分词好的数据向量化为后面 LDA 模型的建立提供处理好的数据集。

由于 TF-IDF 采用权重稀疏矩阵方式建模，因此在同等数据量的情况下，其建模速度较优于 Corpus。本文研究的文本数据量较大，故采用 TF-IDF 算法将分词后的 AIGC 微博博文文本向量化以便于 LDA 主题模型建模。

2.2.3. 确定主题数目方法

确定主题数目的方法是计算困惑度[6]和一致性[7]，困惑度处于较低的水平，则模型的主题结构就趋于稳定，其预期误差值也会相对较小。但是当主题个数越多，模型的困惑度就越低，模型往往会过拟合。主题一致性也是衡量主题质量最有效的方法，也是估计主题数目的重要技术之一，当主题一致性越大时，每一个聚类主题下特征词的连贯性越好[8]。

3. 研究设计与处理过程

3.1. 数据集介绍

本文使用八爪鱼采集器对微博进行关键词搜索，爬取了 2022 年 9 月 1 日至 2023 年 3 月 31 日微博博文中包含“AIGC”关键词的文本数据，数据集中包含用户 ID、发布时间和微博文本内容。本文最终研究的数据集为 17,410 条微博文本。

3.2. 文本数据预处理

将爬取得到的关于 AIGC 微博博文进行数据预处理。

首先设置停用词表。采集的原始文本数据中通常包括数字、标点符号和表情符号等与主题内容无关的字符，因此，本文在“哈工大停用词表”的基础上，结合词频统计，将出现频次高但对主题内容无关的特殊符号或者词语，例如“网页链接”，“收起全文”，“主要”等删去，同时将数词、介词、代词等词也纳入停用词表，最终构建的停用词表共 873 个词。

其次，进行同义词合并。原始数据中存在许多意义相同的一组词语，本文将这类词语进行合并，例如“AIGC 概念”和“AIGC”语义相似，将其合并为“AIGC”。

最后使用 jieba 分词对原始数据进行切分。在分词前为提高分词的准确性，添加自定义词如“独特价

值”、“人工智能生成内容”、“文心一言”等共计 381 个。最后进行 jieba 分词得到结果如下表 1 所示：

Table 1. Situation before and after using jieba segmentation (part)
表 1. 使用 jieba 分词前后的情况(部分)

分词前	分词后
<p>AIGC 的独特价值#李彦宏称过去 10 年百度研发投入超千亿##李彦宏判断 L2 之后率先商用的是 L4#L 青花水墨时的微博视频#李彦宏称过去 10 年百度研发投入超千亿##李彦宏判断 L2 之后率先商用的是 L4#AIGC 背后的技术</p> <p>#5ipr 资讯#【李彦宏：“AI 考生”40 秒写 40 篇高考作文得分超 75%】9 月 1 日消息，今日，主题为“智联世界元生无界”的 2022 年世界人工智能大会在上海开幕。在此次大会上，百度创始人李彦宏表示，百度的 AI 数字人度晓晓今年挑战写高考作文，40 秒写了 40 篇，得分可以排在总考生前 25%。这就是所谓的 AIGC——人工智能自动生产内容。O 李彦宏：“AI 考生”40 秒写 40 篇高考作文得分超 75%收起全文 d</p> <p>【李彦宏：“AI 考生”40 秒写 40 篇高考作文得分超 75%】今日，主题为“智联世界元生无界”的 2022 年世界人工智能大会在上海开幕。在此次大会上，百度创始人李彦宏表示，百度的 AI 数字人度晓晓今年挑战写高考作文，40 秒写了 40 篇，得分可以排在总考生前 25%。这就是所谓的 AIGC——人工智能自动生产内容。L 究竟视频的微博视频收起全文 d</p>	<p>AIGC、独特价值、过去、百度、研发、投入、千亿、判断、之后、率先、商用、青花、水墨过去、百度、研发、投入、千亿、判断、之后、率先、商用、AIGC、背后、技术</p> <p>ipr、资讯、AI、考生、写篇、高考作文、得分、消息、主题、智联、世界、元生、无界、世界人工智能大会、上海、开幕、此次、大会、百度、创始人、百度、AI 数字、人度晓晓、挑战、高考作文、得分、排在、考生、所谓、AIGC、人工智能、自动生产、AI、考生、写篇、高考作文、得分</p>

对分词后的文本制作词云图如下图 3 所示：



Figure 3. Word cloud map after word segmentation
图 3. 分词后的词云图

从分词后的文本的词云图中可以明显看出，“AIGC”、“ChatGPT”、“AI”、“科技”、“公司”、“板块”、“技术”等关键词所占的比例较大。

3.3. LDA 主题模型建模

在使用 TF-IDF 算法进行特征提取后，需要通过困惑度和一致性确定最优主题数目，计算得到不同主题数目下的困惑度和一致性的变化情况如图 4 和图 5 所示。

结果显示, 困惑都随着主题数目的增加而下降, 当主题数目超过 9 个时, 困惑度的数值下降幅度较大, 模型复杂度变高, 出现过拟合, 因此应在 1~8 的主题个数中确定最终的主题个数。再结合一致性变化情况, 由于本文研究文本数量较多, 当主题数选取 2~4 个时对文本主题划分不够细致, 最终确定主题数目为 7。

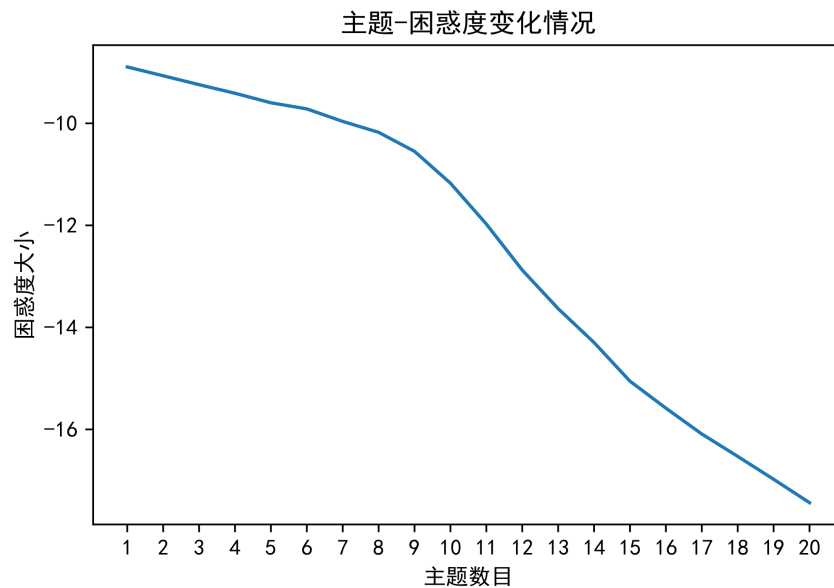


Figure 4. Topic-confusion degree change

图 4. 主题 - 困惑度变化情况

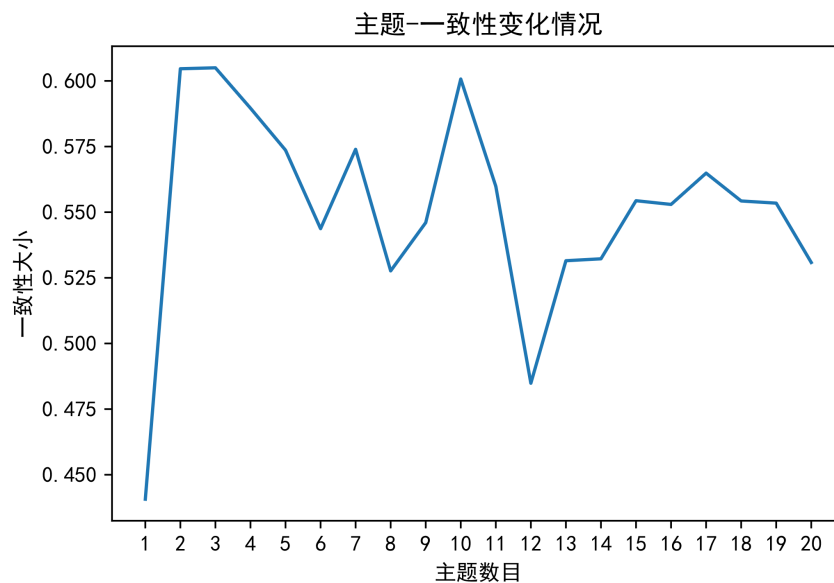


Figure 5. Topic-consistency change

图 5. 主题 - 一致性变化情况

设置 LDA 模型参数 $\alpha = 5$, $\eta = 0.01$, 将文本迭代 100 次, 抽取每个主题下概率最大的 20 个词汇, 各词汇按频率从大到小排序, 作为描述主题含义的代表, 结合这些词汇对该主题进行标识[9]。训练结果如表 2 所示:

Table 2. Seven subject categories obtained by training
表 2. 训练得到的 7 个主题类别

主题编号	主题类别	主题关键词
Topic0	人工智能概念股	创业板、深成指、成交额、半导体、上证指数、产业链、交易日、元宇宙、概念股、毫米波、一体化、跌多涨少、计算机、净流入、从业者、中国电信、特斯拉、工程机械、涨停板、开门红
Topic1	智能数字内容创作	AIGC、人工智能、生成式、创作者、UGC、创业者、AI 创作、PGC、想象力、年轻人、有没有、hui、成指涨、大部分、新闻报道、设计师、天下网、创造力、不确定性、HJT
Topic2	数字经济下的科技股投资趋势	数字经济、概念股、Web、拓尔思、年月日、web、房地产、区块链、高开低走、仅供参考、接下来、科技股、智能化、Chiptlet、中国联通、需谨慎、关键词、Midjourney、中国移动、题材股
Topic3	ChatGPT 自然语言技术的突破	ChatGPT、CPO、基础设施、数据中心、数字化、NLP、解决方案、实验室、GPU、越来越、净利润、自然语言处理、大规模、高质量、PICO、人工智能、AI 芯片、研究员、Chat、自动化
Topic4	人工智能在新能源领域的应用	新能源、服务器、大数据、进一步、运营商、云计算、投资者、chatGPT、上市公司、成交量、钙钛矿、元宇宙、子公司、贵金属、发展趋势、科创板、零部件、人民币、NFT、基本面
Topic5	元宇宙虚拟数字人	AIGC、互联网、AI 绘画、文心一言、数字人、元宇宙、生成内容、商业化、ChatGPT、研究院、虚拟人、AI 技术、生产力、持续性、demo、Content、Generated、人工智能、新一轮、爱奇艺 API
Topic6	AIGC 创业机遇	GPT、机器人、OpenAI、AIGC 领域、OAIGC、搜索引擎、CEO、亿美元、创始人、人工智能生成内容、人工智能、第一桶金、开发者、工程师、同花顺、图像识别、马斯克、亚马逊、Diffusion、普通人

对 AIGC 主题挖掘结果进行基于 pyLDAvis 的主题模型可视化。如图 6 所示，左侧部分 7 个聚类主题之间相隔较远，相对互斥，模型聚类效果良好；右侧柱形图展示了对应主题中最相关的 30 个主题特征词语。

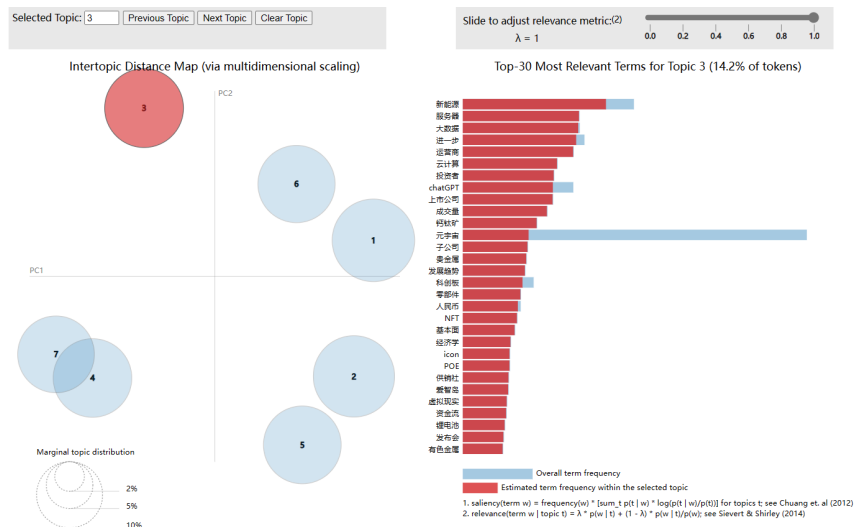


Figure 6. LDA results visualization
图 6. LDA 结果可视化

3.4. 文本主题分析

LDA 模型训练得到 7 个主题类别，分别是人工智能概念股、AIGC 智能数字内容创作、数字经济下的科技股投资趋势、AIGC 自然语言技术的突破、人工智能在新能源领域的应用、元宇宙虚拟数字人、AIGC 创业机遇。

3.4.1. 人工智能概念股

以 ChatGPT 为代表生成式人工智能新技术将持续激发市场需推动了人工智能、量子计算等前沿技术大范围地突破和应用。同时随着以 ChatGPT 为代表的人工智能应用的崛起，包括 AIGC、ChatGPT 等人工智能相关概念股受到市场追捧，具有广阔的市场前景和商业价值。但是，由于人工智能技术的发展还处于初级阶段，人工智能概念股也存在一定的风险和不确定性。因此，投资人需要进行充分的研究和分析，以便做出明智的投资决策。

3.4.2. AIGC 智能数字内容创作

AIGC 智能数字内容创作的核心技术之一是语言生成技术，这种技术可以让机器自动生成文章、新闻、评论等内容，为数字内容创作提供了新的可能性。除了语言生成技术，AIGC 智能数字内容创作还可以利用图像生成技术生成图片、插图、漫画等内容，即 AI 绘画，为数字内容创作提供更加多样化的选择。此外，内容创作者可以从人工智能生成的作品中寻找灵感和思路，有利于拓展创新边界，生产出过去无法想出的杰出创意。

3.4.3. 数字经济下的科技股投资趋势

数字经济是当前经济发展的重要趋势，科技股作为数字经济的重要组成部分，备受市场关注。随着科技的快速发展，科技公司的市值在不断增加。科技公司通常拥有较高的增长潜力、创新能力和竞争优势，能够带来长期的高收益。在科技股未来的赛道上需要投资者非常前沿的跟踪科技产业链，掌握产业情报，判断出投资机会，做出明智的投资决策。

3.4.4. ChatGPT 自然语言技术突破

在模型训练结果中，“ChatGPT”、“NLP(自然语言识别)”、“自然语言处理”、“基础设施”、“数据中心”、“大规模”等词表明，大规模语言训练模型 ChatGPT 的诞生，使得诸多自然语言理解和生成任务上得到了突破性的性能提升，同时算力需求也大规模增长。据 OpenAI 报告，ChatGPT 的总算力消耗约为 3640PF-days(即假如每秒计算一千万亿次，需要计算 3640 天)，需要 7~8 个算力 500P 的数据中心才能支撑运行。ChatGPT 的广泛应用拉动全球智能算力需求爆发[10]。

3.4.5. 人工智能在新能源领域的应用

能源和电力行业正在以多种方式部署人工智能，人工智能正在加速能源转型。例如在电力系统管理上，随着新能源电动汽车的发展，未来将有大量实体设备接入电网，电力和能源系统将变得难以管理。对此，人工智能可以支持规划电网，进行实时的数据分析，识别模式并对电网行为建模，提高电力线路传输容量的利用率[11]。

2019 年，美国能源部宣布支持人工智能领域的创新性研究，并成立了人工智能技术办公室，以促进提升电网韧性，增强环境可持续性，建设智能城市，加速新材料研发等工作的开展。英国政府也在其产业战略挑战基金中频繁提及能源及人工智能的前沿领域，例如机器人在核能、海洋能及深井开采等极端环境能源开发中的应用。中国国家发改委和能源局也强调通过人工智能技术与能源电力行业的深度融合，加速智能电网和能源互联网的建设，并指出这是国家能源基础设施建设的重要环节[12]。毫无疑问，人工智能将在能源领域发挥至关重要的作用，并将有力促进全球经济社会的绿色可持续发展。

3.4.6. 元宇宙虚拟数字人

在该主题的主题词中，“数字人”、“元宇宙”、“商业化”、“虚拟人”、“爱奇艺”等词表明，随着人工智能技术的不断发展，元宇宙概念受到市场热捧，虚拟数字人作为元宇宙的重要基础设施，在元宇宙场景下也有着巨大的应用前景和商业化应用。比如在直播行业，虚拟数字人以真人形象与观众互动；在医疗方面中，虚拟数字人可以应用于疾病诊断、治疗当中、实现与患者实时对话；在影视娱乐行业，虚拟歌姬“洛天依”、虚拟 up 主“柳夜熙”等虚拟偶像深受年轻用户的喜爱。虚拟数字人商业价值释放领域多元，已经成为品牌营销发力的风口。各大品牌纷纷着手打造虚拟偶像或是与之进行合作，助力品牌俘获更多的目标用户。生成式 AI 等技术不断迭代的环境下，中国虚拟人产业高速发展，元宇宙的热潮加速推动虚拟人产业升级。

3.4.7. AIGC 创业机遇

如今，AIGC 领域涌现出非常多的创新创业机遇。在海外市场，AIGC 不只是大厂的机会，也新起了很多创业公司，例如 Jasper.ai 目前应用在营销、电商和自媒体创作等领域。在国内也涌现出新工种，如人工智能训练师[13]、数据标注员、AI 产品经理等等。目前这些职业暂未形成明显的人才缺口，大部分公司近期才设立这些职位，但随着大部分 AI 创业公司的发展，所服务的行业领域越来越多，市场将进入快速增长期，构建数据方面的行业壁垒(积累领域数据、提升数据标注效率等)，将逐渐成为一种趋势，而“人工智能训练师”正好能满足这些需求，在未来几年会得到各家公司更多的重视与关注。

4. 总结和未来展望

4.1. 总结

本文利用爬虫技术爬取关于 AIGC 微博博文数据，再进行 jieba 分词，最后通过 LDA 主题模型建模，对微博平台关于人工智能内容生成的文本内容主题归类，更加直观准确地看出人工智能内容生成地发展现状和未来方向。在投资方面，人工智能具有广阔的市场前景和商业价值，是值得关注的投资机会。在技术层面，AIGC 技术的突破能推进数实融合，加快产业的升级。此外，AIGC 技术的发展也为各行各业带来了更多可能性和新的创业机遇。

4.2. 未来展望

在产业互联网领域，生成式 AI 技术迎来重大发展，合成数据将牵引人工智能的未来。MIT 科技评论将 AI 合成数据列为 2022 年 10 大突破性技术之一；Gartner 也预测称，到 2030 年合成数据将彻底取代真实数据，成为训练 AI 的主要数据来源[14]。未来，“AIGC + 产业”将持续大放异彩，有望带来一场自动化内容生产与交互变革，在各行各业引发巨震[15]。同时，AIGC 的发展也面临许多科技治理问题的挑战，比如知识产权、信息安全、伦理道德等诸多问题有待业界解决[16] [17]。

展望未来，一切皆有可能。无论遇到怎样的挑战和问题，都无法阻挡 AIGC 产业的迅猛发展。在经历一段时间的磨合期后，AIGC 产业将迎来发展更为迅猛的价值增长期，形成更为完善的生态体系，并在需求更加清晰且明确的场景落地。未来整个世界，也将因此变得全面智能化、数字化。

基金项目

项目来源：福建省科技厅；项目名称：基于 ATOT 技术的智能养老系统设计与开发；项目编号：2023350104000282。

参考文献

- [1] 杨欣. 人工智能“智化”教育的内涵、途径和策略——人工智能何以让教育变得更聪明[J]. 中国电化教育, 2020(3): 25-31.
- [2] 张智雄, 曾建勋, 夏翠娟, 王东波, 李白杨, 蔡迎春. 回应 AIGC 的信息资源管理学人思考[J]. 农业图书情报学报, 2023, 35(1): 4-28.
- [3] Sequeira, G. (2023) Will ChatGPT Write the Next Indian Blockbuster?
<https://www.filmcompanion.in/features/bollywood-features/will-chatgpt-write-the-next-big-indian-blockbuster-rohit-shetty-deepika-padukone-ranveer-singh-lokesh-kanagaraj>
- [4] 郭全中, 袁柏林. AIGC 与 WEB3.0 有机融合: 元宇宙内容生产的新范式[J]. 南方传媒研究, 2023(1): 36-47.
- [5] 李倩, 王帅. LDA 模型下我国公共图书馆微信平台阅读推广内容主题研究[J]. 图书情报工作, 2022, 66(8): 72-83.
- [6] 关鹏, 王曰芬. 科技情报分析中 LDA 主题模型最优主题数确定方法研究[J]. 现代图书情报技术, 2016(9): 42-50.
- [7] 蒋倬昱. 基于微博话语的长沙历史文化地名认知与传播效应研究[D]: [硕士学位论文]. 长沙: 中南林业科技大学, 2022.
- [8] 贺亚辉. 基于学术领域影响力的期刊审稿专家精准化推荐[D]: [硕士学位论文]. 太原: 山西医科大学, 2022.
- [9] 孙瑞英, 陈宜泓. 基于 LDA 主题模型的国内智慧阅读研究热点及发展导向研判[J]. 图书馆建设, 2023(3): 82-93+103.
- [10] 林碧涓. 运营商持续发挥云优势促 AI 产业发展[N]. 通信信息报, 2023-04-05(005).
- [11] 谢黎, 唐川, 黄茹. 世界经济论坛关于利用人工智能加速能源转型的建议与启示[J]. 科学观察, 2022, 17(3): 80-83.
- [12] 中华人民共和国国务院新闻办公室. 新时代的中国能源发展[N]. 人民日报, 2020-12-22(010).
- [13] 李隆帆. 人机交融的领航员——“人工智能训练师”新职业介绍[J]. 中国培训, 2022(11): 60-62.
- [14] 胡晓萌. AIGC 技术和产业生态迎来发展快车道[J]. 数字经济, 2023(3): 40-45.
- [15] 曹建峰, 陈楚仪. AIGC 浪潮下, 合成数据关乎人工智能的未来[J]. 新经济导刊, 2022(4): 25-31.
- [16] Verma, P. and Oremus, W. (2023) ChatGPT Invented a Sexual Harassment Scandal and Named a Real Law Prof as the Accused. Washington Post. <https://www.washingtonpost.com/technology/2023/04/05/chatgpt-lies/>
- [17] Murgia, M. and Borrelli, S.S. (2023) Italy Temporarily Bans ChatGPT over Privacy Concerns. Financial Times. <https://www.ft.com/content/3ce7ed9d-df95-4f5f-a3c7-ec8398ce9c50>