基于对偶四元数的协同知识图谱推荐模型

曹宗胜[1),4)] 许倩倩[2)] 李朝鹏[1),4)] 姜阳邦彦[1),4)] 操晓春[1),4)] 黄庆明[2),3),4),5),6)]

[1)](中国科学院信息工程研究所信息安全国家重点实验室 北京 100093)
[2)](中国科学院计算技术研究所智能信息处理重点实验室 北京 100190)
[3)](中国科学院大学计算机科学与技术学院 北京 101408)
[4)](中国科学院大学网络空间安全学院 北京 100049)
[5)](中国科学院大数据挖掘与知识管理重点实验室 北京 101408)
[6)](鹏城实验室 广东 深圳 518055)

摘 要 为减轻用户-商品交互数据稀疏和冷启动问题对协同过滤算法推荐效果的影响，许多研究者将知识图谱引入推荐系统. 然而既往工作对协同知识图谱中关系的建模较为单一，表达能力较弱，不利于建模实体间的复杂关系以及学习用户和商品的嵌入. 为克服这一问题，本文提出基于对偶四元数的协同知识图谱推荐模型(DQKGR)，利用富有表达力的对偶四元数嵌入表示用户和商品，以有效建模实体和关系之间复杂的潜在依赖. 在商品和用户的嵌入学习部分，本文设计了一个新的知识图谱嵌入模型(DQKGE)，可基于莫比乌斯变换捕捉协同知识图谱中实体间多种复杂关系，并基于此设计偏好传播与聚合方法以利用知识图谱中的结构信息进行推荐. 为验证所提 DQKGR 模型的有效性，在公开的 Last-FM、MovieLens-20M 和 Book-Crossing 数据集上进行大量实验，结果表明所提 DQKGR 模型推荐结果在多个评测指标上优于现有方法，可在平均意义上产生 2.83%以上的性能提升.

关键词 知识图谱；推荐系统；对偶四元数；莫比乌斯变换；偏好传播
中图法分类号 TP391 **DOI 号** 10.11897/SP.J.1016.2022.02221

Dual Quaternion Based Collaborative Knowledge Graph Modeling for Recommendation

CAO Zong-Sheng[1),4)] XU Qian-Qian[2)] LI Zhao-Peng[1),4)] JIANG Yangbangyan[1),4)]
CAO Xiao-Chun[1),4)] HUANG Qing-Ming[2),3),4),5),6)]

[1)](*State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, Beijing* 100093)
[2)](*Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing* 100190)
[3)](*School of Computer Science and Technology, University of Chinese Academy of Sciences, Beijing* 101408)
[4)](*School of Cyber Security, University of Chinese Academy of Sciences, Beijing* 100049)
[5)](*Key Laboratory of Big Data Mining and Knowledge Management (BDKM), Chinese Academy of Sciences, Beijing* 101408)
[6)](*Peng Cheng Laboratory, Shenzhen, Guangdong* 518055)

Abstract To alleviate the issues caused by sparse user-item interaction data and cold-start problems, which limit the recommendation performance of collaborative filtering, researchers integrate the knowledge graph (KG) to recommender systems. However, most existing KG-aware recommendation models use the single transformational method for KG embedding, which may limit the modeling of complicated relations existing in real-world data. Besides, these models often represent users

收稿日期：2021-11-01；在线发布日期：2022-06-15. 本课题得到科技创新 2030-“新一代人工智能”重大项目(2018AAA0102000)、国家自然科学基金项目(U21B2038，61931008，6212200758，61976202)、中央高校基本科研业务费专项资金、中科院青促会会员项目、中国科学院战略性先导科技专项(XDB28000000)资助. **曹宗胜**，博士研究生，主要研究方向为推荐系统与知识图谱嵌入算法. E-mail：caozongsheng@iie.ac.cn. **许倩倩**，博士，副研究员，中国计算机学会(CCF)高级会员，主要研究方向为统计机器学习及其在多媒体和计算机视觉领域的应用. **李朝鹏**，硕士，主要研究方向为推荐系统与计算机视觉. **姜阳邦彦**，博士研究生，中国计算机学会(CCF)学生会员，主要研究方向为机器学习与计算机视觉. **操晓春**，博士，研究员，中国计算机学会(CCF)高级会员，主要研究领域为计算机视觉、多媒体分析. **黄庆明**，博士，讲席教授，中国计算机学会(CCF)会士，主要研究领域为多媒体计算、图像处理、计算机视觉、模式识别.

《计算机学报》

and items using real-valued embeddings, which are of less representation capacity. In this paper, we propose Dual Quaternion-based Collaborative Knowledge Graph Modeling for Recommendation (DQKGR), which represents users and items with dual quaternion embeddings in hypercomplex space, so that the latent inter-dependencies between entities and relations could be captured effectively. In the core of our model, we propose a method called Dual Quaternion-based and Mobius Transformation-based Knowledge Graph Embeddings (DQKGE), which can capture multiple complex relations in collaborative KGs. On top of this, those embeddings are updated by a customized preference propagation and aggregation method with structure information concerned. Finally, we apply the proposed DQKGR to three real-world datasets, including Last-FM, MovieLens-20M, and Book-Crossing. Results show that DQKGR outperforms existing methods on several metrics, and achieves a more than 2.83% performance gain in average.

Keywords knowledge graph; recommender system; dual quaternion; Mobius transformation; preference propagation

1 引 言

随着互联网的快速发展,用户可获得的数据量日益增长.由于信息过载,人们很难在众多选择中挑选出真正感兴趣的内容.为了向用户提供个性化的建议,从而改善用户体验,推荐系统(Recommender System)应运而生,在购物网站[1]、多媒体内容(如视频[2]、音乐[3]、新闻[4]等)发布和传播平台、社交软件[5]等现实场景中得到广泛应用.在众多推荐算法中,基于协同过滤(Collaborative Filtering)[6]的推荐算法主要根据历史用户-商品交互数据对用户/商品相似性进行分析,从而进行推荐.由于其不需要额外的用户或商品特征,并且通常可以有效地捕获用户偏好,因此得到广泛应用.然而,当面对历史用户-商品交互数据稀疏和冷启动问题时,这类算法难以取得令人满意的效果.为此,研究人员提出混合的推荐算法,将用户或者商品的辅助信息与协同过滤算法相结合,作为对稀疏历史用户-商品交互数据的补充,从而提高算法的性能.目前,已有多种不同辅助信息被应用于推荐系统,例如商品的图像[7]、评价文本[8-9]、用户的属性[10]等.与上述孤立的辅助信息不同,知识图谱(Knowledge Graph)作为一种结构化的语义网络,能够连接不同的商品实体[11]并为推荐系统提供大量现实世界的知识,因而近年来受到大量关注.

通常而言,知识图谱是由节点表示实体、由边表示关系的有向异构图.现实世界中丰富的事实可以通过知识图谱中形如(头实体,关系,尾实体)的三元组进行结构化表示.例如,电影知识图谱中的三元组(The Truman Show,film.director,Peter Weir)表示"电影《楚门秀》的导演为彼得·威尔"这一事实(Fact).知识图谱的引入可以为推荐系统提供更多的结构信息,有利于进一步挖掘用户和商品之间的关系信息,从而提高推荐结果的精准程度.与此同时,知识图谱上连接的三元组路径也有利于增强推荐结果的可解释性.在过去的几年中,研究人员对知识图谱在推荐系统中的应用进行了大量的研究[11-15],以期利用知识图谱的结构信息学到有效的用户和商品表示.其中,最经典的是基于嵌入的方法,例如,Zhang 等人[12]提出的 CKE 模型使用 TransR[16]方法从知识图谱中学习商品的表示.这种方法通常基于知识图谱嵌入(Knowledge Graph Embedding)[16-18]算法对商品知识图谱进行表示学习,灵活性较好,但得到的商品表示可能更适合原始图谱上的链路预测任务而非推荐任务.第二类是基于路径的方法,利用知识图谱中的链接模式构建元路径,进而辅助推荐任务.具体而言,Yu 等人[13]将知识图谱视为异质信息网络(Heterogeneous Information Network),提取基于元路径的潜在特征,从而对用户和商品在不同关系路径下所具备的连通性进行建模.这类方法可更自然和直观地利用知识图谱的信息,并且可通过元路径对推荐结果进行解释,但是严重依赖预先设置的元路径,灵活性较差.第三类方法是混合的方法,这类方法将基于嵌入的方法和基于路径的方法相结合,从而更充分地挖掘知识图谱中的有效信息.例如,Wang 等人[11]提出 RippleNet 模型,设计一种与水波相似的偏好传播机

《计算机学报》

制，用户的兴趣被沿着知识图谱中三元组的路径显式传播. 随着图神经网络(Graph Neural Networks, GNN)[19]的快速发展，结合 GNN 的混合算法被提出，代表性地，KGCN 模型[14]基于图卷积神经网络处理知识图谱，从而自动捕捉用户的高阶个性化偏好信息. 此外，KGAT 模型[15]将用户-商品交互信息和知识图谱转化为统一的用户-商品图谱，并结合 TransR[16]和 GNN 的思想，在统一的图谱上显式地对图中的高阶关系进行建模. 实验结果表明，混合的方法通常可以更有效地提高个性化推荐的精准程度.

尽管现有结合知识图谱的推荐模型已取得初步成功，但在从协同知识图谱提取信息的过程中，仍存在如下问题：(1) 目前关于协同知识图谱的嵌入算法通常将关系建模为单一的几何变换，不能同时建模协同知识图谱中实体的多种复杂关系模式[20](如对称/反对称模式、反演模式、组合模式以及多关系模式等). 例如，Huang 等人[21]提出的 KSR 模型、Ye 等人[22]提出的 BEM 模型均使用 TransE[17]算法从知识图谱中学习，不能建模对称模式；Cao 等人[23]提出的 KTUP 框架使用 TransH[18]算法，CKE[12]、AKUPM[24]及 KGAT[15]等模型使用 TransR[16]算法提取辅助信息，这两种知识图谱嵌入算法均无法建模反演模式和组合模式，限制了对知识图谱中复杂关系的充分挖掘；(2) 使用实数向量在欧氏空间建模用户和商品，不利于挖掘协同知识图谱中实体间的潜在交互，使模型表达能力欠佳；随着参数量的增加，模型易趋于过拟合，限制性能提升. 考虑到莫比乌斯变换[25]蕴含丰富的几何变换(平移、旋转、反射和位似等)，因此针对问题(1)，可设计一种基于莫比乌斯变换的嵌入学习算法，以建模知识图谱中实体间的复杂关系. 同时，注意到 Clifford[26]提出的对偶四元数(Dual Quaternion)具有直观的几何和物理解释. 且对偶四元数乘法中蕴含的汉密尔顿乘积(Hamilton product)具有较强的表示能力(已成功应用于语音识别[27]、知识图谱嵌入[28]等领域)，可用于捕捉实体间丰富的交互. 以上分析促使本文尝试基于对偶四元数在超复数空间中建模用户、商品、实体和关系，使问题(2)也可同时得到解决.

综上所述，本文主要贡献如下：

(1) 为解决推荐协同知识图谱中对关系建模单一的问题，提出一种基于对偶四元数和莫比乌斯变换的知识图谱嵌入算法(DQKGE). 引入对偶四元数，可在超复数空间中对所构建的协同知识图谱进行表示学习，对偶四元数乘积蕴含的丰富表示能力有利于捕捉协同知识图谱中的潜在交互关系. 此外，莫比乌斯变换将关系建模为丰富的几何变换，可建模协同知识图谱中存在的多种基本关系模式.

(2) 为实现更为精准和高效的推荐，提出基于对偶四元数的知识图谱推荐模型(DQKGR). 针对协同知识图谱嵌入后得到用户、商品和实体的对偶四元数表示，提出一种基于注意力的偏好传播与聚合机制，从而利用知识图谱的结构信息丰富实体的表示.

(3) 在 Last-FM、MovieLens-20M 和 Book-Crossing 数据集上进行的大量实验结果表明，相比于既往工作，本文所提出的 DQKGR 模型在多个数据集上不同值情况下的 Top-K 推荐结果更为精准，从而验证了所提模型的有效性.

本文第 2 节为相关工作；第 3 节介绍与本文相关的基础知识；第 4 节详细介绍提出的 DQKGR 模型；第 5 节为实验结果及分析；第 6 节对本文进行总结.

2　相关工作

在介绍本文主要工作前，本节先对既往工作及其必要技术细节进行简要回顾，2.1 小节主要回顾四元数以及对偶四元数的应用，2.2 小节主要回顾知识图谱表示学习相关方法，随后 2.3 小节中回顾现有结合知识图谱的推荐模型.

2.1　四元数及对偶四元数

四元数最早由数学家汉密尔顿提出：四元数是复数在超复数空间中的一种扩展，由一个实部和三个虚部组成. 与实数和复数相比，四元数的表示能力相对更强，其可以表示三维空间的旋转. 两个四元数之间的汉密尔顿乘积是不可交换的，并具有内在的反对称性质. 对偶四元数由 Clifford[26]提出，是四元数在超复数空间中的一种扩展，同时，对偶四元数乘法中的汉密尔顿乘积可使两个对偶四元数的所有实部和虚部之间进行充分交互，因此具有很强的表达能力. 鉴于以上优势，四元数以及对偶四元数得到了许多学者的关注，并且被成功应用于多个领域，包括机器人学[29]、图像识别[30]、知识图谱嵌入[28,31]、自动语音识别[27]和通用推荐系统[32]等. 此外，Parcollet 等人[33]提出的四元数多层感知机在语言理解上取得了优于标准的实数感知机的表现；在音素识别任务上，该团队提出的四元数循环神经网络[34]通过更小的计算量取得优于传统循环神经网络的性能. 简

计算机学报

而言之，四元数和对偶四元数更有利于捕捉神经网络编码多维输入特征之间潜在的相互依赖和内部依赖，并显示出良好的模型性能.

值得注意的是，目前尚未有工作将对偶四元数表示引入到结合知识图谱的推荐系统中，尽管已有工作如文献[32]等将四元数表示引入推荐系统，并取得了良好的性能，但该工作仅利用四元数实现矩阵分解技术的扩展，无法利用辅助信息提高推荐的精准程度. 因此，本文进一步探索利用对偶四元数实现结合知识图谱的个性化推荐任务，以期从大规模知识图谱中挖掘有效的辅助信息，从而提升个性化推荐的精准程度.

2.2 知识图谱表示学习

知识图谱表示学习，也称作知识图谱嵌入，其目的是将实体和关系映射到低维向量，同时保留其内在信息. 作为知识图谱补全、三元组分类、实体识别和关系提取等具体知识获取任务的基础，如何进行有效的知识图谱表示学习成为一个热点问题. 根据对三元组中实体和关系建模空间的不同，本文将大多数现有的知识图谱嵌入方法分为三类，即基于欧式空间的嵌入模型、基于复数空间的嵌入模型以及基于非欧空间的嵌入模型.

基于欧式空间的嵌入模型通常将关系视为头尾实体之间的平移操作，其中最具代表性的方法是 TransE[17]. 该方法将实体和关系表示为同一个空间中的向量，并将关系 $\boldsymbol{r}$ 视为头实体 $\boldsymbol{h}$ 到尾实体 $\boldsymbol{t}$ 的平移向量. 因此，对于图谱中真实存在的三元组 $(\boldsymbol{h},\boldsymbol{r},\boldsymbol{t})$，其头实体嵌入 $\boldsymbol{h}$、关系嵌入 $\boldsymbol{r}$、尾实体嵌入 $\boldsymbol{t}$ 满足 $\boldsymbol{h}+\boldsymbol{r}\approx\boldsymbol{t}$. 该方法形式简单，但是仅能有效建模 1 对 1 的关系，不能处理 1 对 N、N 对 N 的关系. 为此，Feng 等人[18]提出 TransH 模型，引入关系特定超平面，将关系解释为其特定超平面上的平移向量. Liu 等人[16]提出 TransR 模型，引入关系特定空间，将关系解释为其特定空间内的平移向量. 其他代表性的相关模型还包括 TransD[35]、TransG[36]、KG2E[37]、TranSparese[38]、TorusE[39]模型等. 综上，基于欧式空间的模型相对简单有效，但无法完全捕获知识图谱中的三种基本关系模式，即对称/反对称模式、反演模式以及组合模式.

基于复数空间的嵌入模型主要在复空间中利用旋转操作建立匹配规则，并进行知识图谱嵌入. 早期代表方法如 Yang 等人[40]提出的 DistMult 模型和 Trouillon 等人[41]提出的 ComplEx 模型，首先将嵌入空间扩展到复数空间，并利用复数运算建立三元组匹配规则. 随后，Sun 等人[20]提出 RotatE 模型，利用复空间内只有一个旋转面的关系运算，将三元组中的关系形式化为从头实体到尾实体的旋转操作，从而首次实现对知识图谱中三种基本关系模式的统一建模. 近期，Zhang 等人[28]提出的 QuatE 模型将复数空间扩展到四元数空间，使用两个旋转面的旋转操作取得进一步提升. Cao 等人[31]进一步将对偶四元数引入到知识图谱中，利用对偶四元数的几何性质，将关系建模为旋转和平移的组合，提升了模型性能. 考虑到在基于几何变换的知识图谱模型中，关系常建模为对实体嵌入向量的各种变换以增强模型的表达能力，而以上方法不能利用反射和位似等几何变换，因此对关系的建模丰富度还存在欠缺.

基于非欧空间的嵌入模型将双曲空间引入知识图谱嵌入. 由于知识图谱的结构通常比较复杂，有的包含大量的层次结构. MuRP 模型[42]最早将知识图谱中的关系建模为双曲空间中的平移，利用双曲空间在刻画层次结构的优势，MuRP 在此类数据集上表现优异. ROTH[43]进一步将关系建模为非欧空间中的旋转，增强了模型对基本关系模式的建模能力. 此外，还有一些其他利用非欧几何进行知识图谱嵌入的方法(如文献[44]等)，其主要将关系建模为非欧空间中的平移或者旋转等变换，这就使得模型在多种关系模型的刻画方面存在欠缺.

综上所述，既往嵌入方法的成功表明了无论是欧式空间、复数空间还是非欧空间，丰富的几何操作更有利于实现对知识图谱中多种关系模式的建模，但同时也可以看到既往方法对关系的建模模式还较为单一(主要考虑旋转和平移). 这启发本文探寻包含丰富几何变换的嵌入方法，以期更有效地从现实世界中复杂的商品知识图谱中提取信息，辅助个性化推荐任务.

2.3 结合知识图谱的推荐模型

推荐系统的目的是帮助用户从海量的商品(如图书、电影等)找到其可能感兴趣的部分，从而提升用户体验. 常用的推荐方法通常包括如下两个步骤：(1)利用已有数据学习给定用户和商品的表示；(2)学习一个评分函数，预测用户对商品的喜好程度. 传统的协同过滤推荐算法仅利用历史的用户-商品交互数据，面对数据稀疏和冷启动问题时效果不佳. 为了缓解这一问题，研究人员提出混合的推荐算法，将用户或者商品的多种辅助信息与协同过滤算法相结合，作为对稀疏历史用户-商品交互数据的补

计算机学报

充，从而提高算法的性能. 本文主要针对结合知识图谱的推荐模型展开相关研究.

根据使用知识图谱的不同方式，结合知识图谱的推荐模型大致可分为基于嵌入的方法[12,23,45-47]、基于路径的方法[9,48-51]和混合的方法[11,14-15,24,52]. 基于嵌入的方法通常利用知识图谱嵌入算法对商品知识图谱进行表示学习，从中提取信息进而辅助推荐任务. 例如，Zhang 等人[12]提出的 CKE 模型使用 TransR[16]方法从知识图谱中学习商品的表示；Cao 等人[23]提出的 KTUP 框架使用 TransH[18]方法进行知识图谱嵌入. 这类方法在使用知识图谱辅助推荐任务中展现出较高的灵活性，但是使用的知识图谱嵌入算法本身通常更适合链路预测等任务而非推荐任务；此外，这种方法产生的推荐结果可解释性较差. 基于路径的方法通常对知识图谱中商品之间的各种连接模式进行探索，为推荐任务提供额外的指导信息. 例如，Yu 等人[13]从知识图谱提取基于元路径的潜在特征，从而对用户和商品在不同的关系路径下所具备的连通性进行建模. Hu 等人提出的[9]学习元路径的表示，从而建模用户-商品交互. 这类方法利用知识图谱的方式更为自然和直观，可通过元路径对推荐结果进行解释，但是严重依赖手动设置的元路径，难以在实际应用中进行优化. 为了兼具上述两种方法的优势，近期越来越多的研究人员提出混合的方法，将基于嵌入的方法和基于路径的方法相结合，从而更充分地挖掘知识图谱中的有效信息. 例如，Wang 等人[11]提出 RippleNet 模型，设计一种与水波相似的偏好传播机制，用户的兴趣被沿着知识图谱中三元组的路径显式地传播. 此外，随着图神经网络的迅速发展，图卷积神经网络、图注意力神经网络的思想也在混合方法中得到应用，以期借助知识图谱中的高阶信息辅助偏好建模. 例如，KGCN 模型[14]使用图卷积神经网络处理知识图谱，从而自动捕捉用户的高阶个性化偏好信息. 此外，KGAT 模型[15]则将用户-商品交互信息和知识图谱转化成统一的用户-商品图谱，利用基于平移的 TransR[16]和图注意力神经网络的思想，在统一的图谱上显式地对图中的高阶关系进行建模. QKGN[53]则进一步将 QuatE[28]引入协同知识图谱，在学习实体的嵌入时将关系建模为空间的旋转，提升了模型的表达力. 混合的方法通常可以更有效地提高个性化推荐的精准程度，并在一定程度上保留推荐结果的可解释性.

综上可知，现有结合知识图谱的推荐模型大多在协同知识图谱中将关系建模为单一的几何变换（如平移或旋转）以及使用实数向量度量用户和商品之间的关系，这使得模型表达能力相对较弱，因而在建模协同知识图谱中复杂关系的方面存在欠缺. 基于此，本文引入莫比乌斯变换以建模协同知识图谱中的复杂关系，并利用对偶四元数在超复数空间学习实体和关系的嵌入，以实现更精准的个性化推荐.

3　相关定义

本节主要介绍与知识图谱和推荐系统相关的问题定义，以及所使用的对偶四元数和莫比乌斯变换相关定义.

3.1　问题定义

给定用户集合 $U=\{u_1,u_2,\cdots,u_{|U|}\}$ 和商品（项目）集合 $V=\{v_1,v_2,\cdots,v_{|V|}\}$，根据用户的隐式反馈定义构建用户-商品交互矩阵 $\boldsymbol{Y}\in\mathbb{R}^{|U|\times|V|}$，其中每个元素 y_{uv} 定义如下：

$$y_{uv}=\begin{cases}1, & \text{若已观测到交互}(u,v)\\ 0, & \text{若未观测到交互}(u,v)\end{cases}\tag{1}$$

根据式(1)，当且仅当用户 u 被观测到与商品 v 产生交互行为（如用户观看电影、购买商品、听音乐、点击广告等）时，将 y_{uv} 记为 1，其余情况下 y_{uv} 则均记为 0.

此外，本文采用商品的知识图谱 $\mathcal{G}$ 作为辅助信息，从而辅助个性化推荐任务. 知识图谱 $\mathcal{G}=(\mathcal{E},\mathcal{R})$ 使用节点表示实体、使用边表示关系的有向异构图，其中 $\mathcal{E}$ 表示节点集合，$\mathcal{R}$ 表示边的集合. 知识图谱通过两个相连的节点及其连边构成形如（头实体，关系，尾实体）的三元组，记录现实世界中按照“主语-谓语-宾语”逻辑存储的事实.

如图 1，给定用户-商品交互矩阵 $\boldsymbol{Y}$ 和商品的知识图谱 $\mathcal{G}=(\mathcal{E},\mathcal{R})$，将用户-商品交互矩阵 $\boldsymbol{Y}$ 转化为用户-商品二部图. 此处实际上是将用户-商品交互转化为三元组，然后依靠商品与商品知识图谱内的实体的对齐关系将二者进行合并为统一的有向图 $\mathcal{G}^c$，称之为协同知识图谱（更多细节可参考附录 A.1）.

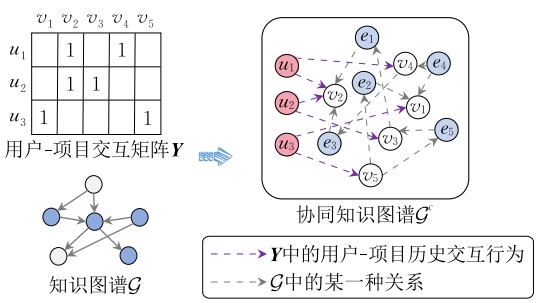


图 1　协同知识图谱构建示意图

基于协同知识图谱$\mathcal{G}^c$，本文的任务是预测每个用户将来可能感兴趣的前 K 个商品. 具体而言，本文旨在学习一个预测函数，对于给定的用户 u 和商品 v 可预测偏好分数 $s(u,v)$，分数越高则表示用户对该商品的潜在感兴趣程度越高.

3.2 符号定义

以上介绍了推荐系统中的问题设置，在上述问题定义的基础上，后续本文还将介绍一些关于知识图谱的模型定义以及符号设置. 这里将本文经常出现的符号总结在表 1 中.

表 1 主要符号定义

符号	含义
$\mathcal{U}$	用户集合
$\mathcal{V}$	商品(项目)集合
u_i	第 i 个用户
v_j	第 j 个商品(项目)
y_{uv}	用户 u 被观测到与商品 v 产生交互
$\mathcal{G}$	知识图谱
$\mathcal{G}^c$	协同知识图谱
$\mathcal{E}$	(协同)知识图谱实体集合
$\mathcal{R}$	(协同)知识图谱关系集合
w_h	头实体的对偶四元数嵌入
w_r	关系的对偶四元数嵌入
w_t	尾实体的对偶四元数嵌入
$\mathcal{N}_h$	头实体的单跳邻居集合
$\odot$	哈达马乘积
$*$	对偶四元数乘积

3.3 协同知识图谱中多种关系模型定义

协同知识图谱(假设其包括实体 $x,y,z\in\mathcal{E}$，关系 $r,r_1,r_2\in\mathcal{R}$)实体间存在丰富的交互关系. 这些关系隶属于不同的关系模式，因而建模这些关系模式对于学习实体的嵌入和挖掘潜在关系至关重要. 协同知识图谱中的多种关系模式的定义如下：

(1) 关系 r 被称为对称或反对称关系，如果：

$$\forall x,y\in\mathcal{E},\ (x,r,y)\Rightarrow(y,r,x)\ \text{or}\ (x,r,y)\Rightarrow\neg(y,r,x) \tag{2}$$

(2) 关系 r_2 是 r_1 的反演关系，如果有：

$$\forall x,y\in\mathcal{E},\ (x,r_1,y)\Rightarrow(y,r_2,x) \tag{3}$$

(3) 关系 r_3 被称为关系 r_1 和 r_2 的组合关系，如果满足：

$$\forall x,y,z\in\mathcal{E},\ r_1(x,y)\wedge r_2(y,z)\Rightarrow r_3(x,z) \tag{4}$$

3.4 对偶四元数相关定义

四元数是超复数系统的一种代表，已被成功应用于多个领域，首先介绍其基本定义.

定义 1. 四元数. 四元数是一种超复数系统，一个四元数 $q\in\mathbb{H}$ 由一个实部和三个虚部组成，定义为

$$q=a+b\boldsymbol{i}+c\boldsymbol{j}+d\boldsymbol{k} \tag{5}$$

其中 a,b,c,d 是实数，$\boldsymbol{i},\boldsymbol{j},\boldsymbol{k}$ 是满足如下汉密尔顿规则的虚单位数：

$$\boldsymbol{i}^2=\boldsymbol{j}^2=\boldsymbol{k}^2=\boldsymbol{ijk}=-1 \tag{6}$$

根据定义 1 所示汉密尔顿规则，可推导得到 $\boldsymbol{ij}=\boldsymbol{k},\boldsymbol{ji}=-\boldsymbol{k},\boldsymbol{jk}=\boldsymbol{i},\boldsymbol{kj}=-\boldsymbol{i},\boldsymbol{ki}=\boldsymbol{j}$ 及 $\boldsymbol{ik}=-\boldsymbol{j}$. 根据以上规则，虚数单元之间的乘积实现不同虚轴之间的转换. 基于此，定义 2 所示四元数的汉密尔顿乘积可建模三维空间中的旋转操作.

定义 2. 四元数汉密尔顿乘积. 给定四元数 $q_1=a_1+b_1\boldsymbol{i}+c_1\boldsymbol{j}+d_1\boldsymbol{k}$ 与 $q_2=a_2+b_2\boldsymbol{i}+c_2\boldsymbol{j}+d_2\boldsymbol{k}$，其汉密尔顿乘积由两个四元数中所有元素的标准乘法组成：

$$\begin{aligned}q_1*q_2=&(a_1a_2-b_1b_2-c_1c_2-d_1d_2)+\\&(a_1b_2+b_1a_2+c_1d_2-d_1c_2)\boldsymbol{i}+\\&(a_1c_2-b_1d_2+c_1a_2+d_1b_2)\boldsymbol{j}+\\&(a_1d_2+b_1c_2-c_1b_2+d_1a_2)\boldsymbol{k}\end{aligned} \tag{7}$$

此外，在四元数的基础上，本文引入对偶数(定义 3)及结合二者实现的对偶四元数(定义 4)，从而同时建模平移和旋转操作.

定义 3. 对偶数. 对偶数 z_d 可定义为

$$z_d=a+\epsilon a_\epsilon \tag{8}$$

其中 a 和 a_ϵ 是实数(或其他代数域的元素)，ϵ 是满足 $\epsilon^2=0$ 的对偶单位. 在上式中，a 和 a_ϵ 分别代表对偶数 z_d 的实部和对偶部.

定义 4. 对偶四元数. 给定四元数 $q_1=a_1+b_1\boldsymbol{i}+c_1\boldsymbol{j}+d_1\boldsymbol{k}$ 与 $q_2=a_2+b_2\boldsymbol{i}+c_2\boldsymbol{j}+d_2\boldsymbol{k}$，对偶四元数 Q 定义为

$$Q=q_1+\epsilon q_2 \tag{9}$$

其中四元数 q_1 和 q_2 分别为对偶四元数 Q 的实部和对偶部分.

3.5 莫比乌斯变换

莫比乌斯变换是定义在扩充复平面 $\hat{\mathbb{C}}=\mathbb{C}\cup\{\infty\}$ 上的几何变换，其中 $\mathbb{C}$ 为复平面，∞ 为无穷远点，此类扩充复平面也被称为黎曼球面. 每个莫比乌斯变换都是黎曼球面到自身的映射. 假设 $[x,y]$ 为 $\hat{\mathbb{C}}$ 中点的齐次坐标，则 $\hat{\mathbb{C}}$ 中的莫比乌斯变换 $\tau:\hat{\mathbb{C}}\rightarrow\hat{\mathbb{C}}$ 为

$$\tau([x,y])=\boldsymbol{S}\begin{bmatrix}x\\y\end{bmatrix},\ \boldsymbol{S}=\begin{bmatrix}a&b\\c&d\end{bmatrix} \tag{10}$$

其中矩阵 $\boldsymbol{S}$ 是可逆的($\det(\boldsymbol{S})\neq0$). 对于空间中的点 x，可以扩充其齐次坐标为 $[x,1]$，则其莫比乌斯变换可以表示为

$$x\rightarrow\begin{bmatrix}x\\1\end{bmatrix}\rightarrow\begin{bmatrix}a&b\\c&d\end{bmatrix}\begin{bmatrix}x\\1\end{bmatrix}\rightarrow\begin{bmatrix}ax+b\\cx+d\end{bmatrix}\rightarrow\frac{ax+b}{cx+d} \tag{11}$$

《计算机学报》

其中，可定义映射 $f:\hat{\mathbb{C}}\rightarrow\hat{\mathbb{C}}$ 如下：

$$f(x)=\frac{ax+b}{cx+d},\ ad-bc\neq 0 \tag{12}$$

式(12)可以表示所有的莫比乌斯变换.

所有莫比乌斯变换的集合在复数域下构成一个群，称为“莫比乌斯群”，记为 $\mathcal{M}(\hat{\mathbb{C}})$. 该群是黎曼球面 $\hat{\mathbb{C}}$ 上的自同构群 $Aut(\hat{\mathbb{C}})$，具体而言，它可以表示为一个在射影空间中由所有配备矩阵乘法的 2×2 矩阵构成的群，记为 $PGL(2,\mathbb{C})$. 鉴于莫比乌斯变换可同构于三维双曲空间中的保向等距同构群，因此本文引入莫比乌斯变换以期更好地建模协同知识图谱中的复杂关系以及学习实体的嵌入.

4　DQKGR 模型

本节主要介绍所提出的方法，包括模型总览和各部分的具体设计，以及模型的学习方法.

4.1　总　览

注意到对偶四元数在图像分类、语音识别等领域的成功应用，本文提出了基于对偶四元数的协同知识图谱推荐模型，其基本框架如图 2 所示. 首先使用用户-商品交互数据和商品知识图谱构建用户-商品协同知识图谱，从而在统一的图谱上实现推荐算法. 为了进行协同知识图谱表示学习，利用对偶四元数乘积可捕捉实体间丰富的潜在交互，增强模型的表达能力；引入莫比乌斯变换，从而可有效地建模知识图谱中的可能存在的多种基本关系模式；为利用知识图谱路径上的邻居信息，基于注意力机制设计了适用于对偶四元数向量的偏好传播与聚合算法；此外，针对已观测到的用户-商品交互数据设计成对比较的损失函数，从而对用户-商品偏好预测进行优化.

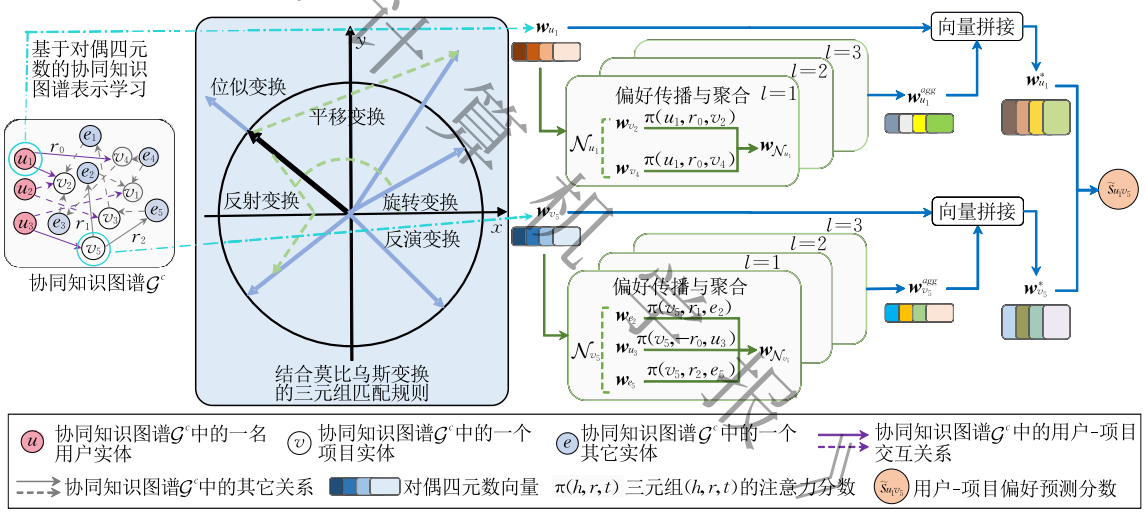


图 2　基于对偶四元数的协同知识图谱推荐模型框架图

4.2　基于对偶四元数和莫比乌斯变换的协同知识图谱嵌入模型(DQKGE)

4.2.1　对偶四元数嵌入

首先，利用对偶四元数对协同知识图谱中的用户、商品、实体和关系以向量形式进行表示. 根据定义 4，可将对偶四元数的向量形式表示为

$$\boldsymbol{w}=\boldsymbol{q}_1+\epsilon\boldsymbol{q}_2 \tag{13}$$

其中 $\boldsymbol{q}_1,\boldsymbol{q}_2$ 为满足 $\boldsymbol{q}_1,\boldsymbol{q}_2\in\mathbb{H}^k$ 的四元数向量：

$$\begin{aligned}\boldsymbol{q}_1&=\boldsymbol{a}_1+\boldsymbol{b}_1\boldsymbol{i}+\boldsymbol{c}_1\boldsymbol{j}+\boldsymbol{d}_1\boldsymbol{k},\\ \boldsymbol{q}_2&=\boldsymbol{a}_2+\boldsymbol{b}_2\boldsymbol{i}+\boldsymbol{c}_2\boldsymbol{j}+\boldsymbol{d}_2\boldsymbol{k}\end{aligned} \tag{14}$$

$\boldsymbol{a}_1,\boldsymbol{b}_1,\boldsymbol{c}_1,\boldsymbol{d}_1,\boldsymbol{a}_2,\boldsymbol{b}_2,\boldsymbol{c}_2,\boldsymbol{d}_2\in\mathbb{R}^k$ 均为维度为 k 的实数向量. 同时，可得 $\boldsymbol{w}$ 的四元数形式 $\boldsymbol{w}(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c},\boldsymbol{d})=\boldsymbol{a}+\boldsymbol{b}\boldsymbol{i}+\boldsymbol{c}\boldsymbol{j}+\boldsymbol{d}\boldsymbol{k}$，其中 $\boldsymbol{a}=\boldsymbol{a}_1+\epsilon\boldsymbol{a}_2$，$\boldsymbol{b}=\boldsymbol{b}_1+\epsilon\boldsymbol{b}_2$，$\boldsymbol{c}=\boldsymbol{c}_1+\epsilon\boldsymbol{c}_2$ 且 $\boldsymbol{d}=\boldsymbol{d}_1+\epsilon\boldsymbol{d}_2$.

因此，对于协同知识图谱 $\mathcal{G}^c$ 中的三元组 $(\boldsymbol{h},\boldsymbol{r},\boldsymbol{t})$，头实体 $\boldsymbol{h}$、关系 $\boldsymbol{r}$、尾实体 $\boldsymbol{t}$ 可分别使用对偶四元数向量表示为 $\boldsymbol{w}_h$，$\boldsymbol{w}_r$ 及 $\boldsymbol{w}_t$，其四元数形式分别记作 $\boldsymbol{w}_h(\boldsymbol{a}_h,\boldsymbol{b}_h,\boldsymbol{c}_h,\boldsymbol{d}_h)$、$\boldsymbol{w}_r(\boldsymbol{a}_r,\boldsymbol{b}_r,\boldsymbol{c}_r,\boldsymbol{d}_r)$ 及 $\boldsymbol{w}_t(\boldsymbol{a}_t,\boldsymbol{b}_t,\boldsymbol{c}_t,\boldsymbol{d}_t)$. 后文将利用此对偶四元数向量对用户、商品、实体和关系进行建模，并实现个性化推荐任务.

4.2.2　三元组可信度评分

为获得任一三元组 $(\boldsymbol{h},\boldsymbol{r},\boldsymbol{t})$ 存在于协同知识图谱中的可信度分数，假定协同知识图谱中的三元组满足如下的语义匹配原则：头实体 $\boldsymbol{h}$ 和尾实体 $\boldsymbol{t}$ 可以通过基于关系 $\boldsymbol{r}$ 的莫比乌斯变换进行匹配. 具体而言，如果该三元组存在于协同知识图谱中，即该三元组的可信度分数高，则期望变换后的头实体与尾实体嵌入之间的内积会尽可能大；反之，若图谱中不

存在该三元组，即该三元组的可信度分数低，则期望其得分函数会尽可能小. 需要注意的是，这里并非对于正负样本存在不同的变换标准，这实际是由后续的损失函数决定. 即损失函数希望增大正样本的可信度分数，同时减少负样本的分数. 注意到关系对偶四元数矩阵的表示为

$$\boldsymbol{w}_r=\begin{bmatrix}\boldsymbol{w}_{r_a} & \boldsymbol{w}_{r_b}\\ \boldsymbol{w}_{r_c} & \boldsymbol{w}_{r_d}\end{bmatrix} \tag{15}$$

其中，$\boldsymbol{w}_{r_a}=\boldsymbol{w}_{r_a}(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c},\boldsymbol{d})=\boldsymbol{a}_{r_a}+\boldsymbol{b}_{r_a}\boldsymbol{i}+\boldsymbol{c}_{r_a}\boldsymbol{j}+\boldsymbol{d}_{r_a}\boldsymbol{k}$，$\boldsymbol{w}_{r_b}$、$\boldsymbol{w}_{r_c}$ 和 $\boldsymbol{w}_{r_d}$ 与 $\boldsymbol{w}_{r_a}$ 形式类似. 接下来，根据以上对偶四元数的运算，与关系交互后的头实体 $\boldsymbol{w}_{h'}$ 可表示为

$$\begin{aligned}&\boldsymbol{w}_h'=f(\boldsymbol{w}_h,\boldsymbol{w}_r)=\frac{\boldsymbol{w}_{r_a}*\boldsymbol{w}_h+\boldsymbol{w}_{r_b}}{\boldsymbol{w}_{r_c}*\boldsymbol{w}_h+\boldsymbol{w}_{r_d}},\\ &\boldsymbol{w}_{r_a}\boldsymbol{w}_{r_d}-\boldsymbol{w}_{r_b}\boldsymbol{w}_{r_c}\neq 0,\ i=1,\cdots,d\end{aligned} \tag{16}$$

其中 f 为莫比乌斯变换，$\frac{x}{y}$ 为变量 x 和 y 之间的对偶四元数除法[26]. 首先计算头实体的表示 $\boldsymbol{w}_h$ 和 $\boldsymbol{w}_{r_a}$ 之间的对偶四元数乘积，用以实现实体和关系的语义匹配：

$$\begin{aligned}\boldsymbol{w}_{r_a}*\boldsymbol{w}_h=&(\boldsymbol{a}_{r_a}\odot\boldsymbol{a}_h-\boldsymbol{b}_{r_a}\odot\boldsymbol{b}_h-\boldsymbol{c}_{r_d}\odot\boldsymbol{c}_h-\boldsymbol{d}_{r_a}\odot\boldsymbol{d}_h)+\\ &(\boldsymbol{a}_{r_a}\odot\boldsymbol{b}_h+\boldsymbol{b}_{r_a}\odot\boldsymbol{a}_h+\boldsymbol{c}_{r_a}\odot\boldsymbol{d}_h-\boldsymbol{d}_{r_a}\odot\boldsymbol{c}_h)\boldsymbol{i}+\\ &(\boldsymbol{a}_{r_a}\odot\boldsymbol{c}_h-\boldsymbol{b}_{r_a}\odot\boldsymbol{d}_h+\boldsymbol{c}_{r_a}\odot\boldsymbol{a}_h+\boldsymbol{d}_{r_a}\odot\boldsymbol{b}_h)\boldsymbol{j}+\\ &(\boldsymbol{a}_{r_d}\odot\boldsymbol{d}_h+\boldsymbol{b}_{r_a}\odot\boldsymbol{c}_h-\boldsymbol{c}_{r_a}\odot\boldsymbol{b}_h+\boldsymbol{d}_{r_a}\odot\boldsymbol{a}_h)\boldsymbol{k}\end{aligned} \tag{17}$$

其中，“$\odot$”表示向量间的哈达马乘积（Hadamard Product）. 根据定义 1 及定义 4 中相应的运算规则将式(17)逐项展开并整理，对偶四元数之间的乘积可表示为 $\boldsymbol{w}_{h,r_a}$：

$$\begin{aligned}\boldsymbol{w}_{h,r_a}=&(\boldsymbol{a}_{h_1}'+\epsilon\boldsymbol{a}_{h_2}')+(\boldsymbol{b}_{h_1}'+\epsilon\boldsymbol{b}_{h_2}')\boldsymbol{i}+\\ &(\boldsymbol{c}_{h_1}'+\epsilon\boldsymbol{c}_{h_2}')\boldsymbol{j}+(\boldsymbol{d}_{h_1}'+\epsilon\boldsymbol{d}_{h_2}')\boldsymbol{k}\\ =&\boldsymbol{a}_h'+\boldsymbol{b}_h'\boldsymbol{i}+\boldsymbol{c}_h'\boldsymbol{j}+\boldsymbol{d}_h'\boldsymbol{k}\end{aligned} \tag{18}$$

完成头实体 $\boldsymbol{w}_h$ 与关系中 $\boldsymbol{w}_{r_a}$ 部分的交互，然后将其与关系的 $\boldsymbol{w}_{r_b}$ 部分做对偶四元数加法，以上运算可使得协同知识图谱中实体和关系的对偶四元数表示得到更充分的交互. 式(16)中的 $\boldsymbol{w}_{r_c}*\boldsymbol{w}_h+\boldsymbol{w}_{r_d}$ 可同理计算. 通过将关系建模为莫比乌斯变换，可以对实体进行平移、旋转、反射、位似等多种几何变换，从而增强了模型的表达能力，以便更好地学习实体的嵌入.

将经过莫比乌斯变换后的头实体记为 $\boldsymbol{w}_h'$，以 $\boldsymbol{w}_h'$ 与尾实体 $\boldsymbol{w}_t$ 之间的扩展内积作为三元组评分函数 $g(\cdot)$，用以度量三元组在该协同知识图谱中是否真实存在的可信度分数. 变换后的头实体和尾实体越接近，则二者的可信度分数越高. 具体地，使用如下方式计算对偶四元数向量之间的扩展内积：

$$\begin{aligned}g(\boldsymbol{h},\boldsymbol{r},\boldsymbol{t})=&\langle\boldsymbol{w}_h',\boldsymbol{w}_t\rangle\\ =&\langle\boldsymbol{a}_h',\boldsymbol{a}_t\rangle+\langle\boldsymbol{b}_h',\boldsymbol{b}_t\rangle+\langle\boldsymbol{c}_h',\boldsymbol{c}_t\rangle+\langle\boldsymbol{d}_h',\boldsymbol{d}_t\rangle\end{aligned} \tag{19}$$

相应地，$g(\boldsymbol{h},\boldsymbol{r},\boldsymbol{t})$ 即为三元组 $(\boldsymbol{h},\boldsymbol{r},\boldsymbol{t})$ 的可信度分数. 理想情况下，$g(\boldsymbol{h},\boldsymbol{r},\boldsymbol{t})$ 越高表明该三元组越有可能真实存在于协同知识图谱中；反之，则表明三元组存在于协同知识图谱中的可信度较低. 考虑到存在于协同知识图谱中的三元组应该比不存在的三元组有更高的可信度分数，采用如下交叉熵损失函数：

$$\mathcal{L}_{\mathrm{KG}}=\sum_{(\boldsymbol{h},\boldsymbol{r},\boldsymbol{t},\boldsymbol{t}')\in\mathcal{T}_{G^c}}-\ln\sigma(g(\boldsymbol{h},\boldsymbol{r},\boldsymbol{t})-g(\boldsymbol{h},\boldsymbol{r},\boldsymbol{t}')) \tag{20}$$

其中 $\mathcal{T}_{G^c}=\{(\boldsymbol{h},\boldsymbol{r},\boldsymbol{t},\boldsymbol{t}')\mid(\boldsymbol{h},\boldsymbol{r},\boldsymbol{t})\in\mathcal{G}^c,(\boldsymbol{h},\boldsymbol{r},\boldsymbol{t}')\notin\mathcal{C}^c\}$ 为协同知识图谱学习对偶四元数嵌入的训练集，$\sigma(\cdot)$ 表示 Sigmoid 激活函数. 上述基于对偶四元数和莫比乌斯变换的知识图谱嵌入（DQKGE）可作为一个新的知识图谱嵌入方法使用. 在本文的推荐中，该方法可引导每个用户和商品利用知识图谱中的信息丰富自身嵌入，进而提升推荐质量.

4.3 基于注意力机制的偏好传播与聚合

经过基于对偶四元数的协同知识图谱表示学习，利用协同知识图谱中三元组的信息对实体和关系的表示进行优化，得到用户和商品的对偶四元数嵌入，从而用于实现个性化推荐. 考虑到仅使用协同知识图谱表示学习得到的嵌入缺乏合理的偏好传播机制，更适合图上具体的知识获取任务，本文进一步基于注意力机制设计偏好传播与聚合算法，沿知识图谱路径利用每个实体的邻居信息辅助用户偏好的学习.

4.3.1 偏好传播

注意到在协同知识图谱中，用户偏好可通过以 $\boldsymbol{h}$ 作为头实体的三元组传播. 因此，基于协同知识图谱中三元组构成的传播路径，期望模型能够利用头实体 $\boldsymbol{h}$ 的邻居信息以加强对每个实体偏好信息的建模. 具体地，将头节点 $\boldsymbol{h}$ 的单跳邻居集合表示为 $\mathcal{N}_h=\{(\boldsymbol{h},\boldsymbol{r},\boldsymbol{t})\mid(\boldsymbol{h},\boldsymbol{r},\boldsymbol{t})\in\mathcal{G}^c\}$.

考虑到不同的尾实体 $\boldsymbol{t}$ 对头实体 $\boldsymbol{h}$ 的重要程度不尽相同，在传播过程中使用基于相似性的注意力机制. 首先使用关系 $\boldsymbol{r}$ 的嵌入对头实体 $\boldsymbol{h}$ 的嵌入进行莫比乌斯变换，并使用 $\tanh(\cdot)$ 作为非线性激活函数，再计算变换后头实体嵌入与尾实体嵌入的相似度作为注意力分数 $\pi(\boldsymbol{h},\boldsymbol{r},\boldsymbol{t})$，用以衡量通过关系 $\boldsymbol{r}$ 由尾实体 $\boldsymbol{t}$ 传播至头实体 $\boldsymbol{h}$ 的偏好信息权重：

$$\pi(\boldsymbol{h},\boldsymbol{r},\boldsymbol{t})=\langle\tanh(\boldsymbol{w}_t),\tanh(f(\boldsymbol{w}_h,\boldsymbol{w}_r))\rangle \tag{21}$$

其中 f 为莫比乌斯变换. 上述注意力机制旨在引导模型为超复数空间中得分更为接近的三元组传播更多

《计算机学报》

的偏好信息. 此后，使用归一化指数函数(Softmax)对以上注意力分数进行标准化：

$$\tilde{\pi}(\boldsymbol{h},\boldsymbol{r},\boldsymbol{t})=\frac{\exp(\pi(\boldsymbol{h},\boldsymbol{r},\boldsymbol{t}))}{\sum(\boldsymbol{h},\boldsymbol{r}',\boldsymbol{t}')\in\mathcal{N}_h\exp(\pi(\boldsymbol{h},\boldsymbol{r}',\boldsymbol{t}'))}\tag{22}$$

使用注意力分数计算关于头实体 $\boldsymbol{h}$ 邻居实体表示的线性组合，用以捕获其一阶结构信息，记作 $\boldsymbol{w}_{\mathcal{N}_h}$：

$$\boldsymbol{w}_{\mathcal{N}_h}=\sum_{(\boldsymbol{h},\boldsymbol{r},\boldsymbol{t})\in\mathcal{N}_h}\tilde{\pi}(\boldsymbol{h},\boldsymbol{r},\boldsymbol{t})\boldsymbol{w}_t\tag{23}$$

通过此方式为协同知识图谱中所有实体(即所有用户、商品和其他实体)，计算其相应的邻居传播信息 $\boldsymbol{w}_{\mathcal{N}_h}$，从而实现后续的偏好聚合过程.

4.3.2　偏好聚合

对于实体 $\boldsymbol{h}$，将其表示 $\boldsymbol{w}_h$ 与其邻居信息 $\boldsymbol{w}_{\mathcal{N}_h}$ 聚合为维度 k' 的对偶四元数向量 $\boldsymbol{w}_h^{agg}(\boldsymbol{a}_h^{agg},\boldsymbol{b}_h^{agg},\boldsymbol{c}_h^{agg},\boldsymbol{d}_h^{agg})$. 为了充分地融合这两种表示，此处采用结合加法和乘法的双交互聚合方式：

$$\begin{aligned}\boldsymbol{w}_h^{agg}=&\mathrm{LeakyRelu}(\boldsymbol{P}_1(\boldsymbol{w}_h+\boldsymbol{w}_{\mathcal{N}_h}))+\\&\mathrm{LeakyRelu}(\boldsymbol{P}_2(\boldsymbol{w}_h\odot\boldsymbol{w}_{\mathcal{N}_h}))\end{aligned}\tag{24}$$

其中 $\boldsymbol{P}_1,\boldsymbol{P}_2$ 是可训练的对偶四元数权重矩阵.

以上偏好传播和偏好聚合模块帮助用户、商品和知识实体显式地捕获其局部结构信息. 考虑到对偶四元数丰富的表达能力，可要求模型仅学习一阶邻居信息以辅助个性化推荐任务. 此外，也可通过堆叠以上偏好传播与聚合层，从而捕获协同知识图谱中的高阶结构信息.

4.4　用户-商品偏好预测

通过以上过程，对于每个用户和商品最终可以得到其原始对偶四元数嵌入和通过偏好传播增强的对偶四元数嵌入，即用户 u 具备两种嵌入 $\{\boldsymbol{w}_u,\boldsymbol{w}_u^{agg}\}$，商品 v 具备两种嵌入 $\{\boldsymbol{w}_v,\boldsymbol{w}_v^{agg}\}$. 为了获得用于偏好排序的效用函数，首先拼接用户 u 的两种嵌入作为其最终表示：

$$\begin{aligned}\boldsymbol{w}_u^*(\boldsymbol{a}_u^*,\boldsymbol{b}_u^*,\boldsymbol{c}_u^*,\boldsymbol{d}_u^*)&=\boldsymbol{w}_u\|\boldsymbol{w}_u^{agg}\\&=[\boldsymbol{a}_u,\boldsymbol{a}_u^{agg}]+[\boldsymbol{b}_u,\boldsymbol{b}_u^{agg}]\boldsymbol{i}+\\&\quad[\boldsymbol{c}_u,\boldsymbol{c}_u^{agg}]\boldsymbol{j}+[\boldsymbol{d}_u,\boldsymbol{d}_u^{agg}]\boldsymbol{k}\end{aligned}\tag{25}$$

其中“‖”表示两个对偶四元数之间的连接操作，即分别拼接其对应的实数部分和虚数部分；$[\cdot,\cdot]$ 表示向量之间的拼接操作. 同理，通过如下方式获得商品 v 的最终表示 $\boldsymbol{w}_v^*$：

$$\boldsymbol{w}_v^*(\boldsymbol{a}_v^*,\boldsymbol{b}_v^*,\boldsymbol{c}_v^*,\boldsymbol{d}_v^*)=\boldsymbol{w}_v\|\boldsymbol{w}_v^{agg}\tag{26}$$

由于对偶四元数丰富的表达能力，本文将用户 u 和商品 v 的对偶四元数表示之间的扩展内积作为偏好排序的效用函数，计算用户 u 对商品 v 的偏好分数：

$$\begin{aligned}\hat{s}(u,v)&=\langle\boldsymbol{w}_u^*,\boldsymbol{w}_v^*\rangle\\&=\langle\boldsymbol{a}_u^*,\boldsymbol{a}_v^*\rangle+\langle\boldsymbol{b}_u^*,\boldsymbol{b}_v^*\rangle+\langle\boldsymbol{c}_u^*,\boldsymbol{c}_v^*\rangle+\langle\boldsymbol{d}_u^*,\boldsymbol{d}_v^*\rangle\end{aligned}\tag{27}$$

本文假设与未交互的商品相比用户更偏爱已交互的商品，从而对用户的偏好进行优化. 具体地，针对同一用户的不同商品可设计如下成对排序损失函数：

$$\mathcal{L}_{\boldsymbol{Y}}=\sum_{(u,i,j)\in\mathcal{T}_{\boldsymbol{Y}}}-\ln\sigma(\hat{s}(u,i)-\hat{s}(u,j))\tag{28}$$

其中，$\mathcal{T}_{\boldsymbol{Y}}=\{(u,i,j)\mid y_{ui}=1,y_{uj}=0\}$ 是由已观测到的用户-商品交互构成的成对比较训练集. 协同排序损失函数可使模型从用户偏好出发，学习到保留排序的效用函数.

4.5　模型学习及优化

采用 L_2 正则化预防过拟合，记为 $\mathcal{L}_{\mathrm{reg}}$，其中包含所提模型中所有的对偶四元数嵌入和可训练的参数. 使用 λ_1 和 λ_2 分别控制对偶四元数协同知识图谱嵌入的损失项和正则化项的权重，从而在最终完整的损失函数 $\mathcal{L}_{\mathrm{DQKGR}}$ 中对 $\mathcal{L}_{\mathrm{KG}}$ 和 $\mathcal{L}_{\boldsymbol{Y}}$ 进行联合优化：

$$\mathcal{L}_{\mathrm{DQKGR}}=\mathcal{L}_{\boldsymbol{Y}}+\lambda_1\mathcal{L}_{\mathrm{KG}}+\lambda_2\mathcal{L}_{\mathrm{reg}}\tag{29}$$

受交替优化在已有工作[15]中成功应用的启发，此处对整体损失函数 $\mathcal{L}_{\mathrm{DQKGR}}$ 和协同知识图谱嵌入损失函数 $\mathcal{L}_{\mathrm{KG}}$ 进行联合优化. 在每次迭代中，首先基于 $\mathcal{L}_{\mathrm{DQKGR}}$ 的梯度联合更新由协同知识图谱嵌入和偏好传播聚合产生的所有对偶四元数嵌入，之后基于 $\mathcal{L}_{\mathrm{KG}}$ 的梯度对协同知识图谱中实体和关系再次进行更新，从而更好地利用汉密尔顿乘积丰富的表达能力加强对知识图谱信息的利用.

4.6　模型分析

在上述模型介绍中，DQKGR 模型在学习实体之间嵌入时将关系建模为莫比乌斯变换 $f(\boldsymbol{h},\boldsymbol{r})$，其可看做一系列几何变换的组合：$f(\boldsymbol{h},\boldsymbol{r})=f_4\circ f_3\circ f_2\circ f_1(\boldsymbol{h},\boldsymbol{r})$，其中 $f_1(\boldsymbol{h},\boldsymbol{r})=\boldsymbol{h}+\frac{r_d}{r_c}$，可看做平移 $\frac{r_d}{r_c}$ 个单位；$f_2(\boldsymbol{h})=\frac{1}{\boldsymbol{h}}$，可看做反演或者反射变换；$f_3(\boldsymbol{h},\boldsymbol{r})=\frac{r_b r_c-r_a r_d}{r_c^2}\boldsymbol{h}$，可看做位似或旋转变换；$f_4(\boldsymbol{h},\boldsymbol{r})=\boldsymbol{h}+\frac{r_a}{r_c}$，可看做平移变换. 表 2 统计了几种流行知识图谱嵌入模型支持的空间变换，与其他嵌入模型相比，DQKGR 中的嵌入学习模块 DQKGE 可将关系建模为更为丰富的几何变换. 基于此，DQKGE 可以充分学习协同知识图谱中的各种关系，推理对称、反演以及组合等多种关系模式(证明见附录 A.2)，为后续推荐过程提供更为精准可靠的嵌入.

《计算机学报》

表 2 几种流行知识图谱嵌入模型支持的空间变换

模型	平移	旋转	位似	反演	反射
TransE	√	×	×	×	×
RotatE	×	√	×	×	×
ComplEx	×	√	√	×	×
QuatE	×	√	×	×	×
DualE	√	√	×	×	×
DQKGE	√	√	√	√	√

注意到与较远处节点相比，一个节点在图结构中通常与较近邻域的节点更为相似，也即协同知识图谱中节点的相似性通常是局部的. 莫比乌斯变换是一种双射共形映射，其具有局部保角性[25]. 由于平移、反演、位似与旋转变换都保持角度不变，两个超复数(或向量)之间的幅角差(夹角)在经过莫比乌斯变换后不变. 因此，使用莫比乌斯变换能够通过特定关系的变换来捕获实体之间的相似性. 这有益于偏好传播过程中对于实体的聚合，也有利于利用知识图谱路径上的邻居信息.

进一步地，对于莫比乌斯变换矩阵 $\boldsymbol{S}$，如果 $\det(\boldsymbol{S})=1$，变换后图子结构的体积和方向将保持不变. 在向量空间中，其他知识图谱嵌入模型由于转换能力有限(如不支持反演和反射)，难以捕捉复杂图结构，因此，基于莫比乌斯变换的 DQKGE 在捕捉图谱结构方面更具有优势. 此外，鉴于 DQKGE 具有投影变换的特性，能够较好地捕获异质结构. 因此，在推荐过程中，DQKGR 捕获的子图结构更有利于对知识图谱路径上的邻居信息进行聚合.

5 实 验

为了验证所提方法的有效性，在三个具有代表性的真实数据集上进行实验，包括音乐数据集 Last-FM、电影数据集 MovieLens-20M 和书籍数据集 Book-Crossing，数据集信息如表 3 所示.

表 3 数据集描述

类型	参数	Last-FM	MovieLens-20M	Book-Crossing
用户-商品交互	用户数量	23566	116679	663
	商品数量	48123	16878	11385
	交互数量	3034796	6596097	16744
	密度/%	0.15	0.33	0.22
知识图谱	实体数量	58266	102569	77903
	关系数量	9	32	25
	三元组数量	464567	499474	151500

5.1 对比方法

为验证本文所提出算法的有效性，在实验中采用以下 9 种对比方法与本文算法进行对比：

(1) BPRMF[54] 是一个传统的矩阵分解模型，其使用内积和成对比较排序损失函数进行个性化推荐.

(2) FM[55] 是一种内容感知因子分解模型，采用因子化的特征交互实现推荐任务.

(3) NFM[56] 是一种较强的分解模型，其可同时利用 FM 的线性特征和神经网络的非线性特征进行推荐.

(4) CKE[12] 通过 TransR[16] 算法从知识图谱中学习商品嵌入，作为补充信息进行个性化推荐.

(5) CFKG[45] 将用户-商品交互和商品知识信息集成到统一的图谱中，从而学习用户和商品表示进行推荐.

(6) RippleNet[11] 是一种类似记忆网络的模型，可以在知识图谱中沿路径传播用户的偏好，辅助推荐任务.

(7) KGCN[14] 是一个端到端的模型，其基于图卷积神经网络在知识图谱中聚合商品信息.

(8) KGAT[15] 采用注意力机制区分邻居的重要性，以对知识图谱中的高阶信息进行显式建模.

(9) QKGN[53] 引入超复数嵌入，通过设计特定的偏好传播机制与结构信息相关的注意力机制进行参数更新.

此外，为了验证模型各部分的有效性，本文添加了消融实验. 具体而言，除所提出 DQKGR 模型之外，还实现两种变体模型，因此共提出如下 3 种模型：

(1) w/o DQKGE 在没有使用对偶四元数协同知识图谱表示学习的情况下，只采用所提出的偏好传播与聚合方法对协同知识图谱中所有用户、商品和实体的随机初始化对偶四元数嵌入进行更新，完成推荐任务.

(2) w/o PPL 只使用所提出的对偶四元数协同知识图谱表示学习为用户和商品求取四元数嵌入，不借助偏好传播与聚合机制，直接实现个性化推荐.

(3) DQKGR 为本文提出的完整模型，结合基于对偶四元数的协同知识图谱嵌入及偏好传播与聚合方法，实现最终的推荐任务.

5.2 评价指标

为了对实验结果进行全面的比较，在实验中采用以下评价指标：

(1) 命中率(Hit Ratio，HR)直观地衡量模型预测的 Top-K 推荐列表是否包括用户真正感兴趣的商品：若存在，则该指标为 1；反之，则该指标为 0.

《计算机学报》

(2)精确率(Precision,P)表示在模型预测的 Top-K 推荐商品列表中,用户真正感兴趣的商品所占比例.

(3)召回率(Recall,R)为用户真正感兴趣的商品中,被正确预测到 Top-K 推荐列表中的比例.

(4)归一化折损累计增益(Normalized Discounted Cumulative Gain,NDCG)根据 Top-K 推荐列表中每一项的排序位置以及其是否在用户真正喜爱的商品中为每一项赋予分值,排序在前的命中商品得分更高,因此该指标可衡量排序算法的平均性能.

5.3　Last-FM 数据集的实验

5.3.1　数据集简介

Last-FM 数据集中原始用户-唱片交互信息来自 LFM-1b[3].本课题采用 KGAT 论文[15]中提出的一个子集,使用 10-core 的设置提高数据集的质量(即只保留至少产生过十次交互的用户和唱片).此外,所有的唱片均被映射到 Freebase 知识图谱中的实体,从而提供商品知识图谱辅助个性化推荐.

在实验中,使用 KGAT[15]发布的公共训练集和测试集,其中训练集包含 80%的用户-唱片交互记录,测试集包含其余的交互记录.将每个观察到的用户-唱片交互视为一个正例,并为该用户随机抽取一个未交互过的唱片作为一个负例,以构建成对比较的数据.

5.3.2　实现细节

在一台配置 2 个 Intel Xeon CPU(E5-2620 v4@2.10 GHz)和 128 GB RAM 的 Linux 机器上使用一块 TITAN RTX GPU 进行所有的实验.具体地,令对比模型使用的嵌入维度 $k=64$(由于 RippleNet 计算量过大,按照其论文的建议将其嵌入尺寸设为 16).在所提出的 DQKGR 中,使用一层维度为 $k'=64$ 的偏好传播与聚合层.此外,参考四元数循环神经网络[34]中提出的四元数网络初始化算法,对涉及到的四元数嵌入及网络参数进行初始化.此外,针对所有的方法,统一设置批大小(Batch size)为 1024,统一使用 Xavier 初始化器和 Adam 优化器.在超参数设置上,对于所提方法,设置 λ_1 为 0.1,λ_2 为 1e-4,学习率 l 为 1e-3,训练参数丢弃概率(Dropout)为 0.1.对于对比方法,使用网格搜索调整其超参数.对于 KGAT 方法,选取其论文[15]中性能最佳的结构,即使用维度分别为{64,32,16}的三个注意力嵌入传播层辅助偏好建模.所有实验均进行 5 次并记录平均值,作为最终实验结果.

5.3.3　实验结果

根据表 4 中 Last-FM 数据集上 $K=20$ 时各模型的定量实验结果,有如下发现:

表 4　所有模型在三个数据集上的实验结果(最优结果使用粗体表示,次优结果使用下划线表示)

数据集	评价指标	BPRMF	FM	NFM	CKE	CFKG	RippleNet	KGCN	KGAT	QKGN	w/o DQKGE	w/o PPL	DQKGR
Last-FM	*HR*@20↑	0.3246	0.3256	0.3104	0.3254	0.2909	0.3335	0.3503	0.3559	0.3681	0.3623	<u>0.3691</u>	**0.3873**
	Improve/%	—	+0.31	−4.38	+0.26	−10.36	+2.76	+7.91	+9.65	+13.40	+11.61	13.70	+19.30
	P@20↑	0.0306	0.0297	0.0290	0.0306	0.0273	0.0326	0.0339	0.0344	0.0348	0.0357	<u>0.0359</u>	**0.0376**
	Improve/%	—	−3.07	−5.23	−0.13	−10.75	+6.50	+10.69	+12.39	+13.73	+16.73	+17.45	+22.87
	R@20↑	0.0733	0.0744	0.0737	0.0730	0.0674	0.0797	0.0842	0.0865	<u>0.0968</u>	0.0929	0.0962	**0.1031**
	Improve/%	—	+1.49	+0.60	−0.41	−8.00	+8.75	+14.87	+18.01	+32.06	+24.04	+31.30	+40.65
	NDCG@20↑	0.0628	0.0480	0.0601	0.0629	0.0577	0.0691	0.0720	0.0736	0.0815	0.0778	<u>0.0822</u>	**0.0873**
	Improve/%	—	−23.57	−4.30	+0.24	−8.10	+10.04	+14.74	+17.24	+29.78	+23.94	+30.99	+38.77
MovieLens-20M	*HR*@20↑	0.6498	0.6245	0.6285	0.6547	0.5294	0.6549	0.6636	<u>0.6672</u>	0.6669	0.6434	0.6610	**0.6682**
	Improve/%		−3.89	−3.28	+0.75	−18.53	+0.78	+2.12	+2.68	+2.63	−0.98	+1.72	+2.83
	P@20↑	0.0639	0.0546	0.0609	0.0667	0.0482	0.0674	0.0688	0.0691	0.0693	0.0617	<u>0.0698</u>	**0.0702**
	Improve/%	—	−14.45	−4.70	+4.37	−24.60	+5.48	+7.75	+8.13	+8.45	−3.40	+9.32	+9.86
	R@20↑	0.1668	0.1453	0.1537	0.1703	0.1134	0.1673	0.1711	0.1727	0.1781	0.1746	<u>0.1811</u>	**0.1821**
	Improve/%	—	−12.89	−7.87	+2.07	−32.05	+0.30	+2.55	+3.49	+5.46	+4.68	+8.58	+9.17
	NDCG@20↑	0.1175	0.0969	0.1020	0.1032	0.0873	0.1131	0.1215	0.1219	0.1228	0.1206	<u>0.1230</u>	**0.1277**
	Improve/%	—	−17.55	−13.19	−12.13	−25.67	−3.70	+3.41	+3.75	+4.51	+2.65	+4.72	+8.68
Book-Crossng	*HR*@20↑	0.0980	0.1192	0.1207	0.0982	0.1071	0.1463	0.1750	0.1780	0.1864	0.1448	<u>0.1872</u>	**0.2017**
	Improve/%	—	+21.54	+23.07	+0.20	+9.23	+49.22	+78.46	+81.54	90.20	+47.69	+90.93	+105.81
	P@20↑	0.0059	0.0072	0.0074	0.0060	0.0051	0.0089	0.0102	0.0103	0.0112	0.0081	<u>0.0115</u>	**0.0122**
	Improve/%	—	+23.13	+26.53	+1.36	−12.76	+51.36	+73.13	+75.68	+89.83	+38.44	+95.57	+106.78
	R@20↑	0.0220	0.0269	0.0295	0.0222	0.0258	0.0378	0.0477	0.0483	<u>0.0521</u>	0.0351	0.0505	**0.0609**
	Improve/%	—	+22.35	+34.03	+0.77	+17.26	+71.83	+116.81	+119.49	+136.82	+59.25	+129.25	+176.81
	NDCG@20↑	0.0133	0.0138	0.0133	0.0136	0.0138	0.0205	0.0255	0.0259	0.0288	0.0178	<u>0.0311</u>	**0.0318**
	Improve/%	—	+3.70	+0.15	+2.49	+3.85	+54.52	+92.31	+95.02	+116.54	+34.24	+134.62	+139.09

计算机学报

首先，所提出的基于对偶四元数的协同知识图谱推荐模型 DQKGR 在所有评价指标上取得明显的性能增益. 具体地，虽然 DQKGR 只使用一层的偏好传播层捕获一阶连通性，但由于利用对偶四元数向量及其汉密尔顿乘积进行建模，在所有指标上均优于最佳对比方法 QKGN. QKGN 引入超复数，在性能方面超越了 KGAT；但在相同参数量的情况下，QKGN 的性能不及 DQKGR，这表明对偶四元数更有利于模型学习协同知识图谱中的潜在关系和和嵌入表示，从而彰显了对偶四元数的优势. 将对偶四元数协同知识图谱表示学习方法直接用于实现个性化推荐的 w/o PPL 模型同样明显优于 KGAT 模型，表明所引入的莫比乌斯变换更有利于从知识图谱的复杂关系中挖掘有利于推荐的辅助信息. 与仅使用用户-商品交互信息的 BPRMF 方法相比，所提 DQKGR 模型在四个指标上分别提升 19.30%、22.87%、40.65% 及 38.77%.

第二，对所提出的整体方法 DQKGR 及其两个组成方法的性能进行分析：由于莫比乌斯变换可以将关系建模为丰富的几何变换，且对偶四元数空间有益于捕捉实体间的潜在关系，所提出三个方法的表现一致优于所有的对比方法. 同时，去除对偶四元数协同知识图谱表示学习部分，或者去除基于注意力的偏好传播与聚合部分，均会导致模型性能下降，推荐效果弱于整体模型，这一现象表明了两个组成部分的有效性. 此外，与去除对偶四元数协同知识图谱表示学习部分的模型（w/o DQKGE）相比，仅去除偏好传播与聚合部分的模型（w/o PPL）可达到更好的结果，这表明经过对知识图谱嵌入和个性化推荐任务的协同处理，所提基于对偶四元数协同知识图谱表示学习方法可以得到较高质量的用户、商品嵌入，可直接应用于个性化推荐任务.

第三，在所有的对比方法中，KGAT、KGCN 和 QKGN 取较优的性能，而这三种方法均属于混合的知识图谱推荐方法. 这表明从知识图谱中获取邻居信息从而丰富实体的建模方式，有助于实现更精准的个性化推荐. FM 和 NFM 方法与基线方法 BPRMF 相比无明显提升，表明此类方法在以知识图谱为辅助信息的推荐场景下受到限制.

此外，CKE 和 CFKG 的性能提升也并不明显，因此，简单地使用 TransE 或 TransR 方法得到的嵌入可能不太适合该协同知识图谱上的知识获取任务，该种嵌入不能充分利用知识图谱的信息辅助个性化推荐任务. 此外，图 3(a)以曲线的形式展示了

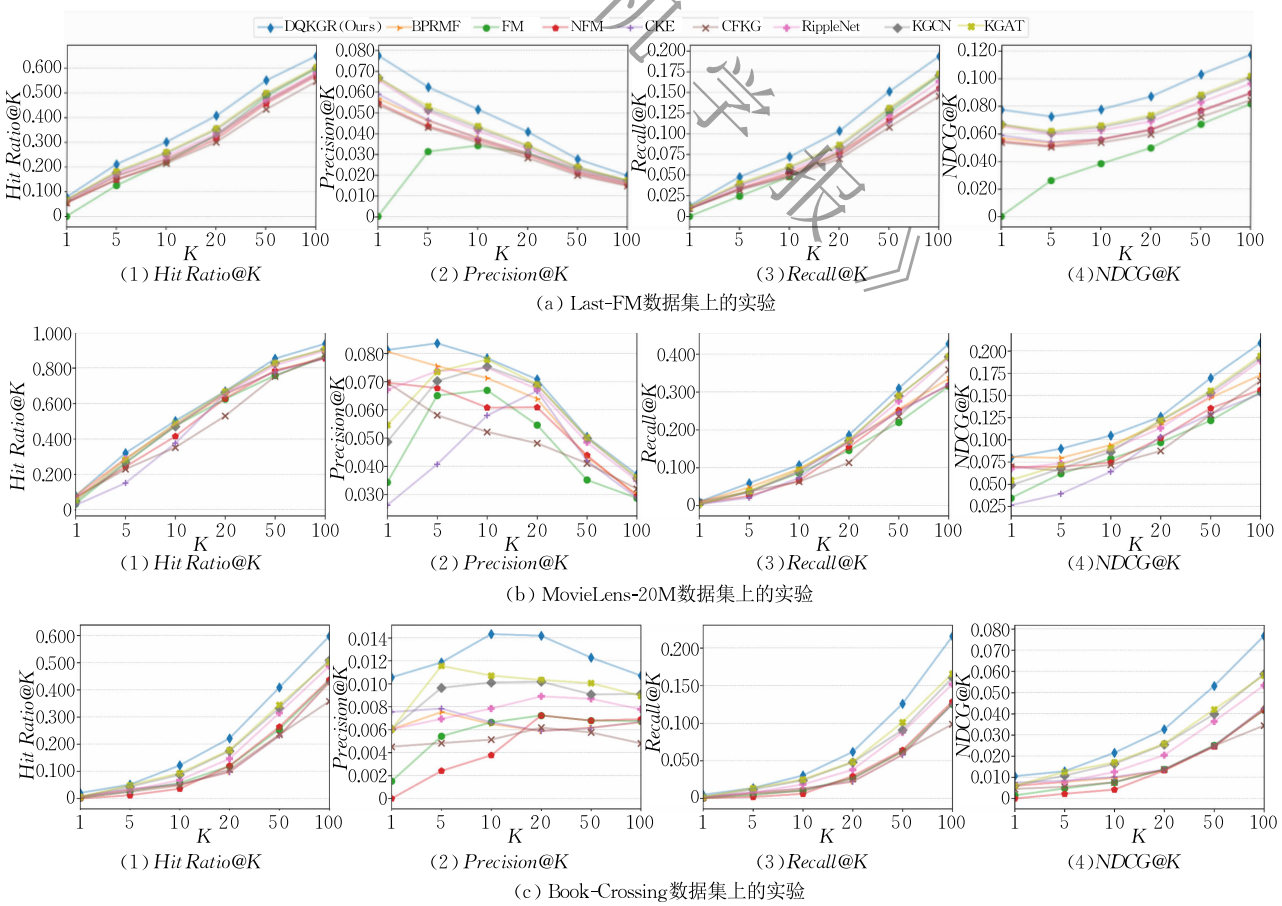


图 3 三个数据集上不同 K 值的 Top-K 推荐实验结果

Last-FM 数据集上不同的 K 值（$K \in \{1,5,10,20,50,100\}$）情况下各模型的性能比较，根据实验结果可知，对于所有不同的 K 值，所提出的 DQKGR 模型推荐性能一致优于所有对比方法，这表明所提的方法可在不同 K 值的推荐场景下提高个性化推荐的精准程度.

5.4　MovieLens-20M 数据集的实验

5.4.1　数据集简介

MovieLens-20M 是完整 MovieLens 数据集的一个子集，包括超过 100 000 用户对大约 27 000 部电影的 20 000 000 次显式评分（分数从 0.5 到 5 不等）. 本文参考文献[14]中的数据处理方式将评分数据转化为隐式反馈，同样采用 10-core 设置以提高数据集的质量. 此外，采用与 Last-FM 相同的方式划分训练集和测试集，并使用 KGCN 论文[14]公开的电影知识图谱作为辅助信息.

5.4.2　实现细节

对于提出的 DQKGR 模型，此处设置 λ_1 为 1e-5，λ_2 为 1e-3，l 为 1e-5. 此外，所有方法的实现细节与 Last-FM 数据集上的实验保持一致. 由于此处使用的知识图谱的数据量较大，考虑到计算设备的限制，实验中使用前 10 000 个用户为所有方法进行训练和测试.

5.4.3　实验结果

由表 4 中 MovieLens-20M 数据集上 $K=20$ 时各模型的定量实验结果可知，所提出的 DQKGR 同样在所有指标上优于所有对比方法. 特别地，DQKGR 在四个指标上分别比基线方法 BPRMF 提升 2.83%、9.86%、9.17% 和 8.68%. 因此，所提方法的有效性再次得到验证.

同时，DQKGR 始终可比单独使用协同知识图谱表示学习和单独使用偏好传播机制取得更好的性能，表明协同知识图谱表示学习和偏好传播聚合机制均为实现高质量 Top-K 推荐的重要组成部分. 仅使用基于对偶四元数的协同知识图谱表示学习方法进行推荐的效果优于仅使用偏好传播与聚合的方法，这表明所设计的基于莫比乌斯变换的三元组匹配规则可有效实现知识图谱嵌入和个性化推荐任务.

此外，可观察到 FM、NFM 和 CFKG 的性能与传统方法 BPRMF 相比并不占优，其原因在于上述方法简单地使用知识图谱嵌入方法学到的商品表示，未能从较大的知识图谱（超过 100 000 个实体）中充分挖掘有利于个性化推荐的辅助信息，导致性能不理想.

类似地，图 3(b) 中 MovieLens-20M 数据集上不同 K 值下的推荐效果评测曲线表明，所提出的 DQKGR 模型在不同 K 值场景下取得最佳性能.

5.5　Book-Crossing 数据集的实验

5.5.1　数据集简介

原始的 Book-Crossing 数据集[57]包含了 Book-Crossing 社区用户对超过 200 000 本图书的超过 1 000 000 个显式评分（分数从 0 到 10 不等）. 考虑到数据集的稀疏性，将所有被评分过的书籍作为已交互样本构建隐式反馈，并通过相同的 10-core 设置对数据集进行清洗. 使用与 MovieLens-20M 相同的方法构造训练集和测试集，并使用 RippleNet 论文[11]公开的图书知识图谱作为辅助信息.

5.5.2　实现细节

此数据集上，将所提的 DQKGR 嵌入维度 k 设为 16，偏好传播与聚合层的维数 k' 同样设为 16. 设置超参数 λ_1 为 1.0，λ_2 为 1e-3，l 为 1e-3. 对于所有对比方法，对其参数使用网格搜索并记录最佳实验结果. 此外，所有方法的实现细节与 Last-FM 数据集上的实验保持一致.

5.5.3　实验结果

根据表 4 中 Book-Crossing 数据集上 $K=20$ 时的定量实验结果，所提出的 DQKGR 模型在所有评价指标上一致优于对比方法. 具体地，DQKGR 模型在四个指标上分别比仅使用用户-商品交互数据的 BPRMF 模型取得超过 1 倍的性能提升.

观察对比方法的定量实验结果可知，与 BPRMF 方法相比，所有其他方法均取得一定提升，其中，CKE 模型和 CFKG 模型在此数据集上提升最不明显，此数据集与商品知识图谱数据相比，用户-商品交互数据规模较小，因此，CKE 模型和 CFKG 模型表现相对较弱的原因在于直接使用知识图谱嵌入的方法对用户和商品进行建模，受该知识图谱规模影响，其用户和商品的表示更适合知识图谱上的链路预测等任务，而不是个性化推荐任务.

对本文的三个模型效果进行分析：仅使用对偶四元数偏好传播与聚合方法的 w/o DQKGE 模型可以取得与最佳对比方法 KGAT 相匹配的性能，其原因在于对偶四元数向量具备较强的表达能力，且设计的偏好传播与聚合机制有利于利用知识图谱中的结构信息辅助个性化推荐任务. 仅使用协同知识图谱表示学习方法的 w/o PPL 模型可取得明显优于 KGAT 模型的推荐效果，其原因在于基于莫比乌

计算机学报

斯变换的三元组匹配规则比单一的几何变换方法在知识图谱实体和关系建模方面更具优势. w/o PPL 模型和整体的 DQKGE 模型之间存在的性能差距再次验证了偏好传播与聚合机制的有效性.

此外,图 3(c)中 Book-Crossing 数据集上不同 K 值场景下的推荐效果评测曲线展现出所提的 DQKGR 模型具备明显优势.

5.6 不同嵌入维度的实验及分析

如图 4 所示,在 Last-FM 数据集上对比了 KGAT 及所提出的 DQKGR 模型在不同嵌入维度 k($k\in\{8,16,32,64,128\}$)设置下的 Top-20 推荐实验结果. 在嵌入维度 k 从 8 增长至 128 过程中,两种方法推荐的精准程度均先上升后下降,其原因在于:维度较大的嵌入可以编码更多的信息,表达能力更强;但是,嵌入维度过高会导致过拟合,以致降低模型性能. 此外,易发现随着嵌入维度的增大,DQKGR 模型的推荐效果始终高于 KGAT 模型,并且在嵌入维度较大的情况下,相对更不易因过拟合而导致性能下降.

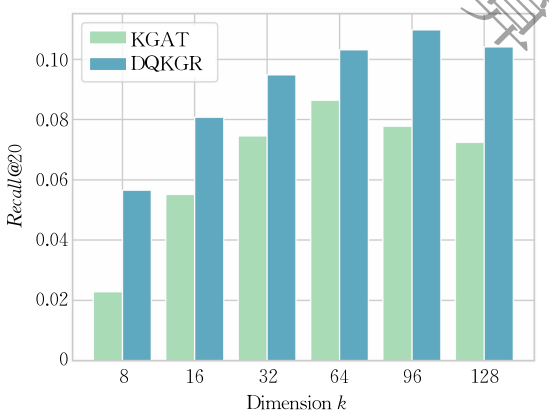


图 4 Last-FM 数据集上 KGAT 与所提 DQKGR 不同嵌入维度的 Top-20 推荐效果比较

5.7 模型组件的消融研究

本文提出的 DQKGR 模型引入了对偶四元数嵌入以及莫比乌斯变换. 为研究这两个组件对模型的贡献,本文进行消融实验. 分别设置:(1)去掉莫比乌斯变换的版本,记作 w/o 莫比乌斯变换;(2)去掉对偶四元数嵌入的版本,这里使用实数数域代替对偶四元数数域,并且保证总参数量不变,记作 w/o 对偶四元数. 消融实验的结果见表 5,可以看到在去除莫比乌斯变换后,模型性能下降,原因在于模型不能很好地建模协同知识图谱中的多种复杂关系. 此外,注意到去除对偶四元数后,尽管模型的参数保持不变,但模型的性能降低,这表明对偶四元数乘法可以有效捕捉实体之间的潜在关系,而仅仅通过在实数域增加嵌入维度并不能实现该效果. 综上所述,以上消融实验证明了莫比乌斯变换和对偶四元数嵌入都是模型的重要组成部分,它们有益于模型为用户和商品学习更为精准的嵌入.

表 5 Book-Crossing 数据集上,$K=20$ 时,模型关于莫比乌斯变换和对偶四元数嵌入的消融实验

方法	$HR@20\uparrow$	$P@20\uparrow$	$R@20\uparrow$	$NDCG@20\uparrow$
w/o 莫比乌斯变换	0.1947	0.0113	0.0550	0.0291
w/o 对偶四元数嵌入	0.1935	0.0111	0.0532	0.0281
DQKGR	**0.2017**	**0.0122**	**0.0609**	**0.0318**

5.8 模型复杂度分析

考虑到本文引入对偶四元数嵌入和莫比乌斯变换,这增强了模型的表达能力,但一定程度上也增加了模型的复杂度,接下来本文将基于时间复杂度和空间复杂度对模型的性能进行分析. 不失一般性,本文基于以上两个角度对 DQKGR 与多个方法进行对比,用以全面性地对模型进行考量.

(1)由表 6,在保持 DQKGR 与其他模型参数相同(34 M)的情况下,DQKGR 可达到最优性能. 本文继续研究了不同模型达到性能最优时的参数量,如表 7 所示,DQKGR 在达到最优性能时,所用的总参数量和其他模型相当. 其原因如下:尽管 DQKGR 的对偶四元数包含多个部分(包括实部和虚部),但这些部分紧密耦合在一起,具有更强的表达能力,减少了实际使用过程中 DQKGR 为每个部分分配的维度,因此参数总量可以和其他模型保持一致. 此外也可以看到,仅仅通过增加其他模型的嵌入维度并不能达到和 DQKGR 相当的效果,这也验证了对偶四元数嵌入的有效性.

表 6 Book-Crossing 数据集上,相同参数量下不同模型性能的比较

方法	$HR@20\uparrow$	$P@20\uparrow$	$R@20\uparrow$	$NDCG@20\uparrow$
KGCN	0.1745	0.0101	0.0471	0.0253
RippleNet	0.1511	0.0092	0.0388	0.0211
KGAT	0.1778	0.0103	0.0482	0.0259
QKGN	0.1866	0.0113	0.0521	0.0291
DQKGR	**0.2017**	**0.0122**	**0.0609**	**0.0318**

表 7 Book-Crossing 数据集上,不同模型达到最优性能时的参数量(其中 k 表示维度,n_e 表示协同知识图谱中实体的数量,n_r 表示协同知识图谱中关系的数量,R-Net 是 RippleNet)

方法	空间复杂度	参数量/M
KGCN	$\mathcal{O}(n_e k+n_r k)$	30.15
R-Net	$\mathcal{O}(n_e k+n_r k)$	33.76
KGAT	$\mathcal{O}(n_e k+n_r k)$	31.05
QKGN	$\mathcal{O}(4n_e k+4n_r k)$	34.18
DQKGR	$\mathcal{O}(8n_e k+8n_r k)$	32.22

计算机学报

(2)此外，本文对不同模型的训练时间进行了比较. 根据表 8 所示，DQKGR 的单次迭代时间比其他模型略长；但注意到 DQKGR 模型可以以较快的速度收敛，因而在总时间方面，比其他模型训练要短. 这表明依托对偶四元数嵌入和莫比乌斯变换，模型可以实现高效学习嵌入.

表 8　Book-Crossing 数据集上，不同模型训练速度的比较（其中 k 表示维度，R-Net 是 RippleNet）

方法	时间复杂度	单次迭代时间/h	收敛迭代次数	总时间/h
KGCN	$\mathcal{O}(k)$	0.65	120	78.0
R-Net	$\mathcal{O}(k\log k)$	0.88	150	132.0
KGAT	$\mathcal{O}(k)$	0.75	150	112.0
QKGN	$\mathcal{O}(k)$	0.80	100	80.0
DQKGR	$\mathcal{O}(k)$	0.82	80	65.5

综上所述，DQKGR 在取得优越性能的同时，在时间和复杂度方面可以和既往模型相当甚至也优于既往模型. 此外，关于更多模型性能的分析可见附录 A.5. 因此 DQKGR 也适用于大规模数据集以及对于算法实时性要求较高的情况.

5.9　嵌入学习部分的对比实验及分析

用户和商品的嵌入学习对推荐系统非常重要. 通常而言，嵌入学习得越精准，推荐的性能会越好. 在前文的模型分析以及实验部分，本文的 DQKGR 推荐模型基于 DQKGE 学习协同知识图谱中商品和用户的嵌入，为了探究 DQKGE 方法学习嵌入的有效性，本文主要基于以下两个方面进行对比探究.

(1)本文将 DQKGR 中 DQKGE 嵌入模块用 TransE[17]、TransR[16]、RotatE[20]、QuatE[28]、MuRP[42]、DualE[31]以及 ROTH[43]等嵌入模型替代，分别记作：TransE-R、TransR-R、RotatE-R、QuatER、MuRP-R、DualE-R 和 ROTH-R. 保持总参数量相等的情况下，对以上模型进行精调并选取最优参数. 如表 9 所示，通过对比，可以看到 DQKGE 在学习协同知识图谱的嵌入时具有优势，这验证了 DQKGE 在 DQKGR 中的有效性.

表 9　Book-Crossing 数据集上，$K=20$ 时，不同知识图谱嵌入模型对实验结果的影响

方法	$HR@20\uparrow$	$P@20\uparrow$	$R@20\uparrow$	$NDCG@20\uparrow$
TransE-R	0.1676	0.0102	0.0372	0.0231
TransR-R	0.1789	0.0104	0.0486	0.0259
RotatE-R	0.1856	0.0108	0.0495	0.0267
QuatE-R	0.1921	0.0110	0.0512	0.0279
MuRP-R	0.1733	0.0108	0.0502	0.0271
DualE-R	0.1949	0.0112	0.0549	0.0289
ROTH-R	0.1951	0.0114	0.0548	0.0291
DQKGR	**0.2017**	**0.0122**	**0.0609**	**0.0318**

(2)探究 DQKGE 模型与其他知识图谱嵌入模型的关系. 在附录 A.3，表 11 对 DQKGE 和其他知识图谱嵌入模型进行了理论上的时空复杂度对比，可以看到 DQKGE 的时间复杂度和空间复杂度都是线性，因而模型复杂度并不高. 此外，本文证明了一些知识图谱嵌入方法（如 TransE、RotatE 等）可看作 DQKGE 的特例，兼之莫比乌斯变换可建模协同知识图谱中多种类型的关系，这从理论上验证了 DQKGE 的建模能力. 在附录 A.4 中，本文在两个知识图谱数据集 WN18RR[58]和 FB15K237[59]上进行了链接预测实验. 通过与其他流行知识图谱嵌入方法进行对比，由表 12 和表 13（附录 A.4 中）可以看到 DQKGE 取得了有竞争力的结果. 值得注意的是，以上数据集包含大量的关系模式，如对称、反对称、反演和组合模式等，上述知识图谱嵌入方法不能完全建模这些关系模式，或者不能有效挖掘实体间的潜在关系. 而 DQKGE 可基于莫比乌斯变换建模多种复杂关系模式，基于对偶四元数可以更为有效地挖掘实体之间的潜在交互，因而 DQKGE 可以更为精准地学习嵌入. 综上所述，以上理论分析和实验结果表明，所提出的 DQKGE 可单独用作一个知识图谱嵌入方法，从而为下游任务提供高效的嵌入.

5.10　多个偏好传播与聚合层的实验及分析

由前文实验结果和分析可知，只使用较少的偏好传播与聚合层可以为所提 DQKGR 模型带来足够的性能提升. 为研究多个偏好传播与聚合层的影响，本文在 Book-Crossing 数据集上进行实验，并将具有 n 个偏好传播与聚合层的模型表示为 DQKGR-n. 实验结果如表 10 所示，使用多个偏好传播与聚合层，分别研究层数为 1、2、3、4、5 时的模型性能. 可以看到，适当增加偏好传播层数，有利于借助图谱中的高阶信息辅助完成推荐任务，从而提升推荐精准程度. 随着层数的增多，模型的推荐性能会得到提升. 但当层数过多时可能因过拟合而导致性能下降.

表 10　Book-Crossing 数据集上，$K=20$ 时，不同偏好传播聚合层数对实验结果的影响

方法	$HR@20\uparrow$	$P@20\uparrow$	$R@20\uparrow$	$NDCG@20\uparrow$
DQKGR-1	0.2017	0.0122	0.0609	0.0318
DQKGR-2	0.2194	0.0137	0.0664	0.0351
DQKGR-3	0.2199	0.0139	0.0668	0.0353
DQKGR-4	0.2184	0.2132	0.0659	0.0346
DQKGR-5	0.2181	0.2130	0.0655	0.0343

6　总　结

本文针对结合知识图谱的个性化推荐问题，提

《计算机学报》

出了基于对偶四元数的协同知识图谱推荐模型DQKGR.该模型引入对偶四元数和莫比乌斯变换，对知识图谱中的三元组和用户-商品交互进行协同处理，从协同知识图谱中得到有效的用户、商品对偶四元数嵌入；同时，采用偏好传播和聚合机制利用每个实体的邻居信息进一步丰富实体表示，从而更精准地刻画用户和商品的偏好情况.此外，本文在三个真实数据集上进行实验，证实所提DQKGR模型推荐结果的多个指标优于现有方法.在后续的工作中，拟考虑以下方向：(1)在用户侧引入社交网络等辅助信息，使得模型能够同时利用用户社交信息和商品知识图谱进行个性化推荐；(2)考虑用户-商品交互发生的时间信息，令用户近期产生的交互行为对偏好建模具备更高的影响力，从而利用时序信息进一步提高下一时刻个性化推荐的精准程度.

参 考 文 献

[1] Lin T, Gao C, Li Y. CROSS: Cross-platform recommendation for social e-commerce//Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. Paris, France, 2019: 515-524

[2] Zhou Y, Wu J, Chan T H, et al. Interpreting video recommendation mechanisms by mining view count traces. IEEE Transactions on Multimedia, 2018, 20(8): 2153-2165

[3] Schedl M. The LFM-1b dataset for music retrieval and recommendation//Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval. New York, USA, 2016: 103-110

[4] Zhu Q, Zhou X, Song Z, et al. DAN: Deep attention neural network for news recommendation//Proceedings of the 33rd AAAI Conference on Artificial Intelligence. Honolulu, USA, 2019: 5973-5980

[5] Liu C, Zhou C, Wu J, et al. Social recommendation with an essential preference space//Proceedings of the 32nd AAAI Conference on Artificial Intelligence. New Orleans, USA, 2018: 346-353

[6] Koren Y, Bell R M, Volinsky C. Matrix factorization techniques for recommender systems. Computer, 2009, 42(8): 30-37

[7] Chen J, Zhang H, He X, et al. Attentive collaborative filtering: Multimedia recommendation with item- and component-level attention//Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. Shinjuku, Japan, 2017: 335-344

[8] Catherine R, Cohen W W. TransNets: Learning to transform for recommendation//Proceedings of the 11th ACM Conference on Recommender Systems. Como, Italy, 2017: 288-296

[9] Hu B, Shi C, Zhao W X, et al. Leveraging meta-path based context for top-*N* recommendation with A neural co-attention model//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. London, UK, 2018: 1531-1540

[10] Wang H, Zhang F, Hou M, et al. SHINE: Signed heterogeneous information network embedding for sentiment link prediction//Proceedings of the 11th ACM International Conference on Web Search and Data Mining. Marina Del Rey, USA, 2018: 592-600

[11] Wang H, Zhang F, Wang J, et al. RippleNet: Propagating user preferences on the knowledge graph for recommender systems//Proceedings of the 27th ACM International Conference on Information and Knowledge Management. Torino, Italy, 2018: 417-426

[12] Zhang F, Yuan N J, Lian D, et al. Collaborative knowledge base embedding for recommender systems//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. San Francisco, USA, 2016: 353-362

[13] Yu X, Ren X, Sun Y, et al. Personalized entity recommendation: A heterogeneous information network approach//Proceedings of the ACM International Conference on Web Search and Data Mining. New York, USA, 2014: 283-292

[14] Wang H, Zhao M, Xie X, et al. Knowledge graph convolutional networks for recommender systems//Proceedings of the World Wide Web Conference. San Francisco, USA, 2019: 3307-3313

[15] Wang X, He X, Cao Y, et al. KGAT: Knowledge graph attention network for recommendation//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Anchorage, USA, 2019: 950-958

[16] Lin Y, Liu Z, Sun M, et al. Learning entity and relation embeddings for knowledge graph completion//Proceedings of the 29th AAAI Conference on Artificial Intelligence. Austin, USA, 2015: 2181-2187

[17] Bordes A, Usunier N, García-Durán A, et al. Translating embeddings for modeling multi-relational data//Proceedings of the Advances in Neural Information Processing Systems 26. Nevada, USA, 2013: 2787-2795

[18] Wang Z, Zhang J, Feng J, et al. Knowledge graph embedding by translating on hyperplanes//Proceedings of the 28th AAAI Conference on Artificial Intelligence. Québec, Canada, 2014: 1112-1119

[19] Schlichtkrull M S, Kipf T N, Bloem P, et al. Modeling relational data with graph convolutional networks//Proceedings of the 15th European Semantic Web Conference. Heraklion, Greece, 2018: 593-607

[20] Sun Z, Deng Z, Nie J, et al. RotatE: Knowledge graph embedding by relational rotation in complex space//Proceedings of the 7th International Conference on Learning Representations. New Orleans, USA, 2019

[21] Huang J, Zhao W X, Dou H, et al. Improving sequential recommendation with knowledge-enhanced memory networks//Proceedings of the 41st International ACM SIGIR Conference

《计算机学报》

on Research and Development in Information Retrieval. Ann Arbor, USA, 2018: 505-514

[22] Ye Y, Wang X, Yao J, et al. Bayes embedding (BEM): Refining representation by integrating knowledge graphs and behavior-specific networks//Proceedings of the 28th ACM International Conference on Information and Knowledge Management. Beijing, China, 2019: 679-688

[23] Cao Y, Wang X, He X, et al. Unifying knowledge graph learning and recommendation: Towards a better understanding of user preferences//Proceedings of the World Wide Web Conference. San Francisco, USA, 2019: 151-161

[24] Tang X, Wang T, Yang H, et al. AKUPM: Attention-enhanced knowledge-aware user preference model for recommendation //Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Anchorage, USA, 2019: 1891-1899

[25] Kisil V V. Geometry of Mobius Transformations: Elliptic, Parabolic and Hyperbolic Actions of SL2 (R) (with DVD-ROM). World Scientific, 2012

[26] Clifford C. Preliminary sketch of biquaternions. Proceedings of the London Mathematical Society, 1871, 1(4): 381-395

[27] Parcollet T, Zhang Y, Morchid M, et al. Quaternion convolutional neural networks for end-to-end automatic speech recognition//Proceedings of the 19th Annual Conference of the International Speech Communication Association. Hyderabad, India, 2018: 22-26

[28] Zhang S, Tay Y, Yao L, et al. Quaternion knowledge graph embeddings//Proceedings of the Advances in Neural Information Processing Systems 32. Vancouver, Canada, 2019: 2731-2741

[29] Lechuga-Gutiérrez L, Medrano-Hermosillo J, Bayro-Corrochano E. Quaternion spiking neural networks control for robotics// Proceedings of the IEEE Latin American Conference on Computational Intelligence. Guadalajara, Mexico, 2018: 1-6

[30] Gaudet C J, Maida A S. Deep quaternion networks// Proceedings of the International Joint Conference on Neural Networks. Riode Janeiro, Brazil, 2018: 1-8

[31] Cao Z, Xu Q, Yang Z, et al. Dual quaternion knowledge graph embeddings//Proceedings of the 35th AAAI Conference on Artificial Intelligence. Virtual Event, 2021: 6894-6902

[32] Zhang S, Yao L, Tran L V, et al. Quaternion collaborative filtering for recommendation//Proceedings of the 28th International Joint Conference on Artificial Intelligence. Macao, China, 2019: 4313-4319

[33] Parcollet T, Morchid M, Bousquet P, et al. Quaternion neural networks for spoken language understanding//Proceedings of the IEEE Spoken Language Technology Workshop. San Diego, USA, 2016: 362-368

[34] Parcollet T, Ravanelli M, Morchid M, et al. Quaternion recurrent neural networks//Proceedings of the 7th International Conference on Learning Representations. New Orleans, USA, 2019

[35] Ji G, He S, Xu L, et al. Knowledge graph embedding via dynamic mapping matrix//Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics. Beijing, China, 2015: 687-696

[36] Xiao H, Huang M, Zhu X. TransG: A generative model for knowledge graph embedding//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Berlin, Germany, 2016: 2316-2325

[37] He S, Liu K, Ji G, et al. Learning to represent knowledge graphs with Gaussian embedding//Proceedings of the 24th ACM International Conference on Information and Knowledge Management. Melbourne, Australia, 2015: 623-632

[38] Ji G, Liu K, He S, et al. Knowledge graph completion with adaptive sparse transfer matrix//Proceedings of the 30th AAAI Conference on Artificial Intelligence. Phoenix, USA, 2016: 985-991

[39] Ebisu T, Ichise R. TorusE: Knowledge graph embedding on a lie group//Proceedings of the 32nd AAAI Conference on Artificial Intelligence. New Orleans, USA, 2018: 1819-1826

[40] Yang B, Yih W, He X, et al. Embedding entities and relations for learning and inference in knowledge bases//Proceedings of the 3rd International Conference on Learning Representations. San Diego, USA, 2015

[41] Trouillon T, Welbl J, Riedel S, et al. Complex embeddings for simple link prediction//Proceedings of the 33rd International Conference on Machine Learning. New York City, USA, 2016: 2071-2080

[42] Balazevic I, Allen C, Hospedales T M. Multi-relational poincaré graph embeddings//Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver, Canada, 2019: 4465-4475

[43] Chami I, Wolf A, Juan D, et al. Low-dimensional hyperbolic knowledge graph embeddings//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Online, 2020: 6901-6914

[44] Wang S, Wei X, Dos Santos C N, et al. Mixed-curvature multi-relational graph neural network for knowledge graph completion//Proceedings of the World Wide Web Conference. Ljubljana, Slovenia, 2021: 1761-1771

[45] Ai Q, Azizi V, Chen X, et al. Learning heterogeneous knowledge base embeddings for explainable recommendation. Algorithms, 2018, 11(9): 137

[46] Wang H, Zhang F, Zhao M, et al. Multi-task feature learning for knowledge graph enhanced recommendation//Proceedings of the World Wide Web Conference. San Francisco, USA, 2019: 2000-2010

[47] Xin X, He X, Zhang Y, et al. Relational collaborative filtering: Modeling multiple item relations for recommendation//Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. Paris, France, 2019: 125-134

[48] Ma W, Zhang M, Cao Y, et al. Jointly learning explainable rules for recommendation with knowledge graph//Proceedings

计算机学报

of the World Wide Web Conference. San Francisco, USA, 2019: 1210-1221

[49] Xian Y, Fu Z, Muthukrishnan S, et al. Reinforcement knowledge graph reasoning for explainable recommendation //Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. Paris, France, 2019: 285-294

[50] Huang X, Fang Q, Qian S, et al. Explainable interaction-driven user modeling over knowledge graph for sequential recommendation//Proceedings of the 27th ACM International Conference on Multimedia. Nice, France, 2019: 548-556

[51] Wang X, Wang D, Xu C, et al. Explainable reasoning over knowledge graphs for recommendation//Proceedings of the 33rd AAAI Conference on Artificial Intelligence. Honolulu, USA, 2019: 5329-5336

[52] Zhao J, Zhou Z, Guan Z, et al. IntentGC: A scalable graph convolution framework fusing heterogeneous information for recommendation//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Anchorage, USA, 2019: 2347-2357

[53] Li Z, Xu Q, Jiang Y, et al. Quaternion-based knowledge graph network for recommendation//Proceedings of the 28th ACM International Conference on Multimedia, USA, 2020: 880-888

[54] Rendle S, Freudenthaler C, Gantner Z, et al. BPR: Bayesian personalized ranking from implicit feedback//Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence. Montreal, Canada, 2009: 452-461

[55] Rendle S, Gantner Z, Freudenthaler C, et al. Fast context-aware recommendations with factorization machines//Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval. Beijing, China, 2011: 635-644

[56] He X, Chua T. Neural factorization machines for sparse predictive analytics//Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. Shinjuku, Tokyo, Japan, 2017: 355-364

[57] Ziegler C, Mcnee S M, Konstan J A, et al. Improving recommendation lists through topic diversification//Proceedings of the 14th International Conference on World Wide Web. Chiba, Japan, 2005: 22-32

[58] Dettmers T, Minervini P, Stenetorp P, et al. Convolutional 2D knowledge graph embeddings//Proceedings of the 33rd AAAI Conference on Artificial Intelligence. New Orleans, USA, 2018: 1811-1818

[59] Toutanova K, Chen D. Observed versus latent features for knowledgebase and text inference//Proceedings of the 3rd Workshop on Continuous Vector Space Models and Their Compositionality. Beijing, China, 2015: 57-66

《计算机学报》

附录 A.

A.1 协同知识图谱构建方法

为实现更为精准的推荐,通常需要借助辅助信息.具体而言,辅助信息由真实世界的实体和它们之间的关系组成.例如,一部电影可以通过其导演、演员和类型来描述.因此辅助信息可由知识图谱的形式呈现,表示为由主体-属性-对象三元组组成的有向图.形式上,给定实体集$\mathcal{E}$和关系集$\mathcal{R}$,知识图谱表示为$(\boldsymbol{h},\boldsymbol{r},\boldsymbol{t})$,其$\boldsymbol{h}\in\mathcal{E}$为头实体,$\boldsymbol{r}\in\mathcal{R}$为关系,$\boldsymbol{t}\in\mathcal{E}$为尾实体.此外,构建商品-实体对齐图$\mathcal{A}=\{(i,e)\mid i\in\mathcal{V},e\in\mathcal{E}\}$,其中$\{i,e\}$表示商品$i$可与知识图谱中的实体$e$进行对齐.

基于以上知识图谱和对齐图,可构建协同知识图谱,它将用户行为与商品的信息编码为一个统一的关系图.首先将每个用户行为表示为三元组$(u,Interact,i)$,其中$\boldsymbol{y}_{ui}=1$表示为用户u和商品i之间存在交互关系.然后基于上述商品-实体对齐图,用户-商品图可以与知识图谱集成为一个统一的协同知识图谱$\mathcal{G}=\{(\boldsymbol{h},\boldsymbol{r},\boldsymbol{t})\mid\boldsymbol{h},\boldsymbol{t}\in\mathcal{E}',\boldsymbol{r}\in\mathcal{R}'\}$,其中$\mathcal{E}'=\mathcal{E}\cup U$和$\mathcal{R}'=\mathcal{R}\cup\{Interact\}$.

A.2 关于 DQKGE 建模多种关系模式的证明

回顾正文 3.3 节部分,本文已经给出了组合、反演和对称关系模式的定义.在证明 DQKGE 可以建模以上多种关系模式之前,首先引入定理 1.

定理 1[25]. 如果$\boldsymbol{S}_1$和$\boldsymbol{S}_2$是两个射影变换τ_1和τ_2对应的矩阵,则它们分别对应两个不同的莫比乌斯变换f_1和f_2.若矩阵$\boldsymbol{S}_3$是由矩阵$\boldsymbol{S}_1$和矩阵$\boldsymbol{S}_2$的乘积产生,那么与矩阵$\boldsymbol{S}_3$相关的τ_3所对应的两个莫比乌斯变换满足$f_3=f_1\circ f_2$.根据定理 1,有如下结论和证明:

(1) DQKGE 可推断组合关系.

证明. 给定实体x,y,z,如果三元组$r_1(x,y)$和$r_2(y,z)$为正样本,则有

$$g_{r_1}(x)=y,\quad g_{r_2}(y)=z \tag{30}$$

需推出$g_{r_3}(x)=z,i=1,\cdots,d$,其中

$$g_{r_i}(h_i)=f(h_i,r_i)=\frac{r_{a_i}h_i+r_{b_i}}{r_{c_i}h_i+r_{d_i}},\ r_{a_i}r_{d_i}-rb_ir_{c_i}\neq 0 \tag{31}$$

从式(30)中,将$g_{r_1}(x)=y$代入$g_{r_2}(y)=z$中,即有$g_{r_2}(g_{r_1}(x))=z$.因此,有

$$g_{r_2}\circ g_{r_1}(x)=z \tag{32}$$

这意味着g_{r_1}和g_{r_2}这两个莫比乌斯变换是可以组合的.考虑以上定理,假设$S_{r_3}=S_{r_1}S_{r_2}$,有$g_{r_2}\circ g_{r_1}(x)=g_{r_3}(x)=z$.这意味着三元组$(x,r_3,z)$存在,因此 DQKGE 可推断组合关系.考虑到实数、虚数、四元数都是对偶四元数的特殊情形,因此可以推断出如果当x,y,z为实数、虚数或者四元数的时候也是成立的,这时只需将对偶四元数相应的虚数部分变为 0 即可. 证毕.

(2) DQKGE 可推断反演关系.

证明. 如果(e_1,r_1,e_2)成立,则有

$$g_{r_1}(x)=y \tag{33}$$

然后需要证明有 $g_{r_2}(y)=x$. 在式(3)中有 $y=g_{r_1}(x)$. 因为 g_{r_1} 是可逆的莫比乌斯变换，因而有 $S_{r_1}=S^{-1}(r_2)$，有如下证明：

$$y=g_{r_2}^{-1}(x) \tag{34}$$

然后有

$$x=g_{r_2}(y) \tag{35}$$

这意味着三元组(y,r_2,x)正确，因此 DQKGE 可推断反演关系. 同理可证，如果将数域从对偶四元数空间退化为四元数、虚数或者实数，以上结论也成立.　　证毕.

(3) DQKGE 可推断对称关系.

证明.　如果三元组(x,r,y)成立，则有

$$g_r(x)=y \tag{36}$$

根据式(2)中，需证 $g_r(y)=x$.

根据式(36)中，有 $y=g_r(x)$，由于 g_r 是莫比乌斯变换，根据定理可以得到 $S_r=S_r^{-1}$，因此有

$$y=g_r^{-1}(x) \tag{37}$$

因而可得

$$x=g_r(y) \tag{38}$$

这意味着三元组(y,r,x)正确，因而 DQKGE 可推断对称关系.

同理可证，如果将数域从对偶四元数空间退化为四元数、虚数或者实数，以上结论也成立.　　证毕.

A.3　DQKGE 与其他知识图谱嵌入模型的关系

在这一节，本文将证明 DQKGE 的模型框架中包含一些常见的知识图谱嵌入模型，如 TransE、RotatE、QuatE 以及 DualE 等.

首先，$\boldsymbol{h}$ 和 $\boldsymbol{t}$ 均为对偶四元数，$\boldsymbol{r}$ 为对偶四元数矩阵，有

$$\boldsymbol{r}=\begin{bmatrix} r_a & r_b \\ r_c & r_d \end{bmatrix}\cdot r_a=a_0+a_1\boldsymbol{i}+a_2\boldsymbol{j}+a_3\boldsymbol{k}+(a_0'+a_1'\boldsymbol{i}+a_2'\boldsymbol{j}+a_3'\boldsymbol{k})\epsilon,$$

其中 ϵ 部分被称为对偶部分，$\boldsymbol{i},\boldsymbol{j}$ 和 $\boldsymbol{k}$ 分别为三个虚数部分. r_b、r_c 和 r_d 形式与 r_a 类似. 注意到 DQKGE 的得分函数为

$$g(h,r,t)=\mathrm{Re}(\langle h_r,t\rangle) \tag{39}$$

其中

$$h_r=[h_{r_1},\cdots,h_{r_d}],\ h_{r_i}=f_r(h_i)=S_{r_i}[h_i,1]^{\mathrm{T}},\quad i=1,\cdots,d \tag{40}$$

(1) TransE 是 DQKGE 的特例.

证明.　根据式(40)，有

$$h_{r_i}=g_r(h_i)=S_{r_i}[h_i,1]^{\mathrm{T}},\ i=1,\cdots,d,$$

其中 $S=\begin{bmatrix} r_{a_i} & r_{b_i} \\ r_{c_i} & r_{d_i} \end{bmatrix}$. 令 r_{a_i}、r_{b_i} 和 r_{d_i} 对偶四元数的对偶部分以及所有虚数部分为零，r_{c_i} 为 0，且令 r_{a_i} 和 r_{b_i} 的实数部分为 1，于是有

$$h_{r_i}=g_r(h_i)=S_r[h_i,1]^{\mathrm{T}}=\begin{bmatrix} 1 & r_{b_i} \\ 0 & 1 \end{bmatrix}[h_i,1]^{\mathrm{T}}$$
$$=[h_i+r_{b_i},1],\ i=1,\cdots,d \tag{41}$$

通过以上变换，可以得到 $h_{r_i}=h_i+r_{b_i}$. 因此可证得 TransE 是 DQKGE 的特例.　　证毕.

(2) RotatE 是 DQKGE 的特例.

证明.　根据式(40)，有

$$h_{r_i}=g_r(h_i)=S_{r_i}[h_i,1]^{\mathrm{T}},\ i=1,\cdots,d,$$

其中 $S=\begin{bmatrix} r_{a_i} & r_{b_i} \\ r_{c_i} & r_{d_i} \end{bmatrix}$. 令 r_{c_i} 和 r_{b_i} 为零，r_{d_i} 的实数部分为 1，其余部分为零，令$|r_{a_i}|=1$ 且其对偶部分以及 $\boldsymbol{j}$ 和 $\boldsymbol{k}$ 虚数部分为零，有

$$h_{r_i}=g_r(h_i)=S_r[h_i,1]^{\mathrm{T}}=\begin{bmatrix} r_{a_i} & 0 \\ 0 & 1 \end{bmatrix}[h_i,1]^{\mathrm{T}}$$
$$=[h_i r_{b_i},1],\ i=1,\cdots,d \tag{42}$$

因此 $g(h,r,t)=-\|h_r-t\|=\|h\circ r_a-t\|$，因此证得 RotatE 为 DQKGE 的特例.　　证毕.

(3) QuatE 为 DQKGE 的特例.

证明.　根据式(40)，有

$$h_{r_i}=g_r(h_i)=S_{r_i}[h_i,1]^{\mathrm{T}},\ i=1,\cdots,d.$$

令 r_{c_i} 和 r_{b_i} 为零，令$|r_{a_i}|=1$ 且其对偶部分为零，令 r_{d_i} 的实数部分为 1，其余部分为零，有

$$h_{r_i}=g_r(h_i)=S_r[h_i,1]^{\mathrm{T}}=\begin{bmatrix} r_{a_i} & 0 \\ 0 & 1 \end{bmatrix}[h_i,1]^{\mathrm{T}}$$
$$=[h_i r_{b_i},1],\ i=1,\cdots,d \tag{43}$$

因此 $g(h,r,t)=h_r$，$t=\langle h\otimes r,t\rangle$，其中“$\otimes$”为四元数乘法. 因此证得 QuatE 为 DQKGE 的特例.　　证毕.

(4) DualE 为 DQKGE 的特例.

证明.　根据式(40)，有

$$h_{r_i}=g_r(h_i)=S_{r_i}[h_i,1]^{\mathrm{T}},\ i=1,\cdots,d.$$

令 r_{c_i} 和 r_{b_i} 为零，令$|r_{a_i}|=1$，r_d 的实数部分为 1，其余部分为零，有

$$h_{r_i}=g_r(h_i)=S_r[h_i,1]^{\mathrm{T}}=\begin{bmatrix} r_{a_i} & 0 \\ 0 & 1 \end{bmatrix}[h_i,1]^{\mathrm{T}}$$
$$=[h_i r_{b_i},1],\ i=1,\cdots,d \tag{44}$$

因此 $g(h,r,t)=\langle h_r,t\rangle=\langle h*r,t\rangle$，其中“$*$”表示对偶四元数乘法. 因此证得 QuatE 为 DQKGE 的特例.　　证毕.

表 11　几个流行知识图谱嵌入模型的复杂度比较

模型	时间复杂度	空间复杂度
TransE	$\mathcal{O}(k)$	$\mathcal{O}(k)$
HolE	$\mathcal{O}(k\log(k))$	$\mathcal{O}(k)$
ComplEx	$\mathcal{O}(k)$	$\mathcal{O}(k)$
QuatE	$\mathcal{O}(k)$	$\mathcal{O}(k)$
DQKGE	$\mathcal{O}(k)$	$\mathcal{O}(k)$

A.4　链接预测实验

为了验证 DQKGR 推荐模型中 DQKGE 嵌入学习部分的有效性，本文在两个常用的知识图谱嵌入数据集 WN18RR[58] 和 FB15K237[59] 上进行了链接预测实验. WN18RR 是 WN18[17] 数据集的子集，里面包含对称、反对称以及组合关系. FB15K237 是 FB15K[17] 的子集，其主要包含反对称、对称以及组合关系. 遵循知识图谱中常用的评估方法，本文采用了平均倒数排名(MRR)，命中率($Hits@n=1,3,10$)作为评价指标. 本文将 DQKGE 与一些流行的方法如 TransE、RotatE、TuckER、

《计算机学报》

ComplEx、QuatE、DualE、ConvE 以及 MuRP 进行对比.

结果分析:不同知识图谱嵌入方法在 WN18RR 数据集和 FB15K-237 数据集上的结果如表 12 和表 13 所示.可以看到,所提 DQKGE 模型在两个数据集上均获得了最优性能.其中:

(1) WN18RR 数据集中包含较多的层次结构,这对模型的链接预测任务构成了挑战.MuRP 模型基于双曲几何建模,因此它可以较好地对该数据集建模.同时,DQKGE 在该数据集上也取得了最优性能,验证了通过射影几何将关系建模为多种变换的有效性.

(2) FB15K237 数据集包含多种复杂关系,TransE 将知识图谱中的关系建模为平移,RotatE 和 QuatE 将关系分别建模为二维空间和三维空间的旋转,当面对复杂关系时,这类方法的结果显示将关系建模为单一变换会使得模型性能较弱.DualE 将关系建模为平移和旋转的组合,模型的性能得到提升.而 DQKGE 通过射影空间将关系建模为多种变换,进一步增强了模型对数据集中复杂关系的建模能力,因而 DQKGE 模型在该数据集上表现较好.综上所述,DQKGE 方法利用对偶四元数乘法,可以充分地捕捉空间中实体的交互,挖掘潜在关系;其利用射影空间将关系建模为空间中包括旋转、平移、反射、反演以及同态等多种变换,可以提升模型的表达能力,因而在以上数据集上表现优异.通过 DQKGE 对辅助信息进行建模,可以为后续的推荐过程提供更为精准的嵌入,因而使得整个 DQKGR 模型在推荐任务上取得良好的性能.

表 12 WN18RR 数据集上,常见知识图谱嵌入模型与 DQKGE 模型的性能比较(其中最优结果用粗体表示,次优结果用下划线表示)

方法	*MRR*↑	*Hits*@1↑	*Hits*@3↑	*Hits*@10↑
TransE	0.226	—	—	0.501
RotatE	0.476	0.428	0.492	0.571
TuckEr	0.470	0.443	0.482	0.526
ComplEx-N3	0.480	0.435	0.495	0.572
QuatE	0.491	0.441	0.51	0.579
DualE	0.492	0.444	0.513	0.584
ConvE	0.430	0.400	0.440	0.520
MuRP	0.481	0.440	0.495	0.566
DQKGE	**0.495**	**0.445**	**0.515**	**0.589**

表 13 FB15K237 数据集上,常见知识图谱嵌入模型与 DQKGE 模型的性能比较(其中最优结果用粗体表示,次优结果用下划线表示)

方法	*MRR*↑	*Hits*@1↑	*Hits*@3↑	*Hits*@10↑
TransE	0.294	0.165	0.306	0.465
RotatE	0.338	0.241	0.375	0.533
TuckEr	0.358	0.266	0.394	0.544
ComplEx-N3	0.357	0.264	0.392	0.547
QuatE	0.348	0.248	0.382	0.55
DualE	0.361	0.268	0.395	0.558
ConvE	0.325	0.237	0.356	0.501
MuRP	0.335	0.243	0.367	0.518
DQKGE	**0.365**	**0.273**	**0.400**	**0.564**

相对于传统的知识图谱嵌入表示,对偶四元数引入了更高的数据维度,本文对该部分进行了时空复杂度的理论分析以及实验评估.

(1) 表 11 对 DQKGE 与其他知识图谱嵌入模型进行了对比,可以看到其时间复杂度和空间复杂度均为 $\mathcal{O}(k)$.具体使用中,复杂度可能会再乘以一个常系数,但理论上模型的时空复杂度也均为线性.

(2) 在链接预测实验上,本文对 DQKGE 与多种模型所需的参数进行了对比.由表 14 可以看到当各个模型达到最优性能时,DQKGE 在 WN18RR 数据集上所需的参数量最少;在 FB15K237 数据集上所需的参数量也相对较少.这表明 DQKGE 的空间复杂度并不高.本文继续对不同知识图谱模型训练收敛所需的时间进行比较,由表 15,与其他对比方法相比,可以看到 DQKGE 在 WN18RR 数据集和 FB15K237 数据集上均能以较短的时间使得训练收敛.其原因在于 DQKGE 中的对偶四元数运算和莫比乌斯变换可以较好地挖掘知识图谱中实体之间的潜在关系,从而加快模型的收敛.

表 14 DQKGE 与其他知识图谱嵌入模型关于参数量的对比实验

数据集	MuRP/M	RotatE/M	QuatE/M	DQKGE/M
WN18RR	34.52	40.95	16.38	15.26
FB15K237	23.18	29.32	5.82	6.43

表 15 DQKGE 与其他知识图谱嵌入模型关于训练收敛时间的对比实验

数据集	指标	MuRP	RotatE	QuatE	DQKGE
WN18RR	迭代次数	800	1000	500	200
	总时间/h	5.1	3.7	3.2	2.8
FB15K237	迭代次数	1000	1500	1000	500
	总时间/h	8.3	5.6	4.8	3.9

A.5 实验结果补充分析

在实验部分,本文测试了多个模型在三个数据集 Last-FM、MovieLens-20M、Book-Crossing 的性能.可以看到 DQKGR 超越了其他模型,性能均优于基准方法.本文中所用的实验设置,包括语义抽取本身,都与流行的方法[14-15,53]相同.但也可以看到在性能结果中,目前主流研究方法的性能指标在这些数据集上也都不足 50%.其原因主要分为两部分:(1) 推荐问题中所用的数据集都是大数据集,数据存在很大的稀疏性,且目前的研究方法的性能指标也都不足 50%;(2) 由于本文评价指标 $HR@n$、$P@n$、$R@n$、$NDCG@n$ 中,n 选取 20,这对模型的要求比较严格,如果选择 $n=100$,如表 16 和表 17 模型的命中率均可超越 50%,召回率也可较 $n=20$ 时有显著提升.

表 16 $n=100$ 时,不同模型在 Last-FM 数据集的实验结果

模型	*HR*@100	*P*@100	*R*@100	*NDCG*@100
KGCN	0.6451	0.0108	0.1832	0.1145
KGAT	0.6488	0.0154	0.1845	0.1189
QKGN	0.6482	0.0163	0.1855	0.1197
DQKGR	**0.6625**	**0.0214**	**0.2057**	**0.1310**

《计算机学报》

表 17　$n=100$ 时，不同模型在 MovieLens-20M 数据集的实验结果

模型	*HR*@100	*P*@100	*R*@100	*NDCG*@100
KGCN	0.8906	0.0315	0.3825	0.1746
KGAT	0.8978	0.0347	0.3941	0.1795
QKGN	0.8983	0.0358	0.3958	0.1798
DQKGR	0.9150	0.0364	0.4183	0.1932

此外，本文进行显著性实验，以验证模型性能的提升. 对于实验中使用的每个指标，将 KGAT 和 DQKGR 两个模型在三个数据集上的结果分别作为对比变量，采取原假设为两个变量不相关，显著性水平 α 设置为 0.01，结果如表 18 所示. 可以看到 p 值均大于显著性水平 0.01，因此接受原假设，认为 DQKGR 相对 KGAT 性能的提升存在显著性差异. 类似地，本文对 DQKGR 与 QKGN 在四个指标上做显著性检验，结果如表 19 所示. 可以看到 p 值均大于显著性水平 0.01，因此接受原假设，认为 DQKGR 相对 QKGN 性能的提升存在显著性差异.

表 18　DQKGR 与 KGAT 性能的显著性验证实验

指标	p 值
HR@20	0.0244
P@20	0.0161
R@20	0.0211
NDCG@20	0.0596

表 19　DQKGR 与 QKGN 性能的显著性验证实验

指标	p 值
HR@20	0.0144
P@20	0.0188
R@20	0.0126
NDCG@20	0.0128

A.6　模型其他关键参数的研究

为探究本文中其他关键参数对模型性能的影响，本文继续研究了模型学习率对模型性能的影响，以及对偶四元数协同知识图谱嵌入损失项权重 λ_1 和正则化权重 λ_2 对模型性能的影响.

(1) 学习率对模型性能的影响. 在 Last-FM 数据集进行实验，结果如表 20 所示，学习率取{1e-1，5e-2，1e-2，5e-3，1e-3，5e-4，1e-4}，可以看到不同学习率对模型性能的影响不大，在正常范围内波动.

表 20　模型学习率对模型的影响

学习率	*HR*@20	*P*@20	*R*@20	*NDCG*@20
1e-1	0.3741	0.0321	0.0834	0.0785
5e-2	0.3797	0.0339	0.0921	0.0821
1e-2	0.3814	0.0348	0.0978	0.0847
5e-3	0.3854	0.0364	0.1022	0.0861
1e-3	**0.3873**	**0.0376**	**0.1031**	**0.0874**
5e-4	0.3871	0.0374	0.1029	0.0872
1e-4	0.3869	0.0372	0.1027	0.0870

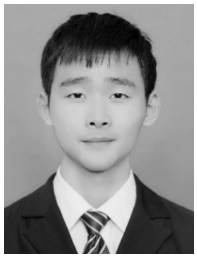
(2) 不同权重对模型性能的影响. 为探究对偶四元数协同知识图谱嵌入项损失权重 λ_1 对模型性能的影响，在 Last-FM 数据集进行实验，分别取 λ_1 为{1e-1，5e-2，1e-2，5e-3，1e-3，5e-4，1e-4}，结果如表 21 所示. 为探究正则化权重 λ_2 对模型性能的影响，在 Last-FM 数据集进行实验，分别取 λ_2 为{1e-1，5e-2，1e-2，5e-3，1e-3，5e-4，1e-4}，结果如表 22 所示. 综上，可以看出相比于 λ_2，λ_1 对模型性能的影响更大，这也在一定程度上说明了利用协同知识图谱进行嵌入学习的重要性.

表 21　λ_1 对模型性能的影响

λ_1	*HR*@20	*P*@20	*R*@20	*NDCG*@20
1e-1	**0.3873**	**0.0376**	**0.1031**	**0.0873**
5e-2	0.3851	0.0373	0.1022	0.0850
1e-2	0.3826	0.0371	0.1008	0.0824
5e-3	0.3786	0.0368	0.0987	0.0803
1e-3	0.3736	0.0365	0.0952	0.0797
5e-4	0.3687	0.0361	0.0943	0.0789
1e-4	0.3645	0.0359	0.0934	0.0782

表 22　λ_2 对模型性能的影响

λ_2	*HR*@20	*P*@20	*R*@20	*NDCG*@20
1e-1	0.3755	0.0348	0.0914	0.0763
5e-2	0.3786	0.0351	0.0934	0.0779
1e-2	0.3814	0.0356	0.0964	0.0821
5e-3	0.3856	0.0365	0.1003	0.0857
1e-3	**0.3873**	**0.0376**	**0.1031**	**0.0873**
5e-4	0.3867	0.0373	0.1023	0.0865
1e-4	0.3862	0.0369	0.1019	0.0861



CAO Zong-Sheng, Ph. D. candidate. His research interests lie in recommender systems and knowledge graphembeddings.

XU Qian-Qian, Ph. D., associated professor. Her research interests include statistical machine learning, and its applications in multimedia and computer vision.

LI Zhao-Peng, M. S. His research interests lie in recommender systems and computer vision.

JIANG Yangbangyan, Ph. D. candidate. Her research interests include machine learning and computer vision.

CAO Xiao-Chun, Ph. D., professor. His main research interests include computer vision and multimedia analysis.

HUANG Qing-Ming, Ph. D., chair professor. His research interests include multimedia computing, image processing, computer vision, and pattern recognition.

《计算机学报》

Background

With the rapid development of the Internet, users are facing a growing amount of online information. To alleviate the information explosion problem and improve user experience, recommender system, as an effective tool to provide users with the personalized suggestion, has been applied in the various online software. However, traditional recommendation algorithms suffer from data sparsity and cold start problems. To alleviate the above issues, researchers combine side information of users or items with recommender algorithms, which effectively improves recommendation performance.

In recent years, recommender algorithms using knowledge graph (KG) as side information have drawn much attention. As a structured semantic knowledge base, knowledge graph brings rich semantic information and structural connectivity information to recommender system, which is conducive to improving the accuracy of the personalized recommendation. Most existing KG-based recommendation models use a single transformation such as a translation-based algorithm to learn representations from KG for the recommendation, which could not model all the three fundamental patterns in KG. Besides, it is common for most models to adopt real-valued embeddings to represent users and items, which are of less representation capacity and prone to overfitting with the growing number of parameters.

To solve the above problems, this paper introduces dual quaternion embeddings to represent users and items and proposes dual quaternion-based collaborative knowledge graph modeling for recommendation (DQKGR). The model applies a semantic matching principle that embraces Mobius transformations to learn expressive representations from the collaborative knowledge graph and adopts a customized preference propagation and aggregation mechanism to attentively update those representations. The experimental results on three public datasets show the effectiveness of our proposed model.

This work was supported in part by the National Key R&D Program of China under Grant (No. 2018AAA0102000), in part by the National Natural Science Foundation of China (Nos. U21B2038, 61931008, 6212200758, 61976202), in part by the Fundamental Research Funds for the Central Universities, in part by the Strategic Priority Research Program of Chinese Academy of Sciences (No. XDB28000000), in part by the Youth Innovation Promotion Association CAS.

《计算机学报》