Statistics and Application 统计学与应用, 2023, 12(4), 833-842
Published Online August 2023 in Hans. https://www.hanspub.org/journal/sa
https://doi.org/10.12677/sa.2023.124087

Hans汉斯

基于LDA主题模型的杭州亚运会微博话题分析

董韶琦，郑　静

杭州电子科技大学经济学院，浙江　杭州

收稿日期：2023年7月4日；录用日期：2023年7月25日；发布日期：2023年8月7日

摘　要

为了探索杭州亚运会预热阶段新兴媒体传播的结构和内容，帮助相关部门更高效地进行舆论监管与引导，本文创新性地对亚运会传播内容进行LDA主题模型的构建。本文在新浪微博爬取与杭州亚运会相关内容，构建亚运会文本的隐含狄利克雷分布(Latent Dirichlet Allocation, LDA)模型，采用困惑度评价指标确定模型最优主题数，然后用框架和语境理论分别从结构和内容挖掘相关文本内涵。结果显示，亚运会预热传播内容主要围绕娱乐宣传、参与人员、基础设施、起止仪式、市场合作、竞技项目6个框架展开，展示出人们对杭州举办此次亚运会的肯定与期待；同时亚运会也为我国经济尤其是杭州经济的发展起到一定的促进作用，也为对内对外经济合作提供了契机。

关键词

杭州亚运会，微博，LDA模型

Analysis of Microblog Topics of Hangzhou Asian Games Based on LDA Topic Model

Shaoqi Dong, Jing Zheng

School of Economics, Hangzhou Dianzi University, Hangzhou Zhejiang

Received: Jul. 4th, 2023; accepted: Jul. 25th, 2023; published: Aug. 7th, 2023

Abstract

In order to explore the structure and content of emerging media communication in the warm-up stage of the Hangzhou Asian Games, and help relevant departments to more effectively supervise and guide public opinion, this paper innovatively constructs the LDA Topic model for the communication content of the Asian Games. This paper builds a Latent Dirichlet Allocation model for the text of the Asian Games by crawling Sina Weibo content related to the Hangzhou Asian Games, uses Perplexity evaluation indicators to determine the optimal number of topics in the model, and then

文章引用: 董韶琦, 郑静. 基于 LDA 主题模型的杭州亚运会微博话题分析[J]. 统计学与应用, 2023, 12(4): 833-842.
DOI: 10.12677/sa.2023.124087

uses the framework and context theory to mine the relevant text content from the structure and content. The results show that the warm-up communication content of the Asian Games mainly revolves around six frameworks: entertainment promotion, participants, infrastructure, start and end ceremonies, market cooperation, and competitive projects, showcasing people's affirmation and expectation of Hangzhou hosting the Asian Games; At the same time, the Asian Games have also played a certain promoting role in the development of China's economy, especially in Hangzhou, and provided opportunities for domestic and foreign economic cooperation.

Keywords

Hangzhou Asian Games, Weibo, LDA Model

1. 引言

2015 年 9 月 16 日，经过亚奥理事会代表大会的投票，杭州获得 2022 年亚运会主办权，成为继北京和广州之后第三个举办亚运会的中国城市。杭州 2022 年亚运会以“中国新时代·杭州新亚运”为定位、“中国特色、浙江风采、杭州韵味、精彩纷呈”为目标，秉持“绿色、智能、节俭、文明”的办会理念，坚持“以杭州为主，全省共享”的办赛原则。亚运会原定于 2022 年 9 月 10 日至 25 日举办，但由于疫情原因，推迟至 2023 年 9 月 23 日至 10 月 8 日举办。此次杭州亚运会是杭州重要外交活动，是展示杭州风貌，提升杭州国际国内形象的良好机会。

随着互联网等科学技术的不断创新发展，社交网络的不断成熟，人们发表看法言论的平台逐渐从传统的报纸等媒体转移到一些新兴媒体平台上，逐渐改变了原有传统媒体占主导地位的情况。新媒体以其自身强大的沟通能力和实时交互性，吸引了越来越多的民众主动参与到公共事件的讨论中，并表达个人观点。其中，新浪微博以其自身强大的时效影响力和具备快速发布、传播信息的独特优势成为我国最主要的互联网社交媒体平台之一[1]。本文在之前一些学者对 LDA 模型研究的基础上，首次对与此次杭州亚运会相关内容进行探究。以微博为平台，从框架和语意内涵两方面探索杭州亚运会预热阶段新兴媒体传播的结构和内容，为完善传播实践体系提供相关方向和建议。

2. 国内外研究现状

隐狄利克雷模型本质上是一种无监督的概率主题模型，具有广泛的适用性，国内外许多学者以此模型为基础展开研究。Aakansha Gupta 等[2]提出了一个基于 LDA 的 PAN-LDA 主题模型，并运用该模型将新冠病例数据和新闻文章合并到通用 LDA 中，以改进时间序列数据的预测。Zhang 等[3]为了解决传统推荐算法存在数据稀疏、不注重推荐结果多样性等问题，使用 LDA 提取有关电影评论的主题，并识别与主题相关的情感倾向。在面对文本数据具有非结构化、特征稀疏时，Wang 等学者[4]在词汇意义共现分析和 LDA 主题模型的基础上，提出了一个 CL-LDA 词意共现主题模型，通过提高主题生成的质量来完成对短文本的主题挖掘任务。2012 年，Ozyurt 等学者[5]针对 LDA 不适合用来处理短文本的缺点，提出了一种适用于短文本的隐狄利克雷分配方法 SS-LDA。为了挖掘出更多能够代表文档的主题信息，宁宁等人[6]通过融合 LDA 主题模型和 Doc2vec 算法得到一种主题向量表示和文档向量表示方法。Sakshi 和

Kukreja Vinay [7]根据提取的研究领域确定突出的识别模型，从提取的研究趋势中绘制发展图表，以指导未来的工作。

针对亚运会各个方面的准备工作，许多学者都进行了研究。基于杭州的地理位置及气象历史数据，任勇等[8]根据赛会核心区内雷电流幅值年最大值的平均值，计算了雷击点与屏蔽空间距离不同时，场馆处无衰减的磁场强度值，为场馆雷电防护及内部电子信息系统设计提供依据和参考。作为一种被“重新发现”的地区文化，良渚文化在中国文明进程中具有极其重要的地位和作用。单凯等学者[9]通过对良渚文化对外传播与2022年杭州亚运会宣传良性互动内容的考察，探讨良渚文化与亚运文化实现合作共赢的协同发展条件，并为杭州亚运会宣传建设以及推动良渚文化深度国际化提供文化传播对策。

但2022年杭州亚运会宣布延期举办，构成了突发性的体育公共事件，对相关宣传工作产生了不利影响。杨柯[10]采用分类统计的方法，按照选题的不同类型，对多家杭州地方媒体微博端的发稿进行归类。通过分析发稿数量和类型的变化，研究危机发生后，媒体调整议程的方法、策略和缺失，文章提出了强化核心议题、强化群众报道两项建议，旨在为突发性体育公共危机中宣传策略的调整提供应对参考。

3. 理论基础

3.1. LDA 模型

本文采用隐狄利克雷模型进行文本聚类分析，LDA模型的优势在于它可以自动将文本编码为一定数量具有实质性意义的主题，以此来提高效率，减少人为干预负担，其模型结构如图1所示。

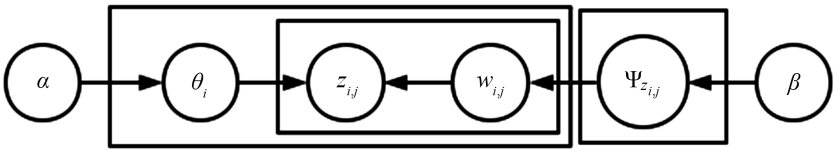


Figure 1. Schematic diagram of LDA model structure
图 1. LDA 模型结构示意图

LDA模型分为文本、主题和词语三层，具体过程如下所示[11]：

1) 按照先验概率 $P(d_i)$ 选择一篇文本 d_i；

2) 从以参数为 α 的的迪利克雷分布中随机生成文本 d_i 对主题的多项分布 θ_i；

3) 从文本 d_i 对应主题的多项式分布 θ_i 中随机生成个 j 个词语主题 $z_{i,j}$；

从以参数为 β 的迪利克雷分布中随机生成主题 $z_{i,j}$ 对应词语的多项式分布 $\Psi_{i,j}$；

4) 综合主题 $z_{i,j}$ 对应词语分布情况 $\Psi_{i,j}$ 生成词语 $w_{i,j}$。

3.2. 困惑度

通常用困惑度来衡量一个概率分布或概率模型预测样本的好坏程度，可以将困惑度看作交叉熵的指数形式，两个模型越接近，交叉熵越小，困惑度也越小[12]，它也可以用来比较两个概率分布或概率模型。它的计算方式为，对于给定测试集 $W = w_1 w_2 \cdots w_N$，将困惑度定义为测试集概率的倒数，并用单词数做归一化。

$$PP(W) = P(w_1 w_2 \cdots w_N)^{-\frac{1}{N}} = \sqrt[N]{\frac{1}{P(w_1 w_2 \cdots w_N)}}$$

使用链式法则来计算 $P(W)$：

$$PP(W)=\sqrt[N]{\prod_{i=1}^{N}\frac{1}{P\left(w_i\middle|w_1w_2\cdots w_{i-1}\right)}}$$

如果使用 bigram 模型，公式为：

$$PP(W)=\sqrt[N]{\prod_{i=1}^{N}\frac{1}{P\left(w_i\middle|w_{i-1}\right)}}$$

词序列的条件概率越高，困惑度越低，也就是说，模型生成能力越强，困惑度值越小。

4. 数据收集与数据处理

4.1. 数据来源

微博作为热点传播的一个重要社交媒体，截止 2022 年四季度末，月活跃用户达到 5.86 亿，日活跃用户达到2.52 亿，成为网络舆情的重要诞生地。本文利用 python 获取微博中关于杭州亚运会的相关博文，以“杭州亚运会”为关键词，选取 2015 年 1 月 1 日~2022 年 9 月 1 日为时间段，共得到 31,459 篇博文。

4.2. 数据预处理

对获取到的数据进行清洗，主要是删除重复以及空白内容等，处理后共得到 28,354 篇博文。再利用 python 软件进行中文分词、去除停用词、词干提取等。

4.3. 确定主题个数

对于预处理之后的文本数据进行困惑度的计算，确定主题数。结果如图 2 所示，随着主题数的增大，困惑度逐渐减小。按原理来说，困惑度应是越低越好，那么就应当选择更多的主题数，但是当主题数过多时，模型会出现过拟合状况。从图 2 中可以看出，几个较为明显的转折点在主题数为 0~2、2~4、6~7 处。当主题数在 0~2 之间时困惑度过高，应当排除。2~4 个主题数与 6~7 相比较，6~7 的困惑度相对较低，因此主题数在 6 和 7 之间考虑。为了避免主题数过多出现过拟合状态，所以本文最终选择 6 个主题数来进行分析。

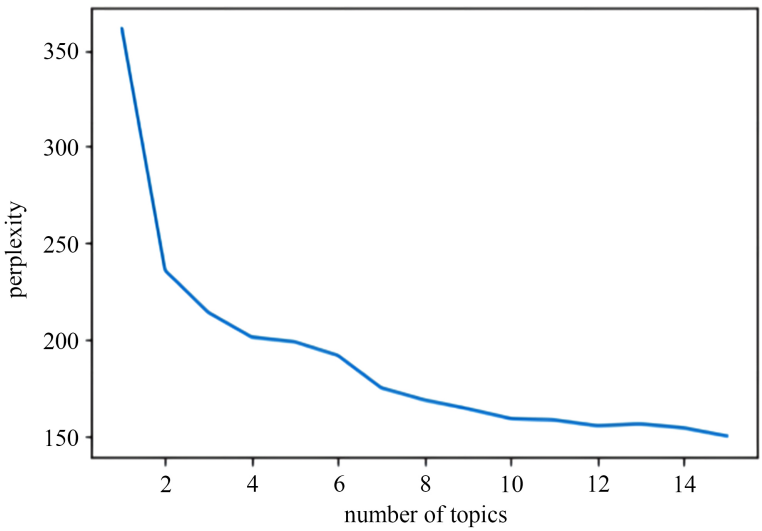


Figure 2. Line chart of perplexity - number of topics
图 2. 困惑度 - 主题数折线图

4.4. 主题可视化

进行 LDA 主题建模之后，将其可视化，便可以知道每个主题出现的频率，同时也可以检测通过困惑度确定主题数的效果。

可视化结果如图 3 所示，气泡的大小编号表示每个主题出现的频率，同时主题之间的位置远近表示各主题之间的接近性，气泡若有重叠则说明两个主题的特征词有重叠交叉部分。从图 3 中可以看出，主题气泡之间均存在一定距离，交叉部分较少，因此主题识别效果较为理想，选择 6 作为主题数是较为合理的。

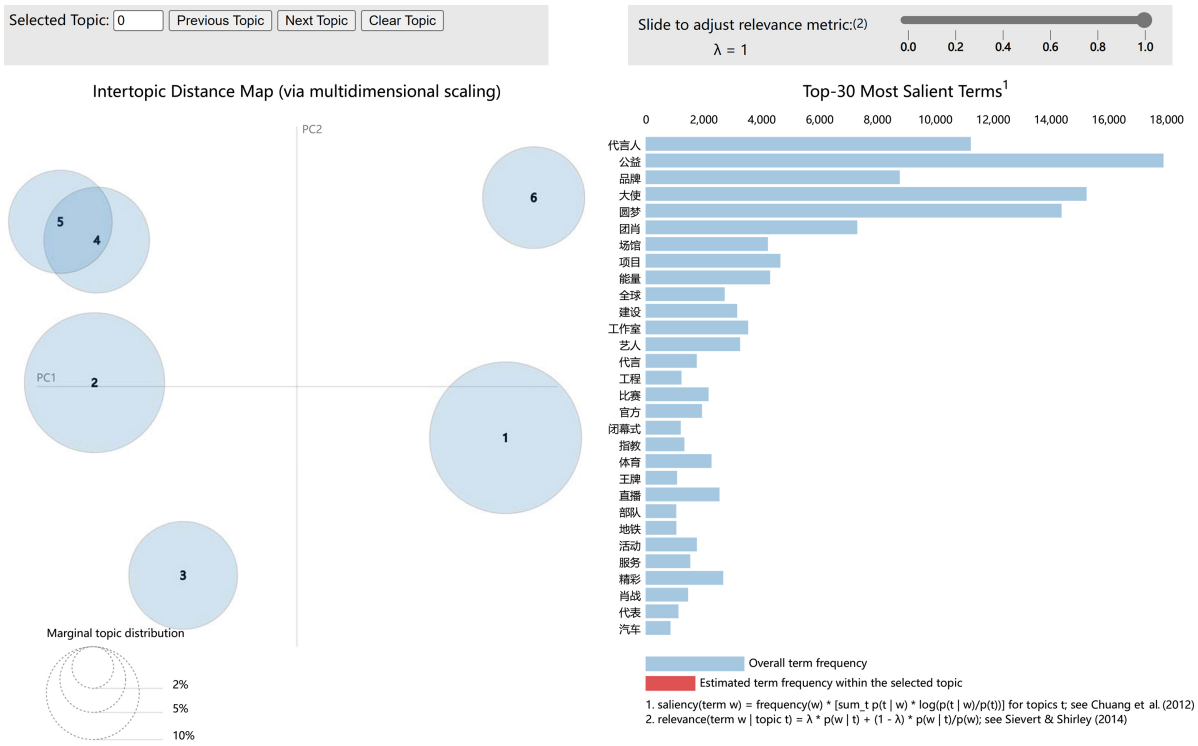


Figure 3. Maximum theme LDA visualization
图 3. 最大主题 LDA 可视化图

5. 结果分析

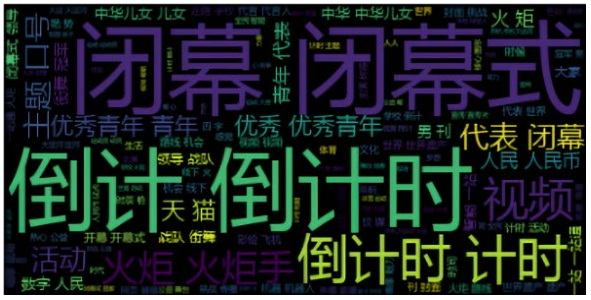
5.1. 模型的话题框架

主题框架重复的频率越高，强度越强，发挥的作用越大。因此，本文采用词云图形式，进一步展示 LDA 主题聚类结果，更加清晰地分析各个框架的频率和强度。图 4 为新浪微博上关于杭州亚运会相关博文中频率和强度最高的 6 个主体框架，分别为 Topic1 娱乐宣传、Topic2 参与人员、Topic3 起止仪式、Topic4 市场合作、Topic5 基础设施、Topic6 竞技项目。各个主题框架的频率从 Topic1 到 Topic6 依次减小，框架中的每个关键词也有不同的权重，由词云图中单词的大小表示。

Topic1 为娱乐宣传，频率为 0.266，是新浪微博中讨论最多的框架，框架频率和强度均排在首位。从词云图中可以看出，此框架主要由代言、代言人、公益、大使、品牌、圆梦、视频、工作室、艺人等话题构成，说明前期宣传中，网民对于明星代言、公益大使等话题比较感兴趣。因此，邀请艺人做亚运会形象大使、公益大使，并让他们担任亚运会代言人，用明星的影响力进行亚运会宣传活动大致可以达到理想中的效果。

Topic1 娱乐宣传(0.266)

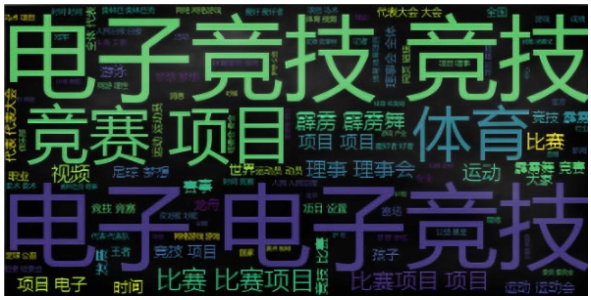
Topic2 参与人员(0.227)



Topic3 起止仪式(0.135)

Topic4 市场合作(0.129)

Topic5 基础设施(0.123)



Topic6 竞技项目(0.120)

Figure 4. The weight and keyword cloud of Weibo themes
图 4. 微博主题权重及主题词云

Topic2 为参与人员，频率为 0.227，包括了主要参与人员和相关主办单位，关键词主要有运动员、志愿者、组委会、理事会等，主要围绕亚运会的主要参与者展开。同时，志愿活动的报名也在火热进行中，从话题强度可以看出，杭州人民对亚运会的参与意向很高，对志愿活动的给予了充分的关注。

Topic3 为起止仪式，频率为 0.135，构成此框架的主要关键词有倒计时、闭幕式、火炬手、主题口号、优秀青年、代表、开幕式、中华儿女、现场等。表现出人们对此次杭州亚运会的开幕式以及闭幕式的仪式的关注，以及对此次活动到来的期待。

Topic4 为市场合作，频率为 0.129，主要挖掘了亚运会的经济价值和商业合作属性，主要是相关经济活动，关键词包括了合作伙伴、合作公司、市场、技术、企业、发展、创业、产品、宣传、服务等。赋予了杭州亚运会一定的经济意义，此次活动有带动杭州体育以及旅游相关产业的发展的可能性。

Topic5 为基础设施，频率为 0.123，主要包括了相关场馆建设和道路等公共设施的建设等，关键词主要有高速公路、体育场、场馆、地铁线路、体育中心、施工、交通、设施等。表明政府等相关部门为保障亚运会顺利召开，在完善杭州各种基础设施的建设方面做出了相应努力。

Topic6 为竞技项目，频率为 0.120，由关键词电子竞技、竞赛项目、体育、比赛、霹雳舞、项目、竞

赛、梦想等构成。这一话题主要围绕运动员们参与的体育赛事，尤其是新增的电子竞技以及霹雳舞项目，格外引人关注。

5.2. 话题内容探索

在框架意义的基础上，以语篇为研究单位，以各个主体中权重最高的微博博文为样本，以各个框架中权重系数前 20 位的博文为参考对象，分析杭州亚运会在语篇中的语境框架、语境主题、语境组合和相关语境因素，语篇语境列表如表 1 所示。

Table 1. List of discourse contexts
表 1. 语篇语境列表

序号	框架	语境主题	语境组合	语境因素
T1	娱乐宣传权重 0.9950	肖战成为亚运会圆梦公益大使。	肖战被评为杭州亚运会圆梦公益大使。 谭松韵献唱杭州亚运会。	来源：新浪微博 评价与预测：在亚运会的宣传方面，主要用明星效应来吸引网民们的关注。
T2	参与人员权重 0.9925	杭州亚运赛会志愿者招募将于 5 月前后启动。	组委会公布了志愿者招募的相关信息，包括要符合什么条件、怎么报名等问题。 各个项目的运动员们也都在为亚运会积极调整状态，积极备战。	来源：新浪微博 评价与预测：亚运会志愿者报名活动受到人们广泛关注，掀起一阵热潮。
T3	起止仪式权重 0.9954	杭州亚运会吉祥物发布，智能小伙伴“江南忆”组合在互联网云端与网友们见面。	吉祥物为一组承载深厚底蕴和充满时代活力的机器人——琮琮、莲莲和宸宸，分别代表了世界遗产良渚古城遗址、世界遗产西湖和世界遗产京杭大运河。 第 18 届亚运会闭幕式“杭州 8 分钟”的演出十分精彩。	来源：新浪微博 评价与预测：人们对亚运会吉祥物以及开幕式等赋予了充分关注。
T4	市场合作权重 0.9940	吉利将在杭州亚运会期间试运营 L4 级别自动驾驶的车辆。	作为杭州 2022 年第十九届亚运会的官方合作伙伴，吉利汽车将以科技赋能亚运，创新未来智能出行新体验。 叠加浙江共同富裕和杭州亚运会概念为主的个股在股市中走势较好。	来源：新浪微博 评价与预测：亚运会的举办将会为经济产业带来相关契机。
T5	基础设施权重 0.9855	沪杭甬高速杭州市区段(乔司枢纽－红垦枢纽)道路封闭施工。	沪杭甬高速公路杭州市区段改建工程是杭州亚运会交通重点保障项目，通过把原有高速公路路段进行抬升改建，成为集快速路、轨道线、地面道路一体的综合交通走廊。 成为亚运会比赛地点的多个场馆等进行重建或翻修。	来源：新浪微博 评价与预测：为了保障亚运会的顺利进行，杭州对相关基础设施进行完善。
T6	竞技项目权重 0.9851	电子竞技、霹雳舞成为杭州亚运会竞赛项目。	在保持 40 个大项不变的前提下，增设电子竞技、霹雳舞两个项目。电子竞技属“智力项目”，霹雳舞属“体育舞蹈”。意味着电子竞技和霹雳舞正式列入杭州亚运会竞赛项目。	来源：新浪微博 评价与预测：新增竞赛项目会吸引更多观众观看，使得观赏性更佳。

Topic1 主要内容是娱乐宣传方面，因为微博的娱乐性功能较强，所以这一框架为强度和频率最大的框架。权重系数前二十位的语篇中，内容主要分布在：肖战杭州亚运会圆梦公益大使、谭松韵献唱杭州亚运会、魏晨亚运会推广大使、杭州亚运会 2023 年 9 月举行、王霜杭州亚运会大使等。这一框架中，主要关注点在亚运会代言人、公益大使等与娱乐明星相关的内容上，突出利用明星艺人等的国民影响力进行宣传。通过此种方法增加亚运会的国民普及度和国民关注度，吸引群众，增加热度。但是亚运会的关注点不应过多地在娱乐方面，运动员应该是主角，此框架知名运动员方面的宣传较少，偏向娱乐化。

Topic2 为参与人员框架，主要关于亚运会志愿者报名和运动员热身准备情况。权重系数前二十位的语篇内容为：2022 年杭州亚运会和亚残会吉祥物全球征集启动、亚运赛会志愿者招募预计 5 月前后正式启动、中国男足国家队以及 U23 国足尽可能多地参加高质量比赛来备战亚运会、中国围棋队启动杭州亚运会选拔力争佳绩、亚运会推迟举办对运动员造成的影响等。在这一框架中，人们对亚运会志愿者报名活动赋予充分关注，志愿者是群众能够亲身参与到亚运会中的最佳途径，因此关注度较高。同时提到运动员的相关内容中，运动员多作为集体的一部分出现，对单独的运动员个人关注还有不足。此外，也有一些亚运会推迟举办，导致很多运动员状态调整方面困难的内容。可以看出疫情对此次亚运会的影响还是较大的，甚至可能造成一些运动员错过职业生涯的黄金时段。

Topic3 为起止仪式框架，主要包括了人们对上届亚运会闭幕式上杭州八分钟以及此次亚运会亚残会吉祥物以及开幕式的关注。Topic3 中权重前二十的语篇有：易烊千玺是参加第 18 届亚运会闭幕式“杭州 8 分钟”唯一中国艺人、杭州亚运会核心图形和色彩系统揭晓、杭州亚运会彩绘飞机群亮相试飞杭州-温州航班、杭州亚运会吉祥物发布、浙江省承办亚运会的 6 个城市纳入新一批数字人民币试点城市、杭州亚运火炬手对全民开放。权重较高的语篇中均传递出人们对亚运会相关周边例如吉祥物火炬手开幕式等赋予了较多关注，组委会通过吉祥物火炬手等其他方面间接吸引群众的注意力聚焦到亚运会起止仪式上。同时易烊千玺参演雅加达亚运会闭幕式杭州八分钟得到网民广泛关注也与 Topic1 娱乐宣传框架有一定联系，同样也是利用明星的个人影响力来加强杭州亚运会的影响力，让更多的人关注到此次赛事。

Topic4 主要立足于与亚运会相关的市场合作方面，展示亚运会的经济价值，是新的合作关系建立的契机。此框架中权重系数前二十位的语篇有：雪天盐业成为亚运会官方供应商，公司现金流和营收利润匹配，良性发展，回款能力较强；浙江板块近期走势很强势；作为杭州 2022 年第 19 届亚运会官方航空客运服务合作伙伴，长龙航空依托浙江及长三角区域的区位优势，将会迎来新的发展机遇；园林工程+杭州亚运会，3 天 2 板，3 个交易日股价上涨了 47.06%；所有板块集体上涨，其中水泥、杭州亚运会、建筑装饰等板块涨幅居前等。这一框架主要分为两个部分，一方面是亚运会在各个行业与不同品牌进行官方合作，促进了行业内的发展和创新，取得更大的进步；另一方面为亚运会板块在股市中持续走高，股市行情较为可观。此次亚运会会为我国经济起到一定的促进作用，尤其是为杭州的经济发展提供了新的机会。杭州的旅游业、服务业等行业会迎来新一轮的发展契机，也会使相关行业内部进行整顿，建立更加完善的市场，促进良性竞争，共同发展。

Topic5 主要围绕亚运相关的基础设施建设展开，从 T5 中可以看出该框架的语意内涵，杭州市政府为保障亚运会的顺利进行，对高速路等相关基础设施进行修缮，保障运动员等的正常活动。Topic5 权重系数前二十位的语篇主要包括了绍兴将实现高速公路网络与城市快速路网的高效衔接,为优化亚运会棒(垒)球场馆周边的交通组织，提升镜湖新区集聚度及交通便利性起到重要推动作用；亚运会重要交通保障工程——杭州西站，正式进入建设冲刺阶段；杭州亚运会竞赛场馆全部竣工；钱塘江上新增过江通道望江隧道等内容。这一框架中的语境组合多为高速公路、高铁、隧道等基础设施，少部分为竞赛场馆等。基础设施框架展现了杭州市政府乃至浙江省为确保亚运会正常进行所付出的努力，同时也方便了居民的日常出行等问题。虽然此次基础设施建设大多为公共性的，并不能直接获得经济收益，但它能通过城市整

体社会效益、环境效益的改善和提高来体现出其投资的价值。调好基础设施系统才能有助于促进城市的经济发展，此次杭州市政府大力加强基础设施建设，将有效地促进杭州经济的进一步增长，此框架与Topic4 市场合作的交叉之处可以在此体现。

Topic6 为竞技项目框架，人们关注的重点主要在新增项目上，尤其是霹雳舞和电竞这种国民度较高的项目，之前仅作为表演项目在上一届亚运会中出现过，因此能否在杭州亚运会中作为正式竞赛项目就让人们尤为关注。Topic6 中权重前 20 的语篇为：杭州亚运会确定举办时间，并确定龙舟并入皮划艇项目；电子竞技成为 2022 年杭州亚运会正式项目，在保持 40 个大项不变的前提下，增设电子竞技、霹雳舞两个项目，电子竞技与棋类项目同属于“智力项目”，霹雳舞则属于“体育舞蹈”；原定于 2022 年 9 月 10 日至 25 日在中国杭州举办的第 19 届亚洲运动会将延期举行；在奥运项目上新增了马术，同时新增克柔术、柔术和板球 3 个非奥项目。从这些语篇中可以看出，此届亚运会与上届相比新增项目较多，主要分为三类：一类是知名度较高的休闲娱乐项目例如电竞、霹雳舞，一类是我国传统文化项目如龙舟，一类综合性运动项目例如板球。组委会在通过增加一些新的项目，尤其是一些国民度比较高的运动来让更多的人关注亚运会，同时也会提高一些趣味性和观赏性。将龙舟并入皮划艇项目，纳入正式竞赛项目，能极大的促进我国传统文化的传播，增强我国的文化自信。

6. 总结与展望

本文通过对新浪微博上与杭州亚运会相关的博文进行研究，发现内容主要围绕娱乐宣传、参与人员、起止仪式、市场合作、基础设施、竞技项目等 6 个话题展开。主要展示出人们对杭州举办此次亚运会的肯定与期待，会吸引人们对体育运动进一步的关注，同时也为我国经济尤其是杭州经济的发展起到一定的促进作用，为对内对外经济合作提供了契机。

但是从这些话题内容中可以看出，现阶段仍存在一些不足之处。Topic1 娱乐宣传框架(0.266)频率最高，其主要内容为亚运会通过明星艺人的影响力来进行宣传活动，比如邀请明星担任公益大使、宣传大使等形式，用其知名度来提高此次亚运会的国民关注度。但是大部分仅局限于娱乐圈，很少涉及到其它领域，仍然有扩展空间，处于被动传播的状态。Topic2 为参与人员框架(0.227)，主体为志愿者和运动员，体现出人们广泛参与的热情，同时也提到了运动员们的艰苦备战。但从权重较高的语篇中可以看出，运动员大多以集体形式出现，具体运动员的表现和态度，以及运动员们因为亚运会延期受到的影响没有得到深入挖掘。Topic3 起止仪式框架(0.135)提及了雅加达亚运会的闭幕式以及人们对此届亚运会的期待，从排名前二十的语篇中可以看出组委会通过宣传为亚运会做出的各种准备，例如杭州亚运会核心图形、吉祥物等来吸引群众更加关注赛事信息。但从词云图中看来，人们对于上届闭幕式的关注大于对本届倒计时的关注，对于此次开幕式的宣传还不够到位。

对于 Topic4 市场合作框架(0.129)，杭州市政府大力筹办此次亚运会，推进各行业向前发展，提供了新的经济发展契机，这也会提高群众的福利待遇和生活水平。杭州亚运会是杭州市甚至是浙江省对外发展的一次重要机会，但从权重较大的语篇中并没有展示出国内外经济协同合作发展。Topic5 基础设施框架(0.123)为政府为筹办亚运会在基建方面做出的相关行动，但过多的建筑工程也会给居民的日常生活带来不便，例如到处修路以及噪声污染。基础设施建设会带动整个城市经济发展，两者是同步增长关系。从长远来看，在基础设施上作出贡献利大于弊。Topic6 竞技项目框架(0.120)频率最低，从其排名靠前的关键词和语篇中可以看出，人们对一些新增项目比较感兴趣。组委会将电竞、霹雳舞等具有一定观赏性的项目确定为正式项目，此举能吸引更多额外群众来关注到此次赛事。人们的关注点都在各种新增竞赛项目上，但是对其项目申请的努力过程和运动员的努力训练并没有太多关注。

根据以上分析结果我们提出以下建议。首先，人们对杭州亚运会的关注多来自明星效应的影响，对

此，主办方可进一步加大宣传力度，还可邀请知名运动员进行代言宣传活动。同时可以提高对努力训练备战的运动员的关注，也可进一步挖掘一些新晋运动员与普通人追梦圆梦的过程，这可能更具宣传意义。其次，杭州市政府也需要更加科学合理地规划基建工程，尽量减少对居民的影响。最后，政府还需要加强经济合作，扩大自己的经济交流面，从宏观角度出发，促进整个浙江省经济的发展。

此前并未出现对亚运会传播内容进行分析的相关文献，因此，本文创新地利用 LDA 主题模型对亚运会相关微博话题进行分析，探索相关热点话题的内在意义，有一定的新颖性和时效性。与之前对 LDA 主题模型进行研究的文献相比，本文在对研究结果进行分析时，既对框架内容进行了挖掘，又进行了内容向度的意义探索。但是以上分析可能存在一定的局限性，本文所选的数据来源仅为新浪微博，具有一定的局限性，而且新浪微博多偏娱乐化，可以进一步扩大数据来源范围，使得结果更加精准。

参考文献

[1] 张晨晨. 基于 LDA 模型的舆情情感主题研究[D]: [硕士学位论文]. 阜阳: 阜阳师范大学, 2022. https://doi.org/10.27846/d.cnki.gfysf.2022.000065

[2] Gupta, A. and Katarya, R. (2021) PAN-LDA: A Latent Dirichlet Allocation Based Novel Feature Extraction Model for COVID-19 Data Using Machine Learning. *Computers in Biology and Medicine*, **138**, Article ID: 104920. https://doi.org/10.1016/j.compbiomed.2021.104920

[3] Zhang, Y.L. and Zhang, L.L. (2022) Movie Recommendation Algorithm Based on Sentiment Analysis and LDA. *Procedia Computer Science*, **199**, 871-878. https://doi.org/10.1016/j.procs.2022.01.109

[4] Wang, J., Wang, L., Xu, J., *et al*. (2021) Information Needs Mining of COVID-19 in Chinese Online Health Communities. *Big Data Research*, **24**, Article ID: 100193. https://doi.org/10.1016/j.bdr.2021.100193

[5] Ozyurt, B. and Akcayol, M.A. (2021) A New Topic Modeling Based Approach for Aspect Extraction in Aspect Based Sentiment Analysis: SS-LDA. *Expert Systems with Applications*, **168**, Article ID: 114231. https://doi.org/10.1016/j.eswa.2020.114231

[6] 宁宁, 莫秀良, 王春东. 基于融合 LDA 和 Doc2vec 算法的文本表示模型的研究[J]. 天津理工大学学报, 2021, 37(2): 55-60.

[7] Sakshi, K.V. (2023) Recent Trends in Mathematical Expressions Recognition: An LDA-Based Analysis. *Expert Systems with Applications*, **213**, Article ID: 119028. https://doi.org/10.1016/j.eswa.2022.119028

[8] 任勇, 邢天放. 杭州亚运会核心区域雷电特征分析和防御建议[J]. 价值工程, 2022, 41(25): 132-135.

[9] 单凯, 黄斐凡. 良渚文化的对外传播与 2022 年杭州亚运会宣传[J]. 浙江体育科学, 2021, 43(5): 1-6+18.

[10] 杨柯. 论杭州亚运会延期对媒体议程影响——以地方主流媒体官方微博为例[J]. 新闻研究导刊, 2022, 13(20): 116-118.

[11] 白健, 洪小娟. 基于弹幕的网络舆情文本挖掘与情感分析[J]. 软件工程, 2022, 25(11): 44-48. https://doi.org/10.19644/j.cnki.issn2096-1472.2022.011.010

[12] 何天文, 王红. 基于语义语法分析的中文语句困惑度评价[J]. 计算机应用研究, 2017, 34(12): 3538-3542, 3546.