

第二节 基于 LDA 模型的主题分析

一、LDA 主题模型

隐含狄利克雷分布 (Latent Dirichlet Allocation, LDA)，由 Blei 等人 2003 年提出。LDA 主题模型是一个三层结构贝叶斯概率模型，由文档、主题和主题词构成。LDA 是一种非监督的机器学习模型。LDA 主题模型的中心思想就是一篇文档是由若干个主题组成的，这些主题会以一定的概率混合成文档，同理每个主题又对应着不同的词，这些词也会以一定的概率混合成主题，并且不用在意这些单词出现的前后顺序，最后求解某个单词在文档中出现的概率，挖掘出文本潜在的主题信息。LDA 模型生成过程如图 3-3 所示，各参数含义如表 3-7 所示。

由图表可看出 LDA 模型的过程为，首先 m 为文档， M 为文档集并且 m 是 M 的一部分， K 为主题， m 是由一定 K 的概率分布构成，从文档中抽出一个主题。由一些关系为概率论分布的单词组成了被抽中的主题，其次从单词分布中抽出一个词；最后一直重复上述步骤。

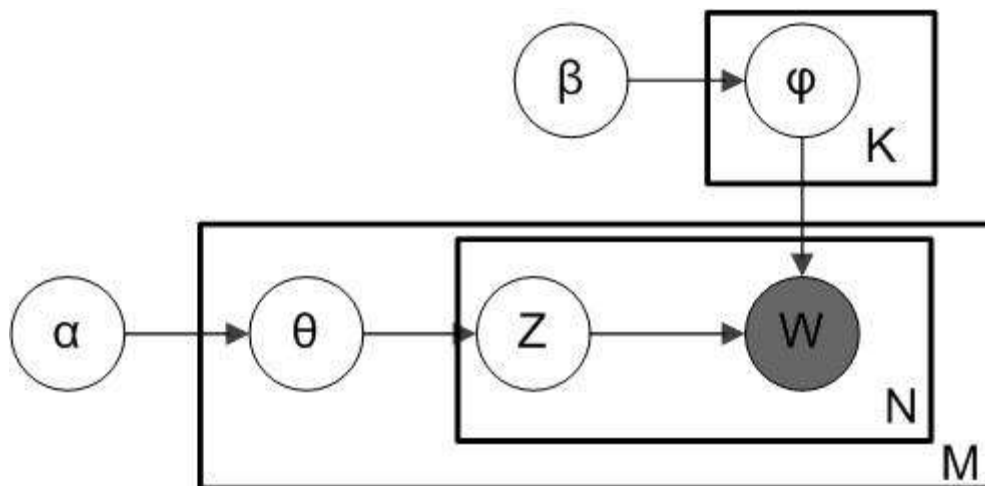


图 3-3 LDA 主题模型

表 3-7 LDA 模型参数含义

参数	含义
K	主题个数
M	文档总数
N_m	第 m 个文档的单词总数
α	每个 Topic 词的多项分布的 Dirichlet 先验参数
β	每个文档下 Topic 词的多项分布的 Dirichlet 先验参数
$Z_{m,n}$	第 m 个文档中第 n 个词的主题
$W_{m,n}$	m 个文档中的第 n 个词
θ_m	第 m 个文档下的 Topic 分布
φ_k	第 k 个 Topic 下词的分布

二、主题挖掘模型建立

（一）主题模型提取目的

经过上文的词云图和语义网络图，初步了解到用户评论中关于 TWS 耳机产品出现频率较高的词，这些信息对于消费者和商家提供的价值还不够，情感分类只解决了评论反映的情感倾向，但无法清楚知道这些高频词的具体指代范围，因此需要用 LDA 主题深入挖掘用户评论中的潜在主题。

（二）主题模型提取步骤

1. 将爬取到的文本数据导入到 Python 中，并进行一系列的数据清洗以及分词等操作。
2. 在 Python 中下载 gensim 模块，基于 gensim 模块构建语料库，并构建语料库中的词与数字一一映射的词典。使用该语料库训练 LDA 主题模型。
3. 将分类好的正、负面评论文本数据分别通过上述步骤来训练主题模型。

（三）确定主题个数

在进行主题模型之前，需要确定主题的个数。这是因为只有确定合适的主题个数，才能得到满意的分类效果，不然主题数过多会导致主题冗杂、过少会导致分类不彻底。因此本文从实际出发，对主题数为 2、3、4 分别进行模型训练，结果发现当主题数为 3 时效果最好。

三、不同品牌 TWS 耳机评论主题挖掘

（一）主题挖掘模型结果分析

对三个品牌 TWS 耳机分别提取 3 个正面评价主题和负面评价主题，每个主题用 7 个特征词汇进行表示。结果展示如下：

表 3-8 苹果 TWS 耳机评论主题提取结果

情感倾向	主题	特征词
正面	1	音质、降噪、模式、不错、清晰、特色、听歌
	2	连接、快、方便、手机、自动、好用、满意
	3	质感、精致、舒服、精细、值得、非常、简约
负面	1	客服、物流、差评、商家、久、时间、发货
	2	续航、掉电、短、电池、充电、一般、垃圾
	3	价格、贵、性价比、不值、有点、买、必须

表 3-9 华为 TWS 耳机评论主题提取结果

情感倾向	主题	特征词
正面	1	颜色、好看、颜值、银色、外观、设计、漂亮
	2	音乐、跑步、能力、买、舒适、推荐、运动
	3	客服、不错、快递、态度、物流、京东、正品
负面	1	音质、音效、降噪、一般、垃圾、杂音、听
	2	续航、能力、电量、快、不好、两天、掉电
	3	价格、有点、降价、差价、波动、比较、性价比

表 3-10 小米 TWS 耳机评论主题提取结果

情感倾向	主题	特征词
正面	1	价格、性价比、推荐、值得、不错、真香、便宜
	2	电量、续航、久、充电、三天、一次、快速
	3	体验、耐用、超级、可以、简单、感觉、不错
负面	1	音质、杂音、降噪、一点、不行、效果、不好
	2	延迟、听歌、体验、垃圾、连接、游戏、有点
	3	做工、缝隙、差、按键、掉漆、二手、退货

从苹果 TWS 耳机的正面评价主题中可以看出, 苹果 TWS 耳机的高频正面特征词主要有是产品的音质、使用感受和外观。从音质角度来说, 苹果 TWS 耳机有降噪模式和通透模式以满足消费者不同需求, 用户无论在工作学习还是休闲娱乐都可以进行相应的切换模式, 并且清晰的音质也深受消费者赞赏。从使用感受来看, 苹果旗下各个产品(手机、平板、耳机、电脑等)之间的互通互联, 当用户佩戴 TWS 耳机的时, 如果用户同时使用手机、平板或者电脑, 那么耳机也会自动迅速地切换到用户所使用的设备, 因此该项功能给予消费者相对满意的用户体验。从外观来看, 苹果 TWS 耳机的外形设计简洁干净、质感精细、大气美观的感觉深受用户的喜爱。从苹果 TWS 耳机的负面评价主题来看, 苹果 TWS 耳机的高频负面特征词主要有产品的售后服务、续航和价格。从售后服务的角度来说, 商家发货慢、客服售后服务体验差; 另外耳机续航时间不能满足一部分消费者的需求。其次由于苹果耳机的价格相比其他两个品牌要高很多, 所以部分用户会觉得价格有点高没有性价比。

华为 TWS 耳机的高频正面特征词主要是外观、应用场景和售后服务。消费者较为看重的是耳机的外观设计, 大部分都认为华为的耳机外观好看, 设计漂亮, 是一款值得购买的耳机。从应用场景来看, 消费者会在运动或者听音乐的时候使用耳机, 并且华为耳机舒适, 体验感非常棒。其次用户对购买耳机的体验比较满意, 比如物流速

度快、正品保障、客服态度礼貌热情等。在华为 TWS 耳机的负面评论中，消费者关注的是产品的音质、续航和价格。在音质上，华为耳机的音质和降噪功能与苹果相比还是有一定的差距的，这是因为苹果 TWS 耳机自带的 H1 芯片性能要优于华为的麒麟 A1 芯片，不过相对于苹果高昂的价格，华为系列耳机显得更有性价比。最后用户反应产品价格经常波动，而且幅度不小，主要原因是京东平台会不定期的进行优惠活动，使得不同时间段耳机的价格不一样。

小米 TWS 耳机的高频正面特征词分别是价格、续航、使用感受。从价格方面来说，小米系列的 TWS 耳机要比苹果和华为整体便宜不少，而且基本功能都应有尽有，只是产品细节和音质降噪功能一般，因此性价比极高，许多消费者表示“真香”。在续航上，小米 TWS 耳机自身电池容量大，并且充电速度快，用户平均充一次电可以使用三天，续航持久。在使用感受方面，用户表示耳机耐用、操作简单、总体体验不错。在小米 TWS 的负面评论中，用户关注的是音质、功能、做工三个方面。小米耳机的音质一般，在使用过程中伴随有杂音，并且降噪效果不好；在日常听歌或者玩游戏的时候会有延迟，影响用户体验；另外部分用户反应耳机容易掉漆、按键缝隙有些大等问题，商家不能因为产量过大而忽视产品的品控。