为了解决信息爆炸问题并增强各种在线应用程序中的用户体验，已经开发了推荐系统来对用户的偏好进行建模。尽管为做出更具个性化的推荐做出了许多努力，但是推荐系统仍然遭受数个挑战，例如数据稀疏和冷启动问题。近年来，以知识图作为辅助信息来生成推荐引起了人们的极大兴趣。这样的方法不仅可以减轻上述问题以获得更准确的推荐，而且可以为推荐项目提供解释。因此, 本文针对基于知识图谱的推荐系统这一领域进行了技术评估。具体地，首先简单介绍推荐系统与知识图谱中的一些基本概念与研究背景。随后，介绍基于图嵌入的推荐方法、基于路径的推荐方法以及基于联合的推荐方法的研究现状与发展趋势。最后，给出了对基于知识图谱的推荐系统技术的评价与建议。

关键词：推荐系统，知识图谱，协同过滤，图嵌入，元路径

随着互联网的飞速发展，数据量呈指数增长。由于信息过多，用户很难在众多选择中挑选出他们感兴趣的东西。推荐系统作为一种有效地解决信息泛滥的方法受到了大量国内、国外研究者的广泛研究。其核心目标是通过分析用户行为、兴趣、需求等信息，在海量的数据中挖掘用户感兴趣的信息, 如商品、新闻等。目前，推荐系统在各种各样的信息系统中都扮演着重要的角色，用来促进商务业务发展以及辅助决策过程，并且广泛地应用于各大电子商务（淘宝、Amazon等）和多媒体（快手、豆瓣）等网站中。

个性化推荐算法是推荐系统的核心，其主要可以被分为 3 类，即基于内容的推荐算法、基于协同过滤的推荐算法和混合推荐算法。其中，协同过滤方法通过利用用户历史的行为偏好数据构建模型，取得了巨大的成功。这种方法的优势在于不需要类似基于内容的推荐算法那样对物品进行复杂的特征提取与建模。尽管这类协同过滤方法通常是有效且普适的，但其依旧存在着多种问题。主要包括用户和物品之间的行为关系数据的稀疏问题和对新用户或者新物品进行推荐时存在的冷启动问题。为此，研究者尝试将协同过滤推荐算法和其他辅助信息相结合（例如, 用户与物品的属性特征、用户社交网络信息等）搭建混合推荐系统来解决以上问题，从而提升推荐效果。

2012年5月17日谷歌正式提出了“知识图谱”这个术语。知识图谱旨在描述真实世界中存在的各种实体或概念，以及他们之间的关联关系。其中，每个实体或概念用一个全局唯一确定的ID来标识，每个（属性一值）对用来刻画实体的内在属性，而关系用来连接两个实体，刻画它们之间的关联。例如三元组（Donald Trump，president of，America）表示 “Donald Trump”是“America”的总统，其中Donald Trump是头实体，America是尾实体， president of是两者的关系名称，（Donald Trump，born in，New York）和（Donald Trump，nationality，America）存储实体Donald Trump的两个属性，即出生地和国籍，这时尾实体也被称作属性值。此外，在特定场景下知识图谱的内容会频繁地动态更改，最近也有学者尝试构建动态知识图谱，将原有的三元组变更为包含时间戳的四元组，从而进一步建模知识的演化性。部分常见的知识图谱, 其中包括诸如YAGO KG，DBpediaKG，Freebase KG和OwnThink KG等通用知识图谱，也有如WordNet，UMLS KG，Douban’s movie KG和MusicBrainz等特定领域的知识图谱。这类特定领域下的知识图谱往往拥有比通用化知识图谱更完整的领域内实体和关系。此外，知识图谱可以融合多种数据源丰富数据语义信息，并且可以结合推理得到的隐含信息为用户提供服务。随着信息检索、智能问答等应用领域的发展要求，将知识图谱应用于这些领域以提高用户体验和系统性能，成为学术界和工业界关注的热点。

（1）基于嵌入的方法：大多数基于嵌入的方法构建具有多种类型的项目边信息的KG来丰富项目的表示形式，并且此类信息可用于更精确地建模用户表示形式。一些模型通过将用户引入图表来构建用户项目图，从而可以直接对用户偏好进行建模。实体嵌入是基于嵌入的方法的核心，一些论文使用GAN 或BEM 完善了嵌入以提供更好的建议。基于嵌入的方法本质上利用了图结构中的信息。文献应用多任务学习策略来联合训练推荐模块以及与图相关的任务，以提高推荐质量。

（2）基于路径的方法：基于路径的方法基于知识图谱生成推荐，并且在过去，这种方法也称为基于HIN的推荐。传统的基于路径的方法通常将MF与HIN中提取的元路径集成在一起。 这些方法利用路径连通性来规范化或丰富用户和/或项目表示。 这些方法的缺点是它们通常需要领域知识来定义元路径的类型和数量。随着深度学习技术的发展，已经提出了不同的模型来对嵌入的路径进行编码。可以使用路径嵌入或通过发现连接用户项对的最显着路径来生成建议。基于路径的方法自然会将可解释性带入了推荐过程。对于传统的基于路径的方法，其动机是在元路径级别上匹配项目或用户的相似性。推荐结果可以从预定义的元路径中找到参考。由于规则和相应的权重是明确的，因此推荐原因也可供用户使用。最近的工作利用深度学习模型为用户-项目对自动挖掘显着路径，这在图中反映了推荐过程。

（3）两者联合的方法：基于嵌入的方法利用KG中用户、项目的语义表示来进行推荐，而基于路径的方法则使用语义连接信息，并且这两种方法都仅利用图形中信息的一个方面。为了充分利用KG中的信息以获得更好的建议，已经提出了将实体和关系的语义表示以及连接性信息集成在一起的统一方法。统一方法基于嵌入传播的思想。这些方法在KG中的连接结构的指导下完善了实体表示。在获得用户和潜在项目的丰富表示之后，可以预测用户的偏好。统一方法既受益于KG的语义嵌入，又受益于语义路径模式。这些方法利用嵌入传播的思想来完善项目或用户在KG中具有多跳邻居的表示。这些工作通常采用自然适合嵌入传播过程的基于GNN的体系结构，并且自RippleNet于2018年提出以来，此类方法已成为一种新的研究趋势。统一方法继承了基于路径的方法的可解释性。传播过程可以视为在KG中发现用户的偏好模式，这类似于在基于路径的方法中发现连接模式。

基于嵌入的方法使用KGE方法预处理知识图谱以获取实体和关系的嵌入，并将其进一步集成到推荐框架中。但是，这种方法忽略了图中的信息连通性模式，很少有工作可以提供推荐结果并给出理由。基于路径的方法通过预定义元路径或自动挖掘连接模式，利用用户项图来发现项目的路径级相似性。基于路径的方法还可以为用户提供结果说明。最近的研究趋势是统一基于嵌入的方法和基于路径的方法，以充分利用双方的信息。而且，统一的方法也可以解释推荐过程。

现有推荐系统技术，尤其是基于协同过滤的推荐方法一直面临着一些严峻的问题，包括用户-物品的交互数据稀疏问题和冷启动问题，会导致推荐结果不准确。知识图谱作为混合推荐系统中一类有效的辅助信息，近年来受到了大量研究者的关注。知识图谱是一种有向信息异构网络。其包含实体与关系，因而蕴含了推荐系统中对于物品的大量背景信息，以及物品之间的关系。并且它可以与用户行为数据构成的用户-物品网络集成起来，从而扩展了用户与商品之间存在的隐藏的关联关系，补充了用户与物品交互数据，因此可以更进一步地提升推荐效果以及缓解、解决上面两个问题。具体地，一些研究通过对知识图谱中的实体和关系进行表征，进而将知识图谱中蕴含的语义信息带入到推荐系统的用户和物品上，从而提升推荐系统的效果。另外，一些研究者通过将知识图谱和原有的用户-物品的交互数据构成一个异质信息网络图，利用挖掘图上用户到物品之间不同的路径，提升推荐效果。这样的异构信息网络图可以帮助我们挖掘大量源数据集中未出现的用户-物品交互的关系，解决数据稀疏与冷启动的问题。

此外，近年来构建可解释性推荐系统成为趋势。可解释性推荐算法的目的是在给用户合适地推荐物品的同时，给出推荐的原因使得用户更易理解系统为什么会做出这样的推荐。研究发现给出推荐的原因会使得用户的信任度和满意度提升，并且更易于使用户选择被推荐的物品。知识图谱作为非常容易理解的外部知识库可以帮助我们将用户、物品之间构建合理的关联，从而可以帮助我们构建可解释的推荐系统，其吸引了研究者的广泛关注。

目前，学术界与工业界的研究者已经构建了一系列知识图谱，例如DBpedia KG、AceKG、Microsoft Satori、百度知识图谱和 OwnThink KG等。并且覆盖了不同的领域，如学术、历史人物、影视、音乐等。因此，为我们在不同应用场景中基于知识图谱构建推荐系统提供了充足的数据基础。用户辅助信息：基于知识图谱的推荐系统大多是通过加入项目辅助信息来构建知识图谱，很少有模型考虑用户辅助信息。然而，如用户社交网络等信息也可以自然地整合到当前基于 KG 的推荐系统框架中。因此，在知识图谱中考虑用户辅助信息可以是另一个研究方向。