(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 118467840 A
(43)申请公布日 2024.08.09

(21)申请号 202410694813.9

(22)申请日 2024.05.31

(71)申请人 北京智谱华章科技有限公司
地址 100086 北京市海淀区中关村东路1号院9号楼10层整层

(72)发明人 刘丁枭 马晋辰

(74)专利代理机构 北京佳信天和知识产权代理事务所(普通合伙) 11939
专利代理师 田英楠

(51)Int.Cl.
G06F 16/9535(2019.01)
G06F 16/36(2019.01)
G06F 16/33(2019.01)
G06N 20/00(2019.01)
G06N 3/006(2023.01)
G06N 5/025(2023.01)
G06F 40/30(2020.01)

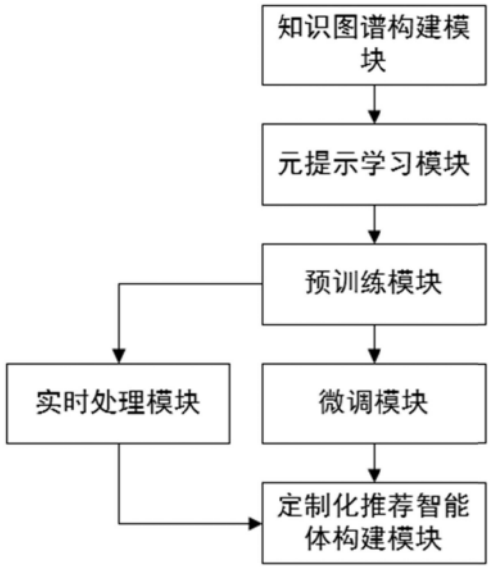
权利要求书2页 说明书9页 附图3页

(54)发明名称

一种基于知识增强和大模型的推荐智能体系统及方法

(57)摘要

本发明属于人工智能技术领域,涉及一种基于知识增强和大模型的推荐智能体系统,其包括:知识图谱构建模块,其用于收集和整合领域相关知识,以形成结构化的知识图谱;元提示学习模块,其用于基于结构化的知识图谱和文本数据,采用元提示学习方法形成一组元提示模板;预训练模块,其用于对大模型进行预训练;微调模块,其用于进行微调;定制化推荐智能体构建模块,其用于形成定制化推荐智能体;实时处理模块,其用于对用户请求进行实时处理并将处理结果输入所述定制化推荐智能体,由定制化推荐智能体生成定制化的推荐结果。其能够更准确地捕捉用户的兴趣和偏好,为用户提供个性化的定制推荐服务,帮助用户发现感兴趣的内容。



CN 118467840 A

1.一种基于知识增强和大模型的推荐智能体系统,其特征在于,包括:

知识图谱构建模块,其用于收集和整合领域相关知识,以形成结构化的知识图谱;

元提示学习模块,其用于基于所述结构化的知识图谱和文本数据,采用元提示学习方法形成一组元提示模板;

预训练模块,其用于使用所述元提示模板对大模型进行预训练;

微调模块,其用于对预训练后的大模型进行微调;

定制化推荐智能体构建模块,其用于使用用户信息、项目信息和用户行为信息对微调后的大模型进行训练,以形成定制化推荐智能体;

实时处理模块,其用于基于所述预训练后的大模型对用户请求进行实时处理并将处理结果输入所述定制化推荐智能体,由所述定制化推荐智能体生成定制化的推荐结果。

2.根据权利要求1所述的基于知识增强和大模型的推荐智能体系统,其特征在于,所述实时处理模块对用户请求进行实时处理具体包括:

使用预训练后的大模型对用户请求进行语义分析和关键词提取;

基于所述语义分析和关键词提取的结果从所述结构化的知识图谱中进行语义关联和信息检索,以获取与用户请求相关的实体、关系和属性;

基于所述实体、关系和属性对用户请求进行增强。

3.根据权利要求1所述的基于知识增强和大模型的推荐智能体系统,其特征在于,所述知识图谱构建模块包括:

数据收集子模块,其用于从多个开源平台收集领域相关知识;

实体识别子模块,其用于从收集到的领域相关知识中识别出实体;

关系抽取子模块,其用于从收集到的领域相关知识中抽取出实体之间的关系;

属性抽取子模块,其用于从收集到的领域相关知识中抽取出能够代表实体特征的属性;

知识融合子模块,其用于将识别出的实体、关系和属性进行融合,构建成所述结构化的知识图谱。

4.根据权利要求3所述的基于知识增强和大模型的推荐智能体系统,其特征在于,所述知识图谱构建模块还包括:

知识更新子模块,其用于将新知识添加到所述结构化的知识图谱中、更新所述结构化的知识图谱中的旧知识和修正所述结构化的知识图谱中的错误知识。

5.根据权利要求4所述的基于知识增强和大模型的推荐智能体系统,其特征在于,所述知识图谱构建模块还包括:

知识存储和管理子模块,其用于存储和管理所述结构化的知识图谱。

6.根据权利要求1所述的基于知识增强和大模型的推荐智能体系统,其特征在于,所述预训练模块采用自监督学习的方式,通过最大化文本序列的概率来调整大模型的参数,并且,其采用的损失函数为交叉熵损失函数、采用的优化方法为Adam优化算法。

7.根据权利要求1所述的基于知识增强和大模型的推荐智能体系统,其特征在于,所述微调模块包括:

加载子模块,其用于从预训练后的大模型中加载初始参数;

冻结子模块,其用于冻结预训练后的大模型的部分层,以保留预训练后的大模型的知

识;

添加子模块,其用于在预训练后的大模型的基础上,添加任务层,以适应推荐智能体系统的具体任务和需求;

训练子模块,其用于使用微调目标函数对预训练后的大模型进行训练;

评估子模块,其用于定期评估训练后的大模型的性能。

8.根据权利要求7所述的基于知识增强和大模型的推荐智能体系统,其特征在于,所述冻结子模块用于冻结预训练后的大模型的底层;所述添加子模块用于在预训练后的大模型中添加嵌入层。

9.一种基于知识增强和大模型的推荐方法,其特征在于,其采用了权利要求1-8中任一项所述的基于知识增强和大模型的推荐智能体系统,并包括以下步骤:

收集和整合领域相关知识,以形成结构化的知识图谱;

基于所述结构化的知识图谱和文本数据,采用元提示学习方法形成一组元提示模板;

使用所述元提示模板对大模型进行预训练;

对预训练后的大模型进行微调;

使用用户信息、项目信息和用户行为信息对微调后的大模型进行训练,以形成定制化推荐智能体;

基于所述预训练后的大模型对用户请求进行实时处理并将处理结果输入所述定制化推荐智能体,由所述定制化推荐智能体生成定制化的推荐结果。

10.根据权利要求9所述的基于知识增强和大模型的推荐方法,其特征在于,对用户请求进行实时处理具体包括:

使用预训练后的大模型对用户请求进行语义分析和关键词提取;

基于所述语义分析和关键词提取的结果从所述结构化的知识图谱中进行语义关联和信息检索,以获取与用户请求相关的实体、关系和属性和关系;

基于所述实体、关系和属性对用户请求进行增强。

一种基于知识增强和大模型的推荐智能体系统及方法

技术领域

[0001] 本发明属于人工智能技术领域,涉及一种推荐智能体系统及方法,尤其涉及一种基于知识增强和大模型的推荐智能体系统及方法。

背景技术

[0002] 随着互联网的普及和信息的爆炸式增长,在海量的信息中,用户难以找到自己感兴趣的内容,而推荐智能体通过分析用户的历史行为和偏好,能够为用户提供个性化的定制推荐,帮助用户发现感兴趣的内容。

[0003] 目前,现有的推荐技术主要包括:

[0004] 1、基于知识图谱的推荐技术。在推荐系统领域,知识图谱的引入是一种常见的知识增强技术。知识图谱通过实体、关系和属性来描述现实世界中的各种概念和它们之间的关系。基于知识图谱的推荐系统利用知识图谱中的结构化知识来提高推荐的准确性和解释性。例如,可以通过知识图谱中的关系路径来发现用户和项目之间的潜在联系,从而提供更加准确的推荐。但是,仅仅依靠知识图谱,无法挖掘用户的真实用途,也无法理解用户问题的语境和用户偏好,从而使得推荐性能不足。

[0005] 2、基于大模型预训练的推荐技术。随着大模型技术的发展,越来越多的研究开始将大规模预训练模型应用于推荐系统。这些大模型通过在大量数据上进行预训练,学习到了丰富的特征表示和用户偏好信息。基于大模型预训练的推荐系统可以通过微调策略来学习用户的特定偏好,从而提高推荐的准确性和多样性。但是,受限于模型结构和训练方法方,无法为用户提供定制化的推荐服务。

[0006] 3、基于内容的推荐技术。它是一种能够根据用户偏好数据和上下文信息提供个性化推荐的系统。基于内容的推荐技术通常需要处理海量的用户行为数据,并能够快速响应用户的偏好变化。为了实现基于内容的推荐技术,现有的推荐系统通常采用了基于协同过滤技术、TF-IDF算法和聚类等技术。但是,其效率和准确性不足,难以提供定制化推荐服务。

[0007] 因此,针对上述现有技术中存在的缺陷,需要研发一种新型的推荐智能体系统及方法。

发明内容

[0008] 为了克服现有技术的缺陷,本发明提出一种基于知识增强和大模型的推荐智能体系统及方法,旨在解决现有推荐系统在处理海量数据、提高推荐准确性和实时性方面的问题,其通过结合知识图谱和大模型技术,具备对用户行为的深度理解和分析能力,能够更准确地捕捉用户的兴趣和偏好,为用户提供个性化的定制推荐服务,帮助用户发现感兴趣的内容。

[0009] 为了实现上述目的,本发明提供如下技术方案:

[0010] 一种基于知识增强和大模型的推荐智能体系统,其特征在于,包括:

[0011] 知识图谱构建模块,其用于收集和整合领域相关知识,以形成结构化的知识图谱;

[0012] 元提示学习模块,其用于基于所述结构化的知识图谱和文本数据,采用元提示学习方法形成一组元提示模板;

[0013] 预训练模块,其用于使用所述元提示模板对大模型进行预训练;

[0014] 微调模块,其用于对预训练后的大模型进行微调;

[0015] 定制化推荐智能体构建模块,其用于使用用户信息、项目信息和用户行为信息对微调后的大模型进行训练,以形成定制化推荐智能体;

[0016] 实时处理模块,其用于基于所述预训练后的大模型对用户请求进行实时处理并将处理结果输入所述定制化推荐智能体,由所述定制化推荐智能体生成定制化的推荐结果。

[0017] 优选地,所述实时处理模块对用户请求进行实时处理具体包括:

[0018] 使用预训练后的大模型对用户请求进行语义分析和关键词提取;

[0019] 基于所述语义分析和关键词提取的结果从所述结构化的知识图谱中进行语义关联和信息检索,以获取与用户请求相关的实体、关系和属性;

[0020] 基于所述实体、关系和属性对用户请求进行增强。

[0021] 优选地,所述知识图谱构建模块包括:

[0022] 数据收集子模块,其用于从多个开源平台收集领域相关知识;

[0023] 实体识别子模块,其用于从收集到的领域相关知识中识别出实体;

[0024] 关系抽取子模块,其用于从收集到的领域相关知识中抽取出实体之间的关系;

[0025] 属性抽取子模块,其用于从收集到的领域相关知识中抽取出能够代表实体特征的属性;

[0026] 知识融合子模块,其用于将识别出的实体、关系和属性进行融合,构建成所述结构化的知识图谱。

[0027] 优选地,所述知识图谱构建模块还包括:

[0028] 知识更新子模块,其用于将新知识添加到所述结构化的知识图谱中、更新所述结构化的知识图谱中的旧知识和修正所述结构化的知识图谱中的错误知识。

[0029] 优选地,所述知识图谱构建模块还包括:

[0030] 知识存储和管理子模块,其用于存储和管理所述结构化的知识图谱。

[0031] 优选地,所述预训练模块采用自监督学习的方式,通过最大化文本序列的概率来调整大模型的参数,并且,其采用的损失函数为交叉熵损失函数、采用的优化方法为Adam优化算法。

[0032] 优选地,所述微调模块包括:

[0033] 加载子模块,其用于从预训练后的大模型中加载初始参数;

[0034] 冻结子模块,其用于冻结预训练后的大模型的部分层,以保留预训练后的大模型的知识;

[0035] 添加子模块,其用于在预训练后的大模型的基础上,添加任务层,以适应推荐智能体系统的具体任务和需求;

[0036] 训练子模块,其用于使用微调目标函数对预训练后的大模型进行训练;

[0037] 评估子模块,其用于定期评估训练后的大模型的性能。

[0038] 优选地,所述冻结子模块用于冻结预训练后的大模型的底层;所述添加子模块用于在预训练后的大模型中添加嵌入层。

[0039] 此外,本发明还提供一种基于知识增强和大模型的推荐方法,其特征在于,其采用了上述基于知识增强和大模型的推荐智能体系统,并包括以下步骤:

[0040] 收集和整合领域相关知识,以形成结构化的知识图谱;

[0041] 基于所述结构化的知识图谱和文本数据,采用元提示学习方法形成一组元提示模板;

[0042] 使用所述元提示模板对大模型进行预训练;

[0043] 对预训练后的大模型进行微调;

[0044] 使用用户信息、项目信息和用户行为信息对微调后的大模型进行训练,以形成定制化推荐智能体;

[0045] 基于所述预训练后的大模型对用户请求进行实时处理并将处理结果输入所述定制化推荐智能体,由所述定制化推荐智能体生成定制化的推荐结果。

[0046] 优选地,对用户请求进行实时处理具体包括:

[0047] 使用预训练后的大模型对用户请求进行语义分析和关键词提取;

[0048] 基于所述语义分析和关键词提取的结果从所述结构化的知识图谱中进行语义关联和信息检索,以获取与用户请求相关的实体、关系和属性和关系;

[0049] 基于所述实体、关系和属性对用户请求进行增强。

[0050] 与现有技术相比,本发明的基于知识增强和大模型的推荐智能体系统及方法具有如下有益技术效果中的一者或多者:

[0051] 1、本发明利用知识增强技术,将领域专家知识、文本类数据等融入到推荐系统中,提升了推荐结果的质量和准确性。

[0052] 2、本发明基于大规模语言模型,对用户画像、项目表示等进行学习和建模,提升了推荐系统的定制化和智能化程度。

[0053] 3、本发明使用元提示学习技术,通过提示模板来分析用户的行为、偏好等信息,提高了用户满意度和使用体验。

[0054] 4、本发明通过实时计算和大规模数据处理技术,实现在用户发起请求时即时生成定制化推荐结果,满足了用户的即时需求。

[0055] 5、本发明通过监控用户反馈和行为,实现推荐智能体的在线学习和优化,不断提升推荐效果和适应用户变化的需求。

附图说明

[0056] 图1是本发明的基于知识增强和大模型的推荐智能体系统的构成示意图。

[0057] 图2是本发明的基于知识增强和大模型的推荐智能体系统的知识图谱构建模块的构成示意图。

[0058] 图3是本发明的一个示例性的元提示模板。

[0059] 图4是本发明的基于知识增强和大模型的推荐智能体系统的微调模块的构成示意图。

[0060] 图5是本发明的基于知识增强和大模型的推荐方法的流程图。

具体实施方式

[0061] 在详细说明本发明的任何实施方式之前,应理解的是,本发明在其应用中并不限于以下描述阐述或以下附图图示的部件的构造和布置细节。本发明能够具有其他实施方式并且能够以各种方式实践或进行。另外,应理解的是,这里使用的措辞和术语出于描述的目的并且不应该被认为是限制性的。本文中使用“包括”或“具有”及其变型意在涵盖下文中陈列的条目及其等同物以及附加条目。除非另有指定或限制,否则术语“安装”、“连接”、“支撑”和“联接”及其变型被广泛地使用并且涵盖直接安装和间接的安装、连接、支撑和联接。此外,“连接”和“联接”不限于物理或机械的连接或联接。

[0062] 并且,第一方面,在本发明的揭露中,术语“纵向”、“横向”、“上”、“下”、“前”、“后”、“左”、“右”、“竖直”、“水平”、“顶”、“底”“内”、“外”等指示的方位或位置关系是基于附图所示的方位或位置关系,其仅是为了便于描述本发明和简化描述,而不是指示或暗示所指的装置或元件必须具有特定的方位、以特定的方位构造和操作,因此上述术语不能理解为对本发明的限制;第二方面,术语“一”应理解为“至少一”或“一个或多个”,即在一个实施例中,一个元件的数量可以为一个,而在另外的实施例中,该元件的数量可以为多个,术语“一”不能理解为对数量的限制。

[0063] 本发明涉及一种基于知识增强和大模型的推荐智能体系统及方法,旨在解决现有推荐系统在处理海量数据、提高推荐准确性和实时性方面的问题,其通过结合知识图谱和大模型技术,具备对用户行为的深度理解和分析能力,能够更准确地捕捉用户的兴趣和偏好,为用户提供个性化的定制推荐服务,帮助用户发现感兴趣的内容

[0064] 图1示出了本发明的基于知识增强和大模型的推荐智能体系统的构成示意图。如图1所示,本发明的基于知识增强和大模型的推荐智能体系统包括:

[0065] 一、知识图谱构建模块。

[0066] 本发明采用的知识增强技术通过所述知识图谱构建模块体现,所述知识图谱构建模块用于收集和整合领域相关知识,以形成结构化的知识图谱。

[0067] 图2示出了本发明的基于知识增强和大模型的推荐智能体系统的知识图谱构建模块的构成示意图。如图2所示,所述知识图谱构建模块包括:

[0068] 1、数据收集子模块。

[0069] 所述数据收集子模块用于从多个开源平台收集领域相关知识,包括公开的知识库(如维基百科、DBpedia等)、专业文献、专家知识、在线论坛等。数据收集的目标是获取尽可能全面和准确的知识信息,为后续的实体识别、关系抽取和属性抽取提供丰富的数据基础。

[0070] 2、实体识别子模块。

[0071] 所述实体识别子模块用于从收集到的领域相关知识中识别出实体,如人名、地名、机构名等。

[0072] 本发明采用命名实体识别(Named Entity Recognition,NER)技术来识别文本数据中特定类型实体并为其分类,其定义如下所示:

[0073] $E=NER(T)$

[0074] 其中,E表示识别出的实体集合,T表示输入的文本数据(也就是,领域相关知识的文本数据)。

[0075] 3、关系抽取子模块。

[0076] 所述关系抽取子模块用于从收集到的领域相关知识中抽取出实体之间的关系,如所属关系、合作关系等。本发明采用Bert模型来识别文本数据中实体之间的特定关系并为其分类,其定义如下所示:

[0077] R=Bert(T,E)

[0078] 其中,R表示识别出的关系集合,T表示输入的文本数据(也就是,领域相关知识的文本数据),E表示已识别出的实体集合。

[0079] 4、属性抽取子模块。

[0080] 所述属性抽取子模块用于从收集到的领域相关知识中抽取出能够代表实体特征的属性。在本发明,所述属性抽取子模块包括用户属性抽取(抽取用户个人信息、偏好、历史行为等,如年龄、性别、购买历史、评分记录等)和项目属性抽取(抽取项目特征,如物品的价格、品牌、类型、规格等)。

[0081] 5、知识融合子模块。

[0082] 所述知识融合子模块用于将识别出的实体、关系和属性进行融合,构建成所述结构化的知识图谱。

[0083] 6、知识更新子模块。

[0084] 所述知识更新子模块用于将新知识添加到所述结构化的知识图谱中、更新所述结构化的知识图谱中的旧知识和修正所述结构化的知识图谱中的错误知识,其被定义如下所示:

[0085] KG_{update}=Update(KG,NewKnowledge)

[0086] 其中,KG_{update}表示更新后的知识图谱,KG表示原有的知识图谱,NewKnowledge表示新增的知识信息。

[0087] 7、知识存储和管理子模块。

[0088] 所述知识存储和管理子模块用于存储和管理所述结构化的知识图谱。知识图谱的存储和管理是确保知识图谱的可用性和可扩展性的关键,本发明根据实际场景,可以灵活采用图数据库(如Neo4j)、关系数据库和NoSQL数据库进行存储;然后知识图谱的管理包括数据的备份、恢复、安全性和访问控制等。

[0089] 二、元提示学习模块。

[0090] 所述元提示学习模块用于基于所述结构化的知识图谱和文本数据,采用元提示学习方法形成一组元提示模板。

[0091] 为了将结构化的知识图谱中的知识和文本数据中的知识同时输入到大模型中,本发明使用元提示学习方法。元提示学习方法可以从用户行为数据中自动学习规则,用于指导大模型进行训练。具体来说,元提示学习采用数据挖掘和机器学习技术,通过分析用户的行为模式、偏好和上下文信息,自动提取出有用的规则;通过聚合的这些规则被称为元提示,它们能够指导推荐智能体系统在处理用户行为数据时做出更准确的定制化推荐服务。

[0092] 元提示是一组与用户行为相关的提示模板,它们能够指导推荐智能体系统在处理用户行为数据时做出更准确和个性化的推荐。例如,一个元提示可以是“如果用户在过去一周内购买了商品A,那么他可能对商品B感兴趣”。元提示学习可以从用户行为数据中自动提取出这些有用的提示规则,捕捉用户行为中的关键信息和模式,为后续的大模型预训练提供准确的输入。图3示出了一个示例性的元提示模板,在该元提示模板中,用户ID、用户属

性、物品ID、物品属性等信息是从所述结构化的知识图谱中获得的,历史行为、物品标签、推荐理由等信息是从文本数据中获得的。

[0093] 三、预训练模块。

[0094] 所述预训练模块用于使用所述元提示模板对大模型进行预训练。有了多个所述元提示模板,即可用所述元提示模板对大模型进行预训练。由于所述元提示模板中集合了结构化的知识图谱中的知识和文本数据中的知识,因此,在本发明中,对大模型进行预训练时同时使用了结构化的知识图谱中的知识和文本数据中的知识,能够学习到丰富的特征表示和用户偏好信息,从而指导推荐智能体做出更准确的定制化推荐服务。

[0095] 本发明可以采用ChatGLM 6B模型作为大模型进行大规模的预训练,它是基于General Language Model(GLM)架构,经过预训练,能够学习到丰富的语言知识和模式,具有强大的语言理解和生成能力。

[0096] 在对大模型进行预训练时,涉及损失函数选择(在本发明中,采用交叉熵损失函数)、优化算法选择和训练过程。具体来说,本发明采用了自监督学习的方式,通过最大化文本序列的概率来调整模型参数,使其能够更好地预测下一个词的出现概率;在训练过程中,通过调整学习率、批量大小等超参数(具体地,首先固定初始学习率和批量大小,然后动态调整学习率和批量大小以帮助大模型在训练过程中更好地适应数据,加快收敛速度,并避免在优化过程中的震荡现象),以及采用特定的优化算法(如Adam优化算法)来优化大模型的训练效果,提高其在文本生成和理解任务上的性能;另外,在训练过程中,对大模型进行定期的验证和评估,以监控其在验证集上的性能表现,并根据评估结果调整训练策略和超参数设置;在大模型预训练完成后,可以对大模型进行压缩和优化(可以采用权重剪枝策略进行模型压缩,即通过删除模型中较小的权重参数,从而减小模型的大小,该方法可以在保持模型性能稳定的同时,显著降低模型的存储和计算需求;另一方面,权重剪枝可能导致模型过拟合,因此需要在压缩过程中加入正则化项来优化模型的泛化性能),以减少其存储和计算资源的消耗,并提高在推理阶段的速度和效率。

[0097] 四、微调模块。

[0098] 所述微调模块用于对预训练后的大模型进行微调。

[0099] 预训练后的大模型在大量数据上进行了训练,能够学习到丰富的特征表示和用户偏好信息。然而,由于预训练后的大模型是在通用任务上训练的,其性能在特定任务上可能不如专门为该任务设计的模型。因此,本发明通过微调预训练后的大模型,可以使其更好地适应推荐智能体的具体任务和需求。

[0100] 本发明在微调时首先需要准备与推荐任务相关的数据,可以采用开源的亚马逊商品数据集(http://jmcauley.ucsd.edu/data/amazon/),包括用户行为数据、项目信息等;然后对数据进行清洗、去噪和格式化,以确保推荐数据的质量和可靠性;接下来定义微调的目标函数,它能够反映推荐服务的具体任务和需求,常用的微调目标函数包括交叉熵损失函数、排序损失函数等。

[0101] 图4示出了本发明的基于知识增强和大模型的推荐智能体系统的微调模块的构成示意图。如图4所示,所述微调模块包括:

[0102] 1、加载子模块。

[0103] 所述记载子模块用于从预训练后的大模型中加载初始参数。

[0104] 2、冻结子模块。

[0105] 所述冻结子模块用于冻结预训练后的大模型的部分层,以保留预训练后的大模型的知识。

[0106] 在本发明中,在微调过程中,可以冻结所述大模型的底层,因为底层通常可以学习到更通用的特征,如边缘和纹理。这有助于大模型在新任务上快速适应,同时保留预训练后的大模型在通用特征学习上的优势。

[0107] 3、添加子模块。

[0108] 如果预训练后的大模型没有嵌入层,或者嵌入层不适合新的推荐任务,可以添加新的嵌入层。所述添加子模块用于在预训练后的大模型的基础上,添加任务特定的层,以适应推荐智能体系统的具体任务和需求。

[0109] 在本发明中,由于是推荐任务,因此,可以为用户和项目添加嵌入层,这些嵌入层将用户和项目的ID映射到低维空间中的向量。

[0110] 4、训练子模块。

[0111] 所述训练子模块用于使用微调目标函数对预训练后的大模型进行训练。在训练过程中,可以调整模型参数,以优化模型的性能。

[0112] 5、评估子模块。

[0113] 所述评估子模块用于定期评估训练后的大模型的性能,并根据评估结果调整训练策略。

[0114] 在本发明中,在评估训练后的大模型的性能时,可以采用RewardModeling模型进行评估。

[0115] 本发明采用的预训练后的大模型微调完成后,得到一个适应推荐智能体系统具体任务和需求的模型,该模型可以用于处理用户行为数据,从而提高定推荐服务的准确性和效率。本发明采用的微调策略可以利用预训练后的大模型的知识,减少训练时间,节省训练资源,提高推荐智能体系统的性能和效率,为用户提供更好的推荐体验。

[0116] 五、定制化推荐智能体构建模块。

[0117] 微调后的大模型虽然是针对推荐任务的大模型,但是,其还是针对通用推荐任务的大模型,无法满足定制化和个性化需求。因此,在本发明中,需要对微调后的大模型进一步进行训练,以形成定制化推荐智能体。所述定制化推荐智能体构建模块用于使用用户信息、项目信息和用户行为信息对微调后的大模型进行训练,以形成定制化推荐智能体。

[0118] 所述定制化推荐智能体也称为推荐引擎,它负责根据用户的需求、行为和上下文信息,从候选项中筛选和排序,生成定制化的推荐结果。推荐引擎的核心任务是对用户和项目进行建模。

[0119] 要构建定制化推荐智能体,所述定制化推荐智能体构建模块首先需要收集和处理各种类型的数据,包括用户信息、项目信息、用户行为等,这些数据可能来自多个来源,如用户浏览历史、购买记录、社交网络等;接着需要对这些数据进行清洗、标准化和特征提取,以便适配定制化推荐智能体的训练和应用。基于收集到的用户数据,对用户进行画像建模,包括用户的兴趣、偏好、行为习惯等方面的分析和建模,更好地理解用户需求和个性化推荐。除了对用户进行画像建模外,还需要对项目进行表示学习,包括对项目的属性、关系等进行建模和学习,以便更好地理解项目之间的相似性和关联性。在数据收集、用户画像建模和项

目表示学习的基础上,定制化推荐智能体构建模块将利用这些数据训练微调后的大模型,以便于形成定制化推荐智能体;定制化推荐智能体将根据用户画像、项目表示和上下文信息,学习用户的偏好和行为模式,从而能够预测用户对项目的喜好程度。当用户发起推荐请求时,定制化推荐智能体会根据用户的特征、上下文信息以及当前环境,生成定制化的推荐结果。此外,定制化推荐智能体还会根据用户的反馈和行为进行实时优化,通过监控用户的点击、购买、评分等行为,可以不断调整和优化,并根据用户的反馈进行在线学习和策略更新。

[0120] 六、实时处理模块。

[0121] 所述实时处理模块用于基于所述预训练后的大模型对用户请求进行实时处理并将处理结果输入所述定制化推荐智能体,由所述定制化推荐智能体生成定制化的推荐结果。

[0122] 所述实时处理模块可以接收用户请求,该用户请求可能是一个查询、一个需求或者一个命令,用于获取推荐结果。用户请求可以包括文字、图像、音频等形式,实时处理模块需要解析和理解这些用户请求以确定用户意图。在接收到用户请求后,实时处理模块会利用所述预训练后的大模型对用户请求进行理解和建模(使用自然语言处理(NLP)、图像识别、声音处理等技术),用于从用户请求中提取关键信息、理解上下文和用户意图。例如,对于文本请求,所述实时处理模块可能会使用所述预训练后的大模型对请求进行语义分析和关键词提取。在理解用户请求的基础上,实时处理模块将从知识图谱中获取相关的知识和信息进行语义关联和信息检索,以获取与用户请求相关的实体、属性和关系。然后,将利用这些信息对用户请求进行知识增强,包括补充缺失的信息、丰富请求的语义内容,或者根据上下文进行信息推断和补全。基于用户请求、上下文理解和知识增强的结果,利用所述定制化推荐智能体进行个性化推荐。最后,将根据定制化推荐智能体的输出生成推荐结果,并将结果反馈给用户;推荐结果可以是文本、图像、视频等形式,展示给用户以满足其需求或兴趣。

[0123] 本发明还提供一种基于知识增强和大模型的推荐方法。如图5所示,所述推荐方法包括以下步骤:

[0124] 1、收集和整合领域相关知识,以形成结构化的知识图谱。

[0125] 2、基于所述结构化的知识图谱和文本数据,采用元提示学习方法形成一组元提示模板。

[0126] 3、使用所述元提示模板对大模型进行预训练。

[0127] 4、对预训练后的大模型进行微调。

[0128] 5、使用用户信息、项目信息和用户行为信息对微调后的大模型进行训练,以形成定制化推荐智能体。

[0129] 其中,对用户请求进行实时处理具体包括:

[0130] 使用预训练后的大模型对用户请求进行语义分析和关键词提取;

[0131] 基于所述语义分析和关键词提取的结果从所述结构化的知识图谱中进行语义关联和信息检索,以获取与用户请求相关的实体、关系和属性和关系;

[0132] 基于所述实体、关系和属性对用户请求进行增强。

[0133] 6、基于所述预训练后的大模型对用户请求进行实时处理并将处理结果输入所述

定制化推荐智能体,由所述定制化推荐智能体生成定制化的推荐结果。

[0134] 最后应当说明的是,以上实施例仅用以说明本发明的技术方案,而非对本发明保护范围的限制。本领域的技术人员,依据本发明的思想,可以对本发明的技术方案进行修改或者等同替换,而不脱离本发明技术方案的实质和范围。

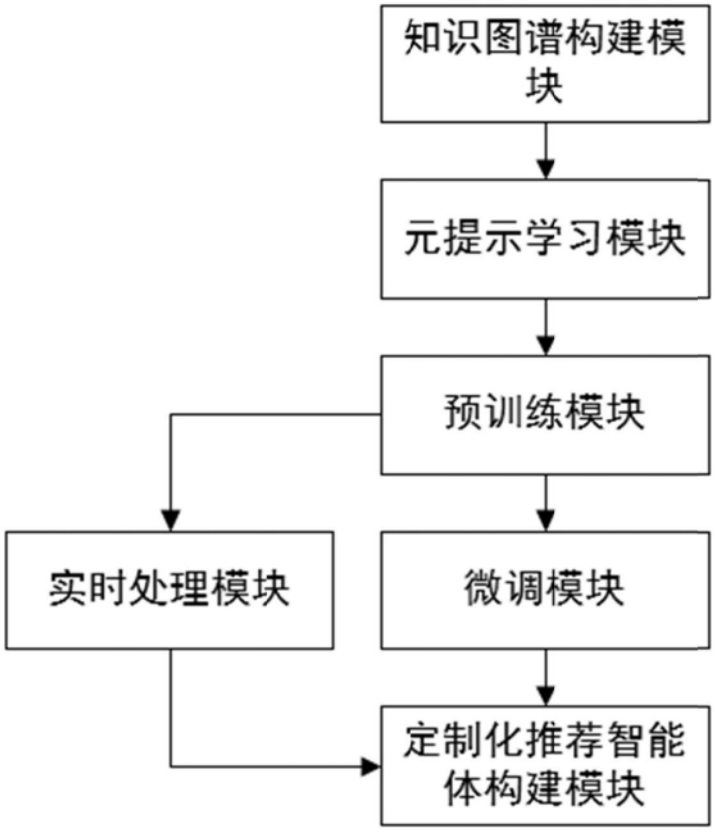

知识图谱构建模块
元提示学习模块
预训练模块
实时处理模块
微调模块
定制化推荐智能体构建模块

图1

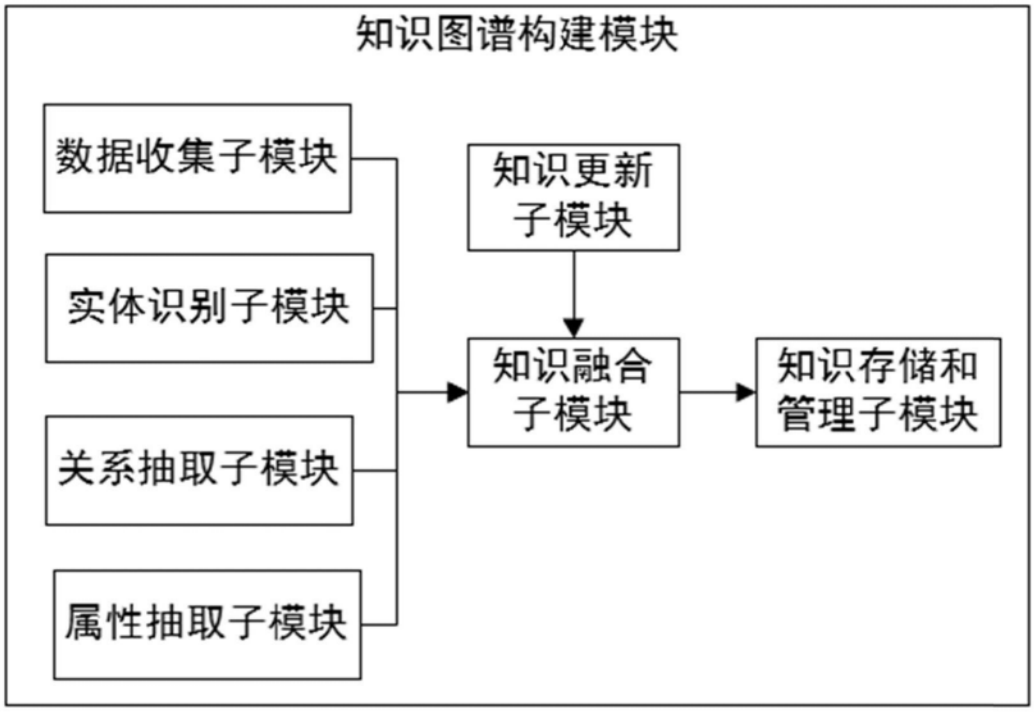

知识图谱构建模块
数据收集子模块
实体识别子模块
关系抽取子模块
属性抽取子模块
知识更新子模块
知识融合子模块
知识存储和管理子模块

图2

问题描述：根据用户信息，生成推荐列表。

输入：

用户信息：

- ✓ 用户ID：user123。
- ✓ 用户属性：年龄25，性别男，地理位置北京，兴趣爱好电影、音乐、旅游。
- ✓ 历史行为：购买记录：电影票、音乐专辑、旅游保险；点击记录：电影预告片、音乐排行榜、旅游攻略；评分记录：电影A评分4星，音乐B评分5星，旅游攻略C评分4.5星。

物品信息：

- ✓ 物品ID：item123。
- ✓ 物品属性：类别书籍，品牌出版社A，价格50元，评价4.5星。
- ✓ 物品标签：关键词旅游、攻略、欧洲。

✓ 输出：

推荐列表：

✓推荐物品ID：item123。

✓推荐物品属性：类别书籍，品牌出版社A，价格50元，评价4.5星。

✓推荐理由：根据用户的历史行为和物品的属性，该物品与用户兴趣高度匹配，评分较高，推荐给用户。

图3

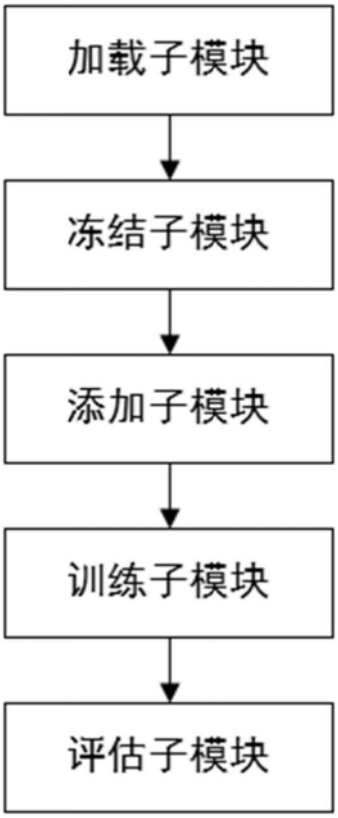


图4


收集和整合领域相关知识，以形成结构化的知识图谱
基于所述结构化的知识图谱和文本数据，采用元提示学习方法形成一组元提示模板
使用所述元提示模板对大模型进行预训练
对预训练后的大模型进行微调
使用用户信息、项目信息和用户行为信息对微调后的大模型进行训练，以形成定制化推荐智能体
基于所述预训练后的大模型对用户请求进行实时处理并将处理结果输入所述定制化推荐智能体，由所述定制化推荐智能体生成定制化的推荐结果

图5