

2019 年（第六届）全国大学生统计建模大赛

基于文本挖掘的智能音箱在线评论分析

参 赛 单 位：广西科技大学

参赛者姓名：张婷婷、许聪、孟晓波

目 录

摘 要.....	I
Abstract.....	II
一、引言.....	1
(一) 研究背景.....	1
(二) 国内外研究现状.....	1
二、数据预处理.....	3
(一) 评论数据采集.....	3
(二) 数据清洗.....	3
1. 文本去重.....	3
2. 机械压缩去词.....	4
3. 评论分词处理.....	4
4. 最终数据形式.....	4
三、情感倾向性分析.....	4
(一) 构建权值词典.....	5
(二) 情感分析.....	5
四、商品评论主题分析.....	7
(一) 语义网络.....	7
1. 语义网络的建立.....	7
2. 结果与讨论.....	8
(二) 基于 LDA 模型的主题分析.....	9
1. LDA 模型介绍.....	9
2. 评论数据的 LDA 主题分析.....	11
3. 结果与讨论.....	11
五、商品属性特征提取.....	12
(一) TF-IDF 的概念和算法.....	12
(二) 高频词汇提取.....	13
1. 高频词汇总.....	13
2. 评论特征因素归纳总结.....	15
(三) TF-IDF 文本属性计算.....	15
(四) 结果与讨论.....	16
六、基于贝叶斯网络的用户满意度模型.....	16
(一) 贝叶斯网络模型介绍.....	16
(二) 建立情感词典及评分标准.....	17

(三) 贝叶斯网络构建.....	18
(四) 满意度节点条件概率.....	19
(五) 满意度影响因素.....	20
(六) 结论与分析.....	21
1. 贝叶斯网络条件概率分析.....	21
2. 特征因素重要性排名分析.....	21
七、结论.....	22
(一) 研究结论.....	22
(二) 相关建议.....	22
(三) 研究创新点.....	23
(四) 不足与展望.....	23
参考文献.....	24
附录一 其余产品语义网络图.....	26
附录二 LDA 主题分析结果.....	34
附录三 贝叶斯网络与用户满意度条件概率分布图.....	38
附录四 贝叶斯网络各节点条件概率.....	47
附录五 其余产品用户满意度影响因素图.....	66
致 谢.....	70

表格和插图清单

表 1	原始数据数量表.....	3
表 2	无效评论示意表.....	3
表 3	预处理后评论数据数量.....	4
表 4	不同产品情感分布情况.....	5
表 5	天猫精灵智能音箱正负面评价潜在主题.....	11
表 6	九款音箱主题分析结果.....	12
表 7	高频词汇统计表.....	14
表 8	九款产品高频词汇统计表.....	14
表 9	新特征因素表.....	15
表 10	不同音箱各属性 TF-IDF 值.....	15
表 11	评分标准表.....	17
表 12	满意度情感倾向及评分规则表.....	17
表 13	总体满意度节点条件概率表.....	20
表 14	各产品满意度影响因素前三位汇总表.....	21
表 15	小度智能音箱正负面评价潜在主题.....	34
表 16	小爱智能音箱正负面评价潜在主题.....	34
表 17	叮咚智能音箱正负面评价潜在主题.....	35
表 18	小艺智能音箱正负面评价潜在主题.....	35
表 19	YOYO 智能音箱正负面评价潜在主题.....	36
表 20	小雅智能音箱正负面评价潜在主题.....	36
表 21	小贝智能音箱正负面评价潜在主题.....	37
表 22	听听智能音箱正负面评价潜在主题.....	37
表 23	性价比条件概率.....	47
表 24	智能条件概率.....	47
表 25	有趣条件概率.....	47
表 26	音质条件概率.....	48
表 27	灵敏条件概率.....	48
表 28	物流条件概率.....	48
表 29	性价比条件概率.....	49
表 30	物流条件概率.....	49
表 31	灵敏条件概率.....	49
表 32	智能条件概率.....	49
表 33	有趣条件概率.....	50
表 34	资源条件概率.....	50
表 35	音质条件概率.....	50
表 36	性价比条件概率.....	50
表 37	物流条件概率.....	51

表 38	有趣条件概率.....	51
表 39	智能条件概率.....	51
表 40	资源条件概率.....	52
表 41	音质条件概率.....	52
表 42	性价比条件概率.....	52
表 43	灵敏条件概率.....	52
表 44	音质条件概率.....	53
表 45	资源条件概率.....	53
表 46	智能条件概率.....	53
表 47	物流条件概率.....	54
表 48	有趣条件概率.....	54
表 49	性价比条件概率.....	54
表 50	智能条件概率.....	54
表 51	有趣条件概率.....	55
表 52	资源条件概率.....	55
表 53	灵敏条件概率.....	55
表 54	物流条件概率.....	55
表 55	音质条件概率.....	56
表 56	性价比条件概率.....	56
表 57	物流条件概率.....	56
表 58	有趣条件概率.....	57
表 59	音质条件概率.....	57
表 60	灵敏条件概率.....	57
表 61	资源条件概率.....	58
表 62	智能条件概率.....	58
表 63	引用条件概率.....	59
表 64	音质条件概率.....	59
表 65	灵敏条件概率.....	59
表 66	性价比条件概率.....	59
表 67	资源条件概率.....	59
表 68	智能条件概率.....	59
表 69	有趣条件概率.....	60
表 70	性价比条件概率.....	60
表 71	有趣条件概率.....	60
表 72	音质条件概率.....	61
表 73	智能条件概率.....	61
表 74	资源条件概率.....	62
表 75	灵敏条件概率.....	62

表 76	物流条件概率.....	63
表 77	性价比条件概率.....	63
表 78	资源条件概率.....	63
表 79	音质条件概率.....	64
表 80	有趣条件概率.....	64
表 81	物流条件概率.....	64
表 82	灵敏条件概率.....	65
表 83	智能条件概率.....	65
图 1	文章脉络图.....	2
图 2	正面情绪分段统计图.....	6
图 3	负面情绪分段统计图.....	6
图 4	天猫精灵智能音箱正面评价网络图.....	7
图 5	天猫精灵智能音箱负面评价网络图.....	8
图 6	正面评价词云图.....	8
图 7	负面评价词云图.....	9
图 8	LDA 模型结构示意图.....	10
图 9	不同音箱各属性比较结果.....	15
图 10	贝叶斯网络建模流程图.....	18
图 11	天猫精灵贝叶斯网络图.....	18
图 12	天猫精灵贝叶斯网络条件概率分布图.....	19
图 13	天猫精灵各因素重要性排序图.....	20
图 14	小度智能音箱正面评价网络图.....	26
图 15	小度智能音箱负面评价网络图.....	26
图 16	小爱智能音箱正面评价网络图.....	27
图 17	小爱智能音箱负面评价网络图.....	27
图 18	叮咚智能音箱正面评价网络图.....	28
图 19	叮咚智能音箱负面评价网络图.....	28
图 20	小艺智能音箱正面评价网络图.....	29
图 21	小艺智能音箱负面评价网络图.....	29
图 22	YOYO 智能音箱正面评价网络图.....	30
图 23	YOYO 智能音箱负面评价网络图.....	30
图 24	小雅智能音箱正面评价网络图.....	31
图 25	小雅智能音箱负面评价网络图.....	31
图 26	小贝智能音箱正面评价网络图.....	32
图 27	小贝智能音箱负面评价网络图.....	32
图 28	听听智能音箱正面评价网络图.....	33
图 29	听听智能音箱负面评价网络图.....	33
图 30	小度智能音箱贝叶斯网络图.....	38

图 31	天猫精灵用户满意度条件概率分布图.....	38
图 32	小度智能音箱贝叶斯网络图.....	39
图 33	小度智能音箱用户满意度条件概率分布图.....	39
图 34	小爱智能音箱贝叶斯网络图.....	40
图 35	小爱智能音箱用户满意度条件概率分布图.....	40
图 36	叮咚贝叶斯网络图.....	41
图 37	叮咚智能音箱用户满意度条件概率分布图.....	41
图 38	小艺贝叶斯网络图.....	42
图 39	小艺智能音箱用户满意度条件概率分布图.....	42
图 40	YOYO 贝叶斯网络图.....	43
图 41	YOYO 智能音箱用户满意度条件概率分布图.....	43
图 42	小雅贝叶斯网络图.....	44
图 43	小雅智能音箱用户满意度条件概率分布图.....	44
图 44	小贝贝叶斯网络图.....	45
图 45	小贝智能音箱用户满意度条件概率分布图.....	45
图 46	听听贝叶斯网络图.....	46
图 47	听听智能音箱用户满意度条件概率分布图.....	46
图 48	小度智能音箱各因素重要性排序图.....	66
图 49	小爱智能音箱各因素重要性排序图.....	66
图 50	叮咚智能音箱各因素重要性排序图.....	67
图 51	小艺智能音箱各因素重要性排序图.....	67
图 52	YOYO 智能音箱各因素重要性排序图.....	68
图 53	小雅智能音箱各因素重要性排序图.....	68
图 54	小贝智能音箱各因素重要性排序图.....	69
图 55	听听智能音箱各因素重要性排序图.....	69

摘要

自 2016 年以来国内智能音箱行业技术逐渐成熟，市场上涌现了多种品牌智能音箱，消费者对这类智能家居产品产生了浓厚兴趣，然而产品质量却参差不齐，难以选择。因较少有对市面上现存产品集中分析的研究，且对智能音箱的评论数据尚未有挖掘探讨，本文旨在通过分析具有代表性的 9 款智能音箱的电商在线用户评论，如天猫精灵、360 小贝、百度小度、华为小艺、京东叮咚、荣耀 YOYO、腾讯听听 9420、喜马拉雅小雅以及小米小爱同学，从而了解市场上现存智能音箱产品的优缺点，为用户的选购及厂商的改进提供参考性的建议。

本文以京东平台作为数据来源，通过抓取在线评论文本，对评论数据进行文本去重、机械压缩去词等初步处理；接着对文本进行情感倾向性分析与语义网络分析，可视化评论特征之间的因果关系；然后以语义网络分析的正负面评价结果为基础进行 LDA 模型分析，挖掘潜在主题；建立特征词频表，通过 TF-IDF 方法提取评论数据的商品属性，发现最优智能音箱。最后采用贝叶斯网络模型探究各产品用户满意度的影响因素。

通过上述分析结果，我们推荐消费者购买小艺和小度两款智能音箱，YOYO 和叮咚智能音箱的表现稳定，也可购买。如果消费者关注性价比，也可购买小贝智能音箱，但此产品的评论两极分化，存在购买风险。天猫精灵的整体表现较差，但有趣属性比较突出，适宜儿童使用。对此现象我们向生产厂商提出如下建议：系列化产品类型，迎合多种顾客类型；自定义语音唤醒识别指令，增加方言等不同语种等语言识别库；扩大音乐库，丰富内容资源；美化外观，定制不同尺寸；提升性价比，改进充电方式。

关键词：智能音箱 情感分析 LDA 模型 TF-IDF 贝叶斯网络

Abstract

The technology of domestic smart speakers has gradually matured since 2016. Many brands have emerged and consumers have shown huge interest in them. However, the quality of different brands varies greatly, making it difficult to choose from them. Because there are few studies on the analysis of existing products on the market, and little research in exploring the online review data of smart speakers, this paper aims to evaluate the e-commerce online reviews of 9 representative smart speakers, including Tmall. Elves, Xiaobei of 360, Xiaodu of Baidu, Xiaoyi of Huawei, Dingdong of Jingdong, YOYO of Glory, TingTing9420 of Tencent, Xiaoya of Himalayan and Xiaoai of Xiaomi, in order to understand their advantages and disadvantages, providing suggestions to the consumers for purchase and the manufacturers for improvement.

The paper uses Jingdong platform as the data source to crawl online comment and performs text deduplication and mechanical compression. Then sentiment analysis and semantic network analysis are conducted to visualize the causal relationship of the comment features. Further, LDA model is applied based on the positive and negative evaluation results of semantic network analysis, to identify potential themes. Later, feature frequency table is established, the commodity attribute of the comment data is extracted by the TF-IDF method, and the optimal smart speaker is found. Lastly, the Bayesian network model is used to explore the influencing factors of user satisfaction of each product.

Through above analysis, we recommend two smart speakers to the customers: Xiaoyi and Xiaodu. Besides, YOYO and Dingdong are stable and are also advisable to purchase. If consumers are more concerned with price/performance ratio, they can also buy Xiaobei, which however may contain some risks as the comments of this product are largely either very positive or very negative. The overall performance of the Tmall Elf is poor, but it is regarded as with outstanding interesting attributes, and thus is suitable for children. Accordingly, we propose the following recommendations to manufacturers: customize the products to cater to the consumers' needs; set the voice wake-up command as self-defined, add languages such as dialects; expand the music library and enrich the content resources; improve the appearance and make the size customized; increase the cost performance and refine the charging methods.

Keywords: Smart speaker; Sentiment analysis; LDA model; TF-IDF; Bayesian network

一、引言

（一）研究背景

随着人工智能技术的飞速发展，各种人工智能产品开始进入普通家庭的生活之中，智能音箱凭借其较高的语音交互日活用户比例（约 60%），成为了各公司研发的新宠儿。自亚马逊于 2014 年发布首款智能音箱 Echo 后，不同品牌的智能音箱层出不穷。2017 年国内市场智能音箱浪潮首次爆发，根据 GfK 全国零售监测数据显示，累计至 2017 年 12 月，智能音箱市场销售量接近 165 万台^[1]。此后智能音箱的火爆程度愈演愈烈，百度、阿里、小米三家公司在 2018 年 3 月份分别发布了自家的战略产品，智能音箱的国内市场销售额达到 3.3 亿元。智能音箱的核心功能虽然类似，但不同品牌也在不断打造自己的差异化优势，如天猫、京东的智能音箱可以跟电子商务相结合，实现语音购物，并且可以在自家公司的购物网站上给予智能音箱产品最好的广告位支持；小米旗下的智能音箱可以跟小米生态链企业的各种智能家具相结合，实现语音控制智能电器；小雅音箱可以直接访问丰富的音频内容等^[2]。

新颖的产品吸引了消费者目光，而多样的功能促进了智能音箱在市场上的火爆，越来越多的消费者希望智能音箱能够使得生活智能简洁化。但在类型众多的智能音箱中，消费者们如何才能找出最适合自己的产品是市场关注的重点。如今电商推销模型盛行，消费者们往往选择网上购物，在商品评论区分享自己的使用感受与评价，方便其他消费者判断的同时也为厂商提供了改进产品的依据。因此，对用户评论数据进行分析，可以帮助生产厂家快速了解市场需求，促进产品销售并提升自身竞争力，也可以为消费者提供更为精准的商品信息。

（二）国内外研究现状

智能音箱内载系统中包括语音识别、语义理解、智能搜索和语音合成四个部分，其中语音识别负责将语音精准地转化为文本；语义理解对文本进行理解处理，结构化查询转达；智能搜索可以整合各领域优质内容，通过查询来返回最符合用户需求的结果；语音合成能够添加语音回答，实现自然顺畅的人机交互。

在智能音箱内在算法方面，大多研究主要集中于语音识别与情感交互方面。随着搜索技术的发展，基于动态规划的语音识别技术被广泛应用，例如马尔科夫模型识别、数字识别和孤立词模块匹配等。而在情感交互领域上，日本学者 Nagamachi 提出感性工学思想，把用户对目标产品所产生的感觉作为设计参考的要素，从用户情感和心理学模型角度探索语音识别交互系统^[3]。

在智能音箱用户市场研究方面，Strategy Analytics 最新发布的研究报告

《智能音箱用户调查—全球结果》^[4]指出，像 Amazon Echo 和 Google Home 这样的智能音箱已经对使用者的生活产生了很大的影响，约 65% 的人表示他们使用科技的方式有了很大的改善，还有大约 70% 的人认为智能音箱比他们预期的更实用。李莹、毛浩地等^[5]在《智能音箱产品及技术研究进展》中对我国市场上的智能音箱的种类及各产品功能、市场份额及受欢迎度、智能音箱关键技术等方面进行了研究分析，同时也对智能音箱发展中可能存在的问题以及未来发展的趋势进行了深入探讨。苏军根、林健等^[6]在《智能音箱技术与产品现状及未来发展趋势分析》中介绍了我国智能音箱市场的情况和产品的发展历程，并对目前我国主流的几款智能音箱产品的业务功能、用户体验及硬件进行了对比分析。王蓉、陈王伟^[7]在《AI 音箱发展现状及发展趋势》中以不同的角度分析了我国智能音箱的发展现状，指出了中国智能音箱市场拥有巨大的市场潜力，但是产品却存在较多问题，并对未来智能音箱发展的趋势进行了预测。

如今对智能音箱的研究大多是对产品的内在算法、语音交互系统或是发展趋势的探讨，较少对智能音箱产品的在线评论进行提取分析。本文在前人研究的基础上，通过爬虫软件抓取京东商城热度前九款智能音箱的在线评论与相关信息，对评论文本数据进行情感分析，利用语义网络和 LDA 模型挖掘其潜在主题，并通过 TF-IDF 方法量化商品属性，最后建立贝叶斯网络模型找出用户满意度的最具影响力因素，为企业改进产品提供建议，具体流程如下图 1 所示。

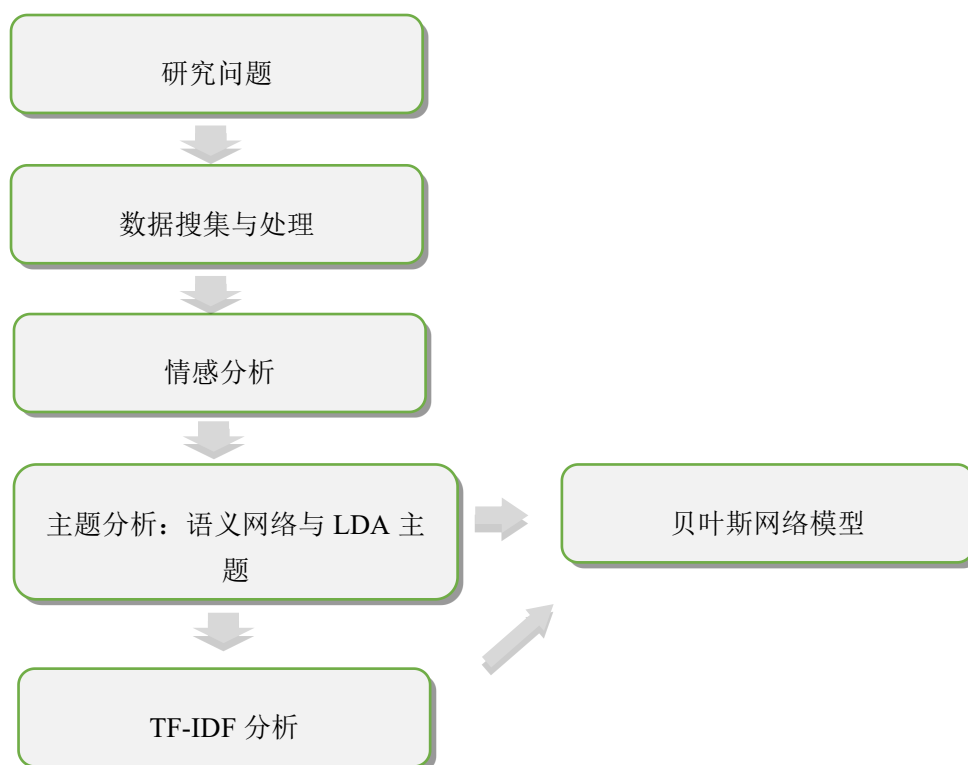


图 1 文章脉络图

二、数据预处理

（一）评论数据采集

我们选取京东平台热度前 9 名产品进行分析，分别为阿里巴巴旗下的天猫精灵、百度旗下的小度、小米旗下的小爱同学，京东旗下的叮咚 mini2 智能音箱，华为旗下的小艺音箱，荣耀旗下的 YOYO，喜马拉雅旗下的小雅，360 旗下的小贝，腾讯旗下的听听 9420 智能音箱。

首先第一步进行在线评论数据的采集，我们采用爬虫软件抓取网页数据，使用工具为八爪鱼采集器。爬取后的数据包括 ID，等级，评论时间，评论内容等信息，共 9744 条，各产品评论数据数量如表 1 所示。

表 1 原始数据数量表

产品	数量
天猫精灵	1072
小度	1245
小爱	1167
叮咚 mini2	1154
小艺	1242
YOYO	1186
小雅	1164
小贝	1202
听听 9420	1183

（二）数据清洗

1. 文本去重

电商平台往往会设置一些自动程序，若用户超过一定时长未评论，平台系统会进行自动好评。而有些用户为了追求经济上的刺激，也会随意复制他人评论，甚至发布一些无关商品的评价内容。这些大量重复的文本评论内容是没有任何参考意义的，需要进行去重处理。本文采用比较删除法进行去重，这种去重方法简单易操作，主要针对完全相同的评论数据，可以保留相似的更多数据。若两两比较后发现完全相同，就予以删除。而针对无关内容评论，目前尚未出现较好的识别方法，只能采用手工识别删除方式进行去重。被视为无效评论的部分剔除数据如表 2 所示。

表 2 无效评论示意表

用户名称	用户等级	评论时间	评论内容
j***7	PLUS 会员	2019-05-20	嘻嘻来了
行***坊	无	2019-05-12	冲着 360
c***8	PLUS 会员	2019-05-10	该用户未作评价
袁***1	PLUS 会员	2019-05-12	就这样

2. 机械压缩去词

虽然重复评论数据已被删除，但是此时的评论数据仍不能直接使用，如“音箱不错不错不错不错”；“奈斯奈斯奈斯奈斯”等句子中存在着大量重复语料、网络谐音与缩写词汇，过于冗余，需要对其进行机械压缩去词的处理。

机械压缩去词可以处理语料中的连续重复部分，我们观察客户评论习惯后发现，连续重复部分多为语句开头或结尾部分，因此判断评论中是否存在连续重复部分，可以通过建立两个放国际字符的列表完成，第一个列表和第二个列表依次放入，然后建立相关的放置判断及压缩规则，将其放入第一或第二个列表或触发压缩判断，若得出重复，则对重复的部分进行压缩去除。

3. 评论分词处理

在中文中，一句话的含义往往通过一段连续的字符进行表达，字符串之间没有明显的标识将其断开。因此我们在对文本处理时，需要进行文本分词，并按照规则重新合成词序列。目前针对中文分词的方法有很多，如KTDictSeg分词器、盘古分词器以及jieba分词器等。本文采用jieba分词工具对产品评论进行处理。经过相关测试，jieba分词精度可达到90%以上^[7]。

4. 最终数据形式

在数据处理环节，本文抓取数据后，使用比较删除法进行文本去重和机械压缩去重，使用jieba分词器对连续语料进行分词。经过预处理后，九种型号的评论数据数量共7781条，具体如表3所示。

表3 预处理后评论数据数量

产品	数量
天猫精灵	999
小度	977
小爱	845
叮咚 mini2	852
小艺	992
YOYO	945
小雅	1000
小贝	973
听听 9420	996

三、情感倾向性分析

情感倾向性分析又称为情感分析，是指对具有情感倾向文本进行分析的过程。情感倾向识别的方法有构建情感词典和使用机器学习方法进行情感分类两种方法，本文选择构建情感词典，通过ROSTContent Mining System(Version 6.0)计算软件，对处理后的评论数据进行情感倾向性分析。在判断褒贬倾向时，首先

采用优化后的情感词典，建立权值词典后再使用 ROST 工具进行分析。

（一）构建权值词典

情感倾向性分析判断语料是否为正面情感或负面情感时，通常采用情感得分的方式对每一条评论进行赋值。若情感得分大于 0，此评论为正面；若情感得分小于 0，此评论为负面；若情感得分等于 0，此评论为中立。

针对电商平台产品评论语料短小、口语化程度高和网络新词频出的特点，本文首先根据评论语料库手动抽取部分情感词汇建立初级情感词典，再通过语义相似度计算基准词与语料库中其余词汇的相似度。设置一个判断阈值，若相似度的值大于阈值，则此词汇可进入情感词典。之后再为情感词典中的词汇设置权重，权值范围为[-20, 20]之间，大于 0 为正面情感词汇，小于 0 为负面情感词汇，等于 0 为中性情感词汇。本文还为不同的程度副词设置不同的权重，并引入否定词和感叹词，否定词数量是双数时代表正面情绪，数量是单数时代表负面情绪。

（二）情感分析

通过 ROSTCM6.0 软件针对九种产品进行情感倾向性分析，分别分为正面情绪、中性情绪和负面情绪三部分，计算每部分所占比例，不同的产品情感分布结果如表 4 所示。

表 4 不同产品情感分布情况

产品	正面情绪	中性情绪	负面情绪
天猫精灵	91.22%	3.93%	4.85%
小度	96.24%	0.91%	2.84%
小爱	96.69%	0.71%	2.60%
叮咚 mini2	95.61%	0.92%	3.46%
小艺	94.76%	1.41%	3.83%
YOYO	91.75%	2.75%	5.50%
小雅	94.70%	1.80%	3.50%
小贝	81.19%	9.35%	9.46%
听听 9420	94.58%	2.31%	3.11%

由情感分析分布情况可以得知，对于这九种品牌智能音箱的评价中正面情绪占据了绝大部分，基本上分布在 90%左右，只有 360 旗下的小贝智能音箱的产品评论正面情绪比例较低，负面情绪比例在九种产品中最高。

为了更为深入地研究这九种品牌的智能音箱的具体情感分析分段情况，进而细致地了解智能音箱的用户体验，我们将正面情绪中得分 0~10 分区间设置为一般积极，11~20 分区间设置为中度积极，21 分以上区间设置为高度积极，将负

面情绪中得分-10~0分区间设置为一般消极，-20~-11分区间设置为中度消极，-20分以下区间设置为高度消极，正、负面情绪分段统计图如图2与图3所示。

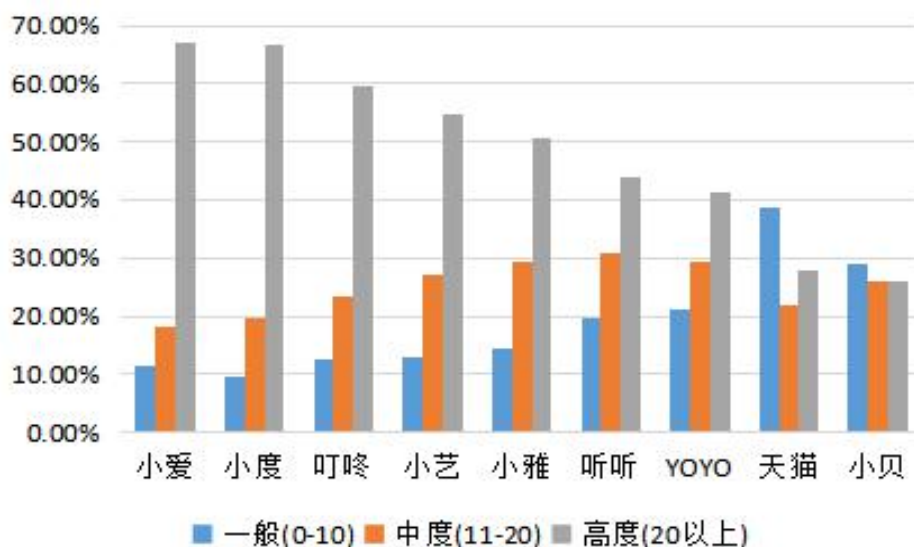


图2 正面情绪分段统计图

从总体表现而言，九种智能音箱的正面情绪评价中高度积极情绪所占的比重较高，小爱、小度、叮咚和小艺这四款智能音箱位列前茅。而针对听听、YOYO和小雅这三款智能音箱，一般积极情绪所占比重较高。

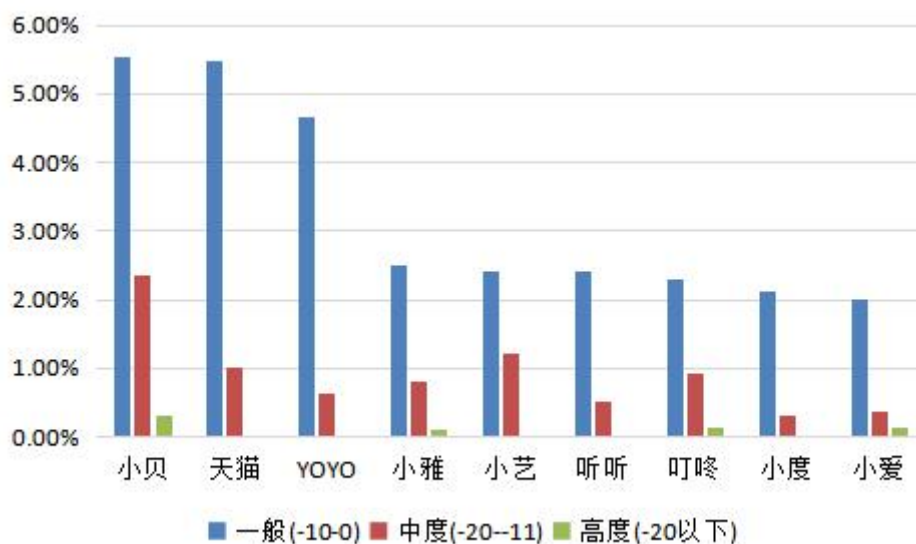


图3 负面情绪分段统计图

负面情绪分段统计图表明，天猫精灵、小贝与YOYO三款智能音箱的一般消极情绪评论比例较高，小贝、小艺和小雅三款智能音箱的中度消极情绪评论比例较高，小贝智能音箱的高度消极情绪评论比例最高。而从总体上观察结果表明，九款智能音箱的高度消极情绪评论较少，中度消极情绪评论比例中等，一般消极

情绪评论比例较高。

情感分析能够直观看出用户评论语句的正负面情感倾向，初步反映国内用户对市场对智能音箱的总体使用感受，但不能给出具体的产品反馈。因此，为了找出各评论词汇之间的语义联系，分析各产品的属性优劣，需要进一步提取精炼评论文本，挖掘评论文本中的主题词汇。

四、商品评论主题分析

（一）语义网络

语义网络是一种知识表达形式，最早是由 Quillian 在博士论文中作为一个显式心理学模型提出，后由 R. F. Simon 正式提出语义网络的概念。语义网络由节点和弧组成，节点代表可以由文字表达的事务，弧代表这些事务之间存在的语言意义上的关系，弧方向表示语言关系的因果指向^[8]。

1. 语义网络的建立

首先对正面情绪词汇和负面情绪词汇重新进行分词处理，引入词典词汇，提取出高频词汇。基于高频词汇，建立词汇之间的语义联系，其余词汇由于缺少具有代表性不予探查，最后抽取行特征构建语义网络。每个产品存在两个语义网络，分别是正面评价语义网络和负面评价语义网络，天猫精灵智能音箱正、负面评价网络如图 4 与图 5 所示，其余产品语义网络见附录一。

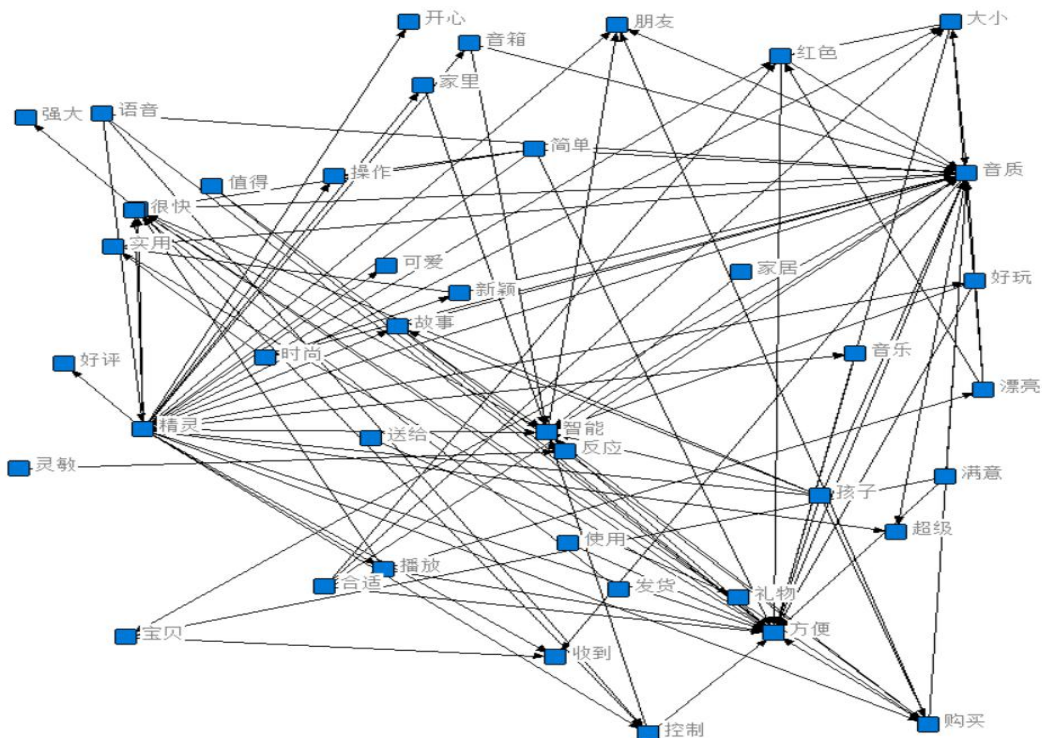


图 4 天猫精灵智能音箱正面评价网络图

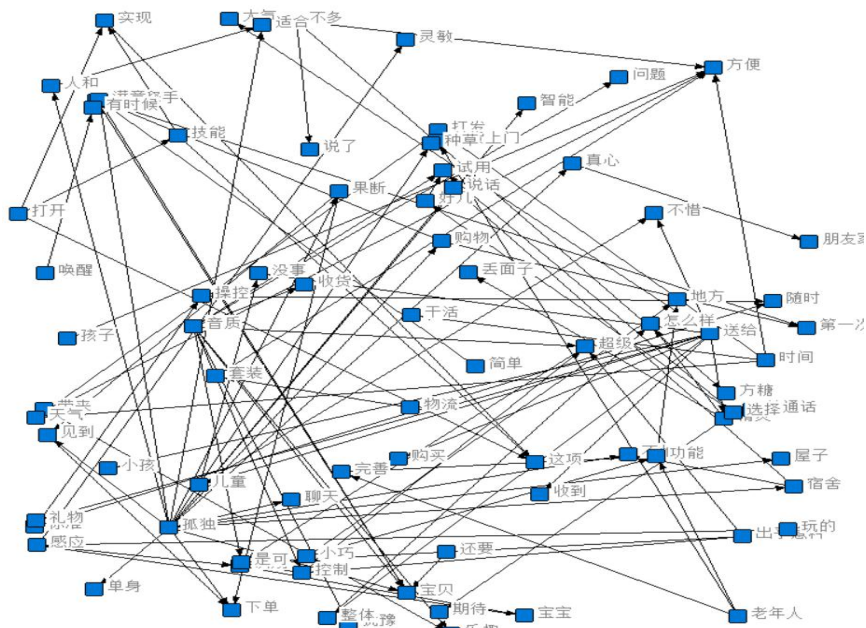


图5 天猫精灵智能音箱负面评价网络图

2. 结果与讨论

2.1 结果分析

我们根据各产品语义网络图结果，将用户对九种智能音箱的正面评价按照以下角度进行归纳：

1. 在用户的交互体验方面，正面评价语义网络中存在唤醒速度快、反应灵敏、使用方便等相似词汇关系；
2. 在产品运营方面，正面评价语义网络中存在物流速度快、发货迅速等相似词汇关系；
3. 在产品外观方面，正面评价语义网络中存在外观小巧、手感舒服等相似词汇关系；
4. 在产品功能方面，正面评价语义网络中存在功能多、智能化、音质好、音乐库庞大等相似词汇关系。

为了更直观看出正面评价的具体范围，我们将其制作为词云图，如图6所示。



图6 正面评价词云图

九种智能音箱的负面评价归纳如下：

1. 在用户交互体验方面，负面评价语义网络中存在不适宜老人使用、唤醒指令死板、存在杂音和高音问题等相似词汇关系；
2. 在产品运营方面，负面评价语义网络中存在服务态度不好、下单发货慢等相似词汇关系；
3. 在产品外观方面，负面评价语义网络中存在外貌普通、样式少等相似词汇关系；存在杂音和高音问题，音效一般；
4. 在产品功能方面，负面评价语义网络中存在缺乏戏曲、音效一般、续航方式不方便等相似词汇关系。

直观化词云图如图 7 所示。



图 7 负面评价词云图

2.2 讨论总结

如今电商购物方式十分流行，由此催生出成熟的物流快递行业，这或许是用户反映产品物流速度较快的主要原因。而智能音箱的语音对话以及内载资源库的数量也受到了用户认可，但是由于国内智能音箱市场仍在发展阶段，目标客户主要是儿童和青年群体，因此针对中老年客户的需求还未开发出相应功能。并且，随着智能音箱的大规模量化生产，音效难以保持稳定，存在杂音等问题。

语义网络把特征词各个节点连接在一起，我们可以看出不同评价词汇之间的关系，但却无法直观提取出产品的优缺点。因此，仍需进行特征主题分析，进一步细化特征词类型，找出产品各自的属性特点。

(二) 基于 LDA 模型的主题分析

1. LDA 模型介绍

LDA 模型即三层贝叶斯概率模型，包含文档 (d)、主题 (z)、词 (w) 三层结构，可以有效对文本进行建模^[9]。通过 LDA 主题模型，可以挖掘评论语料中

的潜在主题，进而分析评论的集中关注点及其相关特征词。

LDA 模型采用词袋模型 (Bag Of Words, BOW)，将每一个文档视为一个词频向量，从而把文本信息转化为易于建模的数字化信息。定义词表大小为 L ，一个 L 维向量 $(1, 0, 0, \dots, 0)$ 表示一个词汇。由 N 个词构成的评论记为 $d = (w_1, w_2, \dots, w_N)^{[10]}$ 。 M 篇评论分布着 K 个主题，记为 $z_i (i=1, 2, \dots, K)$ 。记 α 和 β 为狄利克雷函数的先验参数， θ 为主题在文档中的多项分布的函数，其服从超参数为 α 的 Dirichlet 先验分布， ϕ 为词在主题中的多项分布的参数，其服从超参数 β 的 Dirichlet 先验分布，具体示意图如图 8 所示。

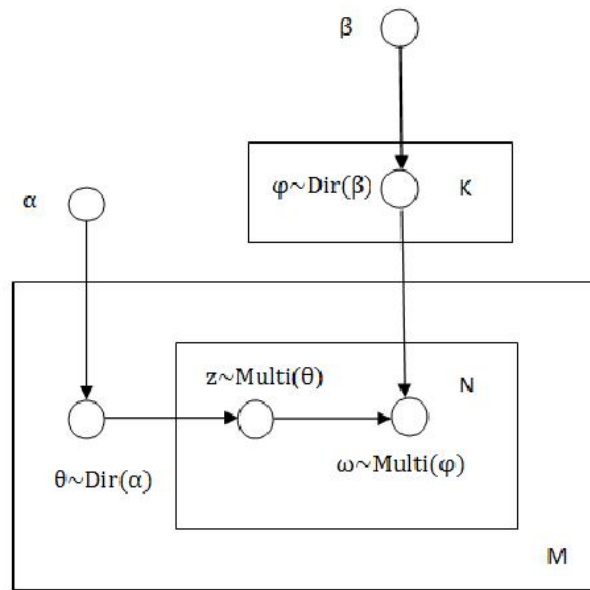


图 8 LDA 模型结构示意图

LDA 模型假定每条评论由每个主题按一定比例组合而成，组合比例服从多项分布^[11]，记为：

$$Z|\theta = \text{Multinomial}(\theta) \quad (1)$$

而每个主题用词典里面各个词语按某个比例组合而成，组合比例也服从多项分布，记为：

$$W|Z, \phi = \text{Multinomial}(\phi) \quad (2)$$

在评论 d_j 条件下生成词 w_i 的概率表示为：

$$P(w_i|d_j) = \sum_{s=1}^K P(w_i|z=s) \times P(z=s|d_j) \quad (3)$$

其中, $P(w_i|z=s)$ 表示词 w_i 属于第 s 个主题的概率; $P(z=s|d_j)$ 表示第 s 个主题在评论 d_j 中的概率^[12]。

2. 评论数据的 LDA 主题分析

首先, 本文在情感倾向性分析的基础上, 对不同情感倾向下的潜在主题分别进行挖掘, 从而得出不同情感倾向下产品的潜在主题与特征词的概率分布情况。

接着, 分别统计整个评论语料库中正负情感倾向的主题分布情况。将主题依次排列, 选择前十个主题作为评论集中的热点, 然后根据潜在主题上特征词的概率分布, 得出相对应的热点评论词。

最后, 通过 LDA 算法中的 Gibbs 抽样方法对主参数进行近似估计^[13], 预设参数 $\alpha=50/K$, $\beta=0.1$, 参数 K 通过困惑度决定, 令 $K=50$ 。

本文使用情感倾向性分析结果产生的正面和负面评价结果作为输入数据集, 删除结果中的前缀评分, 天猫精灵智能音箱潜在主题结果如表 5 所示, 其余产品结果见附录二。

表 5 天猫精灵智能音箱正负面评价潜在主题

正面			负面		
主题一	主题二	主题三	主题一	主题二	主题三
喜欢	的	很	的	很	的
精灵	了	好	非常	少	也
用	买	不错	了	库	买
了	智能	非常	天	不	宝贝
天	给	也	~	给	用
猫	都	音质	我	音乐	还
听	是	功能	是	宝宝	了
可以	挺	方便	这个	别人	就
就	个	好评	贵	起来	不好看
孩子	好	能	没	它	一般

天猫精灵的正面评价特征词挖掘结果中, 第一个主题反映天猫精灵深受儿童喜欢; 第二个主题反映其智能程度高; 第三个主题反映其音质不错, 功能强大、使用方便。负面评价特征词挖掘结果中, 第一个主题反映价格区间较高; 第二个主题反其音乐数量较少; 第三个主题反映外观不够美观。

3. 结果与讨论

3.1 结果分析

通过对九款智能音箱正面评价和负面评价的潜在主题挖掘, 我们发现 LDA 模型能较好地提取主题关键特征词, 但仍存在部分杂词提取现象。

为了方便对比九款智能音箱之间的优缺点，我们把九款智能音箱的主要主题分析结果整理如表 6 所示。

表 6 九款音箱主题分析结果

产品	正面评价	负面评价
天猫精灵	孩子喜欢、智能、音质好、功能多、使用方便	贵、音乐少、不美观
小度	音质好、智能、送货速度快	音响效果不好、音乐少、使用不方便、语音识别不智能
小爱	产品传播度高、反应灵敏、音质好、智能度高	尺寸小、音响一般、音乐少
叮咚	资源多、智能、音质不错	质量一般、音质一般、不灵敏
小艺	品牌好、智能、功能强大	音乐少、打电话不稳定、不美观、手感不好、贵
YOYO	音质好、智能、功能全面	尺寸小、存在杂音
小雅	音乐多、识别灵敏、智能	音质一般、手感一般、播放速度慢
小贝	音质好、音乐多、智能	不美观、贵、迟钝、尺寸小
听听	资源丰富、智能、使用方便、音质好、送货速度快	不美观、音响一般、迟钝

3.2 讨论总结

可以看出，正面评价中智能、音质、方便等词频频出现，说明消费者对市场上现存智能音箱的智能程度和音质效果比较满意，叮咚、小雅和小贝的数据库资源数量多，小度、小爱、小雅的反应灵敏；负面评价中不美观、贵、音乐少和尺寸等词频频出现，说明消费者对智能音箱的外观和尺寸大小仍存在需求，对价格的敏感度较高，其中 YOYO 智能音箱的负面评价中还存在杂音现象。

商品评论的主题分析是从产品评论文本出发，通过找出各特征词之间的语义因果关系，将不同角度的特征词进行属性分类，可以得出九种产品的优劣，但却缺乏产品之间的对比分析。因此，为了得出同一功能之间产品的优劣，需要深入提取产品的具体属性，将其量化转变后进行比较。

五、商品属性特征提取

（一）TF-IDF 的概念和算法

TF-IDF 是一种统计方法，用以评估一个词汇对与一个文件集合或是一个语料库中其中一份文件的重要程度。词汇在文件中出现的频率越高，重要程度越高。同时词汇的重要程度也与在语料库中出现的频率相关，词汇在语料库中出现的频率越高，重要程度就越小^[14]。

TF-IDF 中的 TF 是指词频 (Term Frequency)，是某一个给定词汇在该文件中出现的次数。词频通常会进行归一化，即词频除以文章总词数，以防它偏向长文件。具体公式如下：

$$TF_w = \frac{n_{i,j}}{\sum_k n_{k,j}} \quad (6)$$

其中 $n_{i,j}$ 是该词汇 t_i 在文件 d_j 中出现的次数， $\sum_k n_{k,j}$ 表示在文件 d_j 中所有词出现次数之和。

但是一些通用的词语对主题并没有太大的帮助，有可能出现频率较少的词汇 t_i 才能代表文章的主题，此时使用 TF 值并不合适。因此引入 IDF，强调把集合中的项作为一个整体的意义，平衡权重的设置。

IDF 是逆向文件概率 (Inverse Document Frequency)，主要思想是如果包含词汇的文件越少，IDF 越大，说明词汇具有很好的类别区分能力。某一特定词汇的 IDF，可以由总文件数目除以包含该词汇文件的数目，再把得到的商取对数^[15]。具体公式如下：

$$IDF_i = \log \frac{|D|}{|\{j: t_i \in d_j\}|} \quad (7)$$

其中 $|D|$ 表示语料库中文件个数， $|\{j: t_i \in d_j\}|$ 表示包含词汇的文件个数，如果该词语不在语料库时，会导致被除数为 0，因此本文采用如下公式计算：

$$IDF_i = \log \frac{|D|}{|\{j: t_i \in d_j\}| + 1} \quad (8)$$

然后，TF-IDF 的值为：

$$TF - IDF_{i,j} = TF_{i,j} \times IDF_i \quad (9)$$

(二) 高频词汇提取

1. 高频词汇总

高频词汇是指在评论数据中出现次数较高，能够代表用户对商品质量或服务态度评价的词汇，对商品特征属性提取具有参考价值^[16]。目前各类电商平台对商品评价都有高频词汇展示，方便消费者搜索，且官方数据比较准确，因此本文采用官方提供的高频词汇，统计如表 7 所示。

表 7 高频词汇统计表

商品名称	高频词汇展示及相应频数
天猫精灵	速度快 (354) 流畅 (312) 方便 (306) 性价比高 (265) 音乐多 (175) 智能度高 (79)
小度	有趣好玩 (3152) 功能强大 (1895) 反应灵敏 (1348) 十分好用 (1329) 智能度高 (1059) 音质好 (896) 灵敏度高 (375) 性价比高 (301)
小爱	简单方便 (2773) 十分有趣 (2518) 功能强大 (1765) 十分好用 (1482) 反应灵敏 (1104) 识别功能强 (650) 有趣好玩 (380) 音质极佳 (375)
叮咚	简单方便 (2947) 功能强大 (1793) 反应灵敏 (1216) 识别度高 (648)
小艺	简单方便 (1112) 反应灵敏 (896) 功能强大 (347) 性价比高 (268)
YOYO	简单方便 (776) 反应灵敏 (569) 识别快 (347) 小巧便捷 (210) 性价比高 (153)
小雅	质量不错 (310) 声音悦耳 (27) 标新立异 (50) 简单方便 (169) 好看 (50)
小贝	反应灵敏 (446) 声音洪亮 (417) 功能强大 (318) 音质优良 (200) 智能度高 (156) 音乐丰富 (133)
听听	智能化高 (698) 音质俱佳 (410) 功能强大 (320) 资源多 (248)

上表高频词汇后面的数字表示此词汇在评论中出现的次数。在对高频词加工处理，去除情感修辞部分后，提取其中的主成分即为商品或服务的特征属性。因汉语中对同一属性的描述有多种方式，因此保留主成分的近义词和同义词。在爬取的商品评论数据中搜索相近词汇并统计数量加以汇总，结果如表 8 所示。

表 8 九款产品高频词汇统计表

关键词	天猫	小度	小爱	叮咚	小艺	YOYO	小雅	小贝	听听
音质	1656	5200	3134	2789	6512	2978	2334	6245	5167
音响	446	1556	945	1389	1423	634	745	1245	978
性价比	300	1167	212	512	167	156	123	545	145
音乐	367	2467	1656	1156	1245	489	367	1256	1489
方便	1500	2512	2323	2023	1967	1456	1578	656	1800
功能	1389	2800	2134	2578	2823	1989	856	934	1267
智能	1356	3823	2800	3534	3678	2134	1900	2845	2189
好玩	1156	889	1078	1123	556	567	345	1012	289
价格	662	1912	934	1567	856	856	845	1512	1000
值	567	878	178	789	178	678	89	645	156
漂亮	500	478	289	678	512	334	445	56	400
反应	478	1367	900	912	734	612	678	323	523
很快	423	1478	912	1000	1089	789	845	434	867
物美价廉	123	223	67	189	1089	112	12	78	34
灵敏	112	156	712	1012	134	612	534	489	500
有趣	89	212	212	112	167	612	123	23	67
资源	45	45	145	512	412	156	612	656	1378
物流	334	867	1023	767	1078	745	823	500	789
快递	119	856	578	478	434	412	434	345	478

根据汇总结果，频数最大的6个关键词为音质、智能、功能、方便、音乐和价格，这说明消费者对智能音箱的音质清晰度、智能程度、功能种类、音乐库数量和价格属性具有较高的关注度。

2. 评论特征因素归纳总结

由于上一步中关键词数量过多且存在相近词，因此参考商品宣传要点及京东平台提供的官方高频因素，将相近词及对同一属性刻画的关键词合并成新的特征因素，整理如表9所示。

表9 新特征因素表

高频关键词汇总	新特征因素
性价比、价格、值、物美价廉	性价比
好玩、有趣	有趣
反应、很快、灵敏	灵敏
音乐、资源	资源
音质、音响	音质
智能、方便、功能	智能
物流、快递	物流

(三) TF-IDF 文本属性计算

本文以新的特征因素为切入点，计算不同商品各属性的TF-IDF值，结果如表10所示（红色为排名第一，蓝色为排名倒数第一）。

表10 不同音箱各属性TF-IDF值

关键词	天猫	小度	小爱	叮咚	小艺	YOYO	小雅	小贝	听听
性价比	0.018	0.018	0.009	0.016	0.013	0.017	0.009	0.022	0.010
有趣	0.013	0.005	0.008	0.007	0.004	0.011	0.004	0.008	0.003
灵敏	0.011	0.013	0.015	0.016	0.011	0.019	0.018	0.010	0.014
资源	0.004	0.011	0.011	0.009	0.009	0.006	0.008	0.015	0.021
音质	0.023	0.029	0.025	0.023	0.044	0.034	0.027	0.059	0.044
智能	0.018	0.056	0.039	0.045	0.079	0.048	0.034	0.032	0.038
物流	0.005	0.007	0.010	0.007	0.008	0.011	0.011	0.007	0.009

将其可视化后，按有趣排序如图9所示。

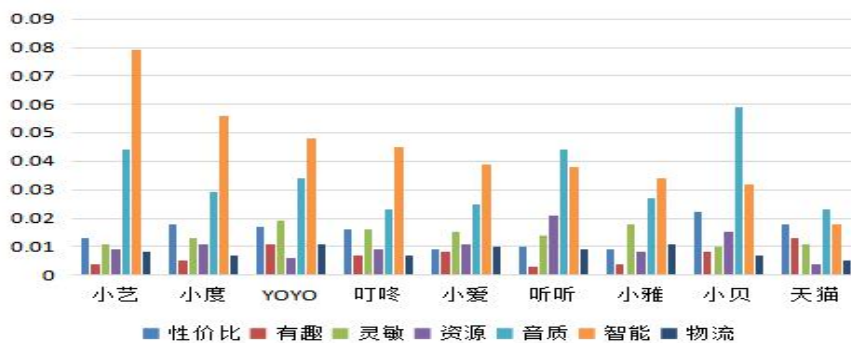


图9 不同音箱各属性比较结果

（四）结果与讨论

综合而言，七个属性中，小艺智能音箱的智能属性排名第一，音质属性排名第二，其余属性处于中等水平；小度智能音箱的性价比属性和智能属性排名第二，其余属性处于中等水平；YOYO 智能音箱在灵敏属性上排名第一，有趣属性上排名第二，其余属性处于中等水平；叮咚智能音箱的智能属性和灵敏属性均排名第三，其余水平处于中等水平。这四款音箱的总体表现稳定，值得购买。

小贝智能音箱虽在性价比和音质属性上均排名第一，但其余属性表现较差，整体表现两极分化，其原因可能在于厂商为打出知名度，追求性价比，采用的内载零件质量一般。天猫智能音箱的整体表现水平较低，各属性排名均处于倒数区间内，其原因可能在于天猫智能音箱作为首批进入国内市场的 AI 音箱，各方面配置较为落后，未跟随潮流变化及时改进产品。因此，消费者在购入这两款音箱时应谨慎考虑。

TF-IDF 算法找出了各产品在国内市场流通的主要特征因素，并根据特征因素比较了各产品的优劣程度，得出排名。但这些产品的“明星”功能是否得到用户认可？用户对不同产品的满意度主要受到哪些因素影响？针对这些问题，我们需要通过定量模型进行深度挖掘。

六、基于贝叶斯网络的用户满意度模型

在前面的分析中，本文通过文本挖掘的方式对九款智能音箱的评论数据进行了深入的探讨，初步得出了在性价比、有趣、灵敏、资源、音质、智能和物流七个特征因素上各产品的表现情况。为了找出影响九款智能音箱用户满意度的影响因素，现针对七个特征因素建立贝叶斯网络模型，从量化角度深度挖掘。

（一）贝叶斯网络模型介绍

贝叶斯网络又称为信度网，是有一个有向无环图和条件概率组成的网络结构图。它主要通过有向无环图（DAG）将变量间复杂的关系以可视化方式呈现出来，并以条件概率和后验概率的形式支持用户进行双向推理。DAG 作为贝叶斯网络的核心内容，由节点和有向弧两部分构成。其中，节点代表要研究的变量，有向弧代表变量间的因果关系，由父节点指向子节点。

贝叶斯网络以贝叶斯公式为数学原理，通过联合概率分布对网络中全体节点的关联程度进行了定义。对于事件 A，设影响其发生的所有事件集合为 $V=(V_1, V_2 \dots V_n)$ ，则相关的贝叶斯公式为：

$$P(V_i|A) = \frac{P(A|V_i)P(V_i)}{P(A)} = \frac{P(A|V_i)P(V_i)}{\sum_{j=1}^n P(A|V_j)P(V_j)} \quad (10)$$

其中， $P(V_i|A)$ 为后验概率， $P(V_i)$ 为先验概率， $P(A|V_i)$ 为时间 V_i 发生条件下事件 A 发生的概率，即条件概率^[17]。

对于已经确定的贝叶斯网络模型而言，其节点的联合概率分布可表示为：

$$P(V) = \prod_{i=1}^n P(V_i | V_{pa(V_i)}) \quad (11)$$

其中， $V_{pa(V_i)}$ 表示节点 V_i 的所有父节点。

(二) 建立情感词典及评分标准

由 TF-IDF 方法中得到的七个变量表达了智能音箱用户的情感关注点，但这些数据并未结构化，需要对因素量化评分。由于京东商城评论系统中，用户需要对整个消费体验进行五分制打分，因此我们同样采用五分制打分制度，并参考李克特量表制定评分标准如表 11 所示。

表 11 评分标准表

评分标准	分值
非常满意	5
比较满意	4
一般满意	3
比较不满	2
非常不满	1

中文表达中习惯用非常、特别等用来表示情感上的起伏，这些修饰词在一定程度上表达了消费者对商品的满意程度。因此本文罗列出各特征因素的修饰词来建立用户满意度情感词典及打分标准，如表 12 所示。

表 12 满意度情感倾向及评分规则表

变量/得分	5	4	3	2	1
性价比	便宜、值、物美价廉	划算、比较便宜	还行、一般、可以	有点贵、不太划算	贵、不值
有趣	好玩、有趣、有意思	娱乐、消遣	一般、能玩	不太好玩	无聊
灵敏	反应快、速度快	省时、灵敏	还行、一般、可以	有点慢、迟钝	慢、等待
资源	丰富、特别多	多、够用	一般、凑合	有点少	没有、少
音质	清晰、响亮	特别清、好听	能听清	模糊、不够大	存在杂音、声音小
智能	非常方便、好用、功能多	挺方便、合适	够用、功能齐全	不太适合、不够用	很不方便
物流	服务好、很快	速度可以	一般	慢、差	太慢

根据上述评分表,制定相应的评分规则,根据不同的特征因素进行量化评分。评论中出现的与评分规则相近的同义词也按照规则进行转换。对于评论中未出现的特征因素,本文统一标识为3分,即情感色彩为一般。

选用京东评论系统中的星级评分作为用户的满意度数据,具有一定可靠性。经过上述规则后得出的数据将作为贝叶斯网络的基础数据进行分析。

(三) 贝叶斯网络构建

本文使用的模型构建工具为Clementine软件,Clementine拥有多种数据挖掘算法,是一款功能强大的数据挖掘软件。具体建模过程如图10所示。

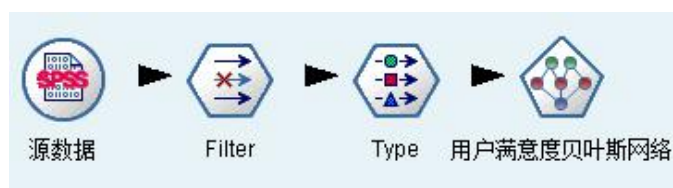


图10 贝叶斯网络建模流程图

首先把数据以数据流的形式导入软件,建立过滤节点与类型节点。将用户满意度作为输出目标,其他七个特征因素作为输入因素,变量类型为set,结构类型为TAN,参数学习方法为最大似然法。

对处理后的数据导入Clementine软件中,通过数据的筛选与类型选择,构建的天猫精灵智能音箱贝叶斯网络模型和条件概率图举例见图11与图12,其余产品结果见附录三。

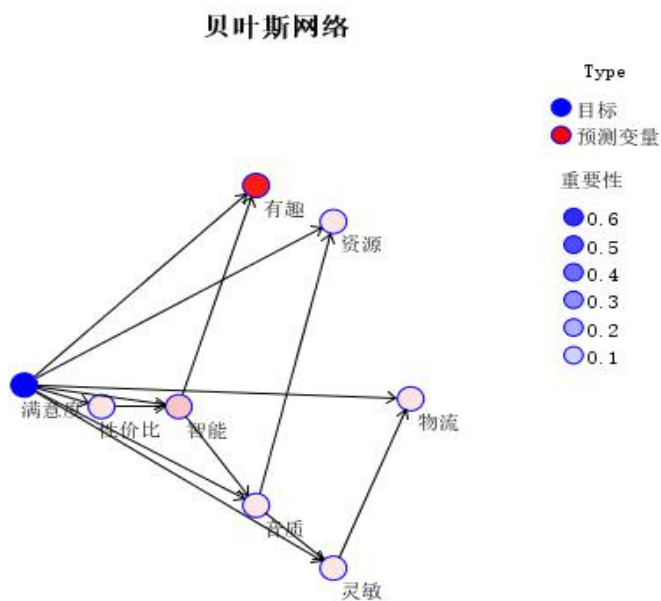


图11 天猫精灵贝叶斯网络图

贝叶斯网络

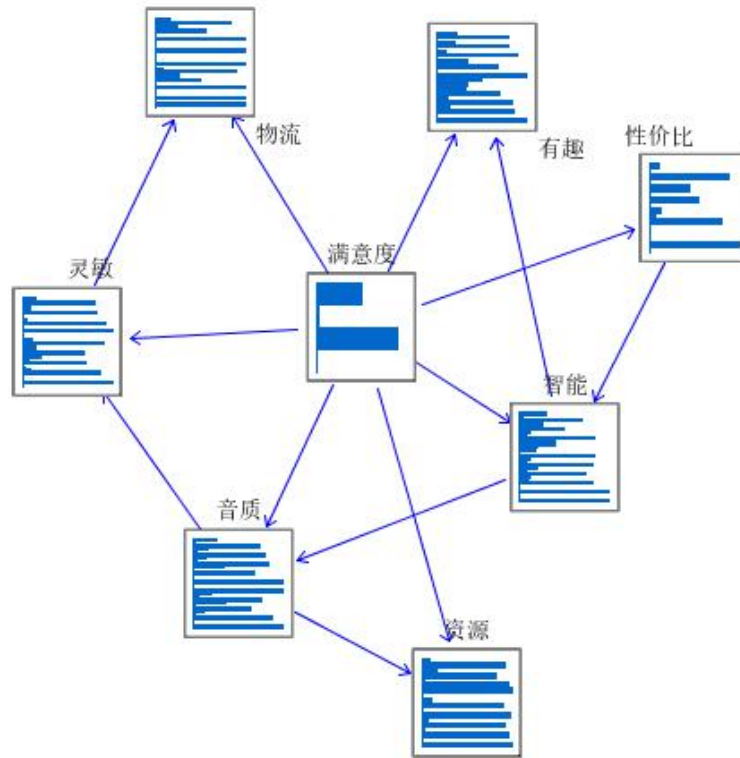


图 12 天猫精灵贝叶斯网络条件概率分布图

通过条件概率分布图可以直观地看出各因素之间的联系。满意度是其他特征因素的父节点，因素之间的箭头指向既说明了因素之间的父子节点关系，也说明了各特征因素之间的联系。。

在贝叶斯网络结构中，红色圆圈代表预测变量，即输入变量；蓝色圆圈代表目标变量，即用户满意度变量；圆圈颜色深浅代表着因素重要程度的高低，颜色越深代表重要性越高。在结构图中，满意度与其他特征因素通过箭头连接，既指明了父节点关系，也说明各因素之间的相互关系。

由贝叶斯网络图可以看出，九款产品的满意度作为目标变量，与七个特征因素均存在关系，七个特征因素之间也存在关联关系。

（四）满意度节点条件概率

满意度节点的条件概率代表着消费者对该款商品的总体满意度分值的分布情况，5分代表非常满意，4分代表比较满意，3分代表一般满意，2分代表不太满意，1分代表不满意。九款智能音箱的各总体满意度节点条件概率表如表 13 所示，具体网络各节点概率见附录四。

表 13 总体满意度节点条件概率表

满意度	1	2	3	4	5
天猫精灵	0	0.003	0.633	0.01	0.351
小度	0	0	0.376	0.009	0.614
小爱	0	0	0.201	0.280	0.518
叮咚	0	0.001	0	0.562	0.436
小艺	0	0	0.399	0.014	0.586
YOYO	0.005	0.009	0.332	0.043	0.609
小雅	0	0	0	0.009	0.99
小贝	0.022	0.019	0.033	0.046	0.877
听听	0	0	0.291	0.275	0.433

天猫精灵最高的总体满意度条件概率对应分数为 3 分，说明用户满意度一般；叮咚最高的总体满意度条件概率对应分数为 4 分，说明用户满意度良好；小度、小爱、小艺、YOYO、小雅、小贝和听听中最高概率对应分值为 5 分，说明用户满意度比较高。

（五）满意度影响因素

为了清晰看出各因素对满意度的影响力大小，将各因素按照重要性排序后，天猫精灵智能音箱重要性排序图见图 13，其余产品排序图见附录五。

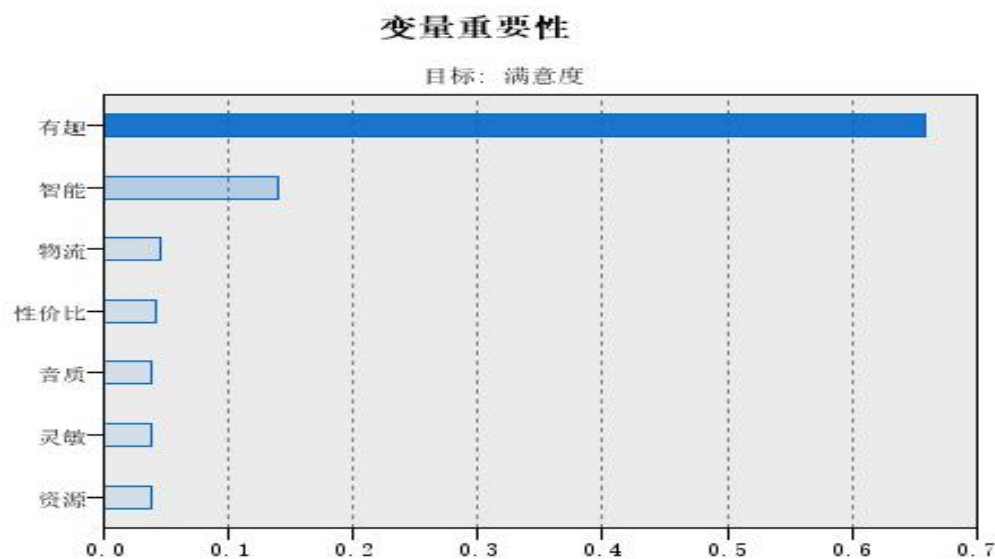


图 13 天猫精灵各因素重要性排序图

观察图列结果，我们发现天猫精灵的有趣性对用户满意度影响最大；小度的灵敏性对用户满意度影响最大；小爱的物流速度对用户满意度影响最大；叮咚的有趣性对用户满意度影响最大；小艺的智能度对用户满意度影响最大；YOYO 的物流速度对用户满意度影响最大；小雅的物流速度对用户满意度影响最大；小贝的音质对用户满意度影响最大；听听的音质对用户满意度影响最大。

（六）结论与分析

1. 贝叶斯网络条件概率分析

在分别以七个特征因素构建而成的九款智能音箱的用户满意度影响因素的贝叶斯网络模型中，分别得到各子节点的父节点，同时可以计算出子节点事件发生的条件概率。

从总体概率分布上来看，在代表父节点事件发生的情况下各子节点发生的条件概率分布具有一定的规律性。其中得分为5分的概率较大，其次为4分，3分。各子节点得分为2分、1分的概率基本很小甚至概率为0，这说明用户对于购买智能音箱的各项特征因素基本上处于满意状态，有些用户倾向于高满意度的优秀评价。

2. 特征因素重要性排名分析

通过建立的贝叶斯网络模型，得出各特征因素的重要性大小及排名，从各特征因素的重要性排名中可以清楚的看出对该款商品满意度影响最大的特征因素。

为了直观看出每种产品对满意度影响最大的特征因素，本文将上述分析结果汇编成表，如表14所示。

表14 各产品满意度影响因素前三位汇总表

产品名称	影响度第一位	影响度第二位	影响度第三位
天猫精灵	有趣	智能	物流
小度	灵敏	性价比	有趣
小爱	物流	性价比	灵敏
叮咚	有趣	音质	资源
小艺	智能	音质	性价比
YOYO	物流	智能	灵敏
小雅	物流	音质	灵敏
小贝	音质	智能	物流
听听	音质	有趣	资源

对于天猫精灵智能音箱，用户主要偏向于其有趣、智能化程度高、物流快；对于小度智能音箱，用户主要偏向于其灵敏性高、性价比高和有趣；对于小爱智能音箱，用户主要偏向于其物流快、性价比高和灵敏性高。

对于叮咚智能音箱，用户主要偏向于其有趣、音质好和资源丰富；对于小艺智能音箱，用户主要偏向于其智能化程度高、音质好和性价比高；对于YOYO智能音箱，用户主要偏向于其物流快、智能化程度高和灵敏性高。

对于小雅智能音箱用户，用户主要偏向于其物流快、音质清晰和智能度高；对于小贝智能音箱，用户主要偏向于其音质好、智能化程度高和物流快；对于听听智能音箱，用户主要偏向于其音质好、有趣和资源丰富。

七、结论

（一）研究结论

本文在对智能音箱产品在线评论的文本挖掘过程中,进行了情感分析、语义网络分析和 LDA 模型主题分析,并在原有传统文本挖掘方法的基础上,提出了基于贝叶斯网络的用户满意度模型,有效找出了影响用户满意度的主要因素。

总体考虑,本文认为小艺智能音箱的综合表现比较优秀,若消费者不介意物流因素,也可考虑小度智能音箱。YOYO 和叮咚表现稳定,消费者也可以考虑购买。如若消费者希望购买到物美价廉的产品,小贝智能音箱可供参考,但存在一定购买风险性;天猫精灵整体表现较差,但受到孩子的喜爱,有趣程度高,如若儿童使用是购买需求,可谨慎考虑。

（二）相关建议

根据上文分析结果与讨论分析,我们向相关企业与生产厂家提出如下几个角度的建议:

1. 在产品规划方面,可系列化产品类型,迎合多种顾客类型。在情感分析与 LDA 模型分析结果中我们发现,智能音箱的用户群体包括儿童、上班族、初高中学生、大学生和老年群体,但目前的产品功能面向群体还是以青壮年为主。我们建议品牌方可以打造不同类型的功能型音箱,如儿童智能音箱、学生知识型智能音箱等。厂家也可打造兼容性音箱,同时满足不同用户群体需求;

2. 在产品交互体验方面,可自定义语音唤醒识别指令,增加方言等不同语种等语言识别库。我们在评论中发现,目前产品的唤醒指令多为产品名字,且识别语言为普通话。因此我们建议可增加自定义唤醒指令功能,开发常见方言识别功能,并且可以面向国际市场添加英语等丰富语种库;厂家也应完善电池功能,如今市面上的智能音箱大多采用插入电源方式,无内置电池,无法脱离电插座使用产品,我们建议可改进充电方式,开发适合移动的充电功能,如:usb 充电,内置锂电池,无线充电等;

3. 在产品核心功能方面,可扩大音乐库,丰富内容资源。我们在产品评论中发现,用户在使用智能音箱时有不同的需求,但目前产品的音乐库过于单一。针对不同的用户群体,我们建议可增加儿歌库、诗歌库、视频语音资源、百科知识库和传统戏曲资源等;

4. 在产品模型设计方面,可美化外观,根据使用场景定制不同尺寸。LDA 模型的分析结果可以表明,产品的负面评价主题中外观和尺寸出现频率过高,用户

希望购买的产品足够美观；而用户不仅在室内使用智能音箱，车厢、户外也有同样的需求。因此我们建议品牌方可以改进音箱外观设计，针对不同地点设计不同尺寸的产品，方便用户选择；

5. 在产品运营方面，可多平台多渠道宣传产品，增加专用服务人员。在研究过程中，我们发现音箱的主要宣传平台多为电视广告及网络营销等线上渠道，在实体超市及门店等线下渠道缺乏营销力度。因此我们建议厂家宣传部门可在大型商场、娱乐场所等地点举办线下促销活动进行产品推广；在平台服务上，可采用专用客服为客户排忧解难，加强态度问题管理。

（三）研究创新点

过往研究多在于对智能音箱中的自然语言处理的研讨或是对少数热门产品的功能对比，本文将文本挖掘技术应用在市场上现存的九款智能音箱产品的在线评论中，具体有以下几点创新：

1. 数据创新。以往的研究数据一般为调查问卷结果或是公司内部数据，市场信服度较低。本研究在用户在线评论数据的基础上，不断进行文本挖掘，数据更贴合市场，更能代表用户感受；

2. 方法创新。文本挖掘的主要研究办法多为情感分析和 LDA 模型探讨，本文在传统方法的基础上，增加贝叶斯网络研究产品的用户满意度影响因素。

（四）不足与展望

本研究从京东商城在线评论出发，基于用户评论数据进行情感分析、潜在主题挖掘、商品特征提取与用户满意度研究，由于作者能力与篇幅限制，本研究仍存在一些不足：

1. 数据量有限。本文爬取的原始数据共 9744 条，经过一系列处理后仅剩 7781 条，且数据来源于一个电商平台，仅代表整个中国市场的一小部分。

2. 模型不够完善。本文在建立情感词典时，未考虑到部分词频较低的特征词，这就导致了模型的准确性与全面性有所不足。并且未进行多种网络模型的对比探讨。

针对以上不足，在以后的研究中可适当增加样本量，提高研究的深度。可根据数据进行贝叶斯网络与人工神经网络的模型准确度对比，找出准确度最高的模型。同时，本研究中的方法还适用于市场上其余商品的探讨，如手机、电脑等。

参考文献

引文文献

- [1] 姚晓霖. 智能音箱:怎么我快成网红了—十款智能音箱大搜查[J]. 家庭影院技术, 2017年: 23-25
- [2] 王颢毅. 基于人工智能技术的智能音箱发展现状与趋势探究[J]. 通讯世界, 2018年:231-232
- [3] 孙梦楠. 智能音箱中自然语言语义理解算法的研究[D]. 湖南大学, 2018: 15-20
- [4] 孙永杰. 智能音箱大战背后:语音识别技术应用仍存挑战[J]. 通信世界, 2017年:17-19
- [5] 李莹, 毛浩地. 智能音箱产品及技术研究进展[J]. 信息与电脑(理论版), 2019年: 144-150
- [6] 苏军根, 林健, 洪博宇. 智能音箱技术与产品现状及未来发展趋势分析[J]. 广东通信技术, 2018年:9-16
- [7] 王蓉, 陈王伟. AI 音箱发展现状及发展趋势[J]. 中国集体经济, 2018年:162-163
- [8] 胡发刚, 谢军. 基于文本挖掘的电商企业评论打分机制有效性验证[J]. 吉林工商学院学报, 2015年:43-46
- [9] 刘磊. 概念内涵属性计算研究[D]. 上海交通大学, 2011年:47-50
- [10] 杨瑞欣. 电商空调产品的评论数据情感分析[D]. 山西大学, 2017年:31-40
- [11] 施乾坤. 基于 LDA 模型的文本主题挖掘和文本静态可视化的研究[D]. 广西大学, 2013年:26-30
- [12] 钟荣飞. 基于主题模型的网络舆情监控系统设计[J]. 电脑知识与技术, 2015年:90-93
- [13] 贾西平, 彭宏, 郑启伦, 基于主题的文档检索模型[J]. 华南理工大学学报(自然科学版), 2008年: 21-23
- [14] Berg B A . Markov Chain Monte Carlo Simulations and Their Statistical Analysis[M]. WORLD SCIENTIFIC, 2004:78-90.
- [15] Salton G , Buckley C . Term-weighting approaches in automatic text retrieval[J].

Information Processing & Management, 1988:513-523.

- [16]刘露. 浅析 TF-IDF 在企业组织风险中的研究[J]. 硅谷, 2011 年: 31-40
- [17]慕春棣, Cn T E , 戴剑彬, et al. 用于数据挖掘的贝叶斯网络[J]. 软件学报, 2000 年:660-666
- [18]徐自跃. 基于京东手机在线评论的客户满意度影响因素研究[D]. 安徽理工大学, 2018: 55-70
- [19]李雷. 农产品电子商务顾客满意度和忠诚度的形成机制研究——基于感知价值和服务质量的视角 [J] . 世界农业, 2017 年: 97-103
- [20]Mouwen, Arnoud. Drivers of customer satisfaction with public transport services[J]. Transportation Research Part A: Policy and Practice, 2015:1-20

阅读型文献

- [21]王伟. 中文在线评论的产品特征与观点识别: 跨领域的比较研究[J]. 管理工程学报, 2017: 73-75
- [22]曹高辉. 消费者持续参与在线评论意愿实证研究[J]. 管理评论, 2017:148-158
- [23]于陶静. 基于在线评论特征的消费者购买决策研究[D]. 哈尔滨理工大学, 2017:65-70
- [24]Chevalier J, Mayzlin D. The effect of word of mouth online: online book reviews[J]. Journal of Marketing Research. 2006:348-354
- [25]Ganu G, Kakodkar Y, Marian A. Improving the Quality of Predictions Using Textual Information in Online User Reviews[J]. Information System, 2012:1-15
- [26]李良强. 基于内容挖掘的在线用户评论时间特征及其影响研究[D]. 电子科技大学, 2016:46-50
- [27]张紫琼. 在线中文评论情感分类问题研究[D]. 合肥工业大学, 2010: 66-70
- [28]马俊伟. 渐进式滑坡多场信息演化特征与数据挖掘研究[D]. 中国地质大学, 2016:57-59
- [29]付斌. 基于 BP 神经网络的数据挖掘方法在需求预测中的应用研究[J]. 情报科学, 2017:132-135
- [30]张倚天. 基于商品特征挖掘的在线评论有用性分类研究[D]. 大连理工大学, 2016:76-80

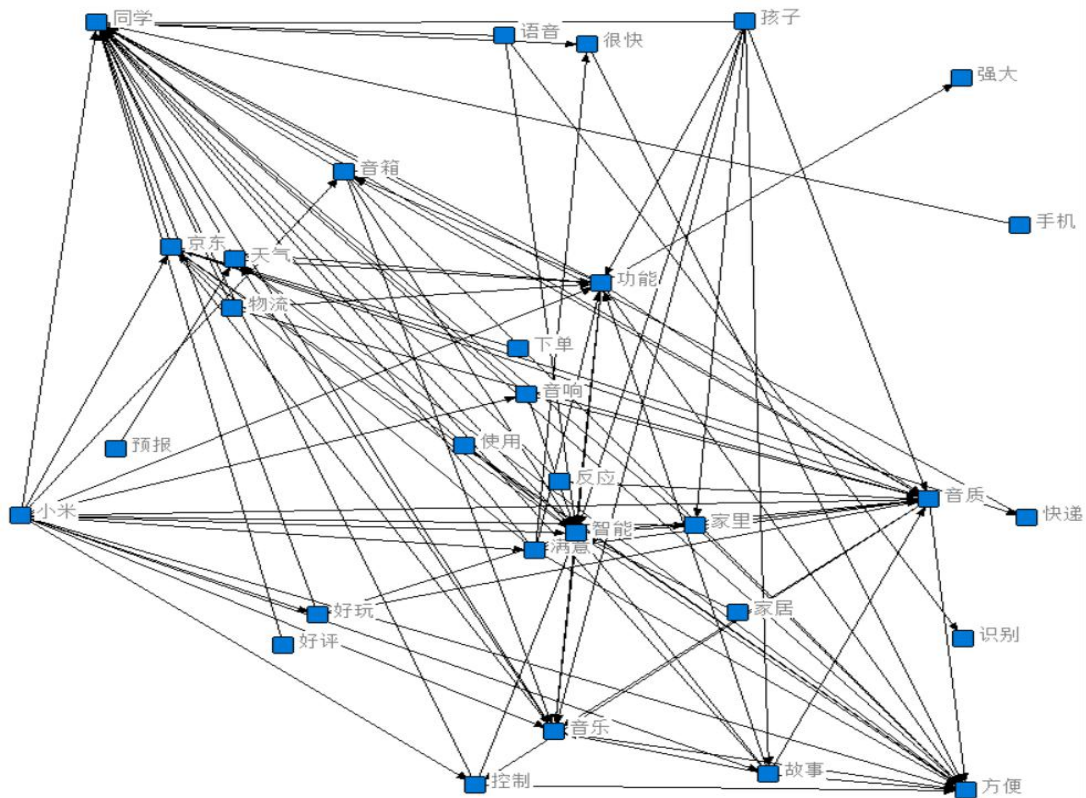


图 16 小爱智能音箱正面评价网络图

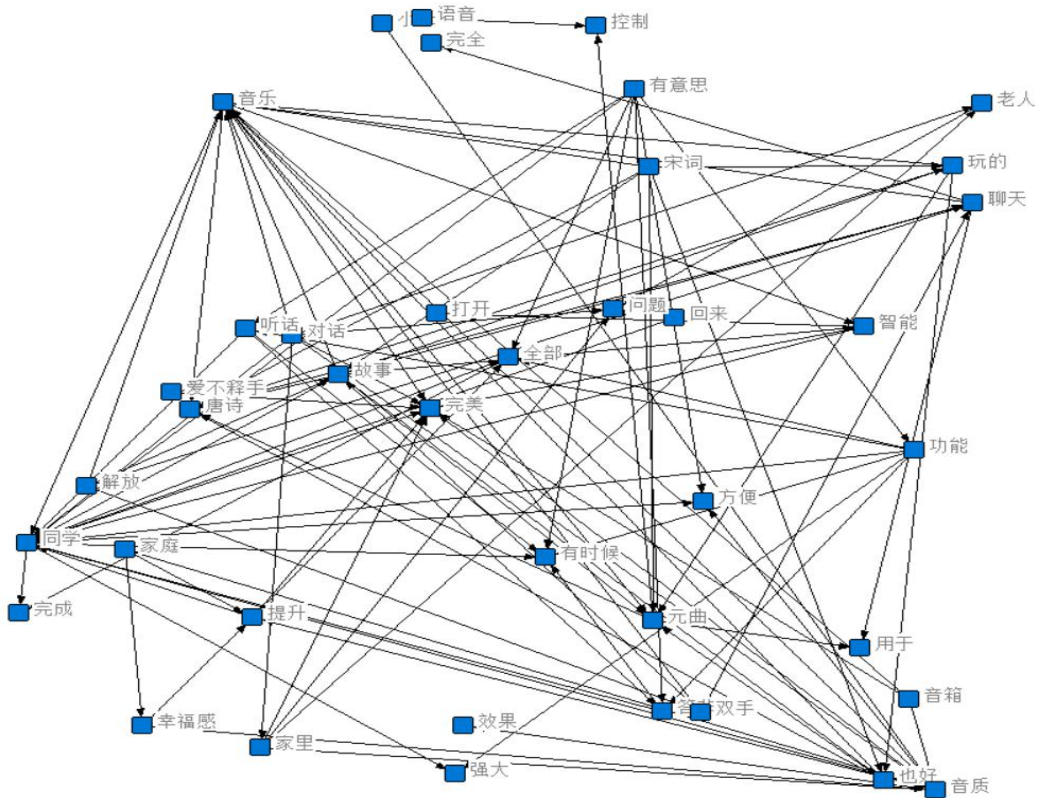


图 17 小爱智能音箱负面评价网络图

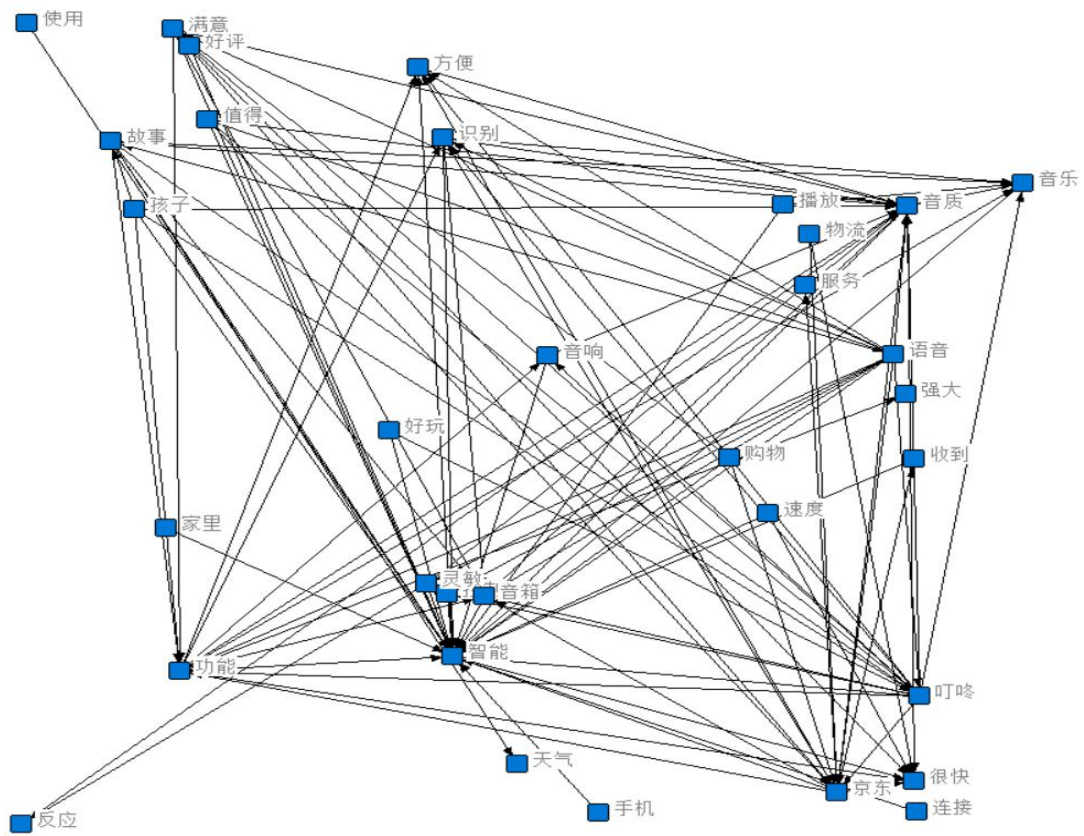


图 18 叮咚智能音箱正面评价网络图

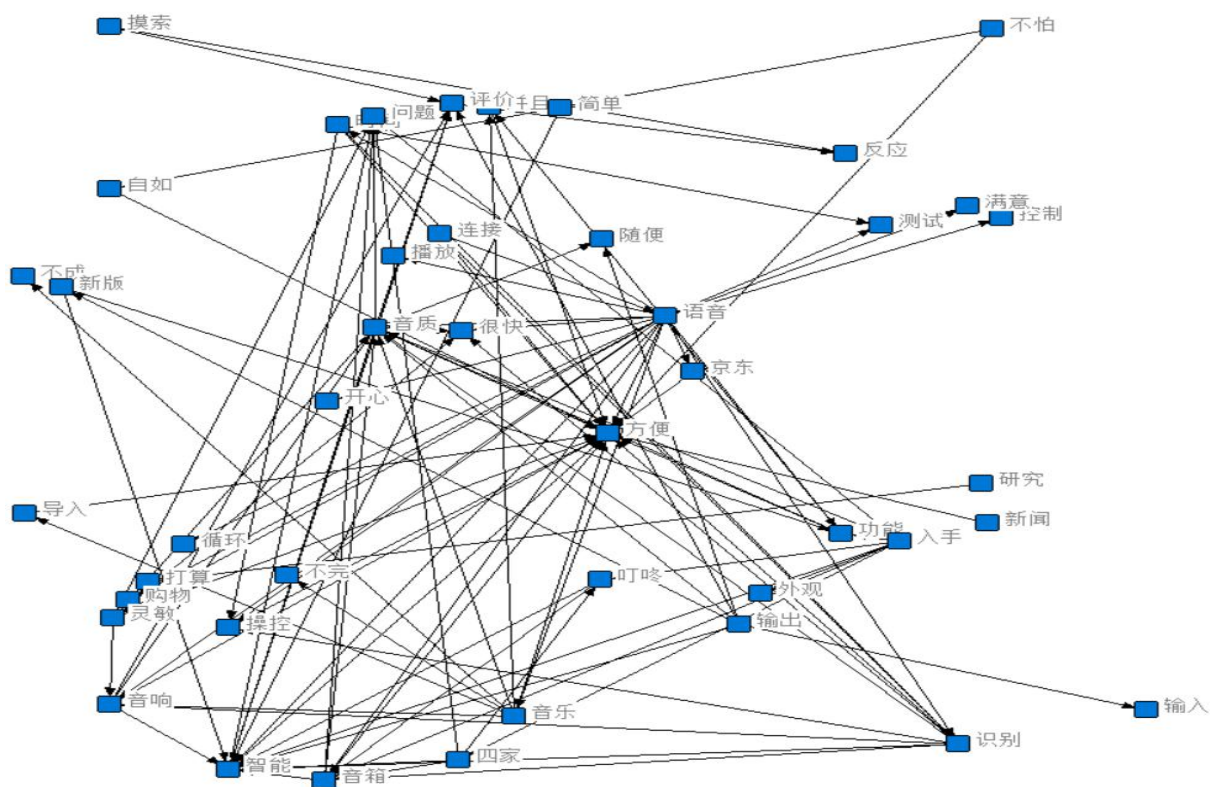


图 19 叮咚智能音箱负面评价网络图

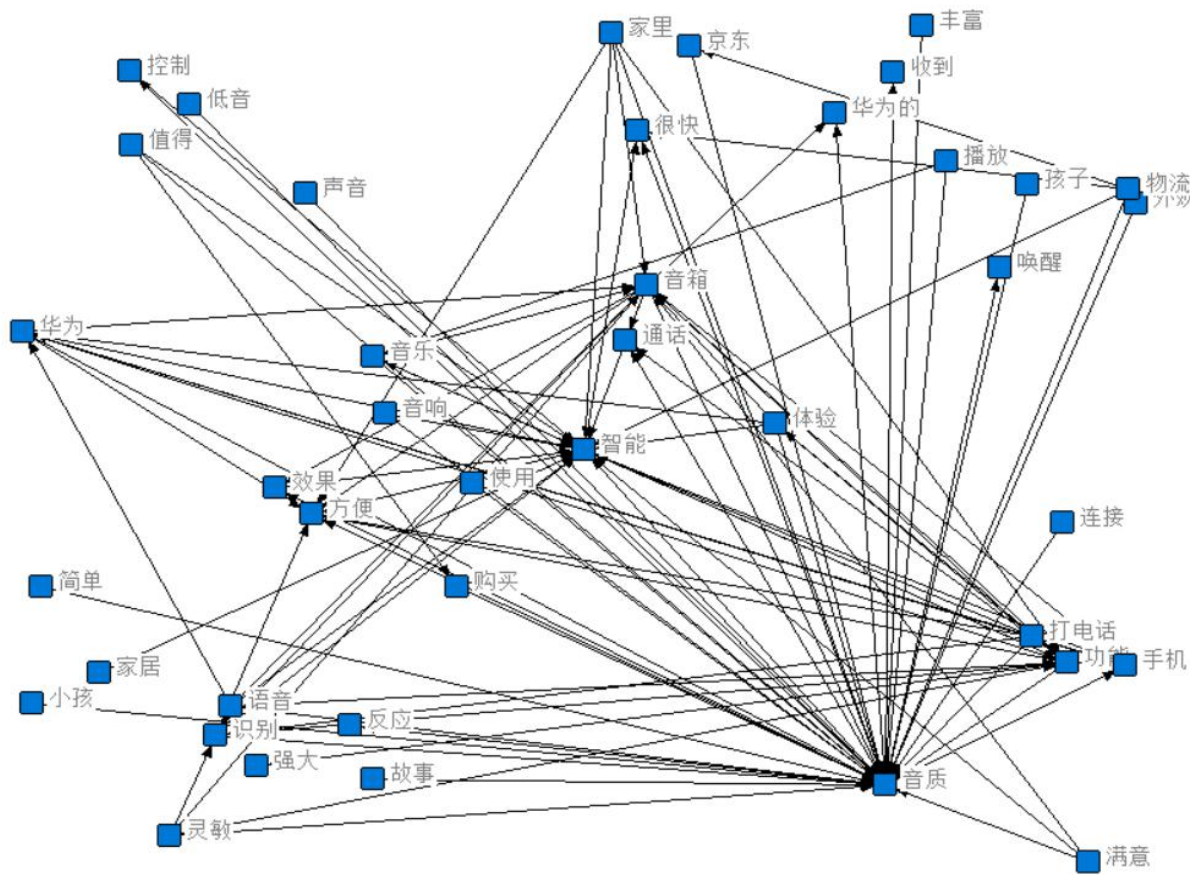


图 20 小艺智能音箱正面评价网络图

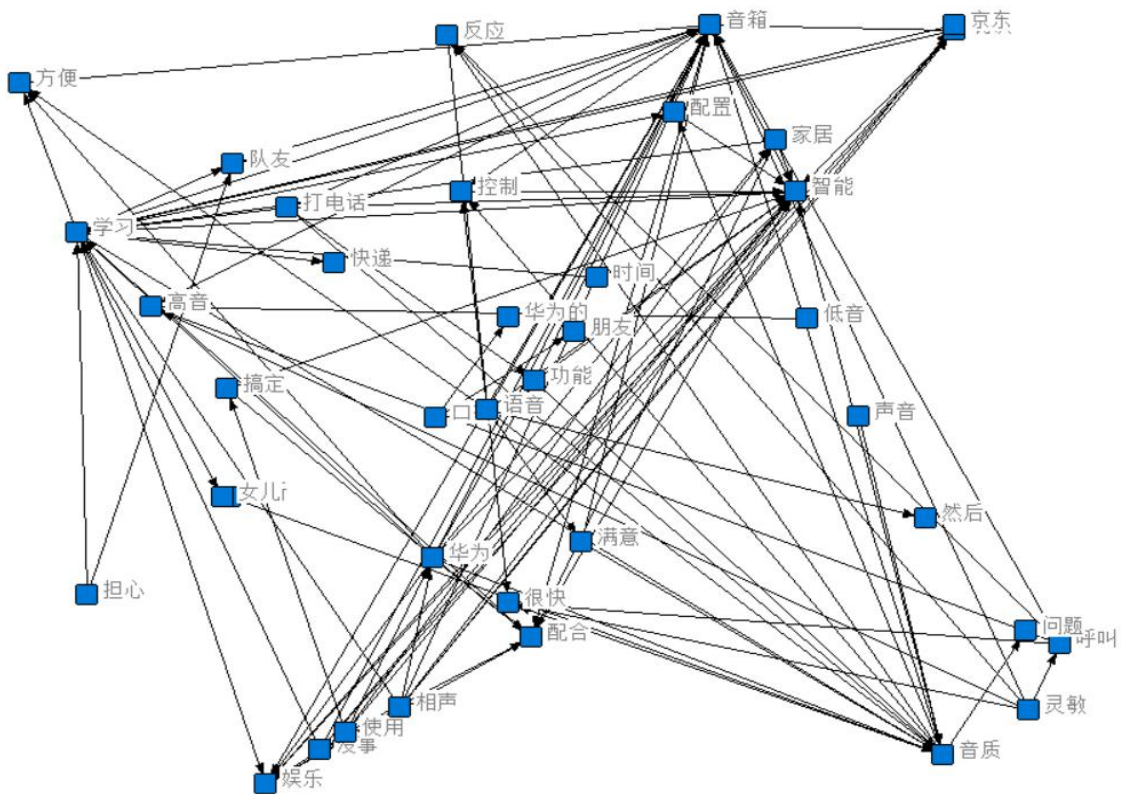


图 21 小艺智能音箱负面评价网络图

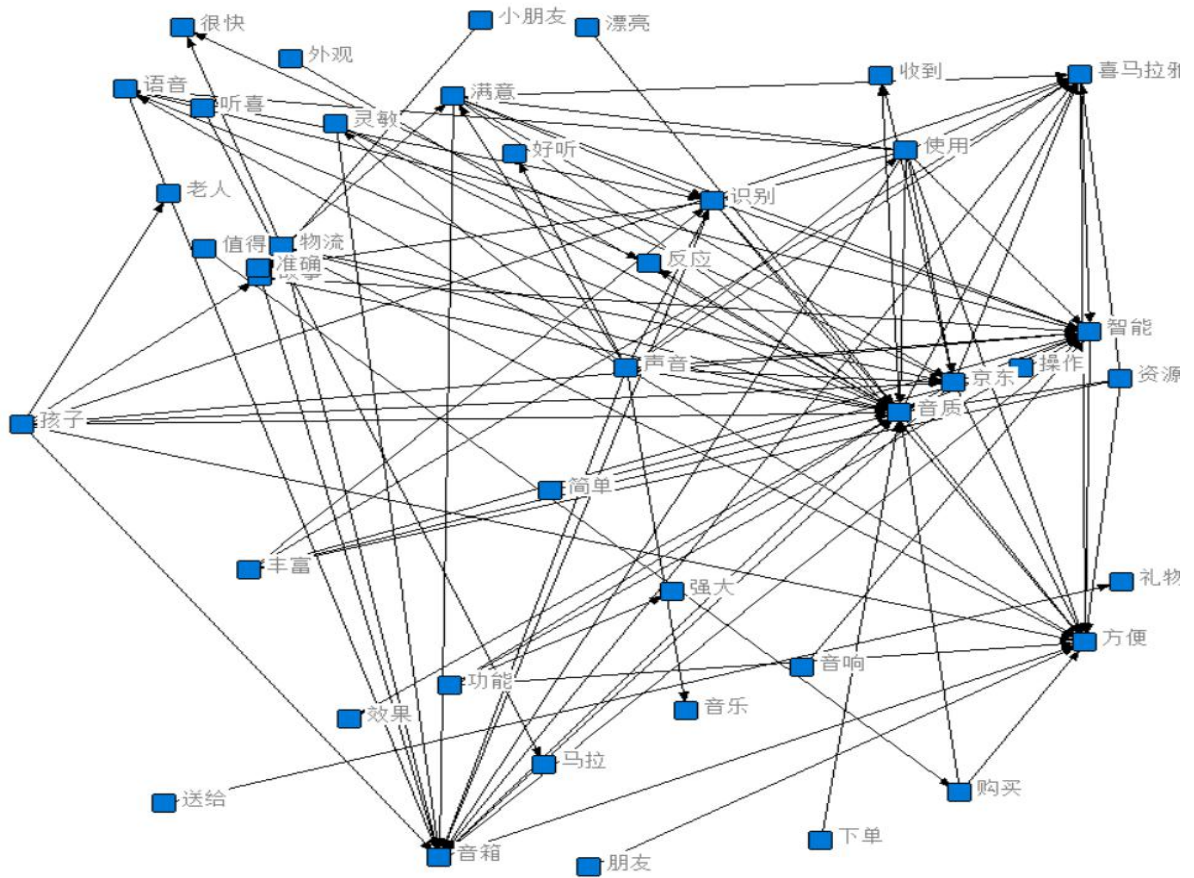


图 24 小雅智能音箱正面评价网络图

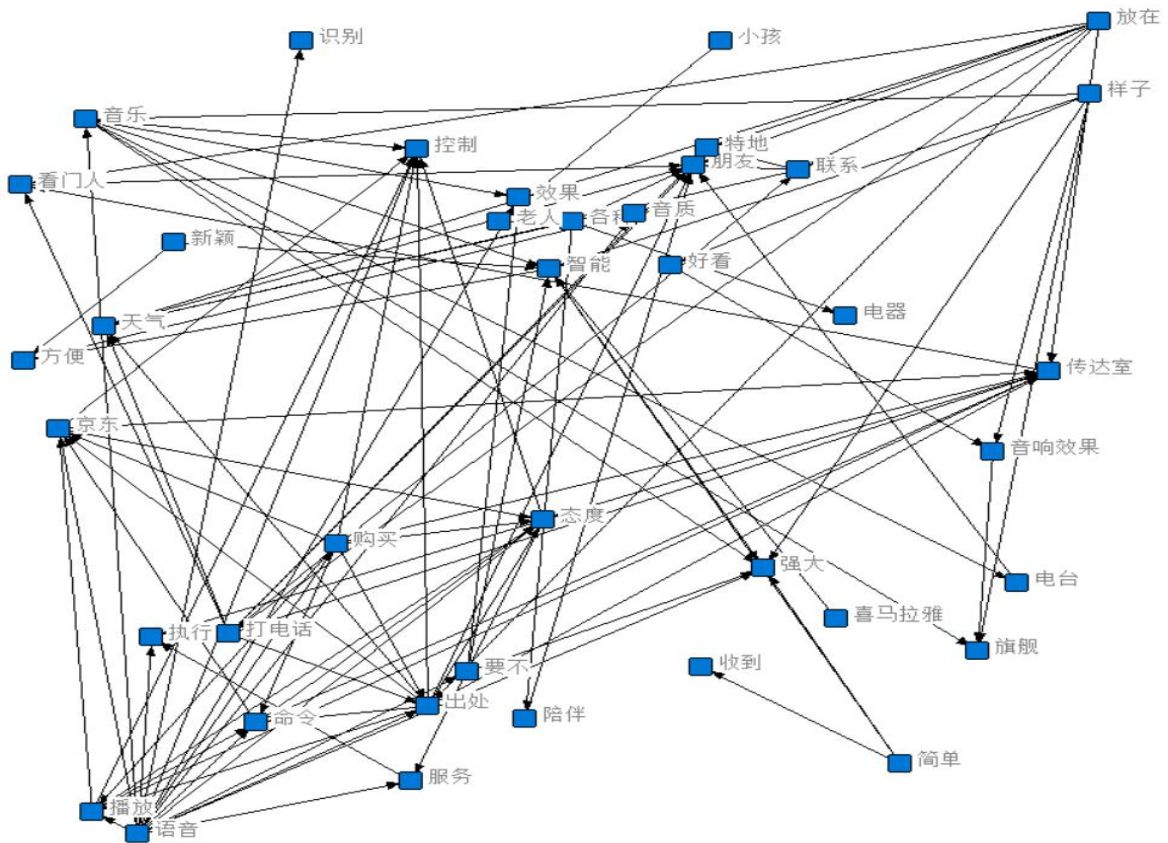


图 25 小雅智能音箱负面评价网络图

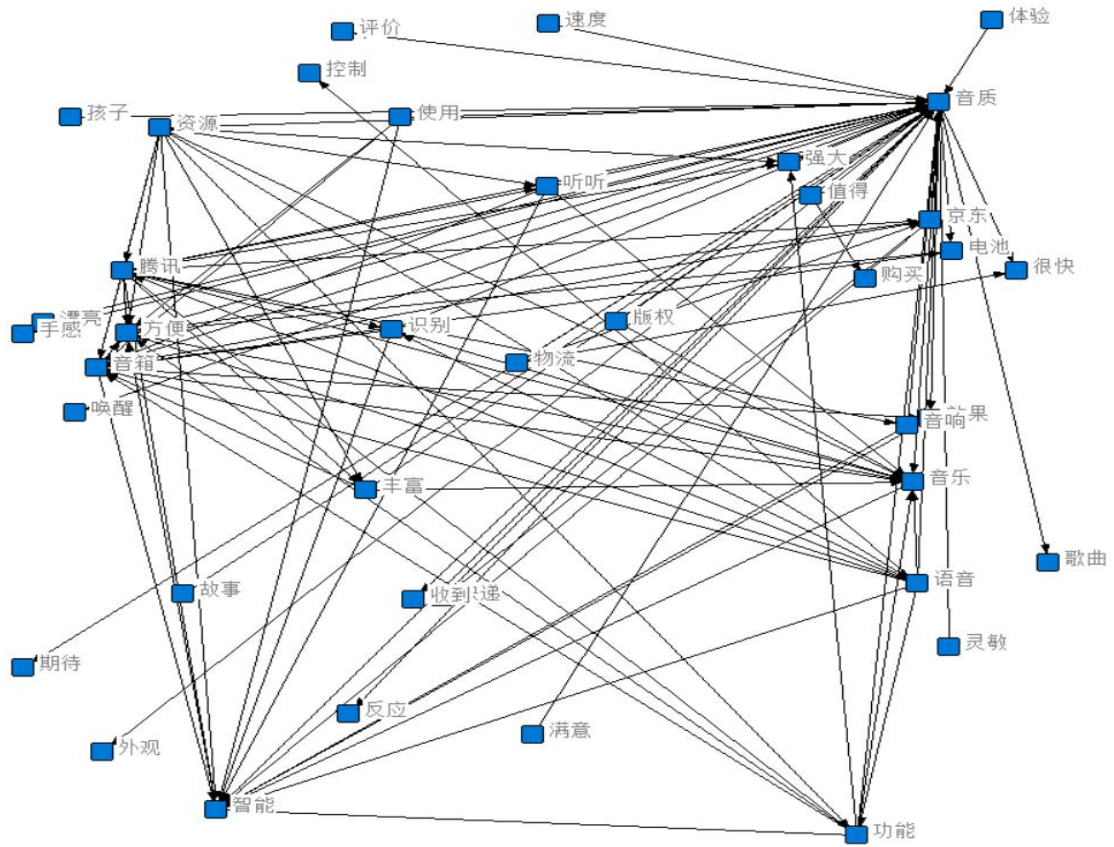


图 28 听听智能音箱正面评价网络图

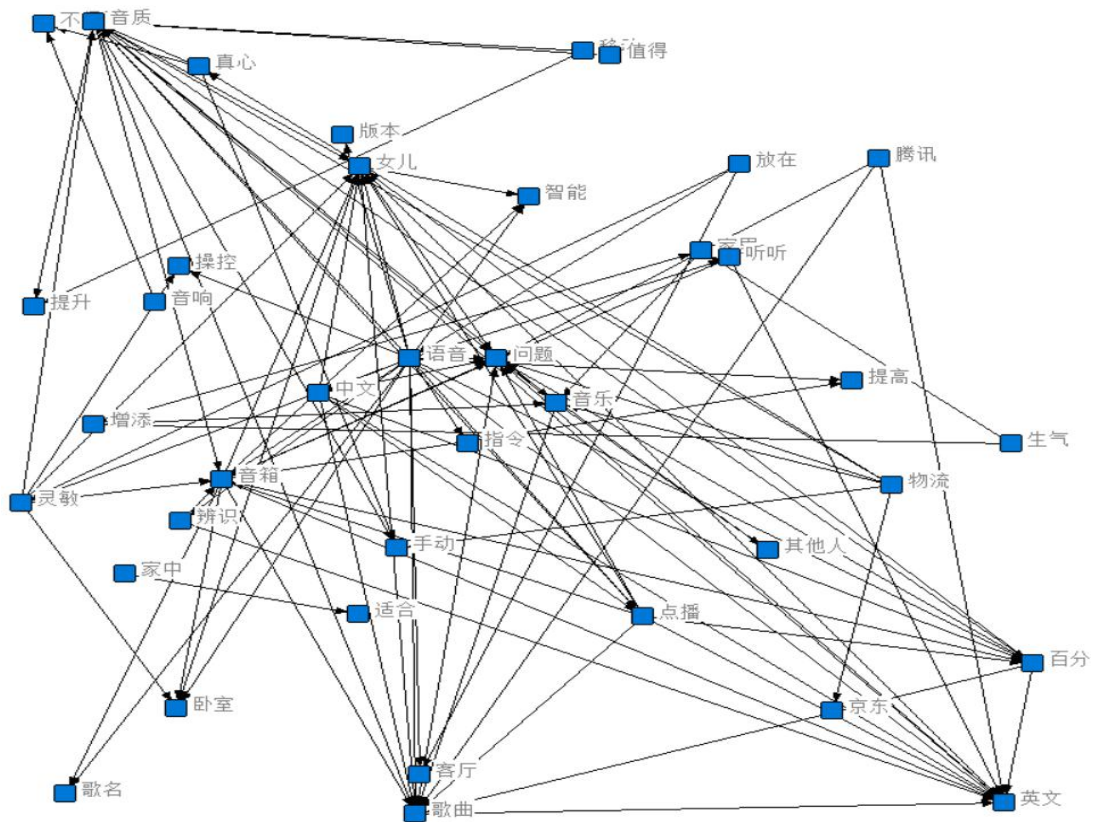


图 29 听听智能音箱负面评价网络图

附录二 LDA 主题分析结果

表 15 小度智能音箱正负面评价潜在主题

正面			负面		
主题一	主题二	主题三	主题一	主题二	主题三
很	的	了	的	很	小
好	可以	小	不好	度	买
也	音箱	度	音响	了	满意
非常	智能	的	功能	也	在
喜欢	音质	快	就	不	不
听	还	就	、	方便	能
都	很	我	音乐	还	智能
不错	是	到	听	可以	我
用	和	个	很少	放	识别
功能	有	速度	使用	使用	音箱

小度智能音箱正面评价特征词挖掘结果中，第一个主题反映小度智能音箱深受消费者喜爱，商品反响不错；第二个主题反映其音质不错，产品智能；第三个主题反映京东电商物流的送货速度快。负面评价特征词挖掘结果中，第一个主题反映其音响效果不好，音乐与功能较少等问题；第二个主题反映使用上不够方便；第三个主题反映语音识别上不够智能。

表 16 小爱智能音箱正负面评价潜在主题

正面			负面		
主题一	主题二	主题三	主题一	主题二	主题三
了	的	很	不好	很	了
小	好	也	小	爱	的
爱	非常	听	同学	也	少
喜欢	是	不错	听	就	音乐
同学	我	可以	一般	能	我
买	小米	音质	买	她	比
给	不	还	还	这样	故事
孩子	京东	智能	音响	电量	都
太	都	用	时候	不	非常
就	灵敏	方便	都	就是	挺

小爱智能音箱正面评价特征词挖掘结果中，第一个主题反映小爱智能音箱在消费者之间的产品传播度高；第二个主题反映其反应灵敏，消费者对京东电商平台和小米品牌的认可度高；第三个主题反映音质清晰，智能度高。负面评价三个主题特征词挖掘结果中，第一个主题反映小爱尺寸较小，音响功能一般；第二个主题反映续航能力一般；第三个主题反映音乐数量较少，故事数量较低。

表 17 叮咚智能音箱正负面评价潜在主题

正面			负面		
主题一	主题二	主题三	主题一	主题二	主题三
的	了	很	了	的	很
是	听	买	也	凑合	不
音箱	买	不错	用	买	价格
智能	就	好	音箱	是	、
可以	音质	也	质量	非常	就
腾讯	用	喜欢	最	音质	真心
有	我	音质	不喜欢	音响	没
方便	还	京东	还	不	能
资源	给	棒	几	在	和
多	歌	语音	垃圾	灵敏	我

叮咚智能音箱正面评价特征词挖掘结果中，第一个主题反映资源库丰富，产品智能，使用方便；第二个主题反映音质不错，听歌效果良好；第三个主题反映消费者对京东电商平台评价较高，语音功能不错。负面评价特征词挖掘结果中，第一个主题反映质量一般，部分消费者不喜欢；第二个主题反映音质一般，音箱效果不佳，反应不够灵敏等问题；第三个主题反映价格不够物美价廉。

表 18 小艺智能音箱正负面评价潜在主题

正面			负面		
主题一	主题二	主题三	主题一	主题二	主题三
了	的	很	小	的	很
喜欢	音箱	好	艺	了	外观
的	是	音质	是	就	价格
品牌	可以	非常	少	不喜欢	手感
我	听	也	我	买	差
买	功能	不错	也	打电话	的
给	都	喜欢	~	听	这个
华为	智能	用	音乐	很	一般
太	还	特别	不	能	华为
就	能	音响	真的	女儿	没有

小艺智能音箱正面评价特征词挖掘结果中，第一个主题反映消费者由于对华为品牌的喜爱而接受小艺智能音箱；第二个主题反映小艺智能音箱产品智能，功能强大；第三个主题反映音质不错，消费者非常喜爱此产品。负面评价特征词挖掘结果中，第一个主题反映音乐数量较少，资源库一般般；第二个主题反映打电话功能需要完善；第三个主题反映外观不够美观，手感不好，价格过高。

表 19 YOYO 智能音箱正负面评价潜在主题

正面			负面		
主题一	主题二	主题三	主题一	主题二	主题三
的	很	也	打电话	尺寸	的
了	好	音质	挺	好	杂
就	用	还	价格	了	用
音响	不错	可以	YOYO	还	东西
YOYO	喜欢	智能	有	非常	给
都	非常	功能	一般	可以	音
是	买	不错	它	小	还是
我	挺	方便	上	买	就
和	给	能	就是	也	乱
荣耀	的	小	反应	我	都

YOYO 智能音箱正面评价特征词挖掘结果中，第一个主题反映音响功能不错；第二个主题反映消费者对 YOYO 智能音响的喜爱；第三个主题反映了音质不错，使用感智能方便，功能全面。负面评价特征词挖掘结果中，反映性价比较低；第二个主题反映尺寸较小；第三个主题反映杂音较多。

表 20 小雅智能音箱正负面评价潜在主题

正面			负面		
主题一	主题二	主题三	主题一	主题二	主题三
音质	还	的	、	很	好
很	的	了	的	不	了
好	用	智能	小	听	雅
不错	可以	小	非常	手感	还
也	没有	音箱	音质	用	慢
这个	音乐	就	一般	不丰富	能
非常	听	我	-	内容	不仅
音响	大	还是	音乐	就	各种
挺	不	是	12	就是	和
喜欢	语音	有	很	没	播放

小雅智能音箱正面评价特征词挖掘结果中，第一个主题反映小雅智能音箱深受消费者喜爱，音响功能不错；第二个主题反映音乐数量符合需求，语音识别灵敏；第三个主题反映使用智能程度较好。负面评价特征词挖掘结果中，第一个主题反映音质一般；第二个主题反映手感一般，内容不够丰富；第三个主题反映小雅智能音箱的播放速度慢。

表 21 小贝智能音箱正负面评价潜在主题

正面			负面		
主题一	主题二	主题三	主题一	主题二	主题三
音质	还	的	外观	还	的
很	的	了	很	的	了
好	用	是	好	用	是
不错	可以	智能	不	可以	慢
也	多	小	也	没有	小
这个	音乐	音箱	这个	价格	音箱
非常	听	就	非常	听	就
音箱	大	我	音响	不	我
挺	不	还是	挺	大	还是
喜欢	语音	有	一般	不好	反应

小贝智能音箱正面评价特征词挖掘结果中，第一个主题反映音质效果较好；第二个主题反映音乐数量较多，语音功能强大；第三个主题反映消费者对小贝智能音箱的智能程度较认可。负面评价特征词挖掘结果中，第一个主题反映小贝智能音箱的外观不够美观，音响播放效果一般；第二个主题反映价格不够物美价廉；第三个主题反映反应速度慢，尺寸较小。

表 22 听听智能音箱正负面评价潜在主题

正面			负面		
主题一	主题二	主题三	主题一	主题二	主题三
的	了	很	了	的	很
是	听	好	也	好	不
音箱	买	的	用	买	慢
智能	就	非常	音箱	是	、
可以	音质	也	可以	一般	就
腾讯	用	喜欢	最	声音	真心
有	我	速度	不喜欢	音响	没
方便	还	京东	还	9420	反应
资源	给	棒	几	在	和
多	歌	语音	外观	凑合	我

听听智能音箱正面评价特征词挖掘结果中，第一个主题反映资源丰富、智能度高，使用方便；第二个主题反映音质较好；第三个主题反映京东物流的速度快。负面评价特征词挖掘结果中，第一个主题反映听听智能音箱的外表不够美观；第二个主题反映音响声音不够符合消费者需求；第三个主题反映听听智能音箱反应较慢。

附录三 贝叶斯网络与用户满意度条件概率分布图

(一) 天猫精灵智能音箱

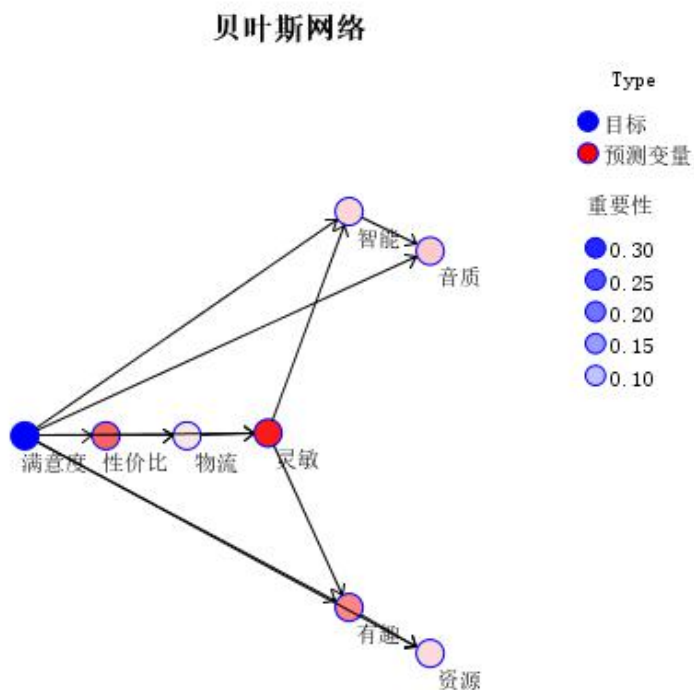


图 30 小度智能音箱贝叶斯网络图

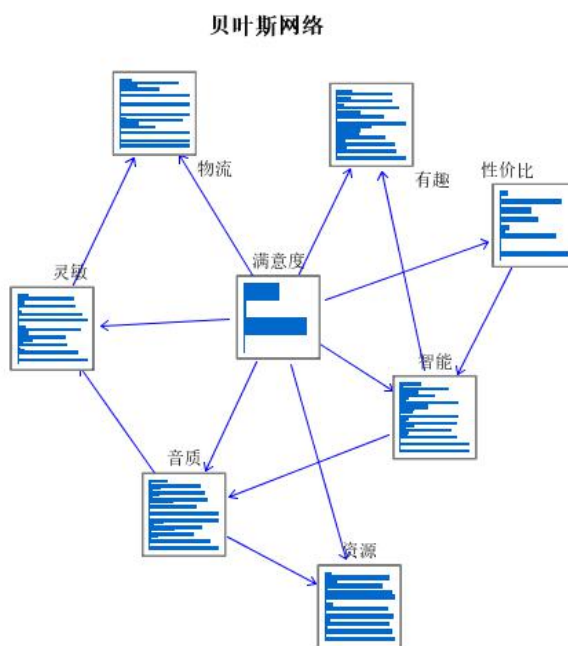


图 31 天猫精灵用户满意度条件概率分布图

(二) 小度智能音箱

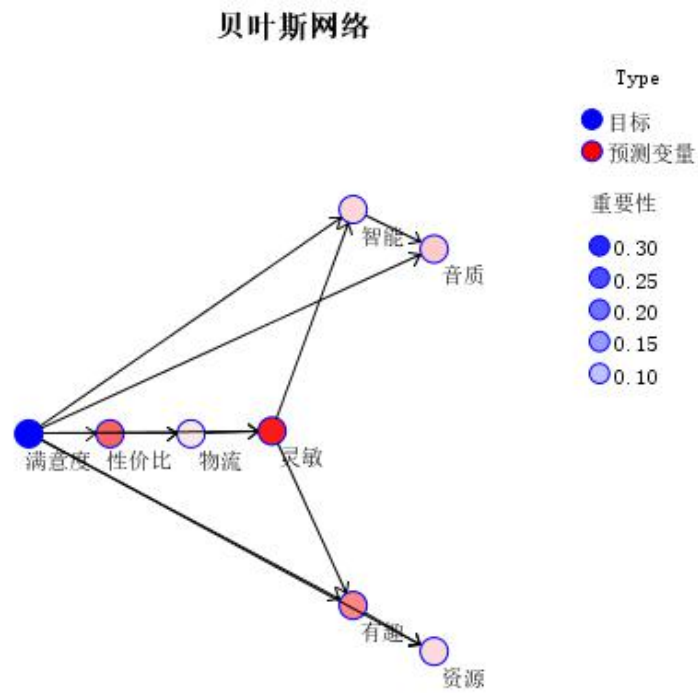


图 32 小度智能音箱贝叶斯网络图

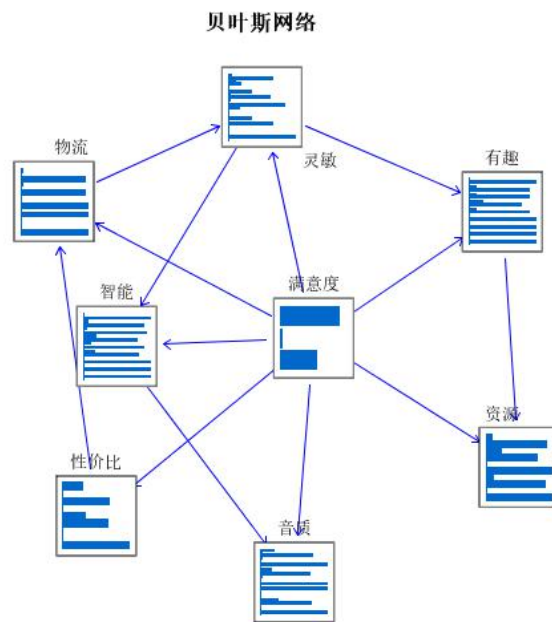


图 33 小度智能音箱用户满意度条件概率分布图

(三) 小爱智能音箱

贝叶斯网络

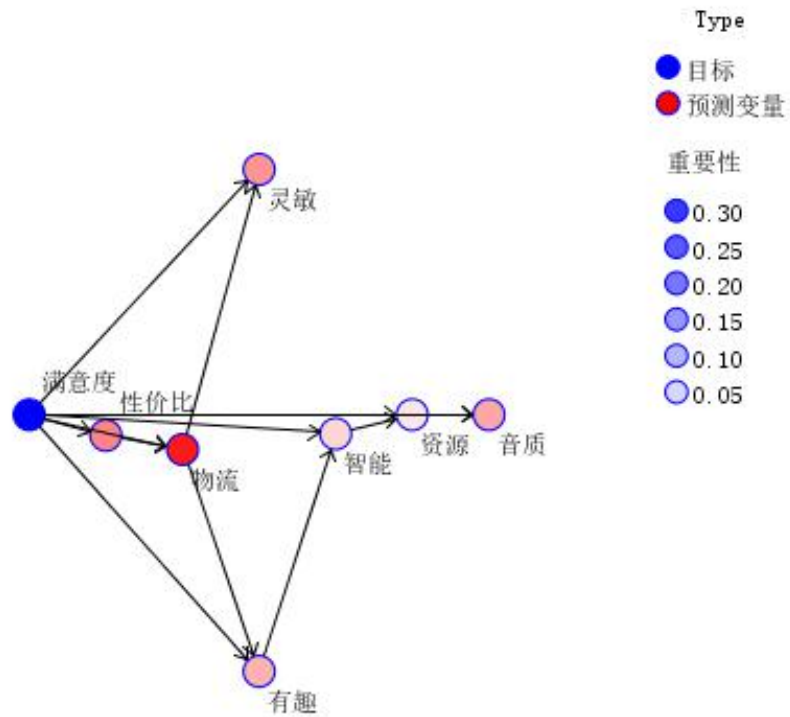


图 34 小爱智能音箱贝叶斯网络图

贝叶斯网络

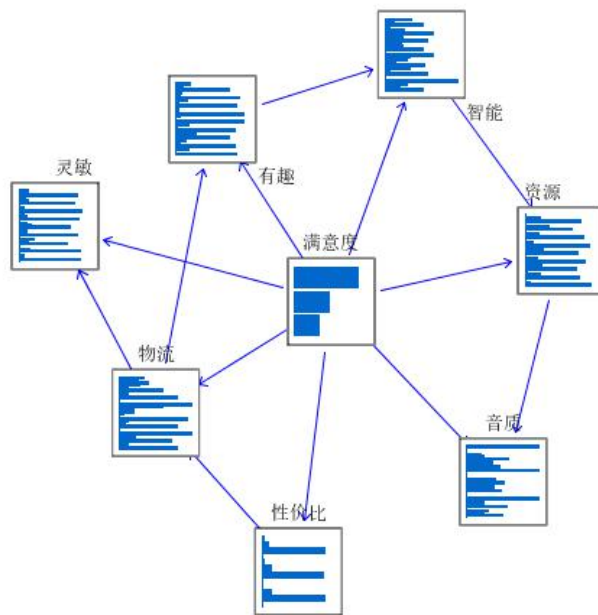


图 35 小爱智能音箱用户满意度条件概率分布图

(四) 叮咚智能音箱

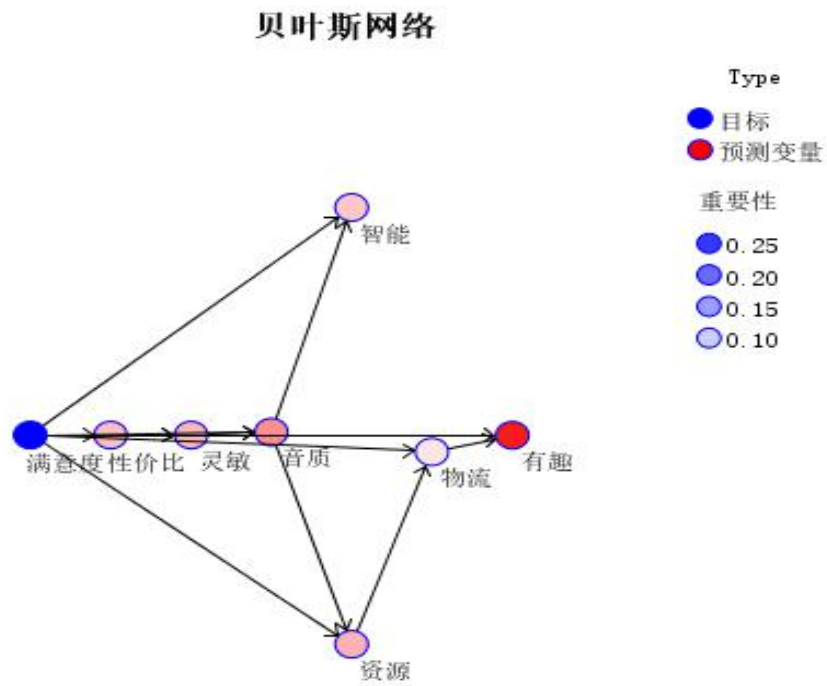


图 36 叮咚贝叶斯网络图

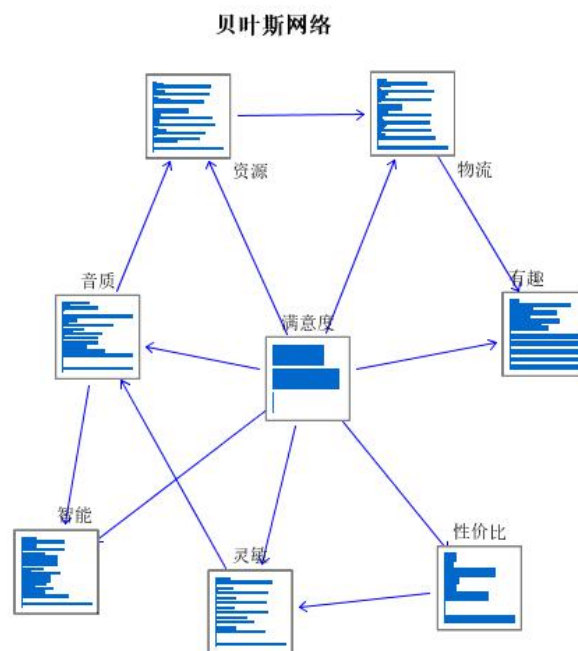
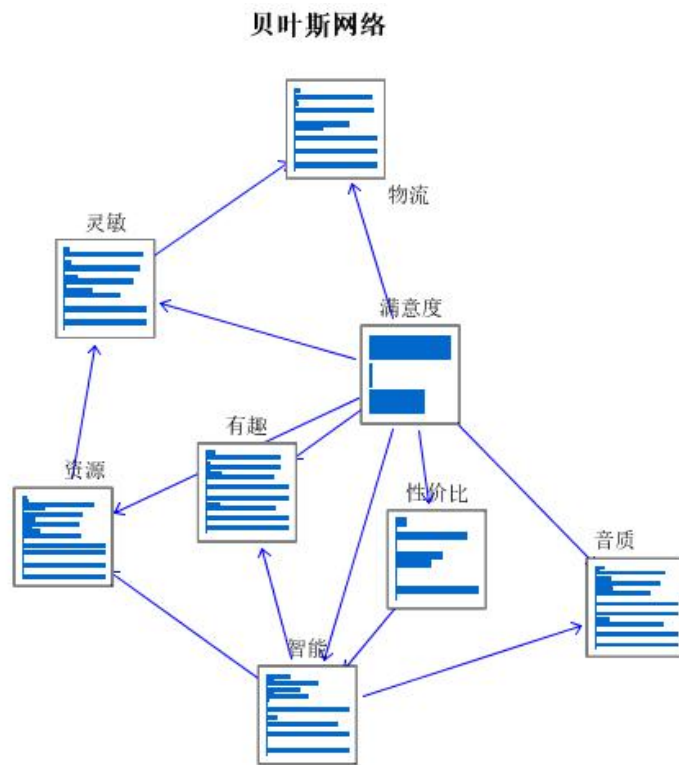
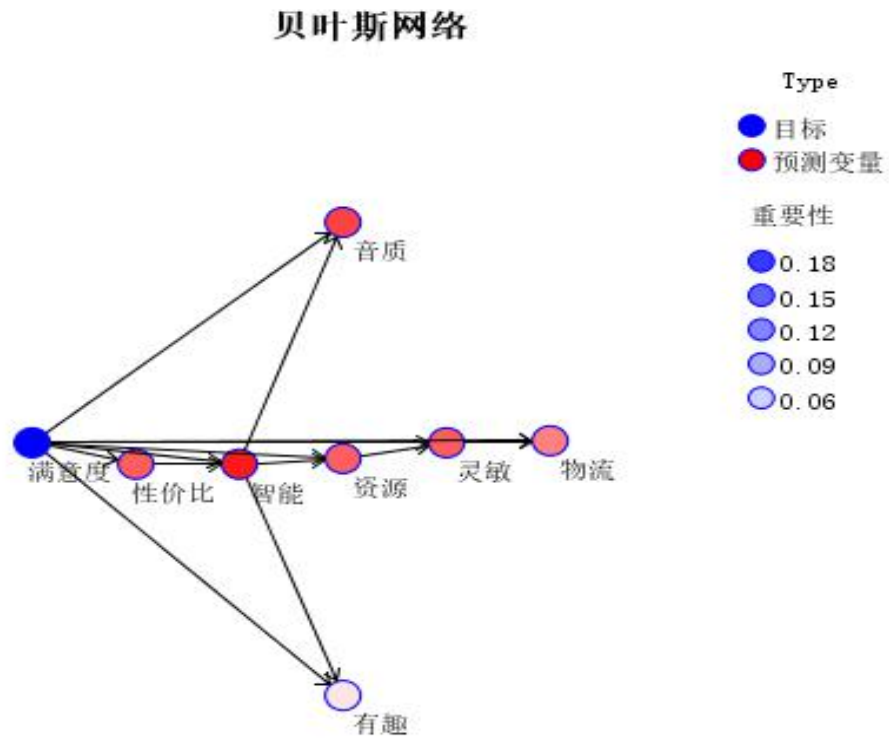


图 37 叮咚智能音箱用户满意度条件概率分布图

(五) 小艺智能音箱



(六) YOYO 智能音箱

贝叶斯网络

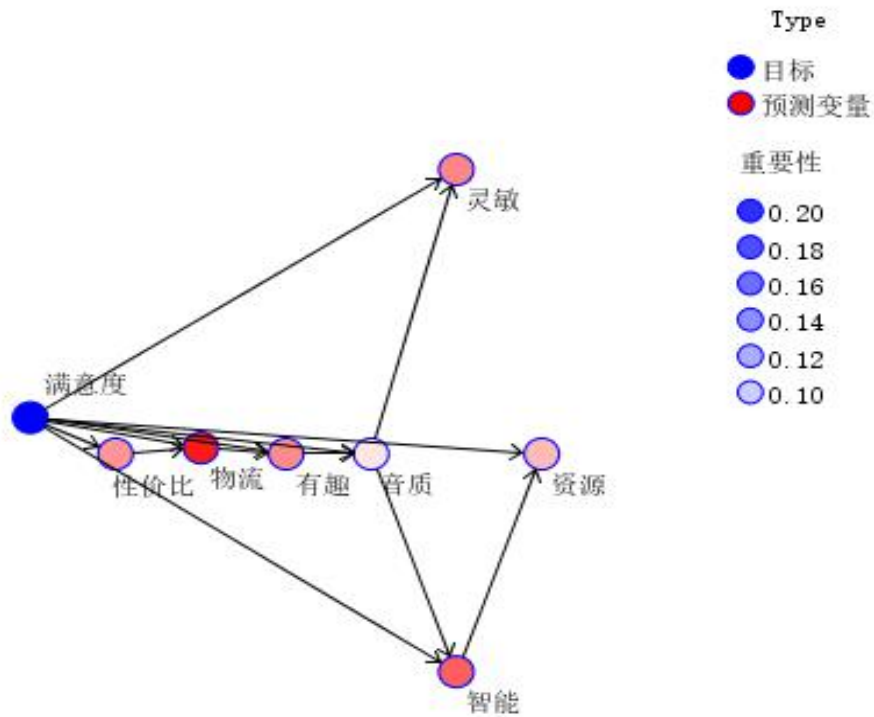


图 40 YOYO 贝叶斯网络图

贝叶斯网络

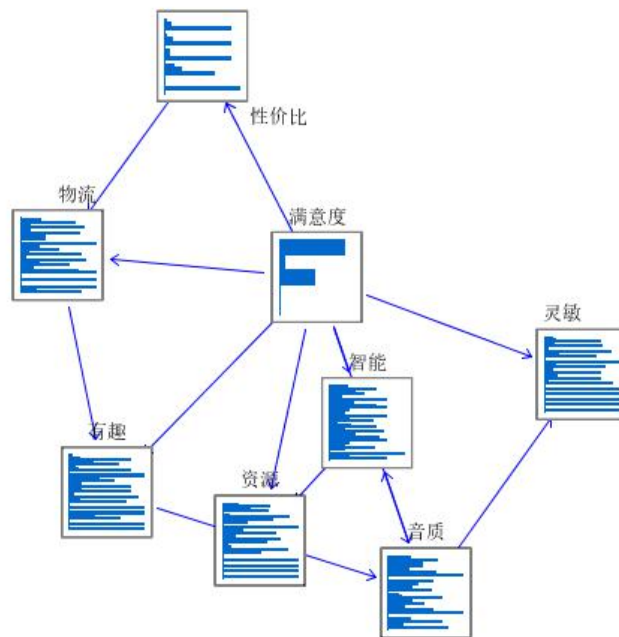


图 41 YOYO 智能音箱用户满意度条件概率分布图

(七) 小雅智能音箱

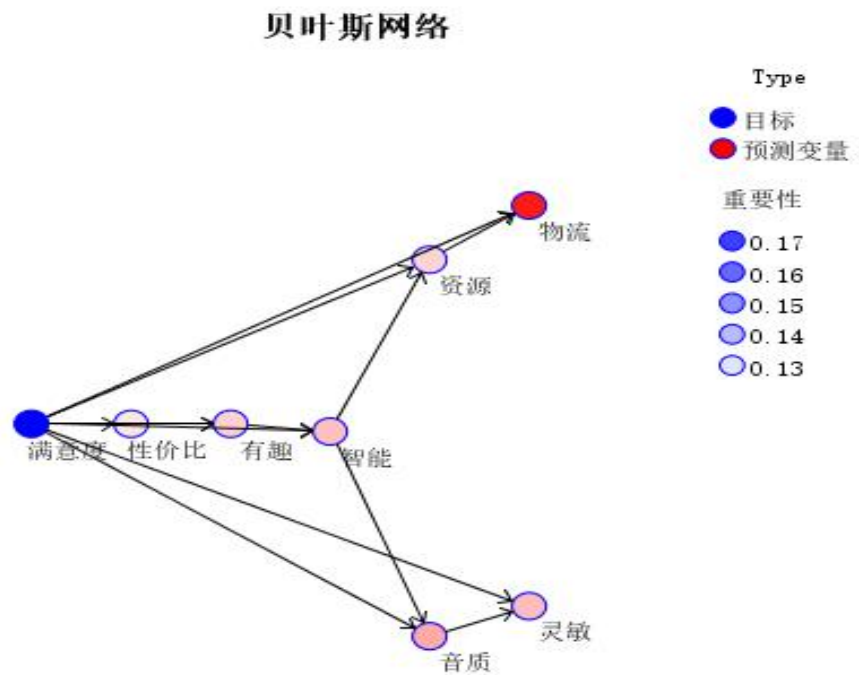


图 42 小雅贝叶斯网络图

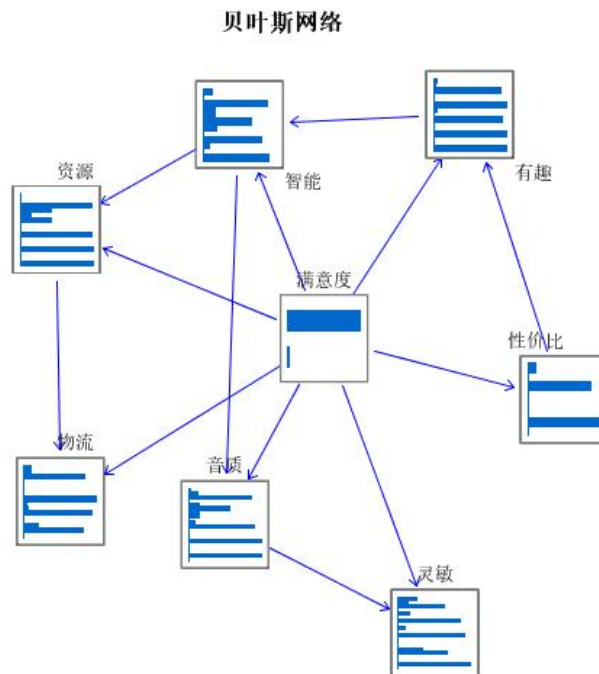


图 43 小雅智能音箱用户满意度条件概率分布图

(八) 小贝智能音箱

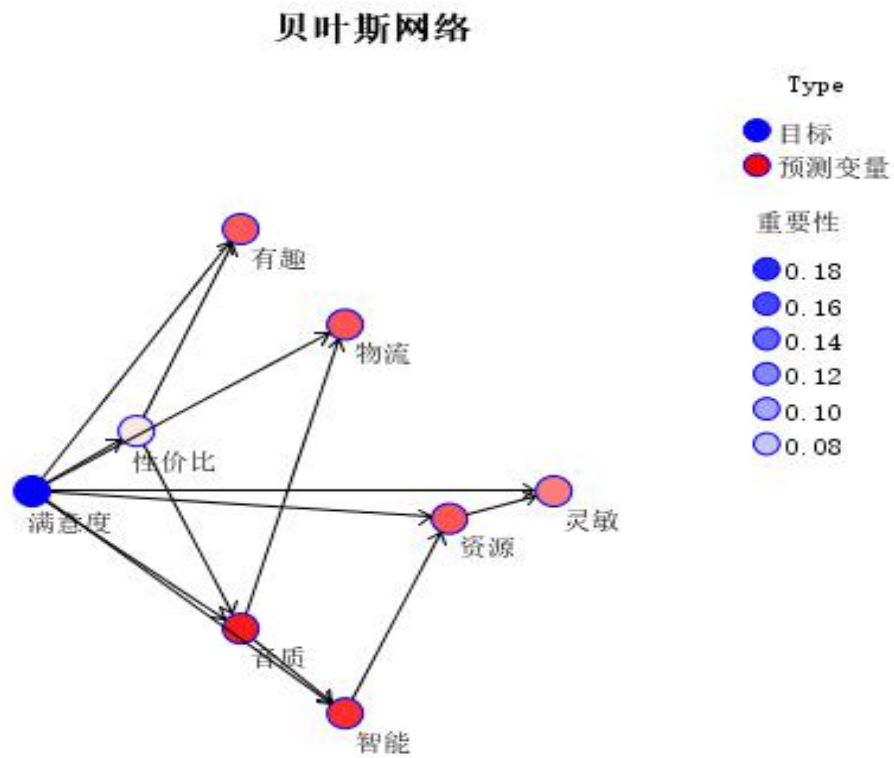


图 44 小贝贝叶斯网络图

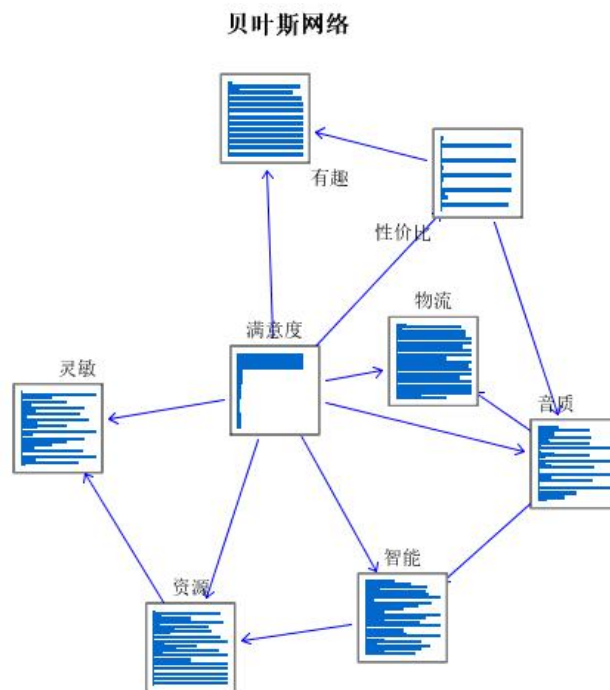


图 45 小贝智能音箱用户满意度条件概率分布图

(九) 听听智能音箱

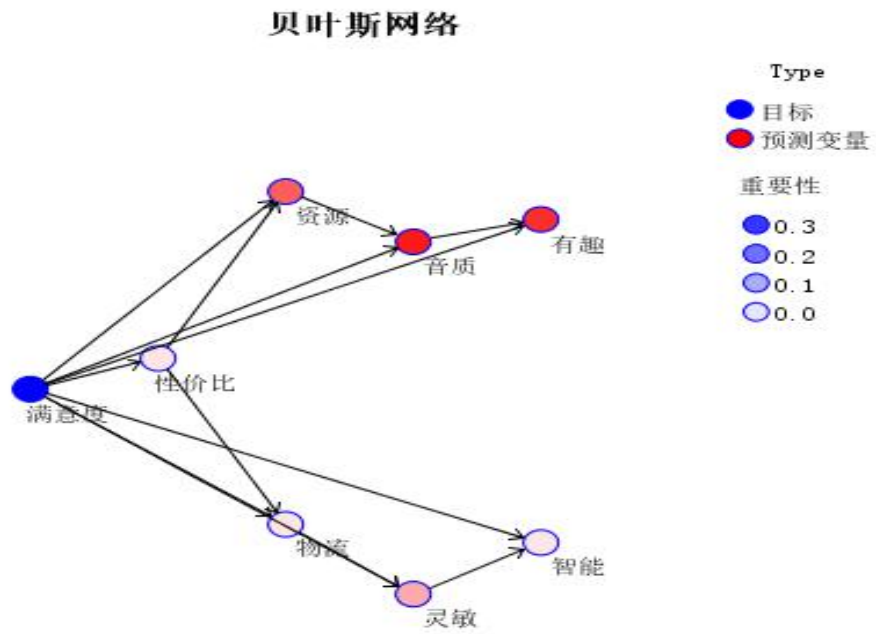


图 46 听听贝叶斯网络图

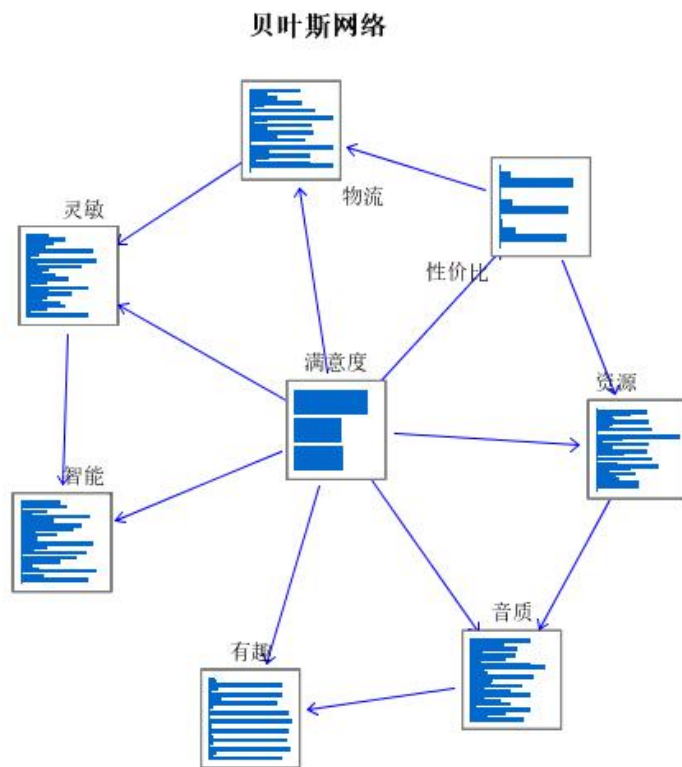


图 47 听听智能音箱用户满意度条件概率分布图

附录四 贝叶斯网络各节点条件概率

(一) 天猫精灵智能音箱

表 23 性价比条件概率

父级		概率			
满意度	2	3	4	5	
2	0	1	0	0	
3	0.001	0.814	0.062	0.121	
4	0	0.545	0	0.454	
5	0	0.88	0.011	0.107	

表 24 智能条件概率

父级		概率			
满意度	性价比	3	4	5	
2	3	1	0	0	
3	2	1	0	0	
3	3	0.804	0.083	0.112	
3	4	0.725	0.075	0.2	
3	5	0.805	0.077	0.116	
4	3	0.833	0	0.166	
4	5	0.2	0.4	0.4	
5	3	0.822	0.074	0.103	
5	4	0.5	0.25	0.25	
5	5	0.684	0.026	0.289	

表 25 有趣条件概率

父级		概率			
满意度	智能	3	4	5	
2	2	1	0	0	
3	3	0.856	0.013	0.129	
3	4	0.826	0.057	0.115	
3	5	0.693	0	0.306	
4	3	0.333	0.166	0.5	
4	4	1	0	0	
4	5	0.666	0	0.333	
5	3	0	0.897	0.102	
5	4	0	0.8	0.2	
5	5	0	0.795	0.204	

表 26 音质条件概率

父级		概率			
满意度	智能	3	4	5	
2	3	1	0	0	
3	3	0.874	0.017	0.108	
3	4	0.634	0.019	0.346	
3	5	0.746	0.053	0.2	
4	3	1	0	0	
4	4	1	0	0	
4	5	0.666	0	0.333	
5	3	0.833	0.035	0.13	
5	4	0.8	0.04	0.16	
5	5	0.727	0.022	0.25	

表 27 灵敏条件概率

父级		概率			
满意度	音质	2	3	4	5
2	3	0	1	0	0
3	3	0.003	0.87	0.091	0.033
3	4	0	0.714	0.071	0.214
3	5	0.011	0.681	0.159	0.147
4	3	0	0.9	0.1	0
4	5	0	0	1	0
5	3	0.003	0.937	0.045	0.013
5	4	0	0.833	0.083	0.083
5	5	0	0.807	0.038	0.153

表 28 物流条件概率

父级		概率			
满意度	灵敏	2	3	4	5
2	3	0	1	0	0
3	2	1	0	0	0
3	3	0	0.998	0	0.001
3	4	0	0.5	0.234	0.265
3	5	0	0.911	0.088	0
4	3	0	1	0	0
4	4	0	0	0	1
5	2	1	0	0	0
5	3	0	1	0	0
5	4	0	0.562	0.187	0.25
5	5	0	0.846	0	0.153

(二) 小度智能音箱

表 29 性价比条件概率

父级		概率		
满意度		3	4	5
3		1	0	0
4		0.666	0.333	0
5		0.698	0	0.301

表 30 物流条件概率

父级		概率		
满意度	性价比	3	4	5
3	3	1	0	0
4	3	0	1	0
4	4	1	0	0
5	3	0.966	0	0.033
5	5	0.961	0	0.038

表 31 灵敏条件概率

父级		概率			
满意度	物流	1	3	4	5
3	3	0	1	0	0
4	3	0	0.666	0	0.333
4	4	0	0	0.166	0.833
5	3	0.017	0.613	0.031	0.338
5	5	0	0.19	0.095	0.666

表 32 智能条件概率

父级		概率			
满意度	灵敏	1	3	4	5
3	3	0	1	0	0
4	3	0	1	0	0
4	4	0	1	0	0
4	5	0	0.833	0	0.166
5	1	0	0.9	0	0.1
5	3	0.002	0.807	0	0.189
5	4	0	0.95	0	0.05
5	5	0	0.914	0.014	0.071

表 33 有趣条件概率

父级		概率			
满意度	灵敏	1	3	5	
3	3	0	1	0	
4	3	0	1	0	
4	4	0	1	0	
4	5	0	1	0	
5	1	0	0.9	0.1	
5	3	0.013	0.779	0.206	
5	4	0	0.9	0.1	
5	5	0	0.9	0.1	

表 34 资源条件概率

父级		概率		
满意度	有趣	3	5	
3	3	1	0	
4	3	0.888	0.111	
5	1	1	0	
5	3	0.762	0.237	
5	5	0.918	0.081	

表 35 音质条件概率

父级		概率			
满意度	智能	1	3	4	5
3	3	0	1	0	0
4	3	0	0.75	0.25	0
4	5	0	0	0	1
5	1	0	1	0	0
5	3	0.007	0.737	0.101	0.152
5	4	0	1	0	0
5	5	0.023	0.776	0.011	0.188

(三) 小爱智能音箱

表 36 性价比条件概率

父级	概率			
满意度	1	3	4	5
3	0	0.858	0.129	0.011
4	0	0.852	0.126	0.021
5	0.002	0.872	0.091	0.034

表 37 物流条件概率

父级		概率		
满意度	性价比	3	4	5
3	3	0.794	0.109	0.095
3	4	0.727	0.227	0.045
3	5	1	0	0
4	3	0.797	0.079	0.123
4	4	0.933	0	0.066
4	5	0.2	0.2	0.6
5	1	1	0	0
5	3	0.795	0.123	0.081
5	4	0.6	0.125	0.275
5	5	0.4	0.266	0.333

表 38 有趣条件概率

父级		概率		
满意度	性价比	3	4	5
3	3	0.835	0	0.164
3	4	0.809	0.047	0.142
3	5	0.733	0	0.266
4	3	0.81	0.01	0.178
4	4	0.941	0	0.058
4	5	0.933	0.033	0.033
5	3	0.829	0.023	0.146
5	4	0.875	0.053	0.071
5	5	0.723	0.085	0.191

表 39 智能条件概率

父级		概率		
满意度	有趣	3	4	5
3	3	0.514	0.192	0.292
3	4	1	0	0
3	5	0.586	0.068	0.344
4	3	0.55	0.146	0.303
4	4	0.333	0.666	0
4	5	0.527	0.25	0.222
5	3	0.59	0.171	0.238
5	4	0.666	0.266	0.066
5	5	0.516	0.112	0.37

表 40 资源条件概率

父级		概率		
满意度	智能	3	4	5
3	3	0.911	0.077	0.011
3	4	0.758	0.206	0.034
3	5	0.705	0.235	0.058
4	3	0.829	0.162	0.007
4	4	0.725	0.25	0.025
4	5	0.882	0.117	0
5	3	0.8	0.196	0.003
5	4	0.657	0.328	0.013
5	5	0.763	0.218	0.018

表 41 音质条件概率

父级		概率			
满意度	资源	2	3	4	5
3	3	0	0.485	0.228	0.285
3	4	0	0.56	0.2	0.24
3	5	0	1	0	0
4	3	0	0.464	0.158	0.377
4	4	0.025	0.512	0.076	0.384
4	5	0	0	0	1
5	3	0.002	0.458	0.187	0.351
5	4	0	0.581	0.224	0.193
5	5	0	0	0	1

(四) 叮咚智能音箱

表 42 性价比条件概率

父级		概率		
满意度		3	4	5
2		1	0	0
4		0.626	0.164	0.208
5		0.715	0.123	0.161

表 43 灵敏条件概率

父级		概率				
满意度	性价比	1	2	3	4	5
2	3	0	1	0	0	0
4	3	0.003	0	0.173	0.003	0.28
4	4	0	0	0.544	0.012	0.443
4	5	0	0	0.74	0	0.26
5	3	0	0	0.721	0	0.278
5	4	0	0	0.739	0.021	0.239
5	5	0	0	0.8	0	0.2

表 44 音质条件概率

父级		概率			
满意度	灵敏	2	3	4	5
2	2	0	1	0	0
4	1	0	0	0	1
4	3	0.003	0.595	0.063	0.338
4	4	0	0.5	0	0.5
4	5	0.013	0.551	0.131	0.303
5	3	0.003	0.726	0.072	0.197
5	4	0	1	0	0
5	5	0.01	0.505	0.103	0.381

表 45 资源条件概率

父级		概率			
满意度	音质	2	3	4	5
2	3	0	1	0	0
4	2	0	0.333	0.666	0
4	3	0.032	0.737	0.172	0.075
4	4	0	0.875	0.05	0.075
4	5	0.012	0.841	0.088	0.056
5	2	0.5	0.5	0	0
5	3	0.004	0.718	0.228	0.048
5	4	0	0.8	0.166	0.033
5	5	0.01	0.813	0.131	0.043

表 46 智能条件概率

父级		概率			
满意度	音质	2	3	4	5
2	3	0	1	0	0
4	2	0	0.666	0	0.333
4	3	0	0.453	0.194	0.352
4	4	0.025	0.4	0.175	0.4
4	5	0	0.556	0.139	0.303
5	2	0	0.5	0	0.5
5	3	0	0.51	0.152	0.337
5	4	0	0.6	0.2	0.2
5	5	0	0.615	0.175	0.208

表 47 物流条件概率

父级		概率			
满意度	资源	2	3	4	5
2	3	1	0	0	0
4	2	0	0.818	0.09	0.09
4	3	0.045	0.759	0.088	0.106
4	4	0.121	0.742	0.075	0.06
4	5	0.035	0.75	0.142	0.071
5	2	0.333	0.333	0.333	0
5	3	0.035	0.758	0.104	0.1
5	4	0.135	0.797	0.04	0.027
5	5	0.176	0.705	0	0.117

表 48 有趣条件概率

父级		概率		
满意度	物流	4	5	
2	2	1	0	
4	2	1	0	
4	3	1	0	
4	4	1	0	
4	5	1	0	
5	2	0.541	0.458	
5	3	0.717	0.282	
5	4	0.666	0.333	
5	5	0.875	0.125	

(五) 小艺智能音箱

表 49 性价比条件概率

父级		概率		
满意度		3	4	5
3	1	1	0	0
4	0	0.428	0.571	0
5	0.006	0.859	0	0.134

表 50 智能条件概率

父级		概率			
满意度	性价比	1	3	4	5
3	3	0	1	0	0
4	3	0	0	1	0
4	4	0	0.875	0	0.125
5	1	0	1	0	0
5	3	0.02	0.496	0.08	0.404
5	5	0	0.628	0.076	0.294

表 51 有趣条件概率

父级		概率		
满意度	智能	1	3	5
3	3	0	1	0
4	3	0	1	0
4	4	0	0.833	0.166
4	5	0	1	0
5	1	0	1	0
5	3	0.003	0.813	0.182
5	4	0.043	0.913	0.043
5	5	0.008	0.893	0.097

表 52 资源条件概率

父级		概率		
满意度	智能	3	4	5
3	3	1	0	0
4	3	1	0	0
4	4	0	1	0
4	5	1	0	0
5	1	0.7	0.2	0.1
5	3	0.697	0.142	0.159
5	4	0.739	0	0.26
5	5	0.875	0.071	0.053

表 53 灵敏条件概率

父级		概率		
满意度	资源	1	3	5
3	3	0	1	0
4	3	0	1	0
4	4	0	0.666	0.333
5	3	0.006	0.837	0.156
5	4	0	0.918	0.081
5	5	0	0.945	0.054

表 54 物流条件概率

父级		概率			
满意度	灵敏	1	3	4	5
3	3	0	1	0	0
4	3	0	1	0	0
4	5	0	1	0	0
5	1	0.333	0.666	0	0
5	3	0	0.96	0.006	0.034
5	5	0.012	0.936	0	0.05

表 55 音质条件概率

父级		概率				
满意度	智能	1	3	4	5	
3	3	0	1	0	0	
4	3	0	1	0	0	
4	4	0	0.833	0	0.166	
4	5	0	1	0	0	
5	1	0	1	0	0	
5	3	0.009	0.664	0.119	0.205	
5	4	0.021	0.782	0	0.195	
5	5	0	0.84	0.048	0.111	

(六) YOYO 智能音箱

表 56 性价比条件概率

父级		概率			
满意度		2	3	4	5
1		0	1	0	0
2		0	0.666	0.222	0.111
3		0	0.885	0.063	0.05
4		0	0.878	0.097	0.024
5		0.005	0.883	0.085	0.026

表 57 物流条件概率

父级		概率			
满意度	性价比	3	4	5	
1	3	0.8	0.2	0	
2	3	1	0	0	
2	4	0	1	0	
2	5	0	0	1	
3	3	0.762	0.064	0.172	
3	4	0.65	0.1	0.25	
3	5	0.875	0.062	0.062	
4	3	0.805	0.111	0.083	
4	4	0.5	0.25	0.25	
4	5	1	0	0	
5	2	0.333	0.333	0.333	
5	3	0.799	0.062	0.137	
5	4	0.857	0.02	0.122	
5	5	0.733	0.266	0	

表 58 有趣条件概率

父级		概率			
满意度	物流	3	4	5	
1	3	1	0	0	
1	4	1	0	0	
2	3	0.666	0.333	0	
2	4	1	0	0	
2	5	1	0	0	
3	3	0.744	0.138	0.117	
3	4	0.904	0.047	0.047	
3	5	0.814	0.148	0.037	
4	3	0.812	0.156	0.031	
4	4	0.6	0	0.4	
4	5	1	0	0	
5	3	0.806	0.117	0.075	
5	4	0.657	0.315	0.026	
5	5	0.818	0.142	0.038	

表 59 音质条件概率

父级		概率			
满意度	有趣	2	3	4	5
1	3	0	0.8	0	0.2
2	3	0	0.714	0	0.285
2	4	0	0	0	1
3	3	0.012	0.593	0.182	0.211
3	4	0	0.476	0.214	0.309
3	5	0	0.709	0.161	0.129
4	3	0	0.575	0.121	0.303
4	4	0	0.4	0	0.6
4	5	0	0	0	1
5	3	0	0.639	0.091	0.269
5	4	0	0.454	0.103	0.441
5	5	0	0.641	0.076	0.282

表 60 灵敏条件概率

父级		概率			
满意度	音质	2	3	4	5
1	3	0	1	0	0
1	5	0	1	0	0
2	3	0	1	0	0
2	5	0	1	0	0
3	2	0	1	0	0
3	3	0	0.902	0.043	0.054
3	4	0.017	0.741	0.068	0.172
3	5	0.029	0.75	0.117	0.102
4	3	0	0.809	0.142	0.047
4	4	0	1	0	0
4	5	0	0.687	0.125	0.187

5	3	0.005	0.892	0.045	0.056
5	4	0	0.773	0.094	0.132
5	5	0	0.757	0.142	0.1

表 61 资源条件概率

父级		概率			
满意度	智能	2	3	4	5
1	3	0	1	0	0
1	5	0	1	0	0
2	2	0	1	0	0
2	3	0	1	0	0
2	4	0	0	0.5	0.5
3	3	0.009	0.876	0.014	0.1
3	4	0	0.078	0.588	0.333
3	5	0.018	0.773	0.037	0.169
4	3	0.037	0.703	0	0.259
4	4	0	0	1	0
4	5	0	0.333	0	0.666
5	3	0.01	0.888	0.007	0.093
5	4	0	0.012	0.615	0.371
5	5	0.009	0.631	0.145	0.213

表 62 智能条件概率

父级		概率			
满意度	音质	2	3	4	5
1	3	0	0.75	0	0.25
1	5	0	1	0	0
2	3	0.2	0.6	0	0.2
2	5	0.25	0	0	0.75
3	2	0	0.666	0	0.333
3	3	0	0.718	0.091	0.189
3	4	0	0.62	0.275	0.103
3	5	0	0.573	0.264	0.161
4	3	0	0.761	0.047	0.19
4	4	0	0.25	0	0.75
4	5	0	0.625	0.062	0.312
5	3	0	0.751	0.098	0.149
5	4	0	0.452	0.339	0.207
5	5	0	0.621	0.147	0.23

(七) 小雅智能音箱

表 63 引用条件概率

父级		概率			
满意度	资源	1	3	4	5
4	3	0	0.818	0.181	0
5	3	0	0.922	0.05	0.027
5	4	1	0	0	0
5	5	0	0.833	0.083	0.083

表 64 音质条件概率

父级		概率			
满意度	智能	2	3	4	5
4	3	0	1	0	0
4	5	0	1	0	0
5	3	0	0.909	0.086	0.004
5	4	0.142	0.142	0.571	0.142
5	5	0.009	0.855	0.12	0.014

表 65 灵敏条件概率

父级		概率			
满意度	音质	1	3	4	5
4	3	0	1	0	0
5	2	0.666	0.333	0	0
5	3	0	0.908	0	0.091
5	4	0	0.845	0	0.154
5	5	0	0.625	0.125	0.25

表 66 性价比条件概率

父级		概率			
满意度	1	3	4	5	
4	0	1	0	0	
5	满意度	0.872	0.013	0.102	

表 67 资源条件概率

父级		概率		
满意度	智能	3	4	5
4	3	1	0	0
4	5	1	0	0
5	3	0.994	0	0.005
5	4	0.428	0.142	0.428
5	5	0.98	0	0.019

表 68 智能条件概率

父级		概率		
满意度	有趣	3	4	5
4	3	0.909	0	0.09
5	3	0.811	0.005	0.183
5	4	0.666	0.166	0.166
5	5	0.88	0	0.12

表 69 有趣条件概率

父级		概率		
满意度	性价比	3	4	5
4	3	1	0	0
5	1	1	0	0
5	3	0.953	0.002	0.043
5	4	1	0	0
5	5	0.923	0.025	0.05

(八) 小贝智能音箱

表 70 性价比条件概率

父级		概率			
满意度		1	3	4	5
1		0	0.909	0	0.09
2		0.052	0.947	0	0
3		0.03	0.939	0	0.03
4		0	1	0	0
5		0.01	0.948	0.016	0.033

表 71 有趣条件概率

父级		概率	
满意度	性价比	3	5
1	3	1	0
1	5	1	0
2	1	1	0
2	3	1	0
3	1	1	0
3	3	1	0
3	5	1	0
4	3	1	0
5	1	1	0
5	3	0.986	0.013
5	4	0.857	0.142
5	5	0.965	0.034

表 72 音质条件概率

父级		概率			
满意度	性价比	1	3	4	5
1	3	0.1	0.35	0.35	0.2
1	5	0.5	0.5	0	0
2	1	1	0	0	0
2	3	0	0.722	0.111	0.166
3	1	0	1	0	0
3	3	0	0.741	0.161	0.096
3	5	0	1	1	0
4	3	0.088	0.688	0.2	0.022
5	1	0	0	1	0
5	3	0.023	0.692	0.158	0.125
5	4	0	0.714	0.071	0.214
5	5	0	0.689	0.034	0.275

表 73 智能条件概率

父级		概率			
满意度	音质	1	3	4	5
1	1	0.333	0.666	0	0
1	3	0.25	0.75	0	0
1	4	0	0.857	0	0.142
1	5	0	0.75	0	0.25
2	1	0	0	0	1
2	3	0.23	0.538	0	0.23
2	4	0.5	0.5	0	0
2	5	0	1	0	0
3	3	0.04	0.72	0	0.224
3	4	0	0.8	0	0.2
4	5	0	1	0	0
4	1	0.25	0.25	0	0.5
4	3	0.129	0.709	0	0.161
4	4	0	0.888	0	0.111
4	5	0	0	0	1
5	1	0	0.842	0.052	0.105
5	3	0.013	0.803	0.001	0.181
5	4	0.015	0.816	0	0.167
5	5	0.008	0.61	0	0.38

表 74 资源条件概率

父级		概率			
满意度	智能	1	3	4	5
1	1	0	1	0	0
1	3	0	1	0	0
1	5	0	1	0	0
2	1	0	1	0	0
2	3	0	1	0	0
2	5	0.5	0.5	0	0
3	1	0	1	0	0
3	3	0.12	0.88	0	0
3	5	0.285	0.571	0.142	0
4	1	0	1	0	0
4	3	0.096	0.903	0	0
4	5	0.222	0.777	0	0
5	1	0.272	0.727	0	0
5	3	0.064	0.926	0.007	0.001
5	4	0.5	0.5	0	0
5	5	0.091	0.902	0	0.005

表 75 灵敏条件概率

父级		概率			
满意度	资源	1	3	4	5
1	3	0.045	0.772	0	0.181
2	1	0	1	0	0
2	3	0	0.823	0	0.176
3	1	0	0.4	0	0.6
3	3	0	0.851	0	0.148
3	4	0	1	0	0
4	1	0.2	0.6	0	0.2
4	3	0	0.9	0.025	0.075
5	1	0	0.793	0.015	0.19
5	3	0.003	0.858	0.006	0.131
5	4	0	0.6	0	0.4
5	5	0	1	0	0

表 76 物流条件概率

父级		概率		
满意度	音质	3	4	5
1	1	0.66		0.333
1	3	1		0
1	4	1		0
1	5	1		0
2	1	1		0
2	3	0.846		0.153
2	4	1		0
2	5	1		0
3	3	0.96		0.04
3	4	1		0
3	5	0.666		0.333
4	1	1		0
4	3	1		0
4	4	0.888		0.111
4	5	1		0
5	1	1		0
5	3	0.925		0.074
5	4	0.931		0.068
5	5	0.867		0.132

(九) 听听智能音箱

表 77 性价比条件概率

父级		概率		
满意度	2	3	4	5
3	0.006	0.806	0.172	0.013
4	0.01	0.832	0.145	0.01
5	0	0.872	0.115	0.011

表 78 资源条件概率

父级		概率		
满意度	性价比	3	4	5
3	2	0	0.5	0.5
3	3	0.504	0.26	0.235
3	4	0.46	0.14	0.4
3	5	0.75	0.25	0
4	2	0.666	0	0.333
4	3	0.614	0.197	0.188
4	4	0.625	0.075	0.3
4	5	1	0	0
5	3	0.668	0.148	0.183
5	4	0.62	0.2	0.18
5	5	0.4	0.6	0

表 79 音质条件概率

父级		概率		
满意度	资源	3	4	5
3	3	0.513	0.152	0.333
3	4	0.571	0.042	0.385
3	5	0.289	0.131	0.578
4	3	0.376	0.123	0.5
4	4	0.187	0.208	0.604
4	5	0.16	0.125	0.714
5	3	0.578	0.091	0.329
5	4	0.434	0.115	0.449
5	5	0.32	0.102	0.576

表 80 有趣条件概率

父级		概率		
满意度	音质	3	4	5
3	3	0.882	0.036	0.03
3	4	0.971	0	0.028
3	5	0.941	0.042	0.016
4	3	0.951	0.024	0.024
4	4	1	0	0
4	5	0.961	0.006	0.032
5	3	0.818	0.131	0.05
5	4	0.88	0.023	0.095
5	5	0.87	0.07	0.058

表 81 物流条件概率

父级		概率		
满意度	性价比	3	4	5
3	2	0	0	1
3	3	0.722	0.064	0.213
3	4	0.72	0.06	0.22
3	5	1	0	0
4	2	0.666	0.333	0
4	3	0.758	0.039	0.201
4	4	0.75	0.05	0.2
4	5	1	0	0
5	3	0.79	0.039	0.169
5	4	0.62	0.06	0.32
5	5	0.2	0.2	0.6

表 82 灵敏条件概率

父级		概率			
满意度	物流	3	4	5	
3	3	0.732	0.038	0.229	
3	4	0.444	0.388	0.166	
3	5	0.222	0.523	0.253	
4	3	0.725	0.038	0.235	
4	4	0.5	0.333	0.166	
4	5	0.24	0.648	0.111	
5	3	0.836	0.024	0.139	
5	4	0.789	0.21	0	
5	5	0.301	0.457	0.24	

表 83 智能条件概率

父级		概率			
满意度	灵敏	2	3	4	5
3	3	0.005	0.64	0.131	0.222
3	4	0	0.729	0.166	0.104
3	5	0	0.373	0.089	0.537
4	3	0	0.635	0.123	0.241
4	4	0	0.702	0.148	0.148
4	5	0	0.35	0.14	0.508
5	3	0	0.582	0.101	0.316
5	4	0	0.66	0.1	0.24
5	5	0	0.439	0.181	0.378

附录五 其余产品用户满意度影响因素图

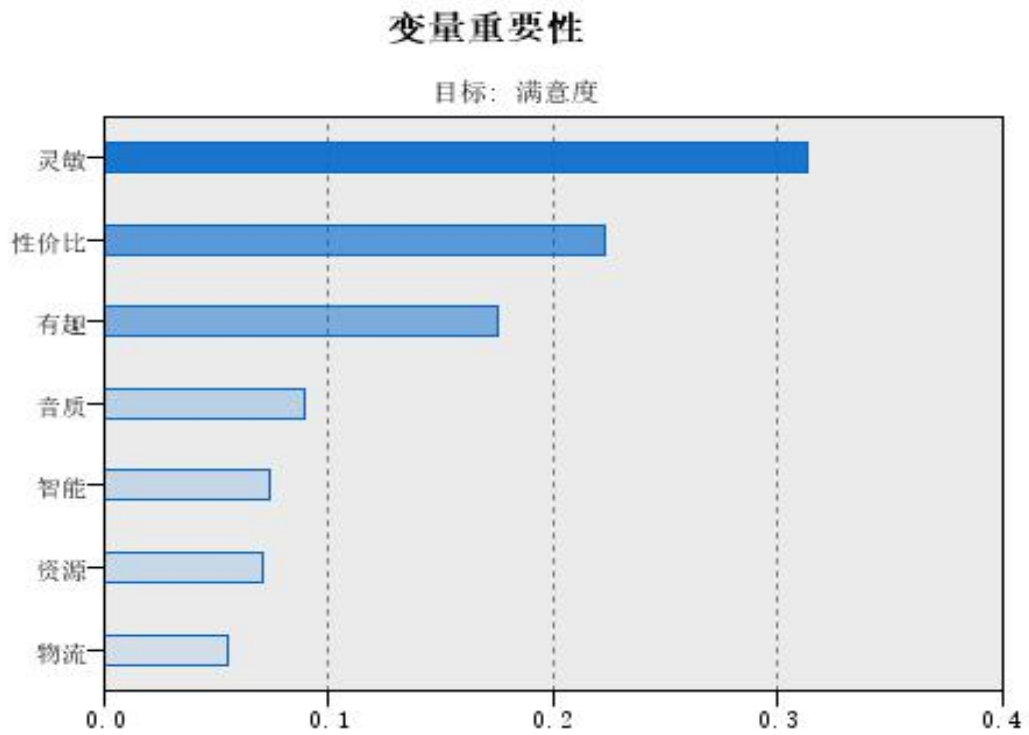


图 48 小度智能音箱各因素重要性排序图

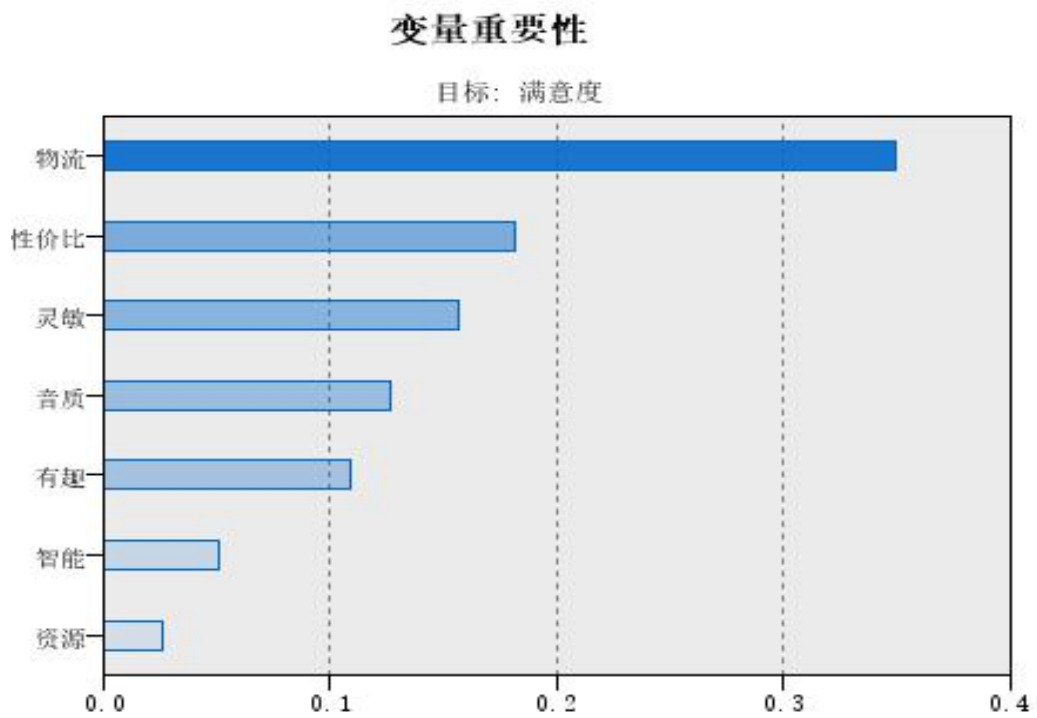


图 49 小爱智能音箱各因素重要性排序图

变量重要性

目标：满意度

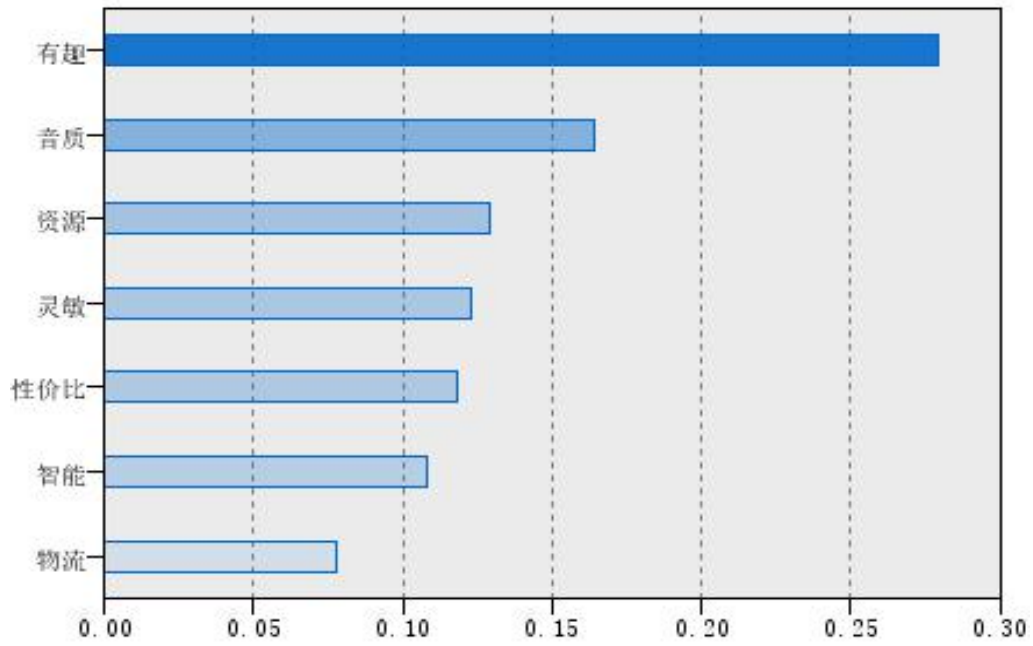


图 50 叮咚智能音箱各因素重要性排序图

变量重要性

目标：满意度

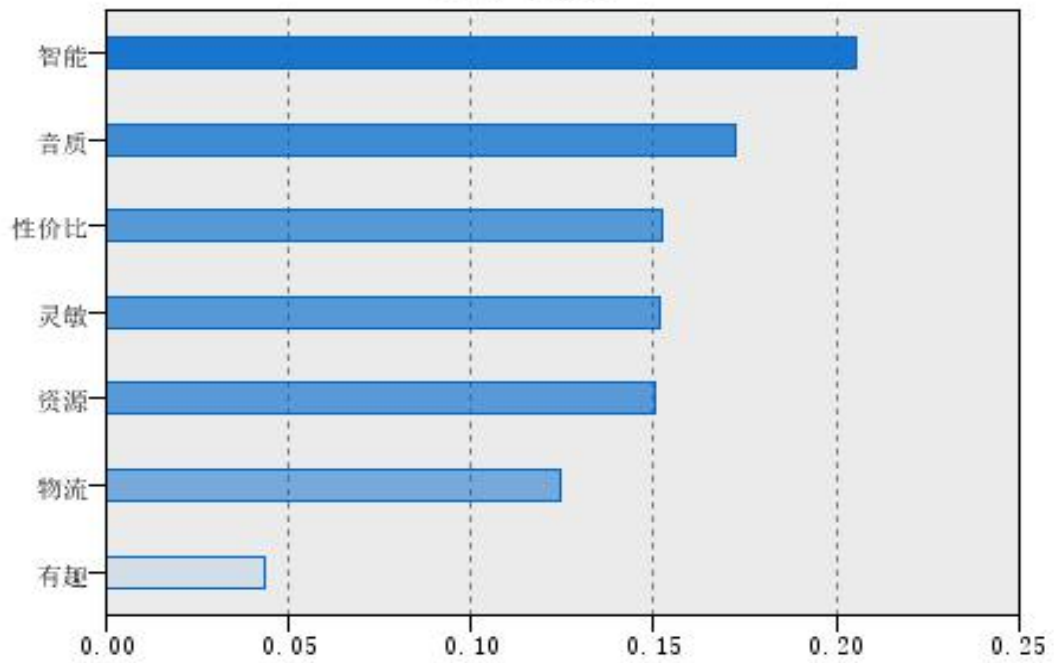


图 51 小艺智能音箱各因素重要性排序图

变量重要性

目标：满意度

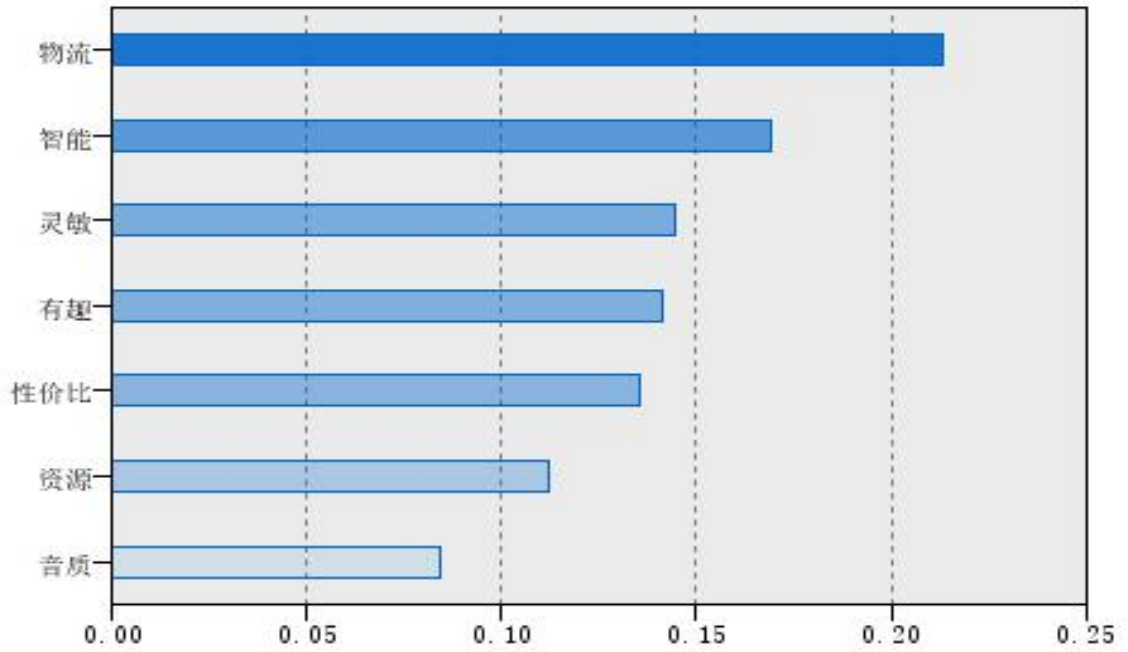


图 52 YOYO 智能音箱各因素重要性排序图

变量重要性

目标：满意度

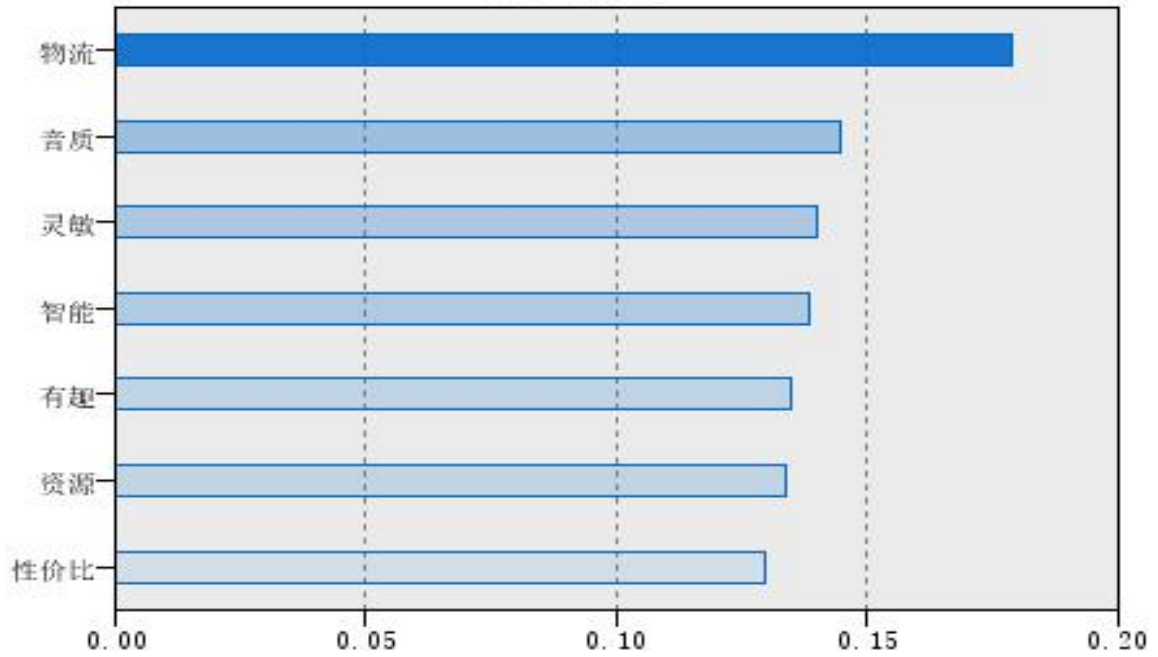


图 53 小雅智能音箱各因素重要性排序图

变量重要性

目标：满意度

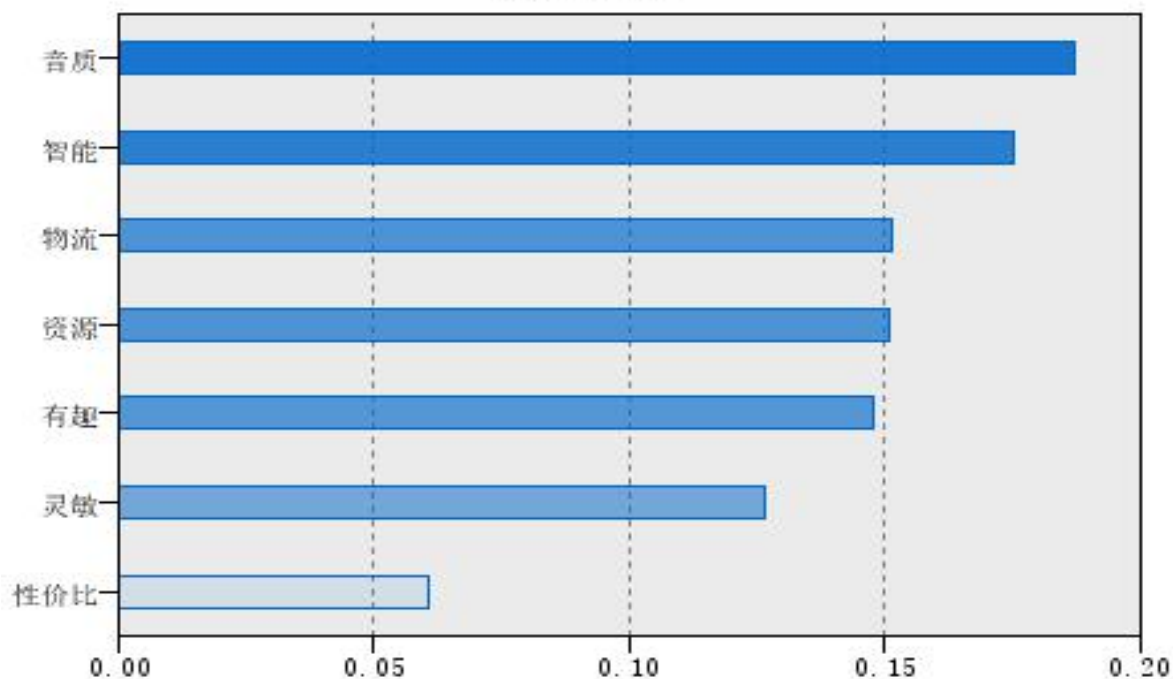


图 54 小贝智能音箱各因素重要性排序图

变量重要性

目标：满意度

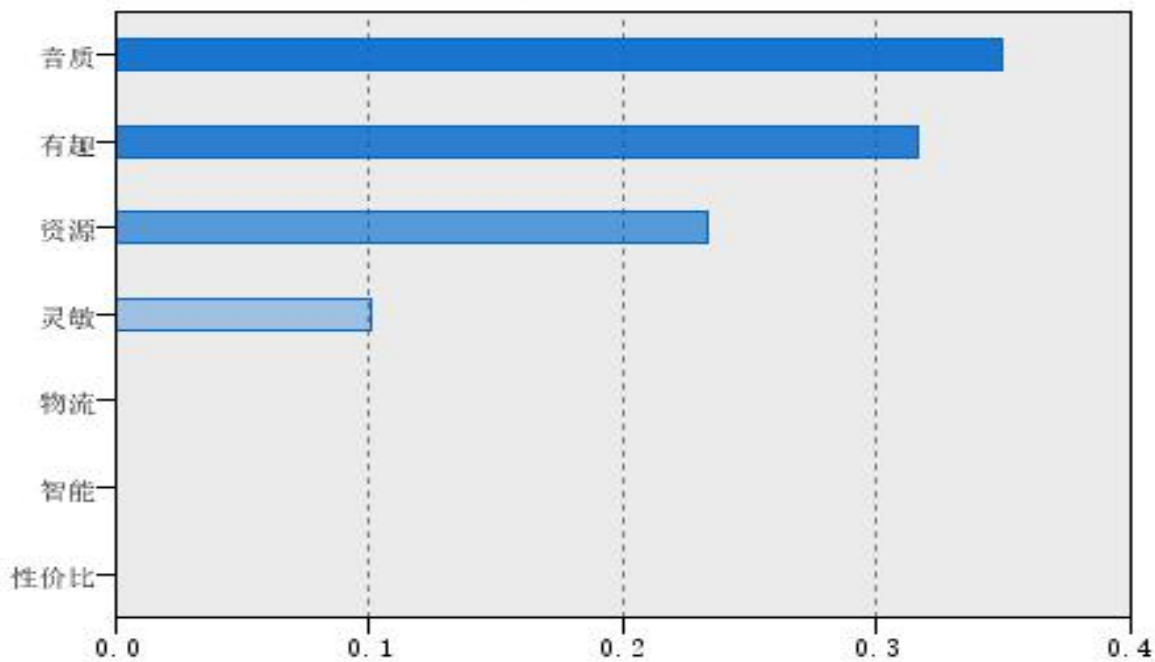


图 55 听听智能音箱各因素重要性排序图

致 谢

本文从论文的选题、写作中，学到了中文文本处理的方法，精进了专业知识，并将所学知识应用到实际问题中。不同模型之间的对比，目的与建模之间的融合改变，都是使我们受益终生的宝贵实践经验。在整个过程中，首先忠心感谢我们的指导老师，感谢老师在比赛中对我们的帮助，让我们小组拓展了专业的知识，学到了更多的实践经验；其次感谢我们的队员，比赛过程中不断鼓励彼此，共同面对困难，不言放弃。