

融合 μ 2Net+ (ViT-L/16)的多模态商品推荐系统

李洪军

上海理工大学光电信息与计算机工程学院, 上海

收稿日期: 2025年3月20日; 录用日期: 2025年6月20日; 发布日期: 2025年6月30日

摘要

随着各领域推荐系统的迅速发展, 多模态推荐已经成为个性化推荐的下一个竞争赛道。在以往的神经网络推荐模型中总是采用单一的推荐模型, 例如Bert4Rec, DLRM等, 它们都存在一些限制, a) 单一架构无法透彻解析用户输入文本信息的意图, b) 总是将输入限制为单一的文本或者图像。为了解决这些限制, 实现推荐系统从单模态向多模态的跨越, 本文提出了KGBM4Rec模型, 这是一种基于知识图谱、图神经网络、Bert序列推荐和 μ 2Net+ (ViT-L/16)的混合神经网络模型。该模型通过构建领域知识图谱, 提高模型扩展性, 利用图神经网络从知识图谱中提取用户及商品之间的关联特征; 采用Bert序列模型对文本特征进行建模; μ 2Net+ (ViT-L/16)利用具有一系列卷积层和池化层的U形网络架构从图像中提取层次特征进行建模。多模态的个性化推荐可以让用户体验得到极大提升, 本文在淘宝直播多模态视频商品检索数据集(Taobao)、来自在线京东购物网站的Products-10K数据集上进行实验, 最终实验结果表明, 所提出的推荐模型在多个数据集上均取得了平均高于基线模型10%的推荐效果。

关键词

Bert, 知识图谱, 图卷积神经网络, 推荐系统, μ 2Net+ (ViT-L/16)

Research on Multimodal Commodity Recommendation System Incorporating μ 2Net+ (ViT-L/16)

Hongjun Li

School of Optical-Electrical and Computer Engineering, Shanghai University of Technology, Shanghai

Received: Mar. 20th, 2025; accepted: Jun. 20th, 2025; published: Jun. 30th, 2025

Abstract

With the rapid development of recommendation systems in various fields, multimodal recommen-

文章引用: 李洪军. 融合 μ 2Net+ (ViT-L/16)的多模态商品推荐系统[J]. 软件工程与应用, 2025, 14(3): 723-735.
DOI: 10.12677/sea.2025.143064

dation has become the next competitive track for personalized recommendation. In the past, neural network recommendation models always used a single recommendation model, such as Bert4Rec, DLRM, etc. They all have some limitations: a) a single architecture cannot thoroughly analyze the intention of the user's input text information, and b) the input is always limited to a single text or image. In order to solve these limitations and realize the transition of the recommendation system from single modality to multimodality, this paper proposes the KGBM4Rec model, which is a hybrid neural network model based on knowledge graph, graph neural network, Bert sequence recommendation and μ 2Net+ (ViT-L/16). This model improves the scalability of the model by constructing a domain knowledge graph, and uses graph neural network to extract the correlation features between users and products from the knowledge graph; the Bert sequence model is used to model text features; μ 2Net+ (ViT-L/16) uses a U-shaped network architecture with a series of convolutional layers and pooling layers to extract hierarchical features from images for modeling. Multimodal personalized recommendations can greatly improve user experience. This paper conducts experiments on Taobao live multimodal video product retrieval dataset (Taobao) and Products-10K dataset from JD.com online shopping website. The final experimental results show that the proposed recommendation model has achieved an average recommendation effect of 10% higher than the baseline model on multiple datasets.

Keywords

Bert, Knowledge Graph, GCN, Recommendation System, μ 2Net+ (ViT-L/16)

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

1.1. 背景及研究动机

随着很多推荐系统的不断修缮,电商平台的推荐算法和模型层出不穷[1]-[8],在针对多模态商品的推荐性能上参差不齐且准确率普遍偏低。当前,多模态推荐系统在商品推荐领域已经展现出重要潜力,并逐渐成为提高推荐系统精度和用户体验的关键技术之一。相比于传统的单模态推荐系统,依赖单一模态(如文本、图像或用户行为)信息往往难以捕捉用户的多元需求,导致推荐结果的多样性和准确性不足。多模态方法通过融合多种模态信息(如文本描述、商品图像、用户评论等),能够更全面地理解商品特征和用户兴趣,从而提高推荐效果。本文在商品推荐上结合知识图谱,图神经网络及 Bert 序列推荐和视觉网络推荐模型 μ 2Net+ (ViT-L/16)融合进行推荐的模型是一种新颖的方法,可以有效提高推荐系统的效率和准确性。知识图谱[9]-[14]是一种表示实体和关系的结构化数据,可以提供丰富的语义信息和背景知识。图神经网络[2] [15]-[17]是一种基于图结构的深度学习模型,可以有效地捕捉实体和关系的复杂特征和依赖性。Bert 序列推荐[18]是一种基于自然语言处理的技术,可以利用预训练的语言模型来生成高质量的商品描述和评价。 μ 2Net+ (ViT-L/16) [19]-[21]是最优秀的视觉模型之一。将这四种技术巧妙融合起来,可以实现对商品的个性化推荐,满足用户的不同需求和偏好。并且模型能够做到充分的知识划分,可以避免多任务模型中的常见问题,如灾难性遗忘[1]。

1.2. 研究目的和主要贡献

本研究的多模态模型通过结合视觉、文本、结构化数据等信息,解决了单模态模型信息单一、表征

能力有限的问题。例如，单纯依赖商品图像或文本描述的推荐系统，往往在处理具有复杂属性的商品时会遇到困难，尤其是无法准确捕捉用户的潜在兴趣。多模态模型通过引入多模态信息融合机制，能够有效缓解这些问题。

具体而言，本文的模型采用了多模态架构构建深度学习推荐模型，在信息融合过程中，通过优化注意力机制、模态对齐方法等，进一步增强了不同模态之间的互补性。相比现有的多模态方法，本研究在训练策略、推理效率和特征提取方面进行了创新，从而提升了模型的推荐性能。

本研究提出了一种结合知识图谱、图神经网络和 Bert 序列推荐及 $\mu 2\text{Net}+$ (ViT-L/16)模型的多模态商品推荐方法。该方法充分利用了图神经网络学习知识图谱中的结构化信息，Bert 模型在自然语言处理领域的优势，以及 $\mu 2\text{Net}+$ (ViT-L/16)在视觉处理上的高精度，从而提高了商品的推荐准确性。实验证明，与现有的推荐方法相比，本文的方法在多个评价指标上取得了更好的效果和性能。

本文的主要研究贡献如下：

- 1) 采用了 $\mu 2\text{Net}+$ (ViT-L/16)作为视觉特征提取模型，比传统 CNN 具有更强的视觉建模能力。
- 2) 结合 BERT 进行文本嵌入处理，使得商品的语义信息得到更好的利用。
- 3) 通过 GNN 传播结构信息，有效建模用户 - 商品的关系网络，提升推荐系统的可解释性。

1.3. 论文结构

本文的其余部分安排如下，在第二部分对商品推荐的发展和相关技术研究进行介绍；在第三部分对模型设计进行详细介绍并探讨其创新性；第四部分完成相关实验并对模型实验结果进行了评价；最后在第五部分对本文进行了总结并探讨了多模态商品推荐系统的未来发展方向。

2. 相关工作

目前，多模态推荐系统主要面临以下挑战：

- 1) 模态信息不对齐问题：不同模态的数据(文本、图像、结构化数据)具有不同的分布，如何进行特征对齐仍然是一个难点。
- 2) 模态间信息冗余与噪声问题：不同模态可能包含重复或不相关的信息，如何去除冗余特征，提高有效特征的权重？
- 3) 计算复杂度问题：融合多个深度模型(GNN, BERT, ViT)会增加计算开销，如何优化计算资源以提高推荐效率？
- 4) 领域泛化能力：模型在不同的商品类别和平台上的适应性如何提升？

2.1. 知识图谱与推荐系统

将知识图谱和推荐系统结合在一起，可以极大地提高了推荐的准确性和个性化程度。知识图谱可以提供丰富的背景信息和上下文，帮助推荐系统更好地理解用户的需求和兴趣，从而提供更精确的推荐。此外，知识图谱还可以帮助推荐系统理解商品或服务的复杂属性和关系，从而提供更深层次、更具针对性的推荐。知识图谱的多模态化是实现人类水平机器智能的必然步骤[12]。

例如，一个电影推荐系统可以利用知识图谱来理解电影之间的复杂关系，如导演、演员、类型、主题等因素，然后将这些信息用于生成更准确的推荐。同样，一个商品推荐系统也可以利用知识图谱来理解商品的属性和类别，以及商品之间的关系，从而提供更具针对性的推荐。

图 1 展示了一种在知识图谱上的双层随机游走算法 TRWA [12]。TRWA 的主要思想是结合两种不同的特征建模方法，将知识图谱的拓扑结构细分为全局图和局部子图，分别进行特征提取。最后对全局模块和局部模块进行加权合并，得到完整的逻辑规则推理算法。知识图谱和推荐系统的结合为提供更准确、

更个性化的推荐开辟了新的可能,对于改善用户体验和提高商业效益都有重要的价值。

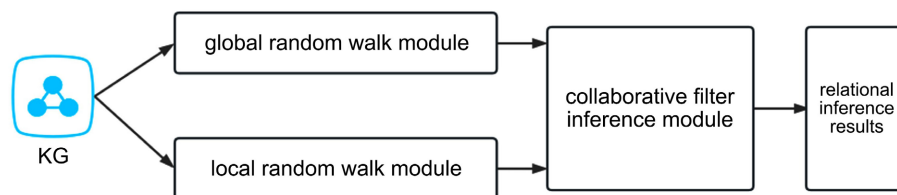


Figure 1. Dual layer randomized wandering algorithm TRWA

图 1. 双层随机游走算法 TRWA

2.2. 图神经网络与推荐系统

传统的推荐系统通常使用协同过滤或基于内容的推荐方法[7][8],但这些方法可能无法充分利用用户和项目之间的复杂交互和潜在关联。但是,图神经网络[22]可以通过将用户和项目的交互视为图结构(其中用户和项目是节点,交互是边),利用图神经网络来学习这个图中的复杂模式,从而提供更准确的推荐。

图 2 展示了采用 GNN 模型的端到端预测任务。图 3 为 GNN 层结构详细示意图。

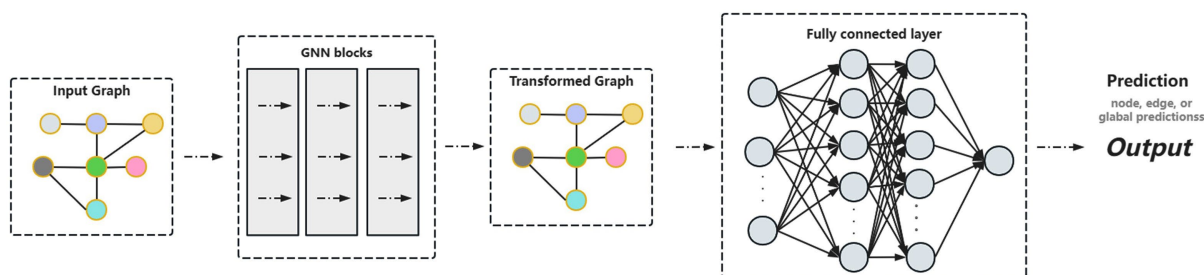


Figure 2. An end-to-end prediction task with a GNN model

图 2. 采用 GNN 模型的端到端预测任务

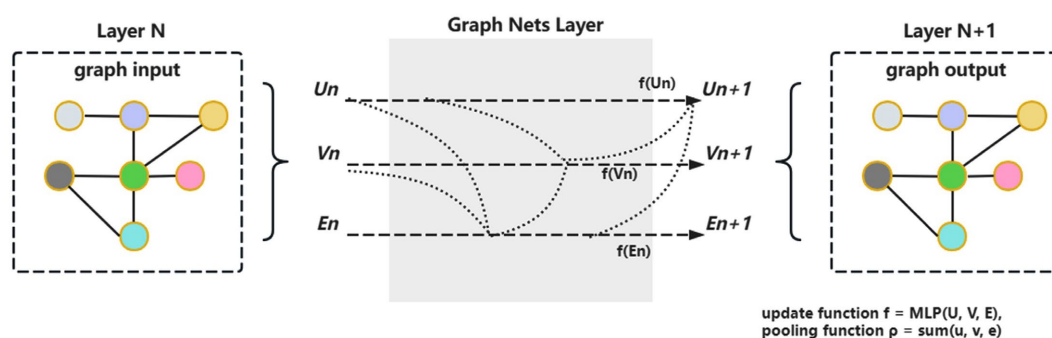


Figure 3. Schematic of a Graph Nets architecture leveraging global representations

图 3. 利用全局表征的 GNN 架构示意图

图神经网络(Graph Neural Networks, GNNs)的基本数学模型可以概括为以下两个步骤:信息聚合(Message Aggregation)和信息更新(State Update)。这两步骤通常是交替进行的,直到达到一个固定的迭代次数,或者网络状态稳定为止。

假设有一个图 G , 其中 V 是节点集合, E 是边集合。每个节点 v 都有一个特征向量 h_v 。以下是这两个步骤的基本数学表示:

- 1) 信息聚合: 在这一步, 每个节点 v 收集其邻居节点的信息。具体来说, 节点 v 的新的特征向量是

其所有邻居节点的特征向量的聚合。这可以用以下的公式表示：

$$m_v = AGGREGATE(h_u : (u, v) \in E) \quad (1)$$

这里，*AGGREGATE* 是一个聚合函数，可以是求和、平均、最大值等，本文实验中采用求和。

2) 信息更新：在这一步，每个节点 v 根据其当前的特征向量和聚合后的信息来更新其特征向量。这可以用以下的数学公式表示：

$$h'_v = UPDATE(h_v, m_v) \quad (2)$$

这里，*UPDATE* 是一个更新函数，通常是一个非线性函数，本文实验中采用 *ReLU*。注意 h'_v 是节点 v 在信息更新后的新特征向量。

然后，这两个步骤会在图中的所有节点上并行执行，并迭代执行若干轮。

对于推荐系统中的应用，有一个目标函数(例如预测用户对商品的评分)，然后通过优化这个目标函数来训练图神经网络的参数。具体的目标函数会根据具体的推荐任务而变化。例如，对于评分预测，一个常见的目标函数是均方误差：

$$L = \sum (r_{uv} - f(u, v))^2 \quad (3)$$

这里， r_{uv} 是用户 u 对商品 v 的真实评分， $f(u, v)$ 是模型预测的评分， $\sum$ 是对所有用户-商品对的求和。模型训练的目标是通过优化网络参数来最小化这个损失函数。

2.3. Bert 模型与推荐系统

Bert 的基本数学模型是基于 Transformer 的自注意力机制[23]。在自注意力机制中，输入是一系列向量 $h_1, h_2, \dots, h_n$ ，输出是一系列新的向量 $h'_1, h'_2, \dots, h'_n$ ，其中每个新向量是输入向量的加权和：

$$h'_i = \sum \text{Attention}(Q(h_i), K(h_j), V(h_j)) \quad (4)$$

这里， Q ， K ， V 分别是查询、键和值函数，它们通常是线性变换；*Attention* 是自注意力函数，通常定义为：

$$\text{Attention}(Q, K, V) = \text{softmax}(QK^T / \sqrt{d_k})V \quad (5)$$

这里， d_k 是键向量的维度， T 表示矩阵转置，*softmax* 是对每一行进行的归一化函数，使得每一行的元素非负且和为 1。此外，为了捕捉序列中的位置信息，Bert 还引入了位置编码。

2.4. μ 2Net+ (ViT-L/16)与推荐系统

受 Transformer 在自然语言处理(NLP)领域取得巨大成功的启发[23]，研究人员尝试将 Transformer 引入视觉领域[24]。在文献[25]中，提出了 ViT (Vision Transformer)来执行图像识别任务。以具有位置嵌入的二维图像块作为输入并在大型数据集上进行预训练，ViT 实现了与基于 CNN 方法相当的性能。

图 4 展示了 ViT 模型的架构，ViT 是一种基于 Transformer 的图像分类模型[26]。ViT 模型的核心思想是将图像分割成一系列的图像块，然后将这些图像块转换为向量序列，再通过 Transformer 模型进行处理，最终输出图像的分类结果。

ViT 模型的输入是一张大小为 $H \times W$ 的图像，首先将图像分割成大小为 $P \times P$ 的图像块，然后将每个图像块通过一个线性变换映射到一个 d 维的向量空间中，得到一个序列 $X = [x_1, x_2, \dots, x_n]$ ，其中 $x_i \in R_d$ 。这个序列 X 就是 ViT 模型的输入。

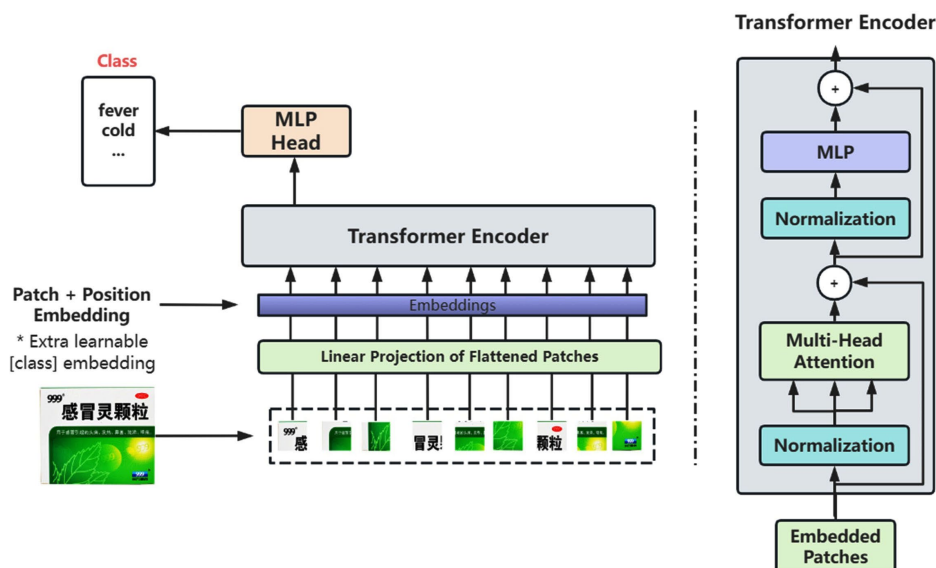


Figure 4. ViT model diagram
图 4. ViT 模型图

接下来, ViT 模型使用 Transformer 模型对序列 X 进行处理。Transformer 模型由多个 Transformer 块组成, 每个 Transformer 块包含多头自注意力机制和前馈神经网络。在 ViT 模型中, 每个 Transformer 块都会对序列 X 进行处理, 得到一个新的序列 $X' = [x'_1, x'_2, \dots, x'_n]$, 其中 $x'_i \in R_d$ 。

最后, ViT 模型将序列 X' 中的所有向量进行平均池化, 得到一个 d 维的向量, 然后通过一个全连接层将这个向量映射到类别空间中, 得到图像的分类结果。ViT 模型通过将图像分割成图像块, 并使用 Transformer 模型对图像块进行处理, 实现了对图像的分类。相比于传统的卷积神经网络, ViT 模型具有更好的可扩展性和泛化能力, 可以处理更大的图像和更复杂的任务。

$\mu 2\text{Net}+$ (ViT-L/16) [19] [20] 是一种更加高级的图像分割模型, 结合了 $\mu 2\text{Net}$ 模型和 Vision Transformer (ViT) 模型的特征。它专为高质量和准确的图像分割任务而设计。 $\mu 2\text{Net}$ 模型是专门为显着目标检测和图像分割而开发的深度神经网络。它利用具有一系列卷积层和池化层的 U 形架构从图像中提取层次特征。该网络[21]结合了低级和高级功能以生成精确的对象边界。

与传统的卷积神经网络(CNN)不同, ViT 模型用自我注意机制取代了卷积, 使其能够捕获图像中的全局依赖关系。通过结合 $\mu 2\text{Net}$ 和 ViT 模型的优势, $\mu 2\text{Net}+$ (ViT-L/16) 模型利用了 $\mu 2\text{Net}$ 模型强大的特征提取能力以及对 ViT 模型的全局上下文理解。这种混合架构增强了模型准确分割图像中对象的能力, 使其非常适合语义分割、实例分割和边界检测等任务。

以上的架构 $\mu 2\text{Net}+$ (ViT-L/16) 指定了使用的 ViT 模型变体, 即补丁大小为 16×16 的 ViT-L。ViT-L 是指具有更多层数的 Vision Transformer 模型的更大变体, 使其能够捕获图像中更复杂的模式和依赖关系。补丁大小决定了图像被划分为 ViT 模型处理的粒度。

基于以上相关工作的研究可以了解到, 知识图谱(KG)能够提供结构化的商品信息, 帮助构建商品与用户之间的关系网络。图神经网络(GNN)能有效学习商品和用户的交互模式, 特别适用于复杂关系建模。BERT 处理商品的文本描述, 能够捕捉语义信息。 $\mu 2\text{Net}+$ (ViT-L/16) 解析商品图像, 补充视觉信息, 有助于提升推荐质量。

融合方式: 通过特征嵌入对齐和注意力机制, 确保不同模态特征的互补性, 从而提升推荐的准确性和鲁棒性。

特征交互方式：在 GNN 传播过程中，增加跨模态注意力(Cross-Modal Attention)机制，使文本、视觉和图谱特征进行深度交互。

融合策略优化：

Early Fusion (早期融合)：在 GNN 计算邻居信息时，直接融合 BERT 和 ViT-L/16 特征，使其在邻居关系传播时发挥作用。

Late Fusion (后期融合)：分别计算 GNN、BERT、ViT-L/16 的表示向量，最后通过 MLP 进行最终推荐。

3. 方法论

本文提出的模型架构如图 5 所示。该框架整合了三类通用模型：图神经网络(GNN)、Bert 和 μ Net，分别针对不同的输入模态进行处理。在模态特征提取后，这些模型将特征嵌入对齐到统一的向量空间，有效实现了推荐系统从单模态向多模态的转变。

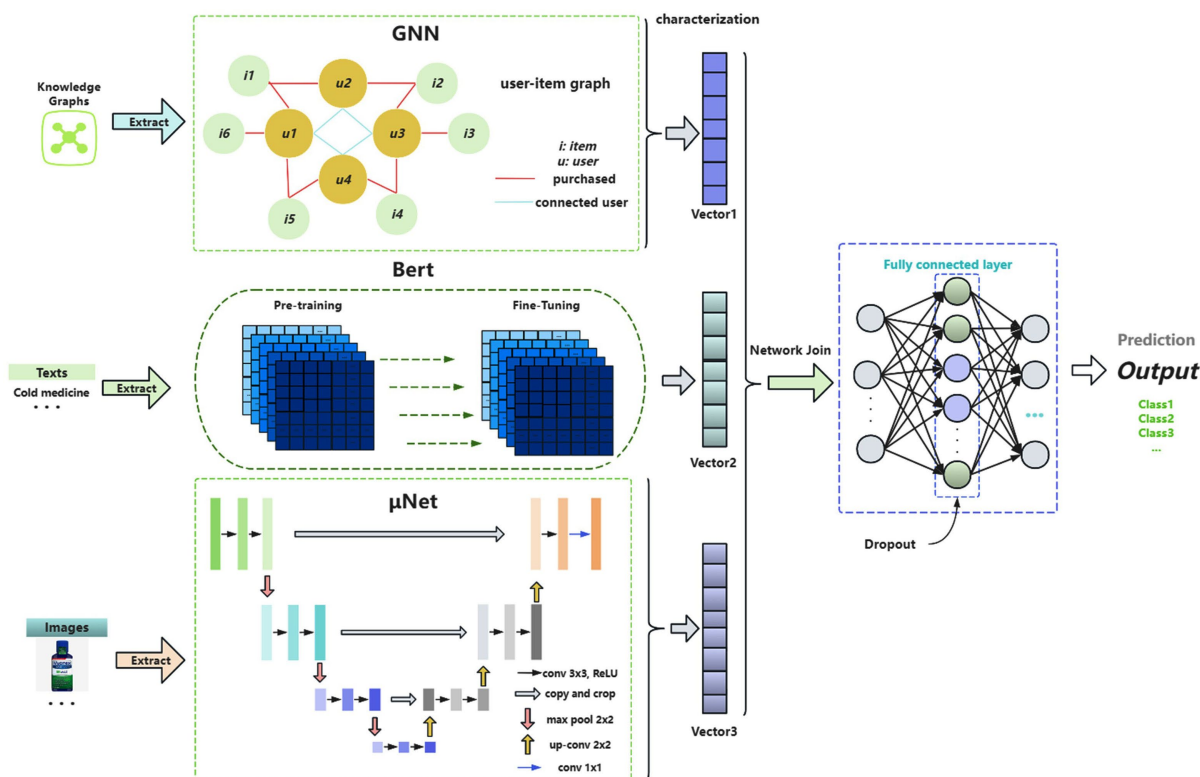


Figure 5. General framework diagram of the KGBM4Rec model

图 5. KGBM4Rec 模型总体框架图

通过结合 GNN 的图结构信息、Bert 的文本理解能力以及 μ Net 的图像处理特性，模型能够充分挖掘多种模态的信息，进而提升推荐的准确性和个性化水平。具体方法包括以下几个步骤：

- 1) 特征提取：分别对不同模态的数据(如图像、文本和结构化数据)进行特征提取。
- 2) 特征嵌入对齐：将提取的特征通过对齐方法映射到同一向量空间，使得不同模态的特征能够进行有效融合。
- 3) 模型融合：利用注意力机制等方法加强不同模态之间的互补性，提升模型的学习能力和推荐性能。
- 4) 训练与优化：通过设定合适的损失函数和优化策略，训练模型以达到最佳的推荐效果。

本研究为多模态推荐系统的设计提供了一个强有力的框架，能够更好地适应复杂的商品推荐任务。

为了使用知识图谱、图神经网络、Bert 模型和 $\mu 2\text{Net}+$ (ViT-L/16)模型来构建推荐系统，本文以对各种商品数据集进行分类推荐为例，按照以下步骤进行。

1) 数据收集和预处理

收集商品的相关数据，如商品描述、功能、成分等。对这些数据进行预处理，包括去除停用词、特殊字符等，并进行分词处理。

2) 构建知识图谱

在知识图谱中，实体通常表示为节点，关系则表示为连接节点的边。基于收集的数据构建知识图谱，其中的实体可以包括商品、用户、标签等，而关系可以包括商品与用户之间的关系、商品与标签之间的关系等。

3) 图神经网络(GNN)

使用图神经网络对知识图谱进行信息传播和表示学习，以捕捉实体和关系之间的复杂关联。将学到的实体嵌入用于后续的分类和推荐任务。为了进一步增加模型的泛化能力，因此在处理知识图谱的网络层中引入卷积核，构建图卷积神经网络[27]-[29]。

4) 文本分类和表示

使用 Bert 模型对商品的文本数据进行分类和表示。Bert 模型可以生成商品文本的上下文表示，该模型基于 BiLSTM-Attention 可以捕捉长距离依赖关系并生成有针对性的文本表示。

5) $\mu 2\text{Net}+$ (ViT-L/16)模型

结合 GNN 生成的实体嵌入和 Bert 生成的文本表示，使用 $\mu 2\text{Net}+$ (ViT-L/16)模型进行图像分割。这样可以充分利用多种类型的信息，实现多模态[30]-[36]并提高推荐的准确性。

6) 分类和推荐

全连接层作为最朴素也最万能的神经网络层，基于 GCN、Bert 和 $\mu 2\text{Net}+$ (ViT-L/16)模型的输出，经过全连接层后对商品进行分类。可根据分类结果为用户推荐相应类型的商品，同时可以考虑用户的历史购买记录、偏好等信息来优化推荐结果[37]-[43]。

全连接层的激活函数选用 Swish 函数，Swish 函数是谷歌提出的一种新的激活函数，其函数形式为 $f(x) = x * \text{sigmoid}(x)$ 。根据谷歌的研究，Swish 函数在某些深度学习模型中可以取得比 ReLU 函数更好的性能。

$$f(x) = \frac{x}{1 + e^{-x}} \quad (6)$$

最后的分类输出采用 softmax 函数归一化。

7) 评估与优化

通过交叉验证、调整超参数等方法对推荐系统进行评估与优化，以提高推荐质量。

综上所述，通过整合知识图谱、GCN、Bert 模型和 $\mu 2\text{Net}+$ (ViT-L/16)模型，本文实现了一个高效准确的商品混合推荐系统。

4. 实验

为了评估本文提出模型的有效性，下文分别进行了不同的实验。在本节中，将介绍数据集、实验设置及其结果。本文特意挑选了三个在商品推荐领域热门的开源数据集，淘宝直播多模态视频商品检索数据集(Taobao)、来自在线京东购物网站的 Products-10K 数据集，并进行了各项实验，并与基线模型进行了比较，具体实验如下。

4.1. 数据集

4.1.1. 淘宝直播多模态视频商品检索数据集

与业界开源数据集对比, 该数据集具有如下特点:

大规模: 数据集包括 7 万对视频商品匹配对, 标注图像 1,042,178 张, 标注检测框实例 1,654,780 个, 转录标注视频文本 7 万段。

多模态: 数据集面向实际直播视频场景, 既包括视频画面也包括对应的主播讲解文本, 商品侧则包括商品图和商品标题文本两个模态的数据。

多样性: 框级标注信息丰富多样, 包括商品检测框、类别、视角、展示方式、实例编号等。其中实例编号在一个视频商品匹配对的图像标注框之间起到同款标识的作用。

多功能: 数据标注了 23 类服饰检测类别和检测框位置, 可用于物体检测的算法研究。数据标注了框级实例编号, 构建了约 8 万组同款商品序列, 可用于物体检索识别算法研究。此外, 数据集提供了片段对应文本和商品标题描述文本, 可用于视觉文本多模态检索算法的研究。

4.1.2. Products-10K 数据集

本文实验中采用的第二个数据集为 Products-10K 数据集, 其中的所有图像均来自在线购物网站京东。数据集中包含上万个高频购买的 SKU, 涵盖时尚、3C、食品、医疗保健、家居用品等全品类。此外, 大尺寸的产品标签以图表形式组织, 以表示复杂的层次结构和产品之间的相互依赖。总共有近 15 万张图像。从实际应用场景来看, 图像数量分布不平衡。所有图像均由京东制作专家团队手动检查/标注。每张图像都会由至少三名人类专家检查其标签是否错误。近 44.5% 的噪声客户图像被人类专家过滤掉。整个数据集的噪声率低于 0.5%。Products10K 中为每个 SKU 收集店内照片和客户图像。客户图像由于背景复杂、色彩失真、光线、视角等复杂, 识别难度较大。Products-10K 数据集已可用于非商业研究和教育用途。

4.2. 基线模型

为了评估本文模型的性能, 本文选择了 GPT4Rec [8] 和 InstructRec [7] 及 KAR [21] 来进行比较。它们是最新的神经网络推荐模型, 在该领域目前被称为最先进的模型。下面对以上 3 种基线方法进行了介绍:

GPT4Rec [8]: 通过生成假设的搜索查询并检索相关项目, 结合语言模型和搜索引擎, 实现了用户兴趣的可解释表示, 在提升推荐的相关性、多样性及冷启动问题上取得了显著效果。

InstructRec [7]: 将推荐系统视为遵循用户自然语言指令的任务, 通过指令微调开源 LLM (如 Flan-T5-XL), 实现了基于用户偏好、意图和任务的自然语言推荐, 实验结果表明该模型在多项推荐任务中优于现有基线, 包括 GPT-3.5, 展现了构建更友好和准确的推荐系统的潜力。

KAR [21]: 利用大语言模型(LLMs)的外部知识, 通过推理用户偏好和获取物品的事实知识来增强推荐系统的性能, 并通过混合专家适配器将这些知识转化为兼容推荐任务的向量。

4.3. 评价指标

在神经网络模型评价指标中, 一般用准确率(Precision)、召回率(Recall)和 F1 值来衡量模型优劣, F1 值是准确率和召回率的一个调和平均值, 计算表达式如下。

$$F1 = \frac{2Precision \cdot Recall}{Precision + Recall} \quad (7)$$

在本文的实验中, 使用了均方误差(MSE)来评估本文提出的模型的性能, 因为它是大多数相关工作中用于评估的指标之一。令 N 为测试集中数据点的总数, MSE 计算表达式如下。

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (r_i - \hat{r}_i)^2 \quad (8)$$

其中 r_i 是第 i 个实际值, $\hat{r}_i$ 是第 i 个预测值。

在最后的文本推荐实验中采用了 HR@10 和 NDCG@10 作为评价指标, HR 为命中率, NDCG 为归一化折损累计增益, 它们常作为 Top-10 推荐评价指标, 它们的计算表达式如下。

$$\text{HR@N} = \frac{\text{hits@N}}{|GT|} \quad (9)$$

其中 GT 为测试集。

$$\text{NDCG@N} = Z_n \sum_{i=1}^N \frac{2^{\text{rel}_i} - 1}{\log_2(i+1)} \quad (10)$$

其中, rel_i 是第 i 位文本的相关性分级值, Z_n 是归一化值。在我们的实验中, 设置 $\text{rel}_i \in \{0, 1\}$ 取决于测试用户是否与文本 i 互动。

4.4. 结果分析与讨论

4.4.1. 精度和 F1 值的表现

在第二个实验中本文将提到的模型在各数据集上进行训练, 实验将各数据集中的图片进行分类学习训练。

从表 1 中可以得出, Products-10K 数据集上的精度达到了 83.79%, 这在大规模图像领域是相当高的精度。在 Taobao 数据集上也取得了 83.47% 和 82.42% 的好成绩, 与其他基线模型相比均得到了较好的结果。

Table 1. Accuracy performance of each model on each dataset

表 1. 各模型在各数据集上的精度表现

Metric	Datasets	GPT4Rec	InstructRec	KAR	KGBM4Rec
Accuracy	Taobao	0.7143	0.6992	0.7635	0.8347
	Products-10K	0.7627	0.7012	0.7831	0.8379

表 2 为本文模型在各数据集上的训练结果。本文提出的模型在各数据集上的 f1-score 分别为 0.9447, 0.9379, 均表现最优。

Table 2. F1 value performance of each model on each dataset

表 2. 各模型在各数据集上的 F1 值表现

Metric	Datasets	GPT4Rec	InstructRec	KAR	KGBM4Rec
F1	Taobao	0.9143	0.9192	0.9235	0.9447
	Products-10K	0.8927	0.9012	0.9031	0.9379

4.4.2. MSE 及推荐指标对比

后面的实验测试了各种模型分别在不同数据集上的表现, 其结果如下表 3 所示。最优的实验结果已加粗显示, 最后一列计算出本文 KGBM4Rec 模型较其他模型的提升, 以便能够更加直观地感受到本文模型的优越性。

Table 3. Performance of various models in MSE after experimentation on various datasets**表 3.** 各种模型在各种数据集上进行实验后在 MSE 的表现

Datasets	Metric	GPT4Rec	InstructRec	KAR	KGBM4Rec	Improvement
Taobao	MSE	1.6893	1.6322	1.6145	<u>1.3003</u>	24.087%
Products-10K		1.6442	1.6232	1.5325	<u>1.3012</u>	20.235%

从表 4 的实验结果可以明显看出, KGBM4Rec 模型在各数据集中均表现优异。本文提出的模型在推荐上的准确性和满意度均高于其他基线模型, 在 HR@1 中最高提升了 17.49%; 在 HR@5 中最高提升了 10.64%; HR@10 中最高提升了 3.17%; 在 NDCG@5 中最高提升了 16.96%; 在 NDCG@10 中最高提升了 5.36%。因此本文的模型在商品的推荐上较其他基线模型取得了更好的效果。

Table 4. Performance of various models on HR@N and NDCG@N after experiments on various datasets**表 4.** 各种模型在各种数据集上进行实验后在 HR@N 和 NDCG@N 的表现

Datasets	Metric	GPT4Rec	InstructRec	KAR	KGBM4Rec	Improvement
Taobao	HR@1	0.3657	0.3963	0.3132	<u>0.4373</u>	10.35%
	HR@5	0.6988	0.6107	0.6311	<u>0.7501</u>	7.34%
	HR@10	0.8869	0.8125	0.8221	<u>0.9033</u>	1.85%
	NDCG@5	0.6211	0.6549	0.6677	<u>0.7721</u>	15.64%
	NDCG@10	0.8634	0.8882	0.8986	<u>0.9233</u>	2.75%
Products-10K	HR@1	0.3667	0.3945	0.3032	<u>0.4173</u>	5.78%
	HR@5	0.6798	0.6211	0.6331	<u>0.7521</u>	10.64%
	HR@10	0.8769	0.8925	0.8321	<u>0.9142</u>	2.43%
	NDCG@5	0.6311	0.6609	0.6737	<u>0.7841</u>	16.39%
	NDCG@10	0.8934	0.8162	0.8286	<u>0.9413</u>	5.36%

5. 结论与展望

本文提出了一种基于知识图谱、图神经网络、Bert 序列模型和 μ 2Net+ (ViT-L/16)相结合的商品推荐算法, 该算法通过构建领域知识图谱, 利用图神经网络从知识图谱中提取商品的关联特征; 采用 Bert 序列模型对商品的文本特征进行建模; 联合 μ 2Net+ (ViT-L/16)视觉模型对商品图片进行深度感知学习; 最后根据用户偏好给出个性化推荐。实验结果表明, 所提出的综合推荐算法在多个数据集上均取得了较好的推荐效果。该算法具有较强的可扩展性和适用性, 为电商和互联网医疗平台等相关领域提供了更加精准的商品推荐服务。

参考文献

- [1] Gesmundo, A. and Dean, J. (2022) Munet: Evolving Pretrained deep Neural Networks into Scalable Auto-Tuning Multitask Systems. arXiv: 2205.10937.
- [2] Gao, C., Zheng, Y., Li, N., Li, Y., Qin, Y., Piao, J., *et al.* (2023) A Survey of Graph Neural Networks for Recommender Systems: Challenges, Methods, and Directions. *ACM Transactions on Recommender Systems*, **1**, 1-51. <https://doi.org/10.1145/3568022>
- [3] Yu, J., Yin, H., Xia, X., *et al.* (2023) Self-Supervised Learning for Recommender Systems: A Survey. *IEEE Transactions on Knowledge and Data Engineering*, **36**, 335-355.

- [4] Liu, P., Zhang, L. and Gulla, J.A. (2023) Pre-Train, Prompt, and Recommendation: A Comprehensive Survey of Language Modeling Paradigm Adaptations in Recommender Systems. *Transactions of the Association for Computational Linguistics*, **11**, 1553-1571. https://doi.org/10.1162/tacl_a_00619
- [5] Walek, B. and Fajmon, P. (2023) A Hybrid Recommender System for an Online Store Using a Fuzzy Expert System. *Expert Systems with Applications*, **212**, Article ID: 118565. <https://doi.org/10.1016/j.eswa.2022.118565>
- [6] Lin, Y., Liu, Y., Lin, F., Zou, L., Wu, P., Zeng, W., et al. (2024) A Survey on Reinforcement Learning for Recommender Systems. *IEEE Transactions on Neural Networks and Learning Systems*, **35**, 13164-13184. <https://doi.org/10.1109/tnnls.2023.3280161>
- [7] Sun, F., Liu, J., Wu, J., Pei, C., Lin, X., Ou, W., et al. (2019) BERT4Rec: Sequential Recommendation with Bidirectional Encoder Representations from Transform. *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, Beijing, 3-7 November 2019, 1441-1450. <https://doi.org/10.1145/3357384.3357895>
- [8] Naumov, M., Mudigere, D., Shi, H.J.M., et al. (2019) Deep Learning Recommendation Model for Personalization and Recommendation Systems. arXiv: 1906.00091.
- [9] Peng, C., Xia, F., Naseriparsa, M., et al. (2023) Knowledge Graphs: Opportunities and Challenges. *Artificial Intelligence Review*, **56**, 13071-13102.
- [10] Chandak, P., Huang, K. and Zitnik, M. (2023) Building a Knowledge Graph to Enable Precision Medicine. *Scientific Data*, **10**, Article No. 67. <https://doi.org/10.1038/s41597-023-01960-3>
- [11] Chen, X., Jia, S. and Xiang, Y. (2020) A Review: Knowledge Reasoning over Knowledge Graph. *Expert Systems with Applications*, **141**, Article ID: 112948. <https://doi.org/10.1016/j.eswa.2019.112948>
- [12] Zhu, X., Li, Z., Wang, X., et al. (2022) Multi-Modal Knowledge Graph Construction and Application: A Survey. *IEEE Transactions on Knowledge and Data Engineering*, **36**, 715-735.
- [13] Wu, X., Duan, J., Pan, Y. and Li, M. (2023) Medical Knowledge Graph: Data Sources, Construction, Reasoning, and Applications. *Big Data Mining and Analytics*, **6**, 201-217. <https://doi.org/10.26599/bdma.2022.9020021>
- [14] Zhou, B., Shen, X., Lu, Y., Li, X., Hua, B., Liu, T., et al. (2022) Semantic-aware Event Link Reasoning over Industrial Knowledge Graph Embedding Time Series Data. *International Journal of Production Research*, **61**, 4117-4134. <https://doi.org/10.1080/00207543.2021.2022803>
- [15] Veličković, P. (2023) Everything Is Connected: Graph Neural Networks. *Current Opinion in Structural Biology*, **79**, Article ID: 102538. <https://doi.org/10.1016/j.sbi.2023.102538>
- [16] Wu, L., Chen, Y., Shen, K., Guo, X., Gao, H., Li, S., et al. (2023) Graph Neural Networks for Natural Language Processing: A Survey. *Foundations and Trends in Machine Learning*, **16**, 119-328. <https://doi.org/10.1561/22000000096>
- [17] Cappart, Q., Chételat, D., Khalil, E.B., et al. (2023) Combinatorial Optimization and Reasoning with Graph Neural Networks. *Journal of Machine Learning Research*, **24**, 130:1-130:61.
- [18] Zhou, C., Li, Q., Li, C., Yu, J., Liu, Y., Wang, G., et al. (2024) A Comprehensive Survey on Pretrained Foundation Models: A History from BERT to ChatGPT. *International Journal of Machine Learning and Cybernetics*. <https://doi.org/10.1007/s13042-024-02443-6>
- [19] Zyuzin, V. and Chumarnaya, T. (2019) Comparison of UNet Architectures for Segmentation of the Left Ventricle Endocardial Border on Two-Dimensional Ultrasound Images. 2019 *Ural Symposium on Biomedical Engineering, Radioelectronics and Information Technology (USBEREIT)*, Yekaterinburg, 25-26 April 2019, 110-113. <https://doi.org/10.1109/usbereit.2019.8736616>
- [20] Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., et al. (2023) Swin-Unet: UNet-Like Pure Transformer for Medical Image Segmentation. In: Karlinsky, L., Michaeli, T. and Nishino, K., Eds., *Computer Vision—ECCV 2022 Workshops*, Springer, 205-218. https://doi.org/10.1007/978-3-031-25066-8_9
- [21] Gesmundo, A. (2022) A Continual Development Methodology for Large-Scale Multitask Dynamic ML Systems. arXiv: 2209.07326.
- [22] Sanchez-Lengeling, B., Reif, E., Pearce, A. and Wiltchko, A. (2021) A Gentle Introduction to Graph Neural Networks. *Distill*, **6**, e33. <https://doi.org/10.23915/distill.00033>
- [23] Vaswani, A., et al. (2017) Attention Is All You Need. arXiv: 1706.03762.
- [24] Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A. and Zagoruyko, S. (2020) End-to-End Object Detection with Transformers. In: Vedaldi, A., Bischof, H., Brox, T. and Frahm, J.M., Eds., *Computer Vision—ECCV 2020*, Springer, 213-229. https://doi.org/10.1007/978-3-030-58452-8_13
- [25] Dosovitskiy, A., et al. (2021) An Image Is Worth 16 × 16 Words: Transformers for Image Recognition at Scale. arXiv: 2010.11929.
- [26] Gabeur, V., Sun, C., Alahari, K. and Schmid, C. (2020) Multi-Modal Transformer for Video Retrieval. In: Vedaldi, A.,

- Bischof, H., Brox, T. and Frahm, J.M., Eds., *Computer Vision—ECCV 2020*, Springer, 214-229. https://doi.org/10.1007/978-3-030-58548-8_13
- [27] Zhang, J., Li, C., Yin, Y., Zhang, J. and Grzegorzec, M. (2022) Applications of Artificial Neural Networks in Microorganism Image Analysis: A Comprehensive Review from Conventional Multilayer Perceptron to Popular Convolutional Neural Network and Potential Visual Transformer. *Artificial Intelligence Review*, **56**, 1013-1070. <https://doi.org/10.1007/s10462-022-10192-7>
- [28] Cong, S. and Zhou, Y. (2022) A Review of Convolutional Neural Network Architectures and Their Optimizations. *Artificial Intelligence Review*, **56**, 1905-1969. <https://doi.org/10.1007/s10462-022-10213-5>
- [29] Gaur, L., Bhatia, U., Jhanjhi, N.Z., Muhammad, G. and Masud, M. (2021) Medical Image-Based Detection of COVID-19 Using Deep Convolution Neural Networks. *Multimedia Systems*, **29**, 1729-1738. <https://doi.org/10.1007/s00530-021-00794-6>
- [30] Zhou, X. (2023) MMRec: Simplifying Multimodal Recommendation. *ACM Multimedia Asia Workshops*, Tainan, 6-8 December 2023, 1-2. <https://doi.org/10.1145/3611380.3628561>
- [31] Liu, K., Xue, F., Guo, D., et al. (2023) MEGCF: Multimodal Entity Graph Collaborative Filtering for Personalized Recommendation. *ACM Transactions on Information Systems*, **41**, 1-27.
- [32] Mu, Y. and Wu, Y. (2023) Multimodal Movie Recommendation System Using Deep Learning. *Mathematics*, **11**, Article 895. <https://doi.org/10.3390/math11040895>
- [33] Zhang, L., Zhou, X. and Shen, Z. (2023) Multimodal Pre-Training Framework for Sequential Recommendation via Contrastive Learning. arXiv: 2303.11879.
- [34] Lu, Q., Sun, X., Sutcliffe, R., Xing, Y. and Zhang, H. (2022) Sentiment Interaction and Multi-Graph Perception with Graph Convolutional Networks for Aspect-Based Sentiment Analysis. *Knowledge-Based Systems*, **256**, Article ID: 109840. <https://doi.org/10.1016/j.knosys.2022.109840>
- [35] Huang, S., Dong, L., Wang, W., et al. (2023) Language Is Not All You Need: Aligning Perception with Language Models. arXiv: 2302.14045.
- [36] Abbasi-Moud, Z., Hosseinabadi, S., Kelarestaghi, M. and Eshghi, F. (2022) CAFOB: Context-Aware Fuzzy-Ontology-Based Tourism Recommendation System. *Expert Systems with Applications*, **199**, Article ID: 116877. <https://doi.org/10.1016/j.eswa.2022.116877>
- [37] Gesmundo, A. (2023) Multipath Agents for Modular Multitask ML Systems. arXiv: 2302.02721.
- [38] Urabe, Y., Rzepka, R. and Araki, K. (2021) Find Right Countenance for Your Input—Improving Automatic Emoticon Recommendation System with Distributed Representations. *Information Processing & Management*, **58**, Article ID: 102414. <https://doi.org/10.1016/j.ipm.2020.102414>
- [39] Zhou, Q., Zhuang, W., Ren, H., Chen, Y., Yu, B., Lou, J., et al. (2022) Hybrid Collaborative Filtering Model for Consumer Dynamic Service Recommendation Based on Mobile Cloud Information System. *Information Processing & Management*, **59**, Article ID: 102871. <https://doi.org/10.1016/j.ipm.2022.102871>
- [40] Misztal-Radecka, J. and Indurkha, B. (2021) Bias-Aware Hierarchical Clustering for Detecting the Discriminated Groups of Users in Recommendation Systems. *Information Processing & Management*, **58**, Article ID: 102519. <https://doi.org/10.1016/j.ipm.2021.102519>
- [41] Zhao, Z., Fan, W., Li, J., Liu, Y., Mei, X., Wang, Y., et al. (2024) Recommender Systems in the Era of Large Language Models (LLMs). *IEEE Transactions on Knowledge and Data Engineering*, **36**, 6889-6907. <https://doi.org/10.1109/tkde.2024.3392335>
- [42] Gao, C., Zheng, Y., Wang, W., Feng, F., He, X. and Li, Y. (2024) Causal Inference in Recommender Systems: A Survey and Future Directions. *ACM Transactions on Information Systems*, **42**, 1-32. <https://doi.org/10.1145/3639048>
- [43] Deldjoo, Y., Jannach, D., Bellogin, A., Difonzo, A. and Zanzonelli, D. (2023) Fairness in Recommender Systems: Research Landscape and Future Directions. *User Modeling and User-Adapted Interaction*, **34**, 59-108. <https://doi.org/10.1007/s11257-023-09364-z>