

# 协同过滤算法在编辑业务创新中的应用： 可能与可为

王鹏涛

**摘要：**从逻辑层面论协同过滤算法可以在选题策划、编辑校对、宣传促销、读者服务、市场调查等业务领域发挥独特功能，可以辅助决策者选择创新方案和规避战略风险，可以提升管理层业务统筹和资源整合的绩效。从操作流程角度出发，阐释协同过滤算法在编辑业务创新中应用的机理，探讨应用过程中可能遇到的问题和破解对策。

**关键词：**协同过滤算法；编辑业务创新；读者导向；综合绩效

**作者简介：**王鹏涛，男，副教授，管理学博士。（南京大学信息管理学院，江苏南京，210093）

**中图分类号：**G232      **文献标志码：**A      **文章编号：**1008-6552(2014)04-0079-05

**DOI:**10.13628/j.cnki.zjcmxb.2014.04.014

协同过滤算法在商务领域主要用于产品和服务的智能推荐，在电子商务中被广泛接受且增长前景良好。其原理大致如下：利用某个兴趣相投、拥有共同经验的消费群体的喜好来推荐用户感兴趣的信息，用户借助合作机制对信息做出回应（如评分），这些数据被记录下来以达到过滤的目的，进而帮助其他人筛选信息，回应不限于特别感兴趣的，不感兴趣的也很重要。协同过滤可分为评价（rating）和社会过滤（social filtering）两类。社会过滤（又称合作过滤）是指根据顾客的交易历史，并从具有相似消费经历的顾客群的交易记录去为这个顾客推荐其“可能喜欢的新品种”，亦即借由社交群体的偏好为个人提供信息、商品等推荐服务。除个性化推荐之外，协同过滤系统还可以自动计算喜好的强弱进而去芜存菁，以使过滤的内容更有凭据，虽然基于模型的协同过滤不能保证绝对准确，但由于加入了强弱的评比，从而能使协同过滤的应用更加广泛，除电子商务之外，还在信息检索、音乐影视收藏、书刊订阅等领域得到广泛应用。<sup>[1]</sup>

协同过滤算法是目前最有前途的个性化推荐技术之一，能有效提高顾客的感知价值和消费体验，因此，它引起了管理学、消费者行为学、信息科学、计算数学和认知科学等多个学科的关注，并取得了一系列值得重视的成果。然而，在编辑出版学领域却没有得到应有的关注和重视。笔者认为，协同过滤算法能提高编辑在数字环境下对选题进行动态估值和深度挖掘的效率，基于大量样本统计和数据挖掘的过滤机制在编辑业务中的应用，不仅可以借助群体智慧大幅提高出版物内容的质量，增加读者的满意度，优化编辑工作绩效，而且还可以让出版企业在瞬息万变的亚稳定环境中更好地驾驭市场的复杂性，进而建立和维持独特的竞争优势。本文拟从数字时代编辑工作的实际困境出发，建立基于协同过滤的编辑业务创新框架，以指导编辑业务的转型升级。

## 一、协同过滤算法在编辑业务创新中应用的逻辑原理

编辑业务创新是出版人必须面对的永恒课题，在任何时代编辑业务都需要借助技术创新、个人探索、团队协同等达成理想目标。在数字时代，编辑创新并非一个简单的工作流程信息化的问题，它有

① 基金项目：2012年江苏省社科基金项目《江苏出版业数字化转型问题研究》（12TQC011）；2012年国家社科基金项目《新媒体环境下阅读行为嬗变与国民阅读素养提升策略研究》（12CTQ015）的研究成果。

着更为复杂和深刻的内涵,比如捕捉新的选题和信息源、迅速更新内容、协助开展网络营销、妥善处理读者抱怨、满足用户的个性化体验诉求,等等。协同过滤在编辑业务创新中应用的逻辑原理可以从以下几个方面来认识:

### (一) 协同过滤推荐系统本质上与编辑活动的基本原理相通

协同过滤以信息的筛选优化为目标,这与编辑活动的核心本质相通。虽然学界关于编辑本质的探讨见仁见智,存在“文化缔构说、策划说、审理说、优化说、选择说”等多种流派,但是从编辑发展史的角度来分析,在既定时空环境下,依据特殊的价值标准和审美视角,对人类创造和积累的文化成果进行甄别取舍、择优汰劣,以提高社会整体文化水平,促进人类文明健康发展,这才是编辑活动贯穿始终的内在规律。无论社会经济、科学技术和商业模式如何演变,编辑活动的本质都是一以贯之的。

市场经济兴起后,编辑工作除了考虑文化增值、意识形态导向功能之外,还必须坚持市场导向,考虑读者的兴趣爱好、阅读习惯和其他诉求。在买方市场格局下,读者导向的理念是经营者在竞争环境中必然会做出的不二选择。在实践中,编辑人员越来越主动地与读者对话,借助各种途径理解其真实需求,通过即时通讯工具与读者交互协商,让读者参与到选题策划乃至整个出版流程中来,根据其意见设计、优化产品。这种趋势既与社会民主进步有关,也是经营者盈利的内在需要。在大数据时代,随着数据存储和分析技术的完善,这种面向读者的编辑理念与协同过滤在电子商务中的应用机理高度契合,那就是要利用网络中的读者数据,准确辨认其个性需求,提升读者的参与度,调节读者的消费体验,增加读者的认同度和忠诚度,建立柔韧持久的互惠关系。

就协同过滤算法而言,不论是基于用户的协同过滤、基于项目的协同过滤,还是基于模型的协同过滤,都是以一定数量的顾客资料为依据,以前后因果关联为假设,对读者的兴趣和需求进行量化分析和理性预测,然后将筛选出的内容进行有针对性的智能推荐。两相比较不难发现,协同过滤和编辑工作本质上具有较高程度的重叠领域,因此,将协同过滤算法应用到编辑业务创新中是切实可行的。

### (二) 协同过滤算法的创造性应用可以大幅改进选题策划的效率

协同过滤以代表用户兴趣的信息为基础,通过相似度的计算生成推荐结果,比如豆瓣网推出的基于用户品味的“豆瓣猜”推荐服务,上述机理对编辑选题有极高的参考价值。从脑科学的角度看,选题策划本质上是一种决策。面对复杂的市场环境,考虑到人类理性的局限性,为提高预测精度和决策效率,借助智能辅助系统提高决策者的信息处理能力是必需的。协同过滤的基础是与用户兴趣最近邻的信息记录,包括顾客的人口统计变量、浏览痕迹、交易频次和反馈评价、分享转载等,这些信息通过专业的分类和挖掘,可以呈现出一定规律,用以指导编辑人员开展后续工作,降低经营风险。

编辑策划是出版业务中是最具创意的环节,在传统时代其效率取决于编辑人员的智慧、经验和职业敏感性,为了提高成功几率,编辑人员会通过与发行人员沟通、直接开展市场调查、分析竞争对手表现等方式来撰写和完善选题方案。在大数据时代,协同过滤算法可以大幅改进该业务环节的效率。一方面,协同过滤算法更加客观精确,避免了个体思维定势、视野狭隘造成的不良影响,能关注到读者的价值诉求和情感偏好,减少信息洪流给编辑人员带来的压力,减轻过度离散的信息对人类创造力的抑制,使其从信息超载的困境中解脱出来。另一方面,协同过滤算法的应用又不能离开编辑的主观判断,这是因为虽然协同过滤会给出一定的运算结果,但是上述建议却不一定与企业的资源运筹力相匹配,或者不符合战略发展的需要,所以最终取舍必须依赖人的思考和决断。协同过滤算法在编辑工作中的应用应该以读者的价值诉求为导向,以提高编辑业务的整体绩效为旨归,既要重视技术创新带来的效率改进,又要充分发挥人的创造性和反思能力。

### (三) 协同过滤算法的应用是新形势下实现人文与科技有机结合以应对复杂市场变化的内在需要

自古以来,出版活动始终都是在人文精神与科技创新的耦合驱动下不断进步的,在数字时代这一

规律表现得更加明显。从全局分析,协同过滤仅仅在业务流程的某些方面能给编辑人员提供支持,因此,协同过滤的应用并不必然带来出版业务整体绩效的提升。这是因为协同过滤在产品开发 and 利润增加方面的确有效,但从长远看来,这种改进或许并不符合战略的需要,因此,如何通过批判性的省思避免上述失误,如何协同各职能部门的业务以保证读者获得最大的价值满足才是理论界应该探索的重点。在经营实践中,协同过滤在编辑工作中的应用可能会导致出版企业与竞争对手的产品结构趋同,这一方面需要编辑人员积极创新,对读者的层次结构进行深度细分,赋予产品以差异化的特征,即便是相似主题的出版物也应在内容整合和附加价值上具备独特品质,另一方面不同职能部门的匹配至关重要。任何经营创新都是系统工程,不同部门只有从价值增值最大化的目标出发,与其他部门实现要素、人员、信息、技术的有机协同,才能将选题方案落实到位,为读者提供最佳的阅读服务。简言之,编辑工作创新要实现人文与科技力量的融合,必须在引入协同过滤的过程中,尽力避免该算法可能带来的思维趋同的陷阱,同时又要通过内部有效协作提高价值增值效率,如此以来即便与竞争对手的选题相似,出版企业依然可以通过营销努力提高让渡价值,以个性化的效用满足吸引和留住读者。

## 二、协同过滤算法在编辑业务创新中的应用机理: 基于过程的设计与论证

协同过滤是一种辅助性 IT 技术工具,它在编辑业务创新中的功能发挥程度端赖专业人员的领悟与应变。协同过滤在编辑工作中的应用机理大致如下:

### (一) 搜集、分类和整理与读者兴趣相关的数据

协同过滤算法以大量与用户兴趣相关的数据为量化分析的基础,因此,读者数据的抓取、分类和整理是其在编辑业务中应用的起点。对编辑人员而言,首先要从理念上适应大数据时代精确化分析的趋势,在重视个人知识积累、技能培训的同时,主动回应数字化生存的挑战,关注目标读者群的阅读兴趣及其演变脉络,及时洞悉消费心理和消费行为的变化。在数字时代,读者需求差异加大、竞争对手层出不穷、市场博弈策略多元,编辑业务创新关乎出版物的核心价值,因此,编辑人员必须面向终端读者,以他们的兴趣和诉求为逻辑起点开展工作。对协同过滤算法而言,读者的邮箱地址、手机号码、消费动机、交易金额和购买日期、用后反馈、对产品和品牌的了解程度、口碑传播与推荐情况、媒体接触习惯等资料都是非常重要的数据,需要广泛而持续地收集、整理,同时要从读者、运营和商品三个维度去分析上述数据,以透彻地了解读者需求的细节性特征。

### (二) 甄别协同过滤算法的运算结果,判断选题价值大小,设计可行的出版方案

面对协同过滤算法的结果(若干备选的内容主题、有价值的分析角度、有意义的参照系等),编辑人员应该运用学识、经验和直觉对其估值,同时虚心听取同事的意见,结合组织的战略、资源和实力等制定最终方案。在此过程中,编辑人员必须始终保持思维的独立性与批判性,对来自技术工具的支持既不排除也不迷信,对新生事物足够敏感,同时又要具有整合出版资源的知识、经验和技能。此外,由于竞争企业获取的资料可能与出版企业相似,因此,协同过滤对读者潜在兴趣和需求发展动向的判断或许是趋同的,在这种情况下,如何通过独特的价值组合去最大限度地满足读者就要考验编辑人员的创造力和想象力了。

### (三) 以读者为中心稳健扎实地开展编辑创新

及时高效地与读者沟通,调整选题方案,将读者反馈和意见吸纳到出版物制作过程中,以改进产品和服务的品质。在这个环节中,读者和其他利益相关者对出版企业的认同和情感投入是其参与创新的基础,只有编辑人员熟练掌握沟通技巧、善于引导组织,才能使读者持续处于激发状态,主动发挥主观能动性,对出版物设计、制作和传递提供有益的意见,这样出版企业就可以利用读者贡献的创意提升让渡价值,进而增加读者的黏性和忠诚度。为了动态优化协同过滤的分析结果,编辑人员还应在最

短时间内将读者反馈转换成可以交互的电子数据,快速更新读者资料以保证数据的准确性和时效性,从而为后续工作提供有力支持。

#### (四) 发挥协同过滤算法在新品宣传促销方面的作用

协同过滤算法关注的是用户兴趣之间的关联性,根据社会学和心理学的研究,相似社会群体在价值观念和消费行为上会有相近的特征,因此,编辑人员应该积极运用新兴媒体开展宣传促销。电子商务的高速发展对传统的出版物零售活动构成重大冲击,仅就宣传促销而言,在选题方案确定后,编辑人员就需要统筹营销传播工作。协同过滤算法可以帮助编辑人员拟定营销传播计划,选择媒体渠道、受众范围和促销方式等,从竞争策略的角度分析,可计量、立体化、互动型的营销传播可以获得更多反馈,而且网站转换率也会增加。按照迭代优化的思路,在销售结束后,编辑人员还应根据读者反馈(评价、抱怨和分享等)分析编辑工作的绩效目标有没有达成,同时对于有价值的反馈信息(如阅读兴趣的变化、阅读习惯的演变)进行记录,以便在后续工作中发挥参考作用。

编辑创新意味着要打破原有操作程序,重构业务流程,这种新流程可以包容新的技术工具、新的分工体系、新的协同机制,同时还可以帮助编辑人员与其他部门、终端读者进行沟通,打破思维局限,从海量信息中发现最具价值、最有趣味、最为可行的选题。此外,出版企业不同部门还要在项目实施过程中不断试错、权衡,实现人员与技术的磨合,及时总结经验教训,固化为可重复使用的规章制度。以上分析说明,协同过滤算法在编辑业务创新中可以起到重要作用,可以贯穿出版活动全程,还可以帮助编辑实现其在文化创意方面的追求。

### 三、需要注意的问题和对策

协同过滤的最大优点在于算法简单、易于实现、执行效率高,同时预测的准确性高。协同过滤在编辑创新中的应用必须审慎考虑和解决若干问题。例如,从系统论的视角分析,协同过滤作为一种技术创新,必须与组织战略、资源配置协同演进。从营销功能的角度分析,协同过滤与数据库营销等软件的结合,能有效提高出版企业对目标读者的辨识和服务水平。

#### (一) 协同过滤算法的应用与战略调适的关系问题

长期战略决定了出版企业发展方向的选择、资源配置的调节和业务结构的优化,是出版经营中最关键的要素。协同过滤算法的应用必须服从企业战略大局,同时要为决策者发现新的利润增长点和确认生产线的生命周期提供有益参考。对编辑人员而言,协同过滤算法的分析结论、竞争对手新上市的产品、其他渠道获得的信息都只能作为参考,出版企业的战略愿景才是编辑人员取舍出版项目的唯一标尺。换言之,出版企业的选题策划与经营范围必须有明确边界,尽力避免使用低层次模仿、盲目跟风等非理性竞争手段去追逐短期利益。

#### (二) 协同过滤算法的应用与出版企业资源配置的优化问题

在充分考虑长期战略、市场环境和需求演进等要素的前提下,如果协同过滤算法提供的新品研发思路具有可行性,此时,负责的编辑人员需要及时转变为项目经理(一种跨职能的角色),调度选题落实需要的各类资源,推动项目顺利实施。在一般情况下,出版企业的资源配置需要遵循一定的审批程序,出版资源在不同领域的分配比例大致保持稳定,但是在数字出版时代,资源配置模式必须能灵活地应对起伏动荡的市场,经验丰富的编辑应该具有一定的资源调度权,可以统筹不同部门的工作重点,确保为读者提供高品质的产品与消费体验,否则,协同过滤算法提供的参考信息即便再有价值,也无法对经营创新产生实质影响。

#### (三) 协同过滤算法与数据库营销等工具的融合问题

在商业领域,通过数据的搜集、管理和分析,运用一系列软件、工具和方法,挖掘信息中有价值的

知识、报告和结论,为经营决策和运营管理服务,被称为“商务智能”。协同过滤属于商务智能的范畴,从综合效果的角度考虑,协同过滤必须与其他商务智能软件协同使用,以收取取长补短之效。<sup>[2]</sup>在实践中,编辑人员在参考协同过滤算法的结论同时,还借鉴其他数据挖掘工具提供的报告,唯有综合考虑各种方案,进而运用经验和智慧加以抉择,才能保证决策的正确与可行。数据库营销关注的重点是顾客的个性需求和消费潜能,而协同过滤算法则侧重考虑顾客的兴趣与需求走向,尤其是相似消费群体需求的关联性和倾向性,因此,这两种预测结果会存在一定差异,编辑人员需要融合多种方法、综合多种意见,以形成最终决策。同时,多参考几种数据挖掘的结果,可以增加智能推荐的多样性和新颖性,实现交叉营销,提高读者购买的多样性。<sup>[3]</sup>

#### (四) 协同过滤算法与编辑人员创意发挥的关系问题

大数据时代来临之初,就有学者担心它对人类思想自由会产生阻碍作用,但在笔者看来,数据存储和数据分析技术的进步不会令人类创意枯竭、思维趋同。一方面,面对诸多技术创新成果,人们对各种方案都会持有保留态度,大量统计分析的结果显示的只是一种趋势预见,任何方案都有潜在风险,人们不会盲目相信;另一方面,作者和编辑不会放弃运用想象力、创造性的本能,来自同行的竞争压力和市场的多元需求会驱策编辑人员发挥创意,开展新品研发。在新媒体逐步冲击和替代传统媒体的前提下,编辑人员不仅要在内容创新上持续用力,还要时刻关注新媒体在表现形式和渠道布局方面的进展,继而面向读者进行全面创新。简言之,协同过滤算法给编辑人员提供的只是一个参照系,其结果仅仅是基于历史资料、相似人群偏好和兴趣的智能推荐,因此,它不应该也不可能直接左右编辑人员的决策。

#### (五) 协同过滤算法与读者潜在兴趣的发现问题

在营销学理论中,顾客始终处于经营活动的核心,顾客的潜在兴趣与现实需求对管理决策至关重要。协同过滤算法的基础是历史资料,这就意味着它在顾客潜在兴趣的发掘方面具有明显的局限,因此,无论是编辑人员还是出版企业高管都需要通过其他方式去调查和发现读者的潜在兴趣。由于没有人比读者本人更了解自己,因此,借助通讯设备实时获取反馈,通过其他渠道搜集读者的资料,进而对其进行整合、分类、挖掘,以洞悉读者的潜在兴趣,是出版企业保持竞争优势的必由之路。在此过程中,编辑人员的任务是引导读者更好地表达真实想法,参与产品设计与制作。对经验丰富的编辑而言,长期的实践能帮助他们形成良好的职业触觉,这种直觉加上对技术负面效应的充分警惕,足以弥补上述缺陷。此外,不同出版企业在市场细分和价值定位方面都有一套独特的模型,因此,参考竞争对手尤其是本领域内龙头企业的做法也是发现目标读者潜在兴趣的可行方案。

随着数字媒体、移动技术和社交网络的迅速进步,阅读消费变得日渐复杂,二十四小时全天候购买变为现实,出版机构的竞争对手和竞争策略层出不穷,新媒体对传统出版商的冲击和替代,跨国界、跨领域、跨媒体的竞合博弈变得司空见惯,上述种种变化都对要求编辑人员必须具备敏捷的反应能力、快速的创新能力和深刻的洞察能力。面对这些考验,编辑人员不仅需要尽快提高专业素养,而且还要及时掌握协同过滤等技术工具,唯此方能胜任数字环境下产业转型升级对从业者提出的新的岗位要求,并把握时机掌握竞争的主动权。

#### 参考文献:

[1] 协同过滤[EB/OL]. <http://zh.wikipedia.org/zh-cn/%E5%8D%94%E5%90%8C%E9%81%8E%E6%BF%BE> 2013-3-13.

[2] 陈运文. 智能推荐系统开发中的十个关键注意点[EB/OL]. <http://www.resyschina.com/2013/03/recommendation-keypoint.html>, 2013-03-25.

[3] 周涛. 科普一下协同过滤[EB/OL]. <http://blog.sciencenet.cn/blog-3075-456134.html> 2011-06-17.

rising rates of the overall social consumer goods prices over the same period. The Chinese psychological prices and price expectations of books make an impact on the social holistic cognition of book price.

## **On the Application of Collaborative Filtering Algorithm in the Innovation of Editing Process: Possibility and Feasibility**

Wang Pengtao

This paper analyses the unique function of the collaborative filtering algorithm in the editing process , which includes topic selection , revision , promotion , market research and post – sales service , and the proper use of the collaborative filtering algorithm can help top managers to select strategic creation scheme and avoid operating risks , then the performance of resources integration and business cooperation will be improved sharply. From the perspective of operating process , the author states the mechanism of the use of collaborative filtering algorithm , and discuss the possible problems and the countermeasures.

## **Time and Space of Animation Film in Natural Philosophy**

Xu Zhouchi

The tradition of “long long ago” , is not only the heritage of classical narrative , but also an exposure of specific relationship between time and space on animation film. Natural ideology has been expressed through this sort of relationship. Natural ideology is a social ideal and world view conveyed by western narrative that complies with nature , pursues instinct and returns to soul. It reflects the persistence on outstanding narrative of the contemporary western animation creation. In most of these stories , natural ideology turns out to be the persistence on natural order and human dignity , while the true meaning of which is to emancipate human instinct and return to soul. Analyzed from the angle of natural ideology , the so – called adult interest in animation conveys the repressed childhood mentality and intention. It explains why an outstanding animation creation could obtain the fondness of both adults and children.

## **Dual Interpretation of the USA Newspaper Harvest Phenomena**

Li Saike

In USA , many newspapers took harvest strategy when the development of the industry is in trouble , which is to compress the layout , reducing its frequency and distribution area , large – scale layoffs and other means to reduce the cost to maintain a certain profit space , but the harvest means also bring USA newspaper the news resources erosion and reporting ability contraction. Harvest strategy for development of the newspaper and the press , is the survival , or self destructive? The reality is often not the dichotomy of this simple can be summarized. For the news industry which is given and looked forward to the democratic function based on more cause for a profession concerned , and discus-