

基于相关性模型与大数据分析的关键词 优化方法研究

文继权

(大连海洋大学 应用技术学院, 辽宁 大连 116300)

摘要:随着计算机网络与信息技术的快速发展,电子商务得到了爆发式的发展,电子商务网店运营成为广大电商从业者所追捧的技术。随着大数据技术的普及应用,电子商务网店运营技术也在不断地推陈出新,基于大数据的关键词筛选优化就是人们所关注的重要技术。相关性模型是电商平台商品搜索排序的基础模型,基于相关性模型的搜索引擎优化(SEO)是目前电子商务数据化运营中进行关键词优化筛选的基本方法。该文将原有方法与大数据分析相结合,从关键词采集、筛选、组合、监测、调换等多维度进行关键词的优化筛选,提出并论述关键词设置的三个筛选优化模型:竞争系数、单品位竞争个数、关键词主体价格区间适应性,并对筛选优化后选用的关键词进行了应用有效性实践监测。大量监测数据表明,使用相关性模型与大数据分析相结合的关键词筛选优化方法得到的商品标题关键词,可以有效提升商品的曝光量、点击率与转化率,实现更高效的网店引流与成交量。

关键词:相关性模型;大数据分析;电商运营;关键词优化;点击率

中图分类号: TP391.9; F724.6; N945.12 **文献标识码:** A **文章编号:** 1673-629X(2024)04-0030-05

doi: 10.20165/j.cnki.ISSN1673-629X.2024.0005

Research on Keyword Optimization Method Based on Correlation Model and Big Data Analysis

WEN Ji-quan

(School of Applied Technology, Dalian Ocean University, Dalian 116300, China)

Abstract: With the rapid development of computer networks and information technology, e-commerce has experienced explosive development. E-commerce online store operation has become a technology pursued by a large number of e-commerce practitioners. With the popularization and application of big data technology, e-commerce online store operation technology is also constantly evolving, and keyword selection and optimization based on big data is an important technology that people pay attention to. The correlation model is the basic model for sorting product searches on e-commerce platforms. Search Engine Optimization (SEO) based on the correlation model is currently the basic method for keyword optimization and screening in e-commerce data-driven operations. We combine the original method with big data analysis to optimize and screen keywords from multiple dimensions such as keyword collection, screening, combination, monitoring, and exchange, and propose and discuss three optimization models for keyword selection: competition coefficient, number of single grade competitions, and adaptability of keyword subject price range. Conduct practical monitoring on the effectiveness of the selected keywords after selection optimization. A large amount of monitoring data shows that using the keyword selection optimization method combining correlation model and big data analysis to obtain product title keywords can effectively improve the exposure of the product click through rate and conversion rate, achieving more efficient online store traffic and transaction volume.

Key words: relational model; big data analysis; e-commerce operation; keyword optimization; click through rate

0 引言

最近几年来,随着行为记录数据化的普及和数据价值挖掘能力的提升,大数据在全球范围内受到热捧,大数据应用渗透到服务业、工业、农业等多个经济领

域,在电子商务零售行业中的应用尤为突出。因为目前各大电商零售平台都已处于平稳运营阶段,并支撑着电子商务高速发展,电子商务数据化运营也越来越多地被各大电商平台所重视,并都已开发出自己有代

收稿日期:2023-08-30

修回日期:2023-12-29

基金项目:2020年辽宁省教育科研经费项目(JL202017)

作者简介:文继权(1978-),男(满族),副教授,硕士,研究方向为关联规则挖掘、大数据分析与应用、电商数据化运营。

表性且较成熟的支持数据化运营的应用产品,如淘宝网的生意参谋和京东商城的京东商智等^[1]。在电子商务运营中传统的商品关键词筛选优化方法—搜索引擎优化(SEO),在实践中存在明显的精准度不足、关键词引流效果不佳等问题,所以传统的SEO方法面临着最新的基于大数据的“数据化运营”的升级改造问题^[2]。该文将原有的基于相关性模型的电商产品关键词筛选优化的传统方法与大数据多维分析方法相结合,优化改造创新基于相关性模型与大数据分析的电子商务数据化运营关键词筛选优化方法,以期为广大的电子商务运营从业者或相关领域的研究人员提供更高效率的电子商务网店运营关键词选词筛词方法。

1 相关性模型与关键词优化

1.1 相关性模型

在电商平台中把影响商品排序的数据统称为因子,因子在商品排名维度中起主导作用,把这些影响商品排名维度的各种因子叫做权重模型,目前主要的权重模型有8类,如图1所示。相关性模型就是权重模型中的一种,又可以分为文本相关性和类目相关性^[3]。

相关性模型主要是指依据商品的归属类目、商品标题等关键词与访客搜索词之间的关联程度,来评定商品在访客搜索时的结果排序。相关性模型是商品在电商平台搜索排序优化的关键,也是其他模型的作用基础。如果基础没有做好,例如相关分类选择错误,关键词选择或排列错误,都会导致商品无法被索引,即使其他模型的评分再高,也无法让商品在搜索综合排序中得到好的曝光^[4]。

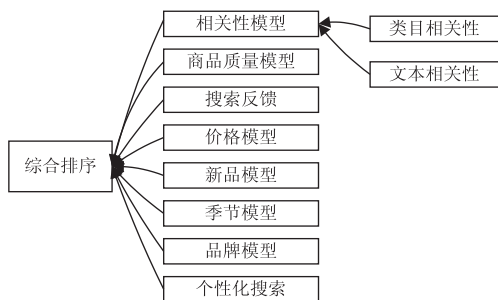


图1 电商平台商品排序权重模型组成

1.2 关键词优化

在相关性模型中,文本相关性是SEO(Search Engine Optimization,搜索引擎优化)工作者最需要熟知的概念。所谓关键词优化就是针对商品标题、标签等使用的关键词,在选词、筛词时,让标题等关键词与买家搜索的关键词最大程度地相关,这样在买家搜索对应的关键词时^[5],卖家包含该关键词的商品在搜索结果中就会有更好的搜索排序,进而商品在网店中就会有更好的曝光率、点击率,甚至是转化率和成交量。

2 基于大数据分析的关键词筛选优化

电子商务运营中产品关键词的筛选优化主要包括五个基本步骤:关键词采集、关键词筛选、关键词组合、关键词监测与调换^[6]。关键词优化流程如图2所示。

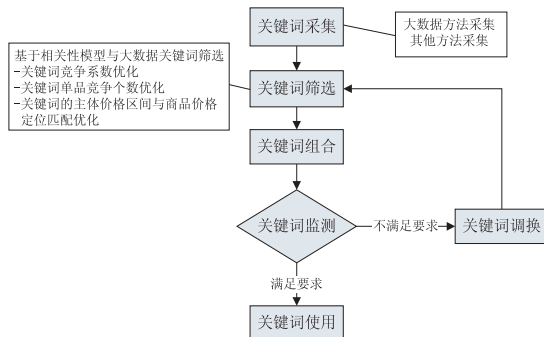


图2 关键词优化流程

2.1 关键词采集

通过特定的方法,收集与本店商品相关的词、词组或长尾词,并形成商品的关键词集合,待设计商品的标题、直通车推广标题或卖点时使用。关键词采集是SEO工作者在关键词优化工作中的第一步,也是非常重要的一步。关键词采集的途径主要有4种,该文将主要研究采用大数据分析的方法进行关键词采集^[7]。

2.1.1 平台前端页面采集

在京东商城和淘宝网的首页都会有搜索框,当买家在搜索框内输入关键词时,搜索框下面就会出现与买家输入关键词相关的推荐搜索关键词,这些关键词都是通过大数据统计并按照搜索热度由上到下以下列表的形式排序显示的^[8]。前端搜索页采集就是要从这些热搜词中根据自己产品的具体情况,从上到下采集和自己产品高度相关的关键词待用。

2.1.2 用户反馈采集

搜索结果页中商品的标题、广告语(卖点)中影响商品转化率的文本,叫做转化率关键词,可以通过查阅用户的反馈,为商品找到可应用在标题或者广告语(卖点)中形成转化率的关键词。获取用户反馈的途径包括:用户评价、标签和即时通信软件(旺旺、京麦等)的反馈信息,本例中以“毛衣”为核心词,从用户反馈中采集到的关键词主要有:百搭、大V领、优雅低调等。这样获取关键词的方法就是用户反馈信息采集。

2.1.3 市场细分词采集

电商平台中很多品类都有买家对商品进行搜索时的统一习惯,这些习惯不是规则,而是通过买家对商品的认知养成的。这些词是商品关键词的有力补充,也是商家的必争词,更准确地说,这些词属于市场细分词,如:小清新、学生装、妈妈装等。不同品类各有不同的市场细分词,它的采集需要运营者对市场有很清晰的洞察,同时可以参考搜索时平台在“相关推荐”处的

细分词。

2.1.4 大数据分析采集

目前各大电商平台基本都有自己配套的大数据分析工具,如淘宝网的生意参谋,京东商城的京东商智等,在这些大数据分析工具软件中都有海量的电商平台交易数据,设有行业关键词的“搜索热度”“点击率”“转化率”等的数据指标的大数据统计分析功能^[9]。在这些大数据分析工具中只要输入商品的核心关键词,工具软件就会把与核心关键词相关的关键词的大数据统计信息(带有全面的数据分析指标)展示给运营者^[10],也可以下载全部统计数据,这样运营者就可以依据这些“统计数据”来采集自己需要的商品关键词。如图 3 所示,就是利用某一平台的大数据工具,输入核心关键词“毛衣”,之后工具软件显示相关搜索词及对应的大数据统计的各数据指标的信息。可以从这样的关键词及数据指标表中,依据“搜索热度”“点击率”“转化率”等数据指标,来采集适合商品的关键词。

搜索词	搜索人气	搜索热度	点击率
毛衣女宽松外穿	477,660	1,208,155	106.89%
毛衣	475,170	1,098,426	118.59%
毛衣女	434,154	1,136,420	119.81%
毛衣男	337,179	772,115	96.60%
女士毛衣	334,917	795,619	84.49%
毛衣女2019新款	323,545	865,144	106.86%
毛衣女秋冬外穿	265,066	669,325	106.05%

图 3 大数据工具关键词采集词表截图

2.2 基于大数据分析的关键词筛选优化

按照前面的关键词采集方法,将采集到的关键词制作成一个关键词采集表。例如,选用“毛衣”作为核心关键词形成关键词采集表,图 3 为大数据采集关键词形成词表的一部分。

关键词采集完成后,需要通过基于大数据分析的关键词筛选优化方法,从此表中筛选出更适合具体商品使用的关键词。关键词筛选优化过程中,对关键词进行筛选的最重要原则是:降低关键词的竞争难度、提高关键词的有效覆盖率、增加销售商品的曝光量。按照这个原则,在关键词筛选时需要做三个方面的优化,即三个关键词筛选优化方法,它们分别是:关键词的竞争系数优化、关键词的单品竞争个数优化、关键词的主体价格区间与商品价格定位匹配优化^[4]。下面分别依据这三个优化方法来进行关键词的优化筛选。

2.2.1 关键词竞争系数优化筛选

关键词竞争系数表示具体关键词下商品竞争

难易程度,计算公式为:

$$G_1 = \frac{S_1}{S_2} \quad (1)$$

其中, G_1 为关键词竞争系数; S_1 为关键词下的商品数量; S_2 为关键词的搜索量(搜索热度)。

竞争系数越小,表示竞争难度越低;竞争系数越大,表示竞争难度越高。因此运营人员要根据竞争系数,找出竞争环境较好(易)的关键词^[11]。如图 4 所示,就是按照关键词竞争系数降序排序筛选出来竞争系数较小的关键词(在表中关键词竞争系数字段已经设置为浅色显示)。

采集方法	关键词	关键词下商品数量	关键词搜索热度	竞争系数
用户反馈	低调优雅	45	279	0.16
大数据平台	jk毛衣	7360	18454	0.40
大数据平台	露背毛衣	5733	11469	0.50
大数据平台	彩虹毛衣	36282	37842	0.96
大数据平台	毛衣	1903579	1305057	1.46
前端热搜词	毛衣外套	912009	354335	2.57
大数据平台	毛衣外套	912009	354335	2.57
前端热搜词	毛衣女宽松	711020	213094	3.34
市场细分	妈妈装	1062046	224057	4.74
大数据平台	儿童毛衣	347964	70426	4.94
大数据平台	手工毛衣	94084	15022	6.26
前端热搜词	毛衣女	1365254	217780	6.27
大数据平台	粉色毛衣	233591	35270	6.62
大数据平台	婴儿毛衣	444028	60782	7.31
前端热搜词	毛衣外套女	927684	116739	7.95
大数据平台	开衫毛衣	782619	89255	8.77
用户反馈	白搭	1231614	136525	9.02
大数据平台	一字领毛衣	87691	9299	9.43
大数据平台	宽松毛衣	692495	68370	10.13
大数据平台	v领毛衣	886347	70147	12.64

图 4 竞争系数优化筛选关键词结果

2.2.2 关键词单品竞争个数优化筛选

关键词单品竞争个数是指平台具体商品在进入具体关键词(关键词搜索结果)下的首页,面对的竞争对手的个数,其公式为:

$$G_2 = \frac{S_3}{S_4} \quad (2)$$

其中, G_2 为关键词单品竞争个数; S_3 为关键词下的商品数量; S_4 为首页坑位数。

通过单品位竞争个数的优化筛选,可对关键词的竞争难易程度进行二次筛选,选出市场竞争力更小的商品关键词。不同的平台首页的坑位数是略有区别的,该是以淘宝网的首页坑位数(48个)为例进行计算的^[12],按照关键词单品竞争个数的计算值对关键词进行升序排序,并获取到单品竞争个数列前 20 的关键词表,结果如图 5 所示(单品竞争个数列也已经设置为浅色显示)。

2.2.3 关键词的主体价格区间与商品价格定位匹配优化

各关键词商品价格定位的计算公式为:

$$G_3 = \{P_1, P_2, \dots, P_n\} (P_n \in G_3, P_{\min} \leq P_n \leq P_{\max}) \quad (3)$$

其中, G_3 为各关键词商品价格定位; P_{\min} 为关键词的主体价格最小值; P_{\max} 为关键词的主体价格最大值。

根据商品的价格模型,当用户搜索关键词时,平台

会优先将商品定价与用户搜索的关键词商品的主体价格区间(60%以上同类产品的价格)相匹配的商品展示^[13]。进行关键词的主体价格区间与商品价格定位匹配优化时,首先在相应的平台查询出所有商品关键词的主体价格区间,之后根据商品的价格定位,去掉关键词主体价格区间不包含商品定位价格的关键词。该

文将核心词为“毛衣”的这个商品的价格定位为80元(这个价格定位是运营者根据自己商品的具体情况来确定的),按照优化规则优化后得到最后的关键词优化结果,如图6所示。商品价格定位与关键词的主体价格区间匹配的,已经将关键词价格区间字段设置为浅色显示。

采集方法	关键词	关键词下商品数量	关键词搜索热度	竞争系数	单品位竞争个数
用户反馈	低调优雅	45	279	0.16	0.94
大数据平台	露背毛衣	5733	11469	0.50	119.44
大数据平台	jk毛衣	7360	18454	0.40	153.33
大数据平台	彩虹毛衣	36282	37842	0.96	755.88
大数据平台	一字领毛衣	87691	9299	9.43	1826.90
大数据平台	手工毛衣	94084	15022	6.26	1960.08
大数据平台	粉色毛衣	233591	35270	6.62	4866.48
大数据平台	儿童毛衣	347964	70426	4.94	7249.25
大数据平台	半高领毛衣	385204	6922	55.65	8025.08
大数据平台	紫色毛衣	400490	28428	14.09	8343.54
大数据平台	条纹毛衣	418528	19412	21.56	8719.33
大数据平台	婴儿毛衣	444028	60782	7.31	9250.58
大数据平台	短款毛衣	645522	45390	14.22	13448.38
用户反馈	大V领	657483	50446	13.03	13697.56
大数据平台	宽松毛衣	692495	68370	10.13	14426.98
大数据平台	大毛衣	702699	13779	51.00	14639.56
前端热搜词	毛衣女宽松	711020	213094	3.34	14812.92
大数据平台	加厚毛衣	772385	3524	219.18	16091.35
大数据平台	镂空毛衣	775126	60902	12.73	16148.46
大数据平台	开衫毛衣	782619	89255	8.77	16304.56

图5 单品位竞争个数优化筛选结果

关键词	关键词下商品数量	关键词搜索热度	竞争系数	单品位竞争个数	关键词价格区间
低调优雅	45	279	0.16	0.94	89-209
露背毛衣	5733	11469	0.50	119.44	44-91
jk毛衣	7360	18454	0.40	153.33	68-138
彩虹毛衣	36282	37842	0.96	755.88	59-168
一字领毛衣	87691	9299	9.43	1826.90	58-188
手工毛衣	94084	15022	6.26	1960.08	169-433
粉色毛衣	233591	35270	6.62	4866.48	69-191
儿童毛衣	347964	70426	4.94	7249.25	40-80
半高领毛衣	385204	6922	55.65	8025.08	49-208
紫色毛衣	400490	28428	14.09	8343.54	60-184
条纹毛衣	418528	19412	21.56	8719.33	50-162
婴儿毛衣	444028	60782	7.31	9250.58	38-76
短款毛衣	645522	45390	14.22	13448.38	40-139
大V领	657483	50446	13.03	13697.56	59-164
宽松毛衣	692495	68370	10.13	14426.98	50-168
大毛衣	702699	13779	51.00	14639.56	80-215
毛衣女宽松	711020	213094	3.34	14812.92	66-159
加厚毛衣	772385	3524	219.18	16091.35	54-136
镂空毛衣	775126	60902	12.73	16148.46	46-172

图6 关键词的主体价格区间与商品价格定位匹配优化结果

图6中“竞争系数”“单品位竞争个数”和“关键词价格区间”三个字段都被设置为浅色的关键词,就是最后选用的关键词,且已经用加粗显示。

2.3 关键词的优化组合

按照基于相关性模型与大数据分析筛选出商品关键词,形成较精准的商品标题设备选关键词词库,但卖家与运营者是否使用以及如何使用,还要根据所卖的具体商品及商品的具体特性决定。同时,如何对所筛选出的关键词排序也是一项重要的工作。具体组合排序规则如下:

关键词组合排序参考公式:中文品牌(英文品牌)+产品名(核心词、热搜词)+产品特点(属性、功能)+类目词+促销词。

关键词组合排序参考规则:

关键词词距:关键字/词间的距离,与访客搜索词精确匹配且字间距越小越好;

关键词顺序:关键字/词间的顺序,与访客搜索词

的顺序相同分高,逆序减分;

关键词长短:关键字/词间的长度(字数或字符数),越精准越短越好。

由如上四个规则可知,在根据备选词进行最后关键字词的选取、组合与排序时,同等优化数据指标条件下,要选择短词,关键词排序要尽量与核心词词序相同,词距要尽量小^[4]。

2.4 关键词监测

运营人员还需要对优化后的关键词的使用效果进行监测。使用平台的“关键词分析”功能,对关键词的到店引流情况、点击率及成交转化率等数据进行分析监测。之后根据监测结果,采取对应的措施。

2.5 关键词调换

根据上一步监测的结果,如在监测周期内这些监测指标都不够理想,就需要考虑重新筛选关键词并对原来的关键词进行调换^[14]。之后再对调换后的关键词进行监测,如此反复进行,直到所有的关键词都达到

了理想的监测结果,才会把关键词应用到实际的网店产品标题或标签中发布使用。

3 实践监测结论

通过关键词监测对比了前文中以毛衣为核心词,使用普通方法进行的关键词优化得到的产品关键词和使用“基于相关性模型与大数据分析”方法进行关键词优化得到的关键词,在网店产品展现量、点击率与转化率的数据表明:通过“基于相关性模型与大数据分析”方法筛选优化得到的商品标题关键词,可有效提高商品的展现量 18% 以上,点击量 10% 以上,转化率 2% 以上,对比如图 7 所示。这些检测数据证明了基于相关性模型与大数据分析的关键词优化方法的科学性与有效性。

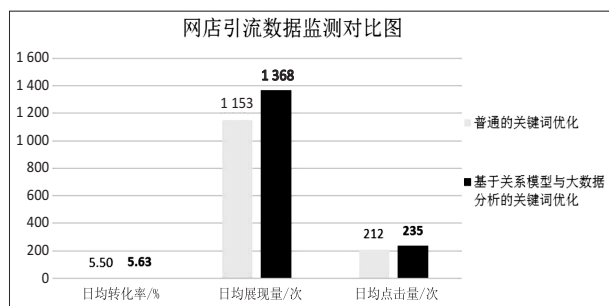


图 7 网店日均引流数据的监测对比

4 结束语

各大电商平台网店的数据化运营的应用越来越深入与广泛,可以说已经延伸到电商运营的各个环节,而各大电商平台推出的大数据分析工具也为电商数据化运营提供了有力的支撑。该文聚焦的“基于相关性模型关键词筛选优化”方法是电商运营中的核心技能之一,而该项技能现已经与大数据分析软件融为一体,变成了“基于相关性模型与大数据分析的关键词筛选优化”。该文正是基于相关性模型^[15],结合大数据分析对关键词筛选优化过程进行改进创新,并做了较详尽的论述,以期为从事电商数据化运营的朋友们提供更科学高效的关键词筛选优化方法,以获取更多和更加精准的访问流量,提高网店的运营效果。

参考文献:

- [1] 李 健,王小蒙. 数据挖掘与分析在网站运营管理中的应用[J]. 科技传播,2017,9(20):96-97.
- [2] 罗燕君. Excel 数据挖掘技术在电商运营预测分析中的应用[J]. 中国管理信息化,2017,20(19):140-143.
- [3] 闵 敏. 电商运营数据分析[M]. 北京:高等教育出版社,2020:198.
- [4] 数据创新组. 京东平台数据化运营[M]. 北京:电子工业出版社,2020:206.
- [5] 姚孝生. 基于大数据技术的电商营销策略研究[J]. 成都工业学院学报,2023,26(4):86-90.
- [6] 许 宁,高 飞. 无线网络电子商务客户流失预测仿真[J]. 计算机仿真,2018,35(9):475-479.
- [7] 武小军,孟苏芳. 基于客户细分和 AdaBoost 的电子商务客户流失预测研究[J]. 工业工程,2017,20(2):99-107.
- [8] 杨 力. 基于在线序列优化极限学习机的电子商务客户流失量预测模型[J]. 南京理工大学学报,2019,43(1):108-114.
- [9] 朱帮助. 基于 SMC-RS-LSSVM 的电子商务客户流失预测模型[J]. 系统工程理论与实践,2010,30(11):1960-1967.
- [10] LI Lei,CHI Ting,HAO Tongtong,et al. Customer demand analysis of the electronic commerce supply chain using big data[J]. Annals of Operations Research,2018,268(1/2):113-128.
- [11] ZHANG Bo, ZHANG Lin. Precise marketing of precision marketing value chain process on the H group line based on big data[J]. Journal of Intelligent & Fuzzy Systems,2018,35(3):2837-2845.
- [12] ZHANG Jianxi,ZHANG Changfeng,YU Huaizhi. Research on e-commerce intelligent service based on data mining [C]//MATEC web of conferences. Nanjing:[s. n.],2018:456-460.
- [13] 韦辉华. 基于搜索相关性的移动 App 排序算法及应用[D]. 哈尔滨:哈尔滨工业大学,2021.
- [14] 成保梅. 电商时代网店的数据化运营策略探讨[J]. 商展经济,2022(23):51-53.
- [15] 于超宁. 大数据视域下期刊关键词的应用分析[J]. 采写编,2022(1):126-127.