2018 年 11 月 第 11 期
Vol. 39 No. 11 2018

一种知识图谱的排序学习个性化推荐算法

杨晋吉,胡 波,王欣明,伍昱燊,赵淦森

(华南师范大学 计算机学院,广州 510631)
E-mail :gzhao@ m. scnu. edu. cn

摘 要: 推荐系统是解决"信息过载"的有效方法,提出一种知识图谱的排序学习个性化推荐算法.本文算法首先构建融合上下文信息的知识图谱,使用基于深度学习的网络表示方法 Node2Vec 抽取知识图谱特征,通过将排序学习模型产生的反馈模型与用户兴趣迁移模型结合,构建混合推荐模型,最终通过排序学习进行 Top-N 推荐.该算法能够将各种不同性质的上下文特征结合在一起,并通过排序学习衡量这些多维特征的权重比例,解决了不同特征的融合问题,并且能够考虑到用户兴趣迁移和长短期偏好.在 Movielens 1M 数据集上的对比实验验证文中算法的有效性,实验表明,该算法能够有效提高推荐的 P@N 和 MAP 值.

关 键 词: 知识图谱;排序学习;兴趣迁移;Node2Vec;上下文信息
中图分类号: TP391 **文献标识码:** A **文 章 编 号:** 1000-1220(2018)11-2419-05

Personalized Recommendation Algorithm for Learning to Rank by Knowledge Graph

YANG Jin-ji, HU Bo, WANG Xin-ming, WU Yu-shen, ZHAO Gan-sen
(School of Computer Science, South China Normal University, Guangzhou 510631, China)

Abstract: The recommendation system is an effective way to solve the problem of "information overload". This paper presents a personalized recommendation algorithm for learning to rank by Knowledge Graph. Firstly, the algorithm constructs the Knowledge Graph of the fusion context information. Secondly, we extract the characteristics of the Knowledge Graph use the network representation method based on the Node2Vec. Thirdly, the feedback model generated by the learning to rank model is combined with the user interest migration model to construct the hybrid recommendation model, and finally the Top-N recommendation is done by learning to rank. The algorithm combines the different contextual characteristics of different properties, measures the weight ratio of these multidimensional features by learning to rank. Through this algorithm, we solve the problem of fusion of different features, and take into account the user interest migration and long and short term preferences. The comparison experiment verifies the validity of the algorithm on the Movielens 1M dataset. The experimental results show that this algorithm can effectively improve the P@N value and MAP value of the proposed system.

Key words: knowledge graph; learning to rank; interest migration; Node2Vec; context information

1 引 言

随着大数据时代的到来,网络上不断涌现的信息呈指数级增长,个性化推荐系统的作用越来越重要.尽管推荐系统的研究和应用均取得了很大的进展,但是它依然面临着很多的挑战,比如数据稀疏性问题、冷启动问题、时效性问题、多样性推荐问题等[1].传统的推荐算法主要分为三类:协同过滤推荐算法、基于内容的推荐算法和混合推荐算法.这类推荐算法在一些应用场景下能取得良好的效果,但它们各自有一些缺陷,如协同过滤主要受冷启动影响,并且难以针对具有特殊喜好的用户进行个性化推荐;基于内容的推荐受物品内容信息提取技术的制约,而且推荐效果比较差;混合推荐难以整合多种推荐算法间的权重.针对传统推荐系统存在的问题,现在很多新的推荐算法被提出来,其中包括排序学习、上下文感知推荐、基于深度学习推荐和社会化推荐等.其中排序学习逐渐成为了推荐系统研究的热点[2],基于排序学习的推荐模型将推荐问题转化为排序问题,构建以排序学习为基础的推荐算法框架,利用排序学习方法的优势去解决多特征维度的推荐问题,可以有效地组织多种推荐模型,并自动优化模型权重参数,提高推荐效果[3].知识图谱是一种基于图的数据结构,由节点和边组成.在知识图谱里,每个节点表示现实世界中存在的"实体",每条边为实体与实体之间的"关系",知识图谱是关系的最有效表示方式,并且能够融合多源异构信息.知识图谱表示学习能够将知识图谱嵌入到一个低维空间,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关系.

围绕上述背景,本文提出了一种知识图谱的排序学习个性化推荐算法.本文算法在文献[4]的基础上,通过排序学习构建反馈特征模型,融合用户兴趣迁移模型,再与基础特征模型构建混合模型,最终通过排序学习进行 Top-N 推荐.本文

收稿日期:2017-11-02 收修改稿日期:2017-12-12 基金项目:教育部博士点基金项目(20124407120017)资助;广东省自然科学基金团队项目(S2012030006242)资助. 作者简介:**杨晋吉**,男,1968 年生,博士,教授,研究方向为模型检测、智能体系统;**胡 波**,男,1990 年生,硕士研究生,研究方向为自然语言处理;**王欣明**(通信作者),男,1980 年生,博士,讲师,研究方向为大数据与软件工程;**伍昱燊**,男,1994 年生,硕士研究生,研究方向为自然语言处理;**赵淦森**(通信作者),男,1977 年生,博士,教授,研究方向为大数据、云计算、信息安全.

算法考虑到用户的长短期偏好和时效性因素,并解决了知识图谱中不同特征的融合问题,提高了个性化推荐效果.

本文的结构描述如下,第 2 节介绍算法的相关理论,第 3 节阐述了本文提出的一种知识图谱的排序学习个性化推荐算法及其实现,最后在第 4 节描述了本算法所使用的数据集,并进行了参数确定和算法比较的实验. 第 5 节对本文工作进行了总结和展望.

2 相关理论

2.1 基于深度学习的网络表示

近几年来,随着知识图谱的发展,业界已经积累了大量的开放本体库,例如 Freebase、DBpedia、WordNet 等. 这些本体库包含了丰富的本体知识,在基于知识图谱的推荐系统中融合实体上下文信息,能够很好的解决冷启动和推荐物品新颖性等问题,有效提高了推荐算法的性能[5].

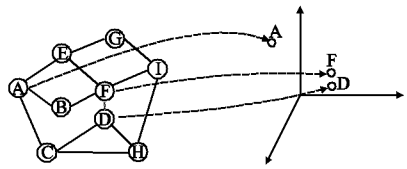

图 1 将知识图谱节点嵌入到 K 维空间

Fig. 1 Embed knowledge graph into K-dimensional space

DeepWalk[6]算法第一次将深度学习的技术引入到网络表示学习领域,DeepWalk 把节点作为自然语言处理中的单词,通过在网络中进行随机游走,获得随机游走路径,把随机游走路径作为句子,这样获得的数据就可直接作为 Word2Vec[7]算法的输入,通过这种方法可以将网络中的节点嵌入到一个 K 维空间,这种分布式表示能够更好地发现实体之间的相关性. Node2Vec[8]通过改变随机游走序列生成的方式进一步扩展了 DeepWalk 算法. 它提出了一种带偏置的随机游走策略,这种策略可以有效地检索分散的相邻节点. Node2Vec 通过参数 p 和 q 能够兼顾深度游走(Depth-First Search)和广度游走(Breadth-First Search).

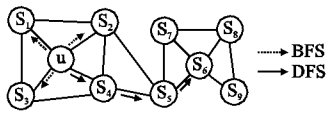

图 2 从节点 u 出发的 BFS 和 DFS 游走

Fig. 2 BFS and DFS walk from node u

广度优先游走侧重邻近的节点并刻画了节点间的同构性,深度优先游走反映了更高层面上的节点间的同质性. 其中游走的条件概率为:

$$P(c_i = x \mid c_{i-1} = v) = \begin{cases} \frac{\pi_{vx}}{Z}, if(v,x) \in E. \\ 0, otherwise. \end{cases} \quad (1)$$

其中 π_{vx}是未归一化的概率,Z 是归一化常数,对于最简单的情况,可以用节点 v 和 x 之间边的权重 ω_{vx}作为未归一化概率 $\pi_{vx} = \omega_{vx}$. t 是上一个节点,v 是当前节点,x 是下一个可能的节点对于 2 阶随机游走,未归一化概率和权重之间的关系为:$\pi_{vx} = \alpha_{pq}(t,x) \cdot \omega_{vx}$,系数 $\alpha_{pq}(t,x)$如下:

$$\alpha_{pq}(t,x) = \begin{cases} \frac{1}{p}, if\ d_{tx} = 0. \\ 1, if\ d_{tx} = 1. \\ \frac{1}{q}, if\ d_{tx} = 2. \end{cases} \quad (2)$$

d_{tx}表示节点 t 和 x 之间的最短距离,p 是 return 参数,q 是 in-out 参数. Node2Vec 能够同时考虑到网络中局部和宏观信息,并且有很高的计算效率和适应性,能够有效提取网络中节点的特征.

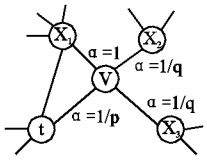

图 3 Node2Vec 算法(修改自文献[8])

Fig. 3 Node2Vec algorithm

2.2 用户兴趣迁移模型

用户对于某些物品的感兴趣程度或关注程度会随着时间推移和交互情况动态增加或减少,然而传统的推荐算法无法反映用户行为数据的动态变化,因此,无法正确解决用户兴趣迁移问题,而融合用户迁移模型的推荐系统能够提高个性化推荐效果.

通过给知识图谱模型中的节点间分配不同的权重可以体现用户对不同物品的兴趣差异. 用户兴趣迁移模型通过用户行为的时间与次数来对知识图谱中的节点间连接权重进行动态修改,能够反映出用户的兴趣迁移. 用户行为越接近当前时间,同种行为的次数越多,节点间分配到的权重越大,用户对该节点的兴趣度越大. 反之,节点之间的连接权重越小.

用户 U_i 和物品 I_j 两个节点间连接权重如公式(3)所示:

$$w_{ij} = \sum_{s=1}^{n} \left(\frac{w}{1 + e^{(t-t_s)-t_0}} + w \right) \quad (3)$$

其中,t 为当前时间,n 为同种行为的次数,t_s 为用户对物品产生反馈的行为时间,t_0 为用户兴趣迁移的时间因子,w 表示权重阈值,即用户随着时间推移,能够提供的推荐能力逐渐减小,最后趋于常量 w,因此可以根据用户兴趣迁移模型动态更新知识图谱.

2.3 基于排序学习的推荐算法

近几年,研究人员发现:如果仅仅依据用户对项目的评分,产生的推荐结果并不能准确地体现用户的偏好[9]. 为了解决传统推荐算法所存在的上述问题,研究人员考虑将排序学习技术[10]融合进推荐算法的推荐过程中. 排序学习方法把用户间推荐得分的计算转化为对一个多特征向量的二分类问题,很好地解决高维特征带来的多参数估计问题. 排序学习推荐模型基本流程如图 4 所示:

基于排序学习的推荐框架中,首先对用户物品对(U_i, I_j)进行标注 $y_{U_iI_j}$,然后抽取用户物品对的特征,得到特征向量 $\vec{x}_{U_iI_j}$,构建为特征向量集合 X:

$$X = \{\vec{x}_{U_1I_1}, \vec{x}_{U_1I_2}, \vec{x}_{U_1I_3}, \cdots, \vec{x}_{U_iI_j}\} \subseteq R^n \quad (4)$$

相应的标注集合 Y 为

$$Y=\{y_{U_1I_1},y_{U_1I_2},y_{U_1I_3},\cdots,y_{U_iI_j}\} \tag{5}$$

其中 n 代表 $\vec{x}_{U_iI_j}$ 的维度,排序学习的最终目的是要通过最优化函数来获得一个决策函数 $f:R_n\rightarrow Y$,使得通过该函数对全体训练实例集合(X,Y)做出的标号预测 Y'能最大限度与它们实际的标号 Y 对应,最终得到权重比例集合 $Z=\{\eta_1,\eta_2,\eta_3,\cdots,\eta_{|features|}\}$.

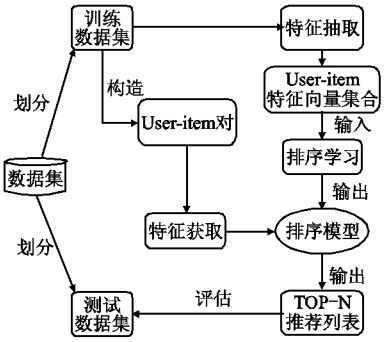

图 4　基于排序学习推荐模型基本流程

Fig. 4　Learning to rank recommendation process

排序学习提出利用机器学习的方法去解决排序问题,是基于机器学习中解决分类与回归问题的思想. 排序学习的目标在于自动地从训练数据中学习得到一个排序函数,使其在文本检索中能够针对文本的相关性、重要性等多种衡量标准对文本进行排序. 排序学习的优势是:整合大量复杂特征并自动进行参数调整,自动学习最优参数,降低了单一考虑排序因素的风险,同时,能够通过众多有效手段规避过拟合问题. 因此,基于排序学习的推荐模型能够提高个性化推荐效果,比较典型的排序学习算法有 Ranking SVM、LambdaMART、RankNet、RankBoost、AdaRank 等.

3　一种知识图谱的排序学习个性化推荐

本文提出了一种知识图谱的排序学习个性化推荐算法,其基本思想是:首先构建基础知识图谱,通过基于深度学习的 Node2Vec 网络表示算法,将知识图谱中的实体嵌入到一个低维空间,然后计算用户物品之间的相似性,构建训练模型作为排序学习的输入,通过目标函数调节不同特征的重要程度来达到最优结果,对基础推荐模型产生的特征比例权重集合,融合用户兴趣迁移模型生成混合知识图谱,通过 Node2Vec 构建反馈特征模型,与基础推荐模型构成混合模型. 最终,在混合模型上进行排序学习,产生 Top-N 推荐列表. 本文算法既能解决知识图谱异构特征间的权重比例,也能够考虑用户的长期、短期偏好和用户兴趣迁移等因素,能够提高个性化推荐效果. 图 5 给出了本文算法的流程图:

3.1　基于带权深度游走的知识图谱特征抽取

传统推荐算法使用邻接矩阵进行数据存储和运算,这种数据表示方法受到计算效率问题的影响. 如邻接矩阵 A 占用了$|V|\times|V|$的存储空间,这在$|V|$增长到百万级时通常是难以计算和处理的. 另一方面,邻接矩阵中绝大多数是 0,数据十分稀疏. 这种数据稀疏性使得快速有效的统计学习方法的应用变得困难[11].

知识图谱能够融合语义、上下文和异构特征信息,通过边的权重能够体现节点间的相互关系. 本文提出的一种知识图谱的排序学习推荐算法在知识图谱上进行深度游走,既能够考虑节点间的同构性也能够考虑节点间的同质性,并且能够很好支持异构信息的融合和考虑用户兴趣迁移情况.

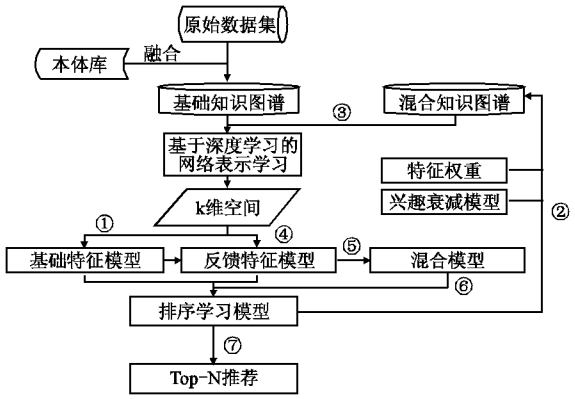

图 5　一种知识图谱的排序学习个性化推荐算法流程

Fig. 5　Learing to rank personalized recommendation based on Knowledge Graph

以电影领域为例,电影实体中主要包括了演员、类型、导演等主要特征. 这些异构特征从一定的程度上概括了这部电影. 利用电影特征,可以得到类似图 6 所示的一个电影知识图谱.

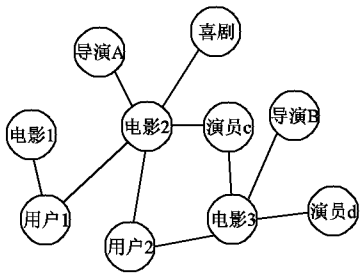

图 6　融合上下文和异构信息,构建电影知识图谱

Fig. 6　KG of context and heterogeneous information fusion

在本文算法中使用 Node2Vec 进行知识图谱网络特征学习,将实体映射到 K 维空间,在 K 维向量空间中,几何上越接近的实体相关性越大,本文算法通过向量的余弦相似度来计算实体 e_i 和 e_j 之间的相关性 $Sim(e_i,e_j)$.

$$Sim(e_i,e_j)=\cos(\vec{e}_i,\vec{e}_j)=\frac{\vec{e}_i\cdot\vec{e}_j}{\|\vec{e}_i\|\times\|\vec{e}_j\|}=\frac{\sum_{t=1}^{n}e_{it}e_{jt}}{\sqrt{\sum_{t=1}^{n}e_{it}^2}\sqrt{\sum_{t=1}^{n}e_{jt}^2}} \tag{6}$$

通过对训练集处理,为用户物品对(U_i,I_j)进行标注 y_{ij},在构建的基础知识图谱上,计算(U_i,I_j)在单一物品上下文特征 *feature* 下的相似性,构建特征向量 $\vec{x}_{U_iI_j}$:

$$\vec{x}_{U_iI_j}=\{Sim(U_i,I_j)_1,Sim(U_i,I_j)_2,\cdots,Sim(U_i,I_j)_{|feature|}\} \tag{7}$$

构建训练集$(y_{ij},U_i,\vec{x}_{U_iI_j})$作为排序模型训练输入,由最优化函数来获得一个决策函数 $f:R_n\rightarrow Y$,通过决策函数产生 Top-N 推荐列表,并得到多维特征对排序结果的权重比例集合 $Z=\{\eta_1,\eta_2,\eta_3,\cdots,\eta_{|feature|}\}$,用以构建反馈模型.

此外,在知识图谱中权重可以代表用户对物品的偏好情况,物品和特征间的相关性. 在文献[4]推荐算法中,把数据集中评分大于4用以表明用户User和物品Item间有关系,设置边的权重为1,没有关系则设置为0. 把边仅仅简单地看作0,1值,因此该算法没有考虑不同特征对推荐结果的重要性是不同的,也没有考虑到用户的偏好因素影响,也没有考虑到人们兴趣随着时间发生偏移的情况.

由于上述原因,本文在该算法基础上,提出了一种融合基础模型,反馈模型和用户兴趣迁移的混合推荐模型.

3.2 融合长短期偏好和兴趣迁移的混合推荐模型

从长期来看,用户的兴趣喜好是相对稳定的,基于用户大量的历史数据进行推荐可以基本反映用户的一般偏好. 但是,从心理学角度看,用户偏好受长期个人兴趣和短期个人兴趣两方面影响,并且存在一定的相互联系.

基于上述原因,本文在基于排序学习的Top-N推荐框架上进行拓展,用基础知识图谱衡量用户的长期偏好. 通过反馈模型和用户兴趣偏移模型构建混合知识图谱模型,用它衡量物品内容的动态变化、用户兴趣的短期波动等时效性因素.

通过3.1节中基于基础知识图谱的排序学习推荐模型能够获得多维特征对推荐结果的权重集合 $Z=\{\eta_1,\eta_2,\eta_3,\cdots,\eta_{|feature|}\}$,通过把影响权重因子集合Z和用户兴趣迁移模型融合,构建混合知识图谱.

混合知识图谱实体间权重更新策略如下:

$$RW_{ij}=\begin{cases}\lambda\times rating\times w_{ij}\times\eta_k, if \quad r_{ij}=k;\\ \eta_{others}, if \quad r_{ij}=others;\end{cases} \tag{8}$$

其中 RW_{ij} 为更新后实体 i 和实体 j 之间边的权重,w_{ij} 是经用户兴趣迁移模型计算后的兴趣度,关系 k 指的是用户 i 和物品 j 之间是评分关系,rating是用户对物品的评分,λ 是归一化因子,使得 $\lambda\times rating$ 归一化于初始权重1,防止评分过大而对随机游走造成影响.

对训练集中的用户物品对 (U_i,I_j),在融合了所有物品特征和用户兴趣迁移的混合知识图谱上进行Node2Vec深度游走,抽取相似性特征 $Sim(U_i,I_j)_{mix}$. 与3.1节公式(7)构建混合特征模型:

$$\vec{x}_{U_iI_j}=\{Sim(U_i,I_j)_1,\cdots,Sim(U_i,I_j)_{|feature|},Sim(U_i,I_j)_{mix}\} \tag{9}$$

构建集合 $(y_{ij},U_i,\vec{x}_{U_iI_j})$ 作为排序模型输入,最终产生Top-N推荐列表. 本文算法能够有效融合多维特征,并能够考虑用户的长短期偏好,提升了个性化推荐效果.

3.3 算法描述

算法步骤如表1所示.

4 实验结果及分析

4.1 数据集

本文所用的基础数据集是Movielens 1M[12],包含6040个用户,3900个电影和100多万条匿名评分组成. 由于本文算法验证用户兴趣随着时间的动态迁移情况,故将融合DBpedia上下文信息后的数据集[13],按照时间戳顺序划分为8:2,分别作为训练集和测试集.

4.2 评价指标

本文实验中所用到的评估指标[14]主要有P@N、MAP,下面分别对其进行说明.

P@N本身是Precision@N的简称,指的是对特定的查询,考虑位置因素,检测前N条结果的准确率. 例如对单次搜索的结果中前100篇,如果有82篇为相关文档,则P@100=82/100=0.82.

表1 算法步骤

Table 1 Algorithm steps

算法:一种知识图谱的排序学习个性化推荐
输入:数据集S、本体库
输出:Top-N推荐列表
1:将数据集S融合本体库上下文信息,构建基础知识图谱
2:通过Node2Vec抽取知识图谱网络特征
3:通过排序学习进行模型训练,产生基础特征模型,获取决策函数 $f:R_n\rightarrow Y$
4:通过决策函数进行反馈,融合用户兴趣迁移模型,产生混合知识图谱
5:重复2步,抽取4产生的混合知识图谱的特征模型
6:将3和5产生的特征模型融合成混合特征模型
7:重复3步骤,产生Top-N推荐列表

测试通常会使用一个查询集合,包含若干条不同的查询词,在实际使用P@N进行评估时,通常使用所有查询的P@N数据,计算算术平均值,用来评判该系统的整体搜索结果质量.

$$P@N=\frac{number\ of\ relvant\ documents\ in\ top\ N}{N} \tag{10}$$

MAP指标通常用于衡量系统在所有排序结果中相关文档的排序质量. 在一个排序结果中,AP(平均查准率)用来计算一个查询的排序结果的精度,MAP是对所有查询的精度取平均值. 其相关定义如下所示:

$$P(j)=\frac{\sum k:\pi_i(k)\leqslant\pi_i(j)y_{i,k}}{\pi_i(j)} \tag{11}$$

$$AP=\frac{\sum_{j=1}^{n_i}P(j)\cdot y_{i,j}}{\sum_{j=1}^{n_i}y_{i,j}} \tag{12}$$

$$MAP=\frac{\sum_i AP}{\sum_i l} \tag{13}$$

其中,$y_{i,j}$ 表示相关度,取二元值0和1. 若第i个查询的第j个文档是相关的,$p(j)$ 计算查询i排序结果中排在文档j前面的相关文档的比例.

对于Top-N的电影推荐结果,电影排得越靠前,被用户浏览到的可能性就越大,因此在评测的过程中,相关度越高的电影,在最终的结果列表中排得位置越靠前,评测函数应该给予更高的权重.

4.3 实验结果分析1

实验硬件,处理器型号为Intel(R)Xeon(R)CPU E5-2620 v2 @ 2.10GHz,内存为80GB;实验软件环境为Python 2.7.12、JDK 1.8.

4.3.1 调整确定特征参数

本文所提出算法的参数参考文献[6]通过在该数据集上实验所确定的最佳参数.

通过表2能够发现在该数据集上,p,q分别为1,4时实

验效果最好,并用 LambdaMark 算法进行排序学习.

表 2　排序学习算法在不同 p,q 下结果

Table 2　Learning to rank result with different p,q

(P,Q)	Algorithm	P@5	P@10	MAP
1,4	**LambdaMark**	**0.0791**	**0.0293**	**0.1717**
4,1	LambdaMark	0.0182	0.0193	0.0570
1,1	LambdaMark	0.0174	0.0188	0.0565
1,4	AdaRank	0.0134	0.0098	0.0278
4,1	AdaRank	0.0078	0.0083	0.0286
1,1	AdaRank	0.0109	0.0098	0.0358

通过表 3 能发现经过排序学习训练完决策函数后,该数据集中不同特征对推荐结果的比例权重是不同的.我们通过决策函数反馈的集合 $Z=\{\eta_1,\eta_2,\eta_3,\cdots,\eta_{|features|}\}$ 和用户兴趣迁移模型构建反馈模型.

表 3　特征权重比例

Table 3　Feature weight ratio

特征	P@5	P@10	MAP
$\rho_{dbo:based\_on}$	0.0529	0.0247	0.0925
$\rho_{cinematography}$	0.0386	0.0290	0.0794
$\rho_{feedback}$	**0.2317**	**0.1708**	**0.3550**
$\rho_{dbo:director}$	0.0219	0.0211	0.0949
$\rho_{dbo:editting}$	0.0294	0.0305	0.0724
$\rho_{dbo:musicComposer}$	0.0077	0.0040	0.0493
$\rho_{dbo:narrator}$	0.0572	0.0389	0.0834
$\rho_{dbo:producer}$	0.0119	0.0241	0.0498
$\rho_{dbo:starring}$	0.0128	0.0372	0.0728
$\rho_{dbo:subject}$	0.0285	0.0326	0.0688
$\rho_{dbo:writer}$	0.0061	0.0216	0.0472

4.3.2　算法比较

最后在该数据集上,使用基于 Python 的开源库 SurpriseLib 实现对比实验,排序算法使用基于 Java 的开源 RankLib 库实现.对比结果如表 4.

表 4　对比实验结果

Table 4　Comparison of experimental results

Model	P@5	P@10	MAP
本文算法	**0.3226**	**0.2374**	**0.4969**
entity2rec	0.2814	0.2127	0.4232
MostPop	0.2154	0.1815	0.2907
NMF	0.1208	0.1150	0.1758
SVD	0.0543	0.0469	0.0888
ItemKNN	0.0463	0.0232	0.0990

通过表 4 能够发现,本文提出的一种知识图谱的排序学习个性化推荐算法在 P@N 和 MAP 上均有所提升,算法能够充分考虑到长短期偏好和用户兴趣的迁移.

5　结束语

本文提出了一种知识图谱的排序学习个性化推荐算法,不但利用了人和物品间的评分信息,也包含了物品的上下文信息,并且考虑长期、短期偏好,因此能够全面地反映出人和物品之间的关系.针对传统推荐算法中多维特征的参数难以估计的痛点,本文算法通过排序学习方法获取不同特征的权重比例进行反馈,构建混合知识图谱,通过基于深度学习的网络特征表示模型构建特征模型,最终与基础特征模型结合,构建混合模型进行推荐.通过在 Movielens 数据集上验证了本文算法的有效性.

本文算法模型没有融合用户的画像特征,而这些特征对推荐系统的意义十分重要.未来将会尝试把用户画像特征融合到本算法模型中,挖掘用户的行为模式,并实现在线学习系统.

References:

[1] Dou Ling-yuan, Wang Xin-hua, Sun Ke, et al. Collaborative filtering fusing label features and time context[J]. Journal of Chinese Computer Systems, 2016, 37(1): 48-52.

[2] Amatriain X. The recommender problem revisited[C]. Proceedings of the 8th ACM Conference on Recommender Systems, ACM, 2014: 397-398.

[3] Weston J, Yee H, Weiss R J. Learning to rank recommendations with the k-order statistic loss[C]. ACM Conference on Recommender Systems, ACM, 2013: 245-248.

[4] Palumbo E, Rizzo G, Troncy R. Entity2rec: learning user-item relatedness from knowledge graphs for Top-N item recommendation [C]. Eleventh ACM Conference on Recommender Systems, ACM, 2017: 32-36.

[5] Lee S, Song S I, Kahng M, et al. Random walk based entity ranking on graph for multidimensional recommendation[C]. ACM Conference on Recommender Systems, ACM, 2011: 93-100.

[6] Perozzi B, Al-Rfou R, Skiena S. DeepWalk: online learning of social representations[C]. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2014: 701-710.

[7] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality[C]. Advances in Neural Information Processing Systems, 2013: 3111-3119.

[8] Grover A, Leskovec J. Node2vec: scalable feature learning for networks[C]. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2016: 855-864.

[9] Pessiot J F, Truong T V, Usunier N, et al. Learning to rank for collaborative filtering[C]. Iceis 2007-Proceedings of the Ninth International Conference on Enterprise Information Systems, Volume Aidss, Funchal, Madeira, Portugal, June. DBLP, 2007: 145-151.

[10] Li H. Learning to rank for information retrieval and natural language processing[J]. Synthesis Lectures on Human Language Technologies, 2014, 7(3): 1-121.

[11] Lin Y, Liu Z, Sun M, et al. Learning entity and relation embeddings for knowledge graph completion[C]. Twenty-Ninth AAAI Conference on Artificial Intelligence, AAAI Press, 2015: 2181-2187.

[12] Harper F M, Konstan J A. The MovieLens datasets: history and context[M]. Association for Computing Machinery, 2016.

[13] Ostuni V C, Noia T D, Sciascio E D, et al. Top-N recommendations from implicit feedback leveraging linked open data[C]. ACM Conference on Recommender Systems, ACM, 2013: 85-92.

[14] Joachims T, Li H, Liu T Y, et al. Learning to rank for information retrieval(LR4IR 2007)[J]. Acm Sigir Forum, 2007, 41(2): 58-62.

附中文参考文献:

[1] 窦羚源,王新华,孙　克.融合标签特征和时间上下文的协同过滤推荐算法[J].小型微型计算机系统,2016,37(1):48-52.