**（一）基于知识图谱的推荐技术研究及应用**

**Research and Application of Recommendation Technology Based on Knowledge Graph**

**--Yiming Wang（中国知网 2018）**

**一、科学问题**

**1.1 本文所涉及科学问题**

文章涉及到知识图谱的构建以及个性化推荐系统建立等问题，提出了一种基于知识图谱的个性化推荐算法，其中包含了知识嵌入、协同过滤以及稀疏性问题等。

**1.2 同行专家如何解决**

近几年，随着知识图谱技术和推荐技术的蓬勃发展，两种领域的相关技术融合已经成为了一种趋势。Blanco通过将语义网与基于内容的推荐算法结合，挖掘了大量关于用户偏好的知识，利用语义推理技术克服了推荐的产品过于相似的问题，从而提高了推荐系统的准确性。Gemmis等基于语义网和用户对物品的标注，利用机器学习技术推断用户兴趣，并在文化遗产的个性化推荐中表现出了较高的预测精度Noia等在推荐技术中利用了DBpedia、Freebas等开放链接数据（Linked Open Data），并通过实验计算其准确性和召回率等指标，验证了这种方法的有效性。

**1.3 本文所解决的问题**

文章对传统的推荐系统面临的稀疏性问题以及冷启动问题进行思考，以基于知识图谱的推荐技术研究及应用为研究课题，重点研究了领域内知识图谱的构建技术，知识图谱中实体和关系的嵌入方法以及基于知识图谱的个性化推荐算法，以此来提高推荐系统的性能，然后提出并实现了一种电影领域知识图谱中实体和关系的向量化方法。

**1.4 本文解决方案效果**

文章首先研究了领域知识图谱的构建方法，设计并实现了一种电影领域的知识图谱构建方法。将知识图谱概念引入到电影领域，实现了领域内实体和关系的划分，在此基础上进行了电影领域本体库的构建，然后抽取了电影领域的实体和关系，并且完成了基于关系数据库的知识图谱的存储，最后利用 Neo4j 改进了图谱的存储方式，改善了知识图谱的性能和实用性。

然后文章提出并实现了一种电影领域知识图谱中实体和关系的向量化方法。通过这种实体和关系的嵌入模型，在保留语义信息的基础上，将实体和关系进行向量化表示。通过实验证明了本文的知识表示方法对知识图谱中实体和关系的向量化表示是成功的。

最后文章提出了一种基于知识图谱的个性化推荐算法。将知识图谱中物品的向量化表示方法和协同过滤算法结合，利用从知识图谱中提取的物品语义信息，弥补基于物品的协同过滤算法没有考虑物品本身内容信息的缺陷。最后文章通过实验证明了本文提出的 KG-CF 模型是可行的。

**二、研究内容**

**2.1 理论与方法介绍**

基于知识图谱的个性化推荐技术主要涉及知识图谱的构建以及个性化推荐技术。

（1）知识图谱构建技术

维基百科对知识图谱给出的解释是，知识图谱是一个谷歌使用的，用从各种来源收集信息用于增强其搜索引擎结果的知识库。相关文献提出，知识图谱是由一系列相互连接的实体及其属性构成。本质上，知识图谱是描述实体和实体关系的一种语义网。根据不同的构建需求和应用场景，知识图谱的构建可能涉及不同的技术，包括知识建模、知识抽取、知识表示、知识融合、知识推理等。

知识建模是知识图谱构建的基石，知识建模通过将知识图谱中的知识实体抽象成概念和关系规范知识图谱中知识的表示；

知识抽取指从不同来源、不同结构的数据中进行知识提取，形成知识的过程，知识抽取往往采用一些自动化的抽取方法从结构化、半结构化和非结构化的信息源中提取出实体、关系、属性等信息，形成三元组关系或者多元关系，知识抽取的关键技术包括实体抽取、关系抽取和属性抽取；

知识表示研究怎样用计算机符号来表示人脑中的知识，以及怎样通过符号之间的运算来模拟人脑的推理过程；

知识图谱中存储的知识是一个规模庞大的关联集合。原始数据中杂乱的信息经过前期的抽取与处理，形成了有序且互相关联的知识。这些知识按照其类别以规范化的形式存储在知识图谱中的不同模块中，且为了智能高效地检索及进行知识的深度挖掘，还需要生成相应的索引。知识比数据的结构更加复杂，知识的存储需要综合考虑图的特点、复杂的知识结构存储、索引和查询的优化等，并且要支持推理。目前，知识图谱的存储结构设计没有统一的标准。典型的知识存储引擎分为基于关系数据库的存储和基于原生图的存储。原生图的存储主要基于图数据库实现。

（2）个性化推荐技术

个性化推荐技术是一门交叉学科，通常来讲，可以将主流的个性化推荐算法分为基于内容的推荐算法和协同过滤的推荐算法。

文章在介绍相关的推荐算法之前，首先介绍相似性度量的方法。在基于协同过滤的推荐算法中，我们需要利用对象相似性度量的方法计算用户相似度以及物品相似度，其中包括基于距离的闵可夫斯基（Minkowski）距离相似性度量方法以及基于相关系数的皮尔森（Pearson）相关系数协方差度量方法。

基于用户或物品的协同过滤算法是一种基于邻域的算法，主要根据所有用户对物品的行为记录来计算用户之间或者物品之间的相似度，然后根据这些相似用户的历史行为将物品推荐给该用户。本质上是采用统计方法，直接根据用户-物品矩阵寻找近邻并进行个性化推荐，主要用于用户兴趣情况比较稳定的情况。

协同过滤算法主要利用用户对物品的行为记录，忽略了用户本身的信息，如用户的年龄、性别、职业、地区等以及物品本身的信息，如物品的类型、生产商等。基于内容的推荐算法首先收集和标注特征信息，从而构建用户和物品的画像（Profile），然后通过用户画像和物品画像的特征匹配算法实现个性化推荐。并且基于内容的推荐算法可以有效解决数据稀疏的问题。

**2.2 验证分析与实验效果**

在设计并实现了一种电影领域的知识图谱构建中，我们已经在本体库和抽取的知识之上完成了图谱的搭建工作。知识图谱中的节点类型及关系类型蕴含了本体库的设计思路，图谱中的实体信息和实体间的语义信息展现了抽取后的电影知识。构建完成的知识图谱节点和关系类型设计合理，知识准确且全面，可用于电影知识的检索。因此，本文所提出的电影领域本体库的构建方法是可行的；基于结构化数据抽取的知识准确性高，覆盖面广；基于 Neo4j构建的知识图谱能形象地展示图谱中的实体和实体间关系信息，并且提供了良好的交互方式。实验结果显示，对于电影领域知识图谱中的实体和关系，嵌入到维度为 200的向量空间是一个较好的选择，并且改进后负采样算法 bern 能够改善模型的性能。因此，本章给出的的知识表示方法在保留了语义知识的基础上实现了实体和关系的较为准确的嵌入，能够为后续的物品推荐打下基础。

而关于一种基于知识图谱的个性化推荐算法中，为了检测推荐模型的的性能，选择简单交叉验证法在原始数据集中随机选取一部分样本作为训练集，其余的部分作为实验的测试集。从实验结果中可以看出，推荐结果的覆盖率随着算法融合度的增加不断上升。在融合度达到1，即代表仅使用基于知识图谱的物品相似度进行推荐时，推荐结果的覆盖率达到最高。实验结果显示，融合度为0.6的KG-CF算法在测试集上也能表现出良好的效果。

**三、论文存在问题及后续研究重点**

**3.1 论文存在问题**

本文在构建电影邻域知识图谱时，电影知识图谱中实体构建不够全面。在对知识图谱进行知识表示时，考虑了电影实体的关系属性，但没有考虑其数据属性如发行时间和剧情简介，这可能会影响推荐系统对用户兴趣的挖掘，在后续的工作中，我们可以改进物品向量化表示的模型，在向量化表示时考虑更多的物品内容要素。

**3.2 后续研究重点**

近几年推荐系统的研究和应用正在蓬勃发展，构建知识图谱的技术已经愈加成熟，而知识图谱中蕴含的丰富的语义信息能极大地帮助推荐系统刻画用户和物品画像，提高推荐系统的性能。因此，作者认为知识图谱和推荐技术的结合将成为推荐系统研究的一个趋势。

**四、该问题相关研究成果**

**4.1 相关论文一**

**（1）题目**：《基于知识图谱表示学习的推荐系统设计与实现》

**（2）作者介绍**：姚胜，东南大学工程硕士。

**（3）摘要**: 随着互联网技术的飞速发展，网络已经成为人们获取信息的重要途径。如今网络上的资源呈指数增长，导致人们面临着信息过载（Information Overload）的问题，推荐便是解决这一问题的有力途径。协同过滤算法在目前的研究和应用中取得了巨大的成功，然而，当用户历史数据较稀疏时，该模型的性能会受到限制。为了解决这一问题，本文提出引入文本信息和知识图谱中结构化知识的推荐算法。以协同过滤算法为基础，以深度学习技术为非结构化信息的提取方法，以表示学习技术为知识图谱中结构化信息的提取方法，提出了基于深度学习的文本摘要算法和知识图谱表示学习算法的协同过滤推荐模型。基于上述工作，设计并实现了一个在线论文推荐系统。

**4.2 相关论文二**

**（1）题目**：《知识图谱研究综述》

**（2）作者介绍**：李涓子（1964－），女，教授，博导，主要研究方向为语义web，新闻与社会挖掘。

**（3）摘要**：知识图谱以结构化的方式描述客观世界中概念、实体及其间的关系，将互联网的信息表达成更接近人类认知世界的形式，提供了一种更好地组织、管理和理解互联网海量信息的能力。知识图谱与大数据和深度学习一起，正在成为推动人工智能发展的核心驱动力。文章概述了当前已有的重要知识资源，对知识图谱关键技术——知识表示、知识图谱构建和知识图谱应用进行了综述，并对知识图谱未来发展的挑战和趋势进行了总结展望。

**4.3 相关论文三**

**（1）题目**：《基于知识图谱表示学习的协同过滤推荐算法》

**（2）作者介绍**：吴玺煜( 1993—) ，男，硕士研究生，主研方向为推荐系统、本体理论；陈启买，教授；刘海( 通信作者) 、贺超波，副教授、博士。

**（3）摘要**：针对协同过滤算法仅使用物品-用户评分矩阵而未考虑语义的问题，提出一种协同过滤推荐算法。使用知识图谱表示学习方法，将业界已有的语义数据嵌入一个低维的语义空间中。通过计算物品之间的语义相似性，将物品自身的语义信息融入协同过滤推荐。算法弥补了协同过滤算法没有考虑物品本身内涵知识的缺陷，在语义层面上增强了协同过滤推荐的效果。实验结果表明，该算法能够有效地提升协同过滤推荐的准确率、召回率和F值。

**（二）基于知识图谱表示学习的协同过滤推荐算法**

**Collaborative Filtering Recommendation Algorithm Based on Representation Learning of Knowledge Graph**

**--WU Xiyu，CHEN Qimai，LIU Hai，HE Chaobo**

**(Computer Engineering 2018 )**

**一、科学问题**

**1.1 本文所涉及科学问题**

关于知识图谱的构建以及知识图谱表示学习，并针对协同过滤算法仅使用物品-用户评分矩阵而未考虑语义的问题，提出一种协同过滤推荐算法。涉及协同过滤、知识图谱、表示学习、语义相似性、推荐系统等问题。

**1.2 同行专家如何解决**

FEＲNÁNDEZ-TOBÍAS将Facebook用户的个性数据融合进协同过滤中，用合并多领域数据集的方式提高了推荐的准确性。杨兴耀在论文《融合奇异性和扩散过程的协同过滤模型》中从评分上下文信息着手，提出了一种融合奇异性和扩散过程的协同过滤模型，解决了信息过载问题。杨武，唐瑞，卢玲在论文《基于内容的推荐与协同过滤融合的新闻推荐方法》中融合了基于内容与协同过滤的推荐算法，提高了新闻推荐的性能。现有各种协同过滤算法都使用了用户隐性反馈数据( 比如购买记录) 、用户显性反馈数据( 比如评分) 等信息，准确地将用户所需要的物品推荐给对方。ZHANG Zijian，GONG Lin，XIE Jian在论文中尝试使用了知识图谱的结构特征，将本体融进协同过滤算法。现有研究表明，知识图谱表示学习方法能将知识图谱嵌入到一个低维语义空间中，可以利用连续数值的向量反映知识图谱的结构特征。这种方法可以高效地计算实体间的语义联系。文献《Collaborative Knowledge Base Embedding for Ｒecom-mender Systems》尝试将知识图谱表示学习算法与基于隐性反馈的协同过滤相结合，把原始数据转化为偏好序列进行参数学习，加强了协同过滤推荐算法的性能。

**1.3 本文所解决的问题**

本文提出了一种基于知识图谱表示学习的协同过滤推荐算法，既利用了物品本身内在的语义信息，又使用了物品外在的评分矩阵，能够更加全面地反映物品的属性。算法通过知识图谱表示学习将实体嵌入到低维空间里，计算实体间的语义相似性，并将其融入到推荐系统中，在语义的层面上增强了协同过滤推荐的效果。

**1.4 本文解决方案效果**

这种种基于知识图谱表示学习的协同过滤推荐算法在语义的层面上增强了协同过滤推荐的效果，解决了协同过滤推荐未考虑语义的问题，提高协同过滤推荐的精度；同时又利用语义的信息，在一定程度上解决了冷启动问题。实验表明，基于知识图谱表示学习的协同过滤推荐算法能够提高推荐算法的效果。

**二、研究内容**

**2.1 理论与方法介绍**

（1）相关理论

基于物品的协同过滤推荐是一种基于最近邻的推荐系统算法，在推荐系统中，算法通过采用余弦相似性度量方法计算物品之间的相似性，从而向用户推荐最合适的物品。此外，针对物品-用户评分矩阵中评分相当稀疏的问题提出加入相似性权重重要性的策略，用于避免数值过分偏倚的问题。而对某一个用户，读取内存中的数据，使用Top-k进行选择推荐。推荐系统中用户和物品往往数量巨大，因此需要预先过滤物品间相似性度量的数量。文章采用2种预过滤技术：Top-N 过滤和阈值过滤。对于预测的评分，文章也使用了Top-k选择和正值选择，使推荐效果提高。

WANG H在论文中将前人思想应用于知识图谱领域，提出了一种知识图谱表示学习算法TransE。基于上述思想，能够通过知识图谱表示学习，对知识图谱中实体和关系进行语义表示学习，通过将富有语义信息的知识图谱三元组嵌进n维语义空间并生成对应向量，从而实现了知识图谱的数值化。通过知识图谱表示学习，可以快速计算两个实体间的语义相似性，并方便地将知识图谱用到其他学习任务中。

（2）基于知识图谱表示学习的协同过滤推荐算法

该算法基本思想是：对于协同过滤计算出来的最近邻，系统推荐给用户；而对于该用户喜欢的物品在语义上相似的物品，系统也可以推荐给用户。利用知识图谱表示学习算法，将推荐的物品嵌入到一个低维空间，然后计算物品之间的语义相似性，生成语义相似性矩阵，最终可以得到物品的语义近邻。同时，通过调节融合比例，对语义近邻和协同过滤潜在物品按比例融合，利用丰富的语义数据一定程度上解决了推荐系统的冷启动问题。过程包括处理基于 TransE 算法的知识图谱表示学习、语义相似性度量、融合语义近邻等。

**2.2 验证分析与实验效果**

文章选用了2012-11-09 版本的Freebase中的电影本体，并采用了编辑距离和字符串规则匹配相结合的方式对电影名称进行了版本号删除等数据预清洗工作，得到实验数据集。通过对评价指标以及混合比例的调整确定，并对表示学习的嵌入维度比较以及算法比较，得出本文提出的基于知识图谱表示学习的协同过滤推荐算法，优于协同过滤推荐算法，利用语义信息可以在一定程度上对协同过滤推荐算法进行改进，在召回率、准确率和F值上均有所提升，可以在语义上弥补基于物品的协同过滤算法的不足。

**三、论文存在问题及后续研究重点**

**3.1 论文存在问题**

该系统仅用于推荐电影本体内容，而对电影本体以外的内容没有进行考虑，电影知识图谱中实体构建不够全面。

**3.2 后续研究重点**

实验表明，基于知识图谱表示学习的协同过滤推荐算法能够提高推荐算法的效果。下一步将尝试把该算法应用于电影本体以外内容进行推荐，并进一步优化推荐性能。

**四、该问题相关研究成果**

**4.1 相关论文一**

**（1）题目**：《基于知识图谱的推荐技术研究及应用》

**（2）作者介绍**：王一鸣，电子科技大学工程硕士

**（3）摘要**: 随着互联网的普及和信息技术的发展，网络信息数据呈现爆炸性的增长，信息过载问题随之凸显。为了帮助用户从大量信息中发现自己感兴趣的信息，以及帮助信息生产者让自己生产的信息受到广大用户的关注，推荐系统应运而生。然而，传统的推荐系统面临的稀疏性问题以及冷启动问题在一定程度上限制了推荐系统的效果。本文通过构建一个领域内的知识图谱，并提取其中物品的语意表示，结合协同过滤算法来提高推荐系统的性能。本文以基于知识图谱的推荐技术研究及应用为研究课题，重点研究了领域内知识图谱的构建技术，知识图谱中实体和关系的嵌入方法以及基于知识图谱的推荐算法。

**4.2 相关论文二**

**（1）题目**：《基于标签主题的协同过滤推荐算法研究》

**（2）作者介绍**：文俊浩( 1969—)，男，教授、博士，主研方向为数据挖掘、服务计算；袁培雷，硕士；曾骏，讲师、博士；王喜宾、周魏，博士。

**（3）摘要**：传统基于标签的推荐算法仅考虑用户的评分信息，导致推荐准确度不高。为解决该问题，提出一种改进的协同过滤推荐算法。对用户-标签矩阵、资源-标签矩阵进行潜在 Dirichlet 分布建模，发掘推荐系统中的潜在语义主题，从语义层面计算用户对各资源的偏好概率，将计算出的偏好概率与协同过滤算法计算出的资源相似度相结合，预测用户偏好值，实现个性化推荐。在 Movielens 数据集上的实验结果表明，与传统基于标签的推荐算法相比，该算法能消除标签中存在的同义词、多义词等语义模糊问题，同时提高推荐准确度。

**4.3 相关论文三**

**（1）题目**：《融合奇异性和扩散过程的协同过滤模型》

**（2）作者介绍**：杨兴耀，于炯，吐尔根·依布拉音，廖彬，钱育蓉

**（3）摘要**：为了进一步提高推荐系统性能，从评分上下文信息着手，通过对项目评分进行分类统计获得评分奇异性，同时借鉴多渠道扩散相似性模型将推荐系统作为用户-项目二分网络的思想，提出了融合奇异性和扩散过程的协同过滤模型。为了表明模型的优越性，比较实验基于MovieLens，NetFlix和Jester这3个不同的数据集展开。实验结果表明，该模型不仅具有良好的扩展性，而且在合理的时间开销下，可以显著提高系统的预测和推荐质量。