

# 基于深度学习的多模态数据融合算法在精准推荐中的应用与优化

许旭莲

杭州久典网络科技有限公司 浙江 杭州 311000

**摘要:** 随着互联网和信息技术的飞速发展,用户产生的数据呈现出爆炸式增长且具有多模态的特点,如文本、图像、音频等。精准推荐系统作为连接用户与海量信息的桥梁,在电商、社交、内容平台等领域得到了广泛应用。它能够根据用户的兴趣和行为,为用户提供个性化的推荐服务,提高用户体验和平台的商业价值。然而,传统的推荐算法在处理多模态数据时存在诸多问题。一方面,单一模态的数据往往无法全面准确地描述用户的兴趣和偏好,导致推荐的准确性和多样性不足。另一方面,现有的多模态数据融合方法大多缺乏深度的语义理解和特征提取能力,难以充分挖掘多模态数据之间的潜在关系。基于此,本文针对基于深度学习的多模态数据融合算法在精准推荐中的应用与优化展开深入分析。首先阐述多模态数据和深度学习的基础理论,然后剖析现有的融合算法,接着进行应用设计并提出优化策略。本文以期通过这些研究,提高多模态数据融合的效果,提升精准推荐系统的性能,为用户提供更加准确、个性化的推荐服务,促进推荐系统领域的发展。

**关键词:** 深度学习; 多模态数据融合; 精准推荐; 算法优化; 个性化服务

## 1 引言

在数字化信息高度繁荣的当下,互联网平台积累了海量的用户数据,这些数据涵盖了多种模态,如用户发表的文本评论、上传的图像、浏览时产生的音频等。精准推荐系统凭借其为用户提供个性化服务的能力,成为了众多互联网企业提升用户体验、增强用户粘性以及实现商业价值的关键工具。从电商平台根据用户的购买历史推荐商品,到社交媒体平台依据用户的兴趣爱好推送内容,精准推荐系统无处不在。然而,当前的推荐系统在处理多模态数据时面临着严峻的挑战。传统的推荐算法大多基于单一模态数据,无法充分利用多模态数据所包含的丰富信息,导致推荐结果的精准度和多样性受限。而且,简单的数据融合方式难以深入挖掘不同模态数据之间的内在联系,使得推荐缺乏深度和全面性。因此,研究基于深度学习的多模态数据融合算法,并将其应用于精准推荐系统,对提高推荐的质量和效果,满足用户日益多样化的需求,推动互联网行业的持续发展具有重要的现实意义。

## 2 多模态数据与深度学习基础理论

### 2.1 多模态数据的概念、类型与特点

多模态数据是指由多种不同数据形态构成的数据集。其类型丰富多样,主要包括文本数据,如用户的评论、帖子、产品描述等,这类数据蕴含着语义信息,能够直接表达用户的观点和需求;图像数据,像用户上传

的照片、产品图片等,包含了直观的视觉特征,可用于识别物体、场景等;音频数据,例如语音指令、音乐等,能传达情感和特定信息<sup>[1]</sup>。多模态数据具有独特的特点。首先是互补性,不同模态的数据可以从不同角度描述同一对象,相互补充,提供更全面的信息。其次是多样性,数据形式和内容差异大,增加了数据处理的复杂性。再者是海量性,随着互联网应用的普及,多模态数据的产生量呈爆发式增长。最后是异质性,不同模态数据在结构、特征表示等方面存在显著差异,给数据融合带来了挑战。

### 2.2 深度学习的基本原理与常见模型架构

深度学习基于人工神经网络,通过构建多层网络结构,自动从大量数据中学习特征和模式。其基本原理是利用神经元之间的连接权重和激活函数,对输入数据进行逐层变换和抽象,从而提取更高级的特征表示。常见的模型架构包括卷积神经网络(CNN),它擅长处理图像数据,通过卷积层、池化层等结构自动提取图像的局部特征,广泛应用于图像识别、目标检测等任务;循环神经网络(RNN)及其变体,如长短期记忆网络(LSTM)和门控循环单元(GRU),适合处理序列数据,如文本和语音,能够捕捉数据中的时间依赖关系;生成对抗网络(GAN),由生成器和判别器组成,通过对抗训练的方式生成逼真的数据样本,在图像生成等领域有重要应用<sup>[2]</sup>。

### 2.3 多模态数据融合在推荐系统中的理论依据

在推荐系统中，多模态数据融合具有坚实的理论依据。一方面，用户的兴趣和偏好是多元的，单一模态数据难以全面准确地刻画。例如，仅依据用户的文本评价不能完全了解其对产品外观的喜好，而结合图像数据能提供更完整的信息。多模态数据融合能够整合不同来源的信息，更全面地描述用户画像和物品特征，从而提高推荐的准确性。另一方面，不同模态数据之间存在内在联系。深度学习模型可以挖掘这些潜在关系，发现隐藏在数据背后的模式。例如，通过融合文本和图像数据，可以找到用户对产品描述与视觉特征之间的关联，进而为用户提供更符合其需求的推荐。这种多模态数据融合的方式能够突破单一模态的局限性，为推荐系统提供更丰富的信息输入，提升推荐的质量和效果。

### 3 基于深度学习的多模态数据融合算法分析

#### 3.1 现有的多模态数据融合算法分类与概述

现有的多模态数据融合算法主要可分为早期融合、晚期融合和混合融合三类。早期融合，也叫数据层融合，是在数据预处理阶段就将不同模态的数据进行整合，例如将图像的像素特征与文本的词向量特征拼接在一起，然后输入到统一的模型中进行处理。这种方式能充分利用各模态数据的原始信息，但对数据的预处理要求较高，且不同模态数据的特征维度差异可能会影响融合效果。晚期融合，即决策层融合，是先对不同模态的数据分别进行处理，得到各自的预测结果，再将这些结果进行融合，比如通过投票、加权平均等方式得出最终决策。该方法灵活性高，可根据不同模态的特点选择合适的模型，但可能会丢失一些模态间的潜在联系。混合融合则结合了早期融合和晚期融合的优点，在不同阶段对不同模态数据进行融合，能够在一定程度上平衡信息利用和模型灵活性。

#### 3.2 基于深度学习的多模态数据融合具体算法介绍

基于深度学习的多模态数据融合算法有多种。如基于注意力机制的融合算法，它通过为不同模态的数据分配不同的权重，突出重要信息，抑制无关信息。例如在文本和图像融合时，注意力机制可以让模型更关注与用户兴趣相关的图像区域和文本关键词。还有基于生成对抗网络（GAN）的融合算法，利用生成器生成融合后的多模态数据，判别器判断数据的真实性和融合效果，通过对抗训练不断优化融合结果。比如在图像和文本融合中，生成器可以生成与文本描述相符的图像。

另外，基于循环神经网络（RNN）和卷积神经网络（CNN）结合的融合算法也较为常见。RNN处理序列数

据，CNN处理图像数据，将两者结合可以有效融合文本和图像模态的数据，用于推荐系统中对用户兴趣的建模<sup>[1]</sup>。

#### 3.3 不同算法在精准推荐中的优势与局限性

早期融合算法在精准推荐中的优势在于能充分利用原始数据信息，挖掘不同模态间的潜在关系，构建更全面的用户和物品表征，提高推荐的准确性。然而，其局限性在于对数据的一致性和兼容性要求高，若不同模态数据特征差异大，融合难度大，且计算成本较高。晚期融合算法的优势是灵活性强，可针对不同模态选择最优模型，计算效率相对较高，还能根据各模态的重要性调整决策融合策略。但它的缺点是各模态独立处理可能会丢失模态间的早期交互信息，导致推荐结果不够精准。混合融合算法综合了前两者的优点，一定程度上平衡了信息利用和模型灵活性，在复杂场景下表现较好。但它也存在模型结构复杂、训练难度大的问题，且需要合理设计融合策略，否则可能无法充分发挥优势，甚至导致性能下降。

### 4 基于深度学习的多模态数据融合在精准推荐中的应用设计

#### 4.1 应用场景分析与需求确定

在电商领域，多模态数据融合可助力精准推荐。用户浏览商品时产生的文本评价、商品图片以及用户的浏览轨迹等多模态数据，能更全面地反映用户对商品的喜好。比如通过分析用户对服装的文本描述评价和模特展示图片的关注，推荐符合其风格的服饰。在内容平台方面，用户阅读的文章文本、观看的视频画面及音频等多模态数据，可挖掘用户兴趣偏好，精准推送新闻、视频等内容。需求在于准确捕捉用户多维度兴趣，提升推荐的相关性和多样性，增强用户粘性。

#### 4.2 融合算法与推荐系统的整合架构设计

将基于深度学习的多模态数据融合算法整合进推荐系统。架构底层为多模态数据采集模块，收集文本、图像、音频等数据。接着是数据预处理层，对各模态数据进行清洗、特征提取等操作。融合层运用选定的融合算法，如注意力机制融合算法，整合不同模态特征。上层的推荐模型基于融合后的特征，为用户生成推荐列表。同时设置反馈模块，根据用户对推荐结果的交互行为，不断优化融合算法和推荐模型，形成闭环优化系统。

#### 4.3 数据预处理与特征工程在应用中的实施

数据预处理时，针对文本数据进行分词、去除停用词、词向量转换等操作，将文本转化为计算机可处理的数值形式。对于图像数据，进行图像缩放、归一化等处理，利用卷积神经网络提取图像特征。音频数据则进行

降噪、特征提取,如梅尔频率倒谱系数(MFCC)提取。特征工程方面,对不同模态特征进行筛选和组合,去除冗余特征<sup>[4]</sup>。比如结合文本的语义特征和图像的视觉特征,构建更具代表性的用户和物品特征向量,为后续的融合和推荐提供高质量的数据基础。

## 5 基于深度学习的多模态数据融合算法的优化策略

### 5.1 算法参数调优与模型选择的方法

算法参数调优是提升基于深度学习的多模态数据融合算法性能的关键环节。对于常见的深度学习模型,如神经网络中的学习率、隐藏层神经元数量等参数,需采用合适的调优方法。例如,可使用网格搜索法,预先设定参数的取值范围,通过遍历所有可能的参数组合,在验证集上评估模型性能,选择最优参数组合。随机搜索法则是在参数空间中随机采样进行评估,相比网格搜索更高效,尤其适用于参数空间较大的情况。在模型选择方面,要综合考虑多模态数据的特点和推荐任务的需求。若数据中图像模态占比较大且对局部特征要求高,卷积神经网络(CNN)可能是较好的选择;若数据具有明显的序列特征,如文本或音频数据,循环神经网络(RNN)及其变体可能更合适。同时,还可以尝试集成多个模型,结合它们的优势,提高算法的泛化能力和推荐准确性。

### 5.2 针对多模态数据特点的融合策略改进

多模态数据具有异质性、互补性等特点,针对这些特点可对融合策略进行改进。由于不同模态数据的特征维度和分布可能差异较大,在融合前可采用特征对齐技术,使各模态数据的特征在同一尺度上。例如,对文本的词向量和图像的特征向量进行归一化处理,以消除量纲的影响。考虑到多模态数据的互补性,可设计更灵活的融合方式。除了传统的早期融合、晚期融合和混合融合,还可以引入层次化融合策略。先在较低层次对相似模态的数据进行融合,提取初步特征,再在较高层次将不同组的特征进行二次融合,以更好地挖掘多模态数据间的复杂关系。此外,利用注意力机制动态调整不同模态数据在融合过程中的权重,突出关键信息,提升融合效果。

### 5.3 提升推荐系统性能与效果的综合优化措施

为提升推荐系统的性能与效果,首先要优化数据质量。持续收集和更新多模态数据,扩大数据规模,同时加强数据清洗和标注工作,确保数据的准确性和完整性。建立数据质量评估机制,定期对数据进行检查和

修正<sup>[5]</sup>。其次,改进推荐模型的评估指标。除了常用的准确率、召回率等指标,引入更能反映用户体验的指标,如平均倒数排名(MRR)、归一化折扣累积增益(NDCG)等,从多个角度评估推荐结果的质量。再者,加强用户反馈机制。鼓励用户对推荐结果进行评价和反馈,根据用户的反馈及时调整推荐策略和模型参数。例如,当用户对某类推荐内容表示不感兴趣时,降低相关特征在模型中的权重,提高推荐的个性化程度。最后,结合实时数据进行实时推荐,根据用户的最新行为和兴趣动态更新推荐列表,提高推荐的及时性和相关性。

## 6 结语

本文围绕基于深度学习的多模态数据融合算法在精准推荐中的应用与优化,进行了系统且深入的研究。先是梳理了多模态数据和深度学习的基础理论,剖析了现有融合算法的优劣势,接着开展应用设计并提出了相应的优化策略。研究表明,借助深度学习实现多模态数据融合,能够有效挖掘不同模态数据间的潜在关系,显著提升精准推荐系统的性能,为用户带来更精准、个性化的推荐服务。但本研究仍存在一定局限,比如在处理超大规模多模态数据时,算法的计算效率还有待提高,模型对复杂语义的理解和融合能力也有提升空间。未来,需进一步探索更高效的多模态数据融合算法和模型架构,提升算法的泛化能力和适应性。同时,加强对多模态数据语义理解的研究,深度挖掘用户需求,以推动精准推荐系统的持续发展,为用户提供更优质的服务体验。

## 参考文献

- [1]周颖玉,柯平,刘海鸥.融合·赋能·共创:区块链嵌入算法推荐的未来发展图景[J].图书情报知识,2024,41(3):144-153.
- [2]王礼,万一帆.基于深度学习的个性化媒体推荐系统算法研究[J].广播电视信息,2024,31(10):109-112.
- [3]范勇,张政波(综述),王晶(审校).基于深度学习的电子病历多模态数据融合研究进展[J].生物医学工程学杂志,2024,41(5):1062-1071.
- [4]周炫余,刘林,陈圆圆,洪嘉玲,卢笑.基于多模态数据融合的大学生心理健康自动评估模型设计与应用研究[J].电化教育研究,2021,42(8):72-78.
- [5]蔡程飞,李军,焦一平,王向学,郭冠辰,徐军.基于深度学习的医学多模态数据融合方法在肿瘤学中的进展和挑战[J].数据与计算发展前沿,2024,6(3):3-14.