

LLMs 推荐发展综述-Representing篇：单模态推荐 & 多模态推荐

原创 方方 方方的算法花园 2024年11月07日 09:02 北京

点击蓝字 关注我们



写在前面

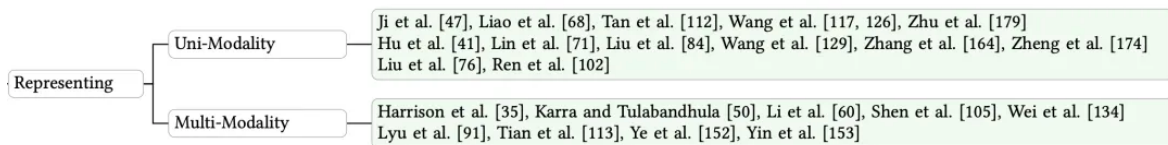
吉林大学、悉尼科技大学、香港理工大学在2024年10月联合发表了一篇论文《Towards Next-Generation LLM-based Recommender Systems: A Survey and Beyond》，论文里探讨了基于大语言模型（LLMs）的推荐系统的发展，包括其在表示与理解、设计与应用、工业部署等方面的应用，以及面临的挑战和机遇。（论文链接：<https://arxiv.org/pdf/2410.19744>）

本文主要介绍论文中关于Representing（表示）部分的内容，其他部分内容请参考此系列其他文章。



REPRESENTING 表示篇

在现代推荐系统中，用户-物品交互的表示形式是生成准确且个性化推荐的基础，其涉及多模态数据的融合与处理，包括文本、图像、音频、视频、元数据等，以全面捕捉用户偏好与物品特征。随着推荐系统的发展以满足用户多样化需求，考虑数据多样性及其所捕捉关系的复杂性至关重要。这使我们探索**单模态推荐**和**多模态推荐**两种方法：**单模态推荐**侧重利用**用户-物品交互**，通过图数据对其建模，偶尔辅以文本信息来完善对用户偏好和物品属性的理解；**多模态方法**则超越单模态界限，整合来自多个来源或模态的数据，如文本、图像、视频等，利用多种数据源的可用性更丰富地了解用户偏好和物品属性，提供更全面灵活的推荐策略。下面就分别对两种方式的研究工作进行介绍。



单模态



单模态推荐：单模态是指利用来自单一模态数据的推荐系统。在这种方法中，用户-物品交互 R 主要通过一种类型的数据来表示，例如用户评分或交互历史，而不整合图像、音频或视频等其他数据源。然而，在某些情况下，会整合文本信息 Ttxt 来增强这种单模态内的表示。

[68] Llara: Largelanguage-recommendation assistant. arXiv preprint arXiv:2312.02445, 2023.

认为仅用基于 ID 或文本的物品序列表示来提示LLM，不能充分发挥其潜力，需让其深入理解序列交互中的行为模式。该研究探索了LLM与序列推荐系统的一致性，提出将“用户序列行

为”作为新模态，使其与语言空间相契合，不再局限于简单提示。

[129] Drdt: Dynamic reflection with divergent thinking for llm-based sequential recommendation. arXiv preprint arXiv:2312.11336, 2023.

聚焦于序列推荐，提出了一种新的推理原则：在检索-重排框架内进行动态反思与发散思维。

[84] Modeling user viewing flow using large language models for article recommendation. In Companion Proceedings of the ACM on Web Conference 2024, pages 83–92, 2024.

提出了一种用户浏览流建模（SINGLE）方法来为用户推荐文章。它对持续和即时的浏览流进行建模，以更好地表示用户兴趣，增强对一般和实时偏好的提取以进行推荐。

[174] Harnessing large language models for text-rich sequential recommendation. In Proceedings of the 33rd International World Wide Web Conference, pages 3207—3216, 2024.

介绍了在富含文本的序列推荐中利用LLM的创新框架（LLM-TRSR）。该方法包括几个关键步骤：首先提取用户的行为历史序列并将其转换为扩展文本格式；然后将该文本分割成多个块，每个块设计为适合LLM的处理限制；最后，使用基于LLM的摘要器，对这些块进行综合，以生成用户偏好的全面总结。

[164] Finerec: Exploring fine-grained sequential recommendation. In Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval, pages 1599–1608, 2024.

引入了处理序列推荐的新视角，探索评论的属性意见以精细揭示用户偏好和物品特征。为了在不同属性下表示用户/物品，这项工作为每个属性创建了一个独特的特定属性用户意见物品图。

[41] Enhancing sequential recommendation via llm-based semantic embedding learning. In Companion Proceedings of the ACM on Web Conference 2024, pages 103–111, 2024

专注于序列推荐，引入了 SAID 框架，旨在利用LLM直接从文本中学习语义对齐的物品 ID 嵌入。对于每个物品，SAID 使用投影模块将物品 ID 转换为嵌入向量，然后由LLM进行处理，以生成与物品相关的精确描述文本标记。生成的物品嵌入经过优化，以捕捉文本描述中包含的详细语义信息。

[71] A multi-facet paradigm to bridge large language model and recommendation. arXiv preprint arXiv:2310.06491, 2023

引入了一种名为 TransRec 的新型多方面范式，将LLM与推荐系统连接起来。具体而言，TransRec 利用多方面标识符，整合 ID、标题和属性，以捕捉独特性和语义含义。

[76] Understanding before recommendation: Semantic aspect-aware review exploitation via large language models. arXiv preprint arXiv:2312.16275, 2023.

引入了一种基于链的提示技术，旨在揭示语义方面感知的交互，在细粒度语义层面提供更详细的用户行为洞察。为了有效利用不同方面的广泛交互，该工作提出了语义方面基础的图卷积网络（SAGCN），这是一种简单而强大的方法。

[102] Representation learning with large language models for recommendation. In Proceedings of the ACM on Web Conference 2024, pages 3464–3475, 2024.

引入了协作档案生成范式和推理驱动的系统提示，强调在生成的输出中整合推理过程。RLMRec 利用对比和生成对齐技术，将协作过滤（CF）关系嵌入与 LLM 语义表示对齐，有效缓解特征噪声。

[126] Llm-enhanced user-item interactions: Leveraging edge information for optimized recommendations. arXiv preprint arXiv:2402.09617, 2024

解决了 LLM 中边缘信息未被充分利用的挑战，特别是在关键注意力机制内。所提出的方法通过纳入用户与物品之间的直接关系以及构建物品之间的二阶关系，增强了模型的理解能力，提供了传统推荐数据中所缺乏的复杂关联。

[47] Genrec: Large language model for generative recommendation. In European Conference on Information Retrieval, pages 494–502. Springer, 2024.

利用物品名称中固有的详细信息，这些信息通常包含适合语义分析的特征，从而有助于更好地理解物品与用户的潜在相关性。所提出的 GenRec 模型通过纳入文本信息来提高生成式推荐的性能。

[112] Idgenrec: Llm-recsys alignment with textual id learning. In Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval, pages 355–364, 2024.

通过引入 IDGenRec 解决了生成式推荐系统中的物品编码问题，它引入了一个 ID 生成器来生成独特、简洁且语义丰富的文本 ID，这些 ID 与平台无关且基于人类词汇。

[179] Collaborative large language model for recommender systems. In Proceedings of the ACM on Web Conference 2024, pages 3162–3172, 2024.

将预训练 LLM 的用户/物品 ID 标记的词汇表进行扩展，以忠实建模用户/物品的协作和内容语义。此外，该工作还提出了一种新颖的“软+硬提示”策略，通过在特定推荐系统语料上进行语言建模，有效地获取用户/物品的协作/内容标记嵌入。它将每个文档分成一个由异构软（用户/物品）标记和硬（词汇）标记组成的提示，以及一个由同质物品标记或词汇标记组成的主要文本。

[117] Multiple key-value strategy in recommendation systems incorporating large language model. arXiv preprint arXiv:2310.16409, 2023.

考虑了多种键值数据，因为这种情况在现实世界应用中很普遍，用户信息（如年龄、职业）和物品细节（如标题、类别）包含多个键。



多模态推荐：多模态是指整合来自多个模态数据的推荐系统，用

$$\mathcal{D} = \{(\mathcal{R}, \mathcal{T}_{\text{txt}}, \mathcal{I}, \mathcal{A}, \mathcal{V}_{\text{vid}}, \mathcal{S}, \mathcal{T})\}$$

表示，以提供对用户和物品更丰富、更全面的表示。通过融合来自文本、图像、音频、视频、序列和时间戳等方面的信息，该系统能够捕捉用户偏好和物品特征的多个方面，从而得出更稳健、更细致的推荐。

[134]Llmrec: Large language models with graph augmentation for recommendation. In Proceedings of the 17th ACM International Conference on Web Search and Data Mining, pages 806–815, 2024.

利用LLM来增强推荐系统中的图，包括增强用户-物品交互边、物品节点属性和用户节点轮廓。它通过授权LLM明确推理用户-物品交互模式来解决隐式反馈信号的稀缺性问题。此外，它还克服了低质量辅助信息的问题，通过生成用户和物品属性并实施去噪增强机制来解决。

[50] Interarec: Interactive recommendations using multimodal large language models. In Pacific-Asia Conference on Knowledge Discovery and Data Mining, pages 32–43. Springer, 2024.

提出了一种创新方法，利用用户互联网浏览活动截图中的信息。这种方法利用多模态LLM，这些模型擅长处理和生成各种模态的内容。通过使用截图而不是网络日志，系统获得了更好的可解释性。截图的视觉性质清楚透明地描绘了用户行为，显著增强了对LLM推断的理解。

[35] Zero-shot recommendations with pre-trained large language models for multimodal nudging. In 2023 IEEE International Conference on Data Mining Workshops (ICDMW), pages 1535–1542. IEEE, 2023.

提出了一种零样本推荐多模态非平稳内容的方法，利用生成式人工智能的最新进展。所提出的方法涉及将不同模态的输入呈现为文本描述，并利用预训练LLM通过计算语义嵌入来获得它们的数值表示。

[60] Large language models for nextpoint-of-interest recommendation. arXiv preprint arXiv:2404.17591, 2024.

专注于增强基于位置的社交网络 (LBSNs) 中的推荐，并引入了一种创新的用于下一个兴趣点 (POI) 推荐的LLM 4POI 框架。该框架与传统的数值方法不同，传统方法通常需要数据转换和使用各种嵌入层。相反，LLM 4POI 保留了基于位置的社交网络数据的原始格式，从而在不损失任何信息的情况下保留了上下文信息。

[153] Heterogeneous knowledge fusion: A novel approach for personalized recommendation via llm. In Proceedings of the 17th ACM Conference on Recommender Systems, page 599–601, 2023

提出了一种用户建模的新范式，从异构用户行为数据中提取和整合多种知识。通过将结构化的用户行为转换为非结构化的异构知识，它有效地捕捉了用户兴趣。在美团外卖的背景下，用户异构行为包括：多个行为主体，如商家和产品；多个行为内容，如曝光、点击和订单；多个行为场景，如 APP 首页和小程序。

[113]Mmrec: Llm based multi-modal recommender system.arXiv preprint arXiv: 2408 .04211, 2024.

利用LLM的推理和总结能力来有效处理多模态信息。该研究利用LLM总结用户评论文本，捕捉微妙的用户行为和偏好，并为图像生成描述性文本，提取关于商家和产品的隐含见解。

[152] Harnessing multimodal large language models for multimodal sequential recommendation. arXiv preprint arXiv:2408.09698, 2024.

引入了一种基于多模态LLM (MLLMs) 的图像总结方法，以循环方式总结用户在多个模态上的偏好。这种方法使我们能够更深入地了解用户随时间的交互和兴趣，增强了捕捉和预测用

户行为的能力。具体而言，这项工作首先利用基于多模态LLM的物品总结器从给定物品中提取图像特征，并将图像转换为文本，使视觉信息更有效地融入推荐过程。

[91] X-reflect:Cross-reflection prompting for multimodal recommendation. arXiv preprint arXiv:2408.15172, 2024.

提出了一种称为交叉反射提示的新颖方法，称为 X-REFLECT。交叉反射提示背后的核心思想是为大型多模态模型（LMMs）提供文本和视觉上下文，同时明确指导模型确定这些信息是支持还是矛盾。然后，大型多模态模型的输出被转换为嵌入，这些嵌入被用作后续推荐模块的输入物品嵌入。

[105] Pmg: Personalized multimodal generation with large language models. In Proceedings of the ACM on Web Conference 2024, pages 3833–3843, 2024.

考虑了两类用户行为：历史点击和对话。输入特征是多模态的，包括文本、图像、音频等。它利用LLM来预处理并将每个物品和对话的文本特征总结为简洁的句子。对于其他模态，它使用标题模型（如 BLIP-2[55]、CLAP[21]）或多模态LLM（如 MiniGPT-4[177]、mPLUG-Owl[151]）将其转换为文本，这些模型能够处理多种类型的输入。这个预处理步骤的目的是浓缩特征，消除冗余，并保持长期上下文。



LLM与推荐 15

LLM与推荐 · 目录

上一篇

【LLM论文阅读】RLRF4Rec：从推荐系统反馈中进行强化学习以增强推荐重排序

下一篇

LLMs 推荐发展综述-生成型推荐 & 非生成型推荐