

基于知识图谱和 Bi-LSTM 的推荐算法

王钰莹, 王 勇

(北京工业大学信息学部, 北京 100124)

摘要: 目前现有基于模型的推荐算法多是将评分数据输入到深度学习模型中进行训练, 得出推荐结果。其缺陷在于无法对预测结果进行可解释性分析。除此之外, 无法有效地解决算法的冷启动问题。因此, 本文提出一种基于知识图谱和 Bi-LSTM 的推荐算法, 来有效解决算法的可解释性和冷启动问题。首先将获取到的数据集进行预处理, 生成预编码向量, 根据数据集结点的连接性, 构建专业领域知识图谱。其次利用知识图谱的元路径提取技术获取到多条用户-物品路径信息, 将其输入到 Bi-LSTM 中, 在路径经过的各结点处加入一层注意力机制, 目的是为了模型能够有效地获取到较远结点的信息。最后将多条路径的训练结果输入到平均池化层中, 用以区分不同路径的重要程度, 利用交叉熵损失函数对模型进行训练, 从而得出预测结果。实验结果表明, 与传统基于循环神经网络模型的推荐算法相比, 该算法可有效地提升算法的可解释性以及预测准确性, 并缓解算法的冷启动问题。

关键词: 知识图谱; 双向循环神经网络; 注意力机制; 可解释性; 冷启动

中图分类号: TP391

文献标志码: A

DOI: 10.3969/j.issn.1006-2475.2021.09.014

Recommendation Algorithm Based on Knowledge Graph and Bi-LSTM

WANG Yu-ying, WANG Yong

(Department of Information Science, Beijing University of Technology, Beijing 100124, China)

Abstract: At present, most of the existing model-based recommendation algorithms input the score data into the deep learning model for training to get the recommendation results. Its defect is that it is unable to analyze the interpretability of the prediction results. In addition, the algorithm can not effectively solve the cold start problem. Therefore, this paper proposes a recommendation algorithm based on knowledge map and Bi-LSTM to effectively solve the problem of interpretability and cold start of the algorithm. Firstly, the data set is preprocessed to generate precoding vector. According to the connectivity of data aggregation points, the domain knowledge map is constructed. Secondly, the meta path extraction technology of knowledge map is used to obtain multiple user item path information, which is input into Bi-LSTM. A layer of attention mechanism is added to each node of the path, so that the model could effectively obtain the information of remote nodes. Finally, the training results of multiple paths are input into the average pooling layer to distinguish the importance of different paths. The cross-entropy loss function is used to train the model and the prediction results are obtained. The experimental results show that, compared with the traditional recommendation algorithm based on the cyclic neural network model, this algorithm can effectively improve the interpretability and prediction accuracy of the algorithm, and alleviate the cold start problem of the algorithm.

Key words: knowledge graph; bidirectional recurrent neural network; attention mechanism; interpretability; cold start

0 引 言

随着信息社会的飞速发展, 人们可以方便地从各种渠道获取到丰富的信息, 但在面对庞杂的数据时, 很难在较短时间内筛选出有效信息, 并做出相应判断。在这种情况下, 推荐算法的产生为人们解决了这一难题^[1]。推荐算法可以利用互联网工具, 在庞大

的数据中个性化地提取出有效信息, 从而为人们提供精准的服务, 以此来满足用户的需求^[2]。推荐系统可以在用户无明确需求的情况下, 根据用户的历史行为做出合理的推荐^[3]。并且在用户使用推荐系统的同时, 及时记录下用户的行为数据, 为后续推荐提供数据来源。

传统的协同过滤推荐系统^[4-6], 主要利用了用户

收稿日期: 2020-12-20; 修回日期: 2020-12-21

作者简介: 王钰莹(1996—), 女, 天津人, 硕士研究生, 研究方向: 深度学习, 推荐系统, E-mail: wangyuying_425@163.com; 王勇(1974—), 男, 副教授, 硕士生导师, 研究方向: 并行与分布式计算, E-mail: wangy@bjut.edu.cn。

和物品的属性信息,并且依赖用户的行为来进行物品推荐,在算法准确性上有显著的提高,但依然存在冷启动和可解释性等问题^[7]。然而,这在提升算法的准确率和召回率方面尤为重要,因此,引入基于模型的推荐算法。基于模型的推荐算法,多是将原始数据信息输入到已构建好的深度学习模型当中,进行模型训练,得出推荐结果。而由于深度学习模型原本就存在一定程度上的不可解释问题,因此引入知识图谱的概念^[8],知识图谱作为先验知识,可以为推荐算法提供语义特征。通过构建专业领域知识图谱,将带有语义的数据输入到模型当中,从而解决模型的可解释性差问题。

目前,基于知识图谱的推荐算法主要是利用知识图谱表示技术或知识图谱元路径提取技术^[9-10]。基于知识图谱表示技术的推荐算法,主要是利用原始数据集构建知识图谱,利用建模的方式将结点和关系在向量空间中进行表示,通过计算从而生成低维稠密向量。Zhang 等人^[11]利用 embedding 将交互信息、文本、图片进行编码,提取出多类型的语义特征,与推荐物品相融合,提出 CKE 框架,充分学习物品的各属性特征,从而提高算法的准确性。Wang 等人^[12]提出了 CDL 模型,利用知识图谱表示学习得到物品结构化信息,并利用去噪编码器网络学习编码层文本表示向量和视觉表示向量,并将这些表示向量与物品的潜在因子向量相结合,利用矩阵分解算法完成推荐。但其问题在于关系类型较少且实体间关系较为稠密,导致推荐的准确性不佳。Wang 等人^[13]提出了 DKN 模型,利用知识图谱表示技术以及卷积神经网络对句子进行学习,并加入注意力机制,实现新闻推荐。但这些模型并未有效解决算法的可解释性以及冷启动等问题。

基于知识图谱元路径提取技术的推荐算法,主要是利用已构建好的知识图谱不同结点之间的连通性,来提取出不同的路径信息,并结合协同过滤模型进行推荐。Qian 等人^[14]提出了一种全新的路径相似度计算方法 PathSim,通过计算不同路径的相似程度得出相似物品,从而进行推荐。Shi 等人^[15]提出了一种基于元路径的策略,利用带权元路径来对路径相似度进行计算,构造带有权重的正则化项损失函数,来进行模型训练。Wang 等人^[16]提出了 RippleNet 模型,将知识图谱中的路径以及结点信息作为额外的知识加入到推荐算法中,将用户对某一物品的喜好理解为水波传播,以某一用户作为中心,不断向外发散,对各层发散结果进行相似度计算,并融合各层结果,得出最终预测值,实现物品推荐,但无法有效地解决算法的

冷启动问题。

因此,本文提出一种基于知识图谱和 Bi-LSTM (双向循环神经网络) 的推荐算法 KG-BiLSTM,将知识图谱嵌入技术和知识图谱元路径提取技术相结合,在提升算法准确率的同时,解决算法的可解释性和冷启动问题。在 MovieLens-1M 和 Yelp 数据集上进行大量实验,实验表明该算法与主流推荐算法相比,在精确率、召回率和 MRR 指标上有较大提升。该算法主要工作有如下几个方面:

1) 以有向图的形式构建专业领域知识图谱,将用户属性信息与物品属性信息作为不同类型的结点加入到知识图谱当中。并利用知识图谱元路径提取技术,提取出知识图谱中不同用户-物品的路径信息,存储于路径信息文件当中。

2) 将不同路径的相应结点信息进行编码,生成嵌入向量,将各路径的各结点嵌入向量输入到 Bi-LSTM 当中,根据前后结点的信息,生成 Bi-LSTM 隐藏层向量,使整条路径具有语义含义。

3) 为充分捕捉到较远距离结点的信息,引入注意力机制,将每条路径中的各结点隐藏层向量输入到注意力机制当中,从而进一步提升预测的准确性。

4) 将各路径信息输入到平均池化层中,用于区分不同路径的重要程度。对不同路径的预测信息进行融合,利用交叉熵损失函数进行模型训练,不断更新模型参数,得出最终预测结果。

5) 将本文提出的 KG-BiLSTM 推荐算法与主流基于知识图谱的推荐算法在 MovieLens-1M 和 Yelp 数据集上进行对比实验,并根据原始数据集,设置用户评分小于 4 的集合为冷启动数据集。根据 Precision、Recall 和 MRR 指标以及可解释性分析,验证该算法在原始数据集和冷启动数据集上均可有效提升推荐算法的准确性,并对推荐结果做出合理解释。

1 相关工作

1.1 知识图谱

知识图谱最早的应用是提升搜索引擎的能力,提高用户的搜索质量,其概念最早是由 Google 于 2012 年 5 月 17 日提出。随后在智能问答、自然语言处理、推荐算法、数据分析等场合被广泛应用。

知识图谱^[17]实质上是一个语义网络,由图的形式来形象地描述不同事物之间的关系,通常由节点和边组成。知识图谱中的节点用于表示实体和概念。实体是出现在知识图谱中的各类具体结点,概念是具有相同属性的某一类结点的集合。知识图谱中的边表示不同结点之间的关系和属性。知识由三元组

(h, r, t) 进行表示,其中 h 为头实体, r 为关系, t 为尾实体,一个完整的知识图谱需要由若干个不同的三元组构成。在知识图谱表示学习中,通常是将知识图谱嵌入到模型中,通过模型训练,使 $h + r \approx t$,从而得到知识的低维向量表示。而知识图谱元路径提取通常是提取出整条路径的结点信息,将路径结点信息输入到模型中进行训练,生成整条路径的低维向量,从而赋予路径特定语义^[18]。

1.2 循环神经网络 LSTM

长短期记忆网络 LSTM 是一种循环神经网络的变形^[19]。主要是为了解决传统循环神经网络中存在的长依赖问题,适用于处理间隔时间较长的事件。LSTM 内部结构如图 1 所示。该结构由 3 个门控单元和 1 个细胞状态组成。其中细胞状态为 c_t , 类似一个链式结构,用于整合和记录信息。而门控单元用于选择信息是否允许通过以及通过量,以此来控制细胞状态。门控单元分别为输入门 i_t 、遗忘门 f_t 和输出门 o_t 。输入门决定需要更新的信息,遗忘门决定需要覆盖哪些信息,而输出门决定哪些信息被输出。由于 LSTM 拥有处理较长时间间隔信息的特性,因此被应用于知识图谱当中,可根据路径的选择来记录结点信息,用于后续模型输入。

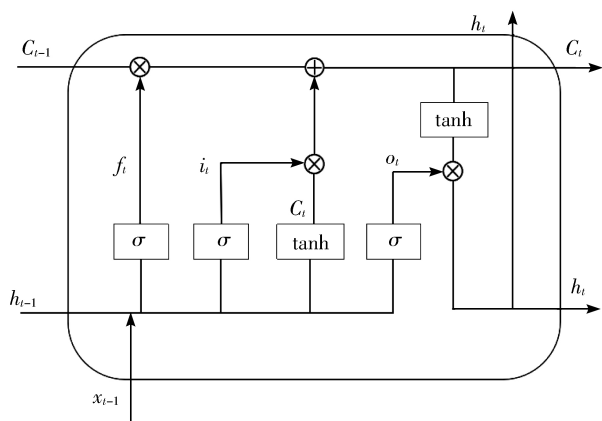


图1 LSTM 内部结构图

1.3 注意力机制

注意力机制最初被应用于机器翻译领域,目前在人工智能领域,注意力机制可融合多种神经网络模型,在自然语言处理、推荐系统、统计学习等方面有较为广泛的应用。

注意力机制^[20]类似人类的视觉机制,在对某事物进行观察时,通常倾向于关注较为重要的信息,而忽视相关度较低的信息,从而有助于做出最终决策。基于此原理,注意力机制可根据输入信息,对其重要性权值进行计算,将得到的重要性权值与输入信息相

乘,再将乘以不同权值的输入信息相加,得出最终的注意力向量。其模型如图 2 所示。

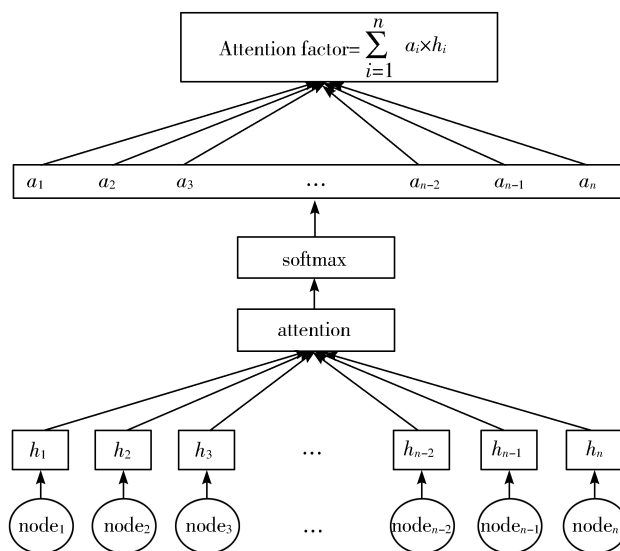


图2 注意力机制模型图

利用该模型可实现将不同结点的特征信息进行有效融合,从而防止因路径过长导致较远结点特征无法完整提取的问题。

2 基于知识图谱和 Bi-LSTM 的推荐算法

本章将详细描述 Bi-LSTM 推荐算法的主要内容。该模型主要分为 5 层: 预处理层、嵌入层、Bi-LSTM 层、注意力层、预测层(包括池化层和全连接层)。基于知识图谱和 Bi-LSTM 推荐算法的模型如图 3 所示。该算法首先将获取到的数据集进行分类预处理,对于每个用户和物品结点赋予唯一的 id 值,并构建一个有向图,将用户结点、物品结点、用户各属性结点以及物品各属性结点添加到有向图中,并用边对各属性和关系相连,接着构建专业领域知识图谱。对已构建好的知识图谱,提取每个用户-物品对的多条路径信息,设置路径条数最大值,将提取出的路径信息统一存储到文件中,作为后续模型的输入。其次,对各用户、物品及其属性信息进行编码,生成嵌入向量。然后,构建双向循环神经网络,将各路径的嵌入向量依次输入到神经网络当中,进行模型训练。同时引入注意力机制,充分获取前端较远结点的特征信息。最后,将各条路径的隐藏层向量输入到平均池化层中,根据路径的重要性程度进行向量整合,并利用交叉熵损失函数完成模型训练。

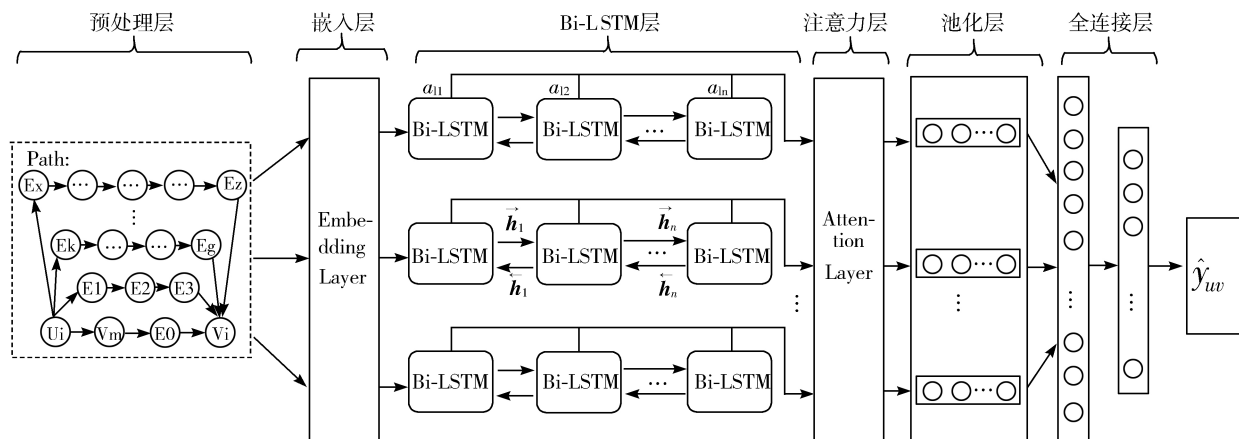


图3 基于知识图谱和 Bi-LSTM 的推荐算法模型

2.1 问题描述

推荐算法通常是根用户根据用户对物品的已评分信息预测未被评分物品的分值,算法的原始数据主要分为用户集合 $U = \{U_1, U_2, \dots, U_i\}$ 和物品集合 $V = \{V_1, V_2, \dots, V_j\}$, 其中 i 为用户总数, j 为物品总数。将用户与物品的交互信息存储于矩阵 $R_{i \times j}$ 中, 若 $r_{ij} = 1$ 则表示用户 i 对物品 j 产生过交互, 若 $r_{ij} = 0$ 则表示用户 i 对物品 j 未产生过交互。

将用户、物品及其属性信息映射到知识图谱 KG 当中, 用有向图的形式来对知识图谱进行存储, 头实体和尾实体对应应有向图中的结点, 关系对应应有向图的边。由于知识图谱是由多个三元组组成, 因此每个用户-物品交互信息可表示为 $(U, \text{interact}, V)$ 三元组的形式, interact 表示用户对物品产生过交互。并且物品-属性信息同样也可表示为三元组形式, 将物品属性信息以结点的形式添加到知识图谱中。

对已构建好的知识图谱提取不同用户-物品对的多条路径信息, 将路径信息进行编码, 输入到模型中训练出相应的用户和物品预测向量, 对预测向量相乘从而得到每个用户-物品的预测值, 生成 top-N 推荐列表^[21], 将预测推荐列表与用户已交互的列表进行对比, 以此来计算出模型的精确率和召回率。

2.2 KG-BiLSTM 算法

2.2.1 预处理层

对已构建好的知识图谱提取不同用户-物品对的路径信息。在该推荐算法当中 (U_i, V_j) 表示一个实体对, 其中 U 代表用户集合, V 代表物品集合。 i 代表某个用户, j 代表某个物品。若存在用户 i 对物品 j 产生过交互, 则利用不同长度的路径来连接该实体对。即存在 $P(U_i, V_j) = \{P_1, P_2, \dots, P_m\}$, m 的值代表用户-物品对路径条数, 对于其中的任意一条路径存在 $p_m = e_0 \xrightarrow{r_1} e_1 \xrightarrow{r_2} e_2 \longrightarrow \dots \xrightarrow{r_n} e_n$ 。其中 $e_0 = U_i, e_n =$

V_j, p_m 代表用户-物品对的第 m 条偏好路径, n 表示经过的结点个数。 r_1, r_2, \dots, r_n 表示连接偏好路径相对应的各个关系。其中所经过的 e_1, e_2 表示经过的物品属性结点。每条路径有其不同的语义表示。将生成的各条路径送入嵌入层, 进行编码。

2.2.2 嵌入层

引入嵌入层的目的是在于将高维稀疏的特征向量转换成低维向量, 便于模型输入。对于预处理层中涉及各个结点 e_i , 首先生成预编码向量。针对已构建好的路径, 提取路径中涉及各个结点, 将其输入嵌入层中, 进行模型训练。利用嵌入层生成每个结点相对应的嵌入向量 $e_n = \{e_{n1}, e_{n2}, e_{n3}, \dots, e_{nr}\}$, r 表示嵌入向量的维数。从而将每个结点表示成低维向量的形式。每个嵌入向量中的各个分量代表结点的特征, 将含有结点特征的嵌入向量输入隐藏层中, 使路径包含一定的语义信息。

2.2.3 Bi-LSTM 层

将每个用户-物品对各条路径信息, 按其结点的先后顺序依次输入到双向循环神经网络当中, 考虑到前端结点与末尾结点距离较远, 本文采用 Bi-LSTM 模型, 可以充分融合前端结点的特征信息, 从而保证特征提取的完整性。其中双向循环神经网络 Bi-LSTM 如图 4 所示。

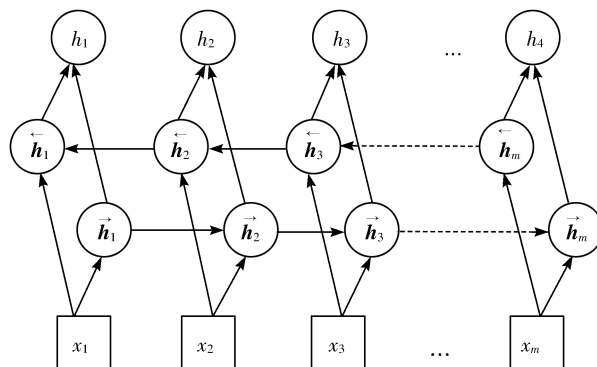


图4 双向循环神经网络 Bi-LSTM 模型图

其中, \vec{h}_m 为模型前向传播隐藏向量, \overleftarrow{h}_m 为模型后向传播隐藏向量。将 \vec{h}_m 和 \overleftarrow{h}_m 相结合, 生成最终的模型隐藏向量。LSTM 具体计算公式如下:

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad (1)$$

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (2)$$

$$\tilde{c}_t = \tanh(W_c[h_{t-1}, x_t] + b_c) \quad (3)$$

$$c_t = f_t \times c_{t-1} + i_t \times \tilde{c}_t \quad (4)$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = o_t \times \tanh(c_t) \quad (6)$$

其中, W_i, W_f, W_c, W_o 分别为输入门权重、遗忘门权重、细胞状态权重和输出门权重。 b_i, b_f, b_c, b_o 分别为输入门偏置、遗忘门偏置、细胞状态偏置和输出门偏置。 x_t 为当前时刻网络的输入值, h_{t-1}, c_t 为上一时刻 LSTM 网络的输出值和细胞状态值。 σ 代表 sigmoid 函数, \tanh 为双曲正切激活函数。通过梯度下降法进行迭代, 更新模型参数。将嵌入层已编码的向量输入到 Bi-LSTM 当中进行训练。

2.2.4 注意力层

对于神经网络已训练出的向量, 引入一层注意力机制, 目的是为了根据路径中不同结点的重要性程度进行向量融合, 注意力层公式定义如下:

$$a_m = \text{softmax}(o_t \times h_t) \quad (7)$$

$$A_m[ij] = a_m \cdot o_t \quad (8)$$

其中, o_t 为 Bi-LSTM 层的输出向量, h_t 为 Bi-LSTM 层的隐藏向量, a_m 为某用户-物品对的第 m 条路径的注意力向量, 将求得的注意力向量与 Bi-LSTM 层输出向量相乘, 得到某用户-物品对的第 m 条路径的注意力层向量 $A_m[ij]$ 。利用注意力层将路径中出现的每个结点信息按其重要性程度充分融合, 保证路径中结点信息的完整性。

2.2.5 预测层

预测层包含池化层和全连接层 2 个部分。

首先将输出的注意力层向量输入到平均池化层, 用于区分某用户-物品对不同路径的重要程度, 使模型对不同语义赋予不同的重要性权值。其具体公式如下:

$$h[ij] = \text{avg}(A_m[ij]) \quad (9)$$

其中, m 表示用户-项目对第 m 条路径, $A_m[ij]$ 为每个 (U_i, V_j) 数对的第 m 条路径注意力层向量。 $h[ij]$ 为经过平均池化层后, 各路径表示的平均值。

然后, 在平均池化层后加入一层全连接层, 对向量进行降维, 其公式如下:

$$O_{ij} = f(h[ij]) = \sigma(W_f \cdot h[ij] + b_f) \quad (10)$$

其中, $f(h[ij])$ 为以 $h[ij]$ 作为自变量的全连接层函数, W_f 为全连接层系数, b_f 为全连接层偏置, 其值根据模型训练结果动态调整。 σ 为 sigmoid 激活函数, O_{ij}

为最终的预测向量。

最后将经过平均池化层后的向量与全连接层系数相乘, 加上全连接层偏置, 最终得到预测向量。

2.3 损失函数

本文利用 BCELoss() 交叉熵损失函数对模型进行训练^[22]。交叉熵损失函数具体公式如下:

$$\text{loss}(o, t) = -\frac{1}{N} \sum_{i=1}^N [t_{ij} \times \log(O_{ij}) + (1 - t_{ij}) \times \log(1 - O_{ij})] \quad (11)$$

其中, i 代表某个用户 i , N 代表用户总数。 O_{ij} 表示 (U_i, V_j) 对的最最终预测向量, t_{ij} 表示 (U_i, V_j) 对的目标向量, 若 (U_i, V_j) 数对存在于正向路径信息文件中, 则 t_{ij} 为 1, 否则, t_{ij} 为 0。 $\log()$ 为对数函数。 $\text{loss}(o, t)$ 为以 o_{ij} 作为输入向量, t_{ij} 作为目标向量的损失函数值。

2.4 本文算法流程

本文算法流程如算法 1 所示。

算法 1 KG-BiLSTM 算法

输入: 用户-物品交互矩阵 R 、物品属性信息、用户、物品和物品属性的预编码向量。

输出: 根据预测值生成用户 U 对物品 V 的 top- N 推荐列表。

1) 构建知识图谱 KG。将用户-物品交互矩阵 R 和物品属性信息中涉及的物品及其属性以结点的形式添加到有向图中并用边对结点相连。

2) 提取知识图谱 KG 中不同用户-物品对的路径信息。

3) 将已提取到的路径信息输入到嵌入层中, 利用 embedding 生成嵌入向量。

4) 将各路径中涉及的结点嵌入向量输入到 Bi-LSTM 层中, 利用式 (1) ~ 式 (6) 计算出各结点的 Bi-LSTM 层隐藏向量。

5) 将每条路径上的各结点隐藏向量, 输入到注意力层中, 用于融合较远结点的属性信息, 利用式 (7) ~ 式 (8) 计算得出整条路径的注意力层向量。

6) 整合各条路径的注意力层向量, 将该向量输入到平均池化层中, 利用式 (9) 对不同路径的重要程度进行区分, 根据重要性权值对多条路径向量进行融合。

7) 将生成的平均池化层向量输入到全连接层中, 用于对向量进行降维, 利用式 (10) 计算得出用户-物品对的最最终预测向量。

8) 利用式 (11) 计算出模型的损失函数值。

9) 利用 SGD 优化器对参数进行梯度更新, 通过不断训练, 更新模型参数值, 以达到使损失函数最小化的目的。

3 实验与分析

3.1 数据集与评价指标

3.1.1 实验数据集

本文选用 MovieLens-1M 和 Yelp 作为实验数据集。MovieLens-1M 数据集是目前推荐算法研究领域应用最为广泛的数据集。它是由 MovieLens 用户提供的电影评分信息组成,其中包含了 756684 条匿名的评分数据,共涉及 6040 个用户以及 3382 个电影。Yelp 数据集中包含了不同的地理位置信息,适用于对位置进行推荐,其中包含了 229178 条评分数据,涉及 37940 个用户和 11516 个物品。数据集具体信息如表 1 所示。

表 1 数据集详细信息

项目		数据集	
		MovieLens-1M	Yelp
用户-物品交互信息	用户数	6040	37940
	物品数	3382	11516
	交互数	756684	229178
	稀疏性/%	3.704	0.052
知识图谱信息	实体数	14578	42302
	实体类型数	10	7
	关系	99975	299872
	关系类型数	13	7
	图稀疏性/%	5.275	1.772

3.1.2 评价指标

本文对上述 2 个数据集分别通过计算模型精确率和召回率来对模型性能进行评价。利用推荐算法对物品进行评分预测,将评分由高到低进行排序,得出某个用户的 Top-N 推荐列表^[23]。精确率是针对预测结果而言,被预测为正样本占真正正样本的比重。召回率是针对原样本而言,被预测为正样本占总测试样本的比重^[24]。其公式如下:

$$\text{Precision} = \frac{\sum_u |R(u) \cap T(u)|}{\sum_u |R(u)|} \quad (12)$$

$$\text{Recall} = \frac{\sum_u |R(u) \cap T(u)|}{\sum_u |T(u)|} \quad (13)$$

其中, $R(u)$ 表示根据用户 u 在训练集上的偏好行为得出的推荐列表, $T(u)$ 表示用户 u 在测试集上的推荐列表。

另外本文还采用了 MRR(Mean Reciprocal Rank) 指标来对该推荐算法性能进行评估。MRR 是指被预测为正样本的结果值在预测结果中的排名。其公式如下:

$$\text{MRR} = \frac{1}{Q} \sum_{i=1}^{|Q|} \frac{1}{\text{rank}_i} \quad (14)$$

其中, $|Q|$ 是用户总数, rank_i 是用户 i 在推荐列表中第一个被推荐物品在用物品列表中的位置。

3.2 实验分析

本节将针对上述提到的算法进行对比实验。实验包括每个用户-物品对路径条数的选取,每个用户-物品对路径长度的选取,以及本文算法与目前较为先进算法的实验对比。本文将输入向量维度设置为 10,隐藏层向量维度设置为 16,输出向量维度设置为 1,学习率 learning_rate 设置为 0.2,模型迭代次数设置为 10。并且利用 SGD 优化器来对模型进行优化。

3.2.1 路径条数选取

本文路径条数在{5,10,20} 中选取。Top-N 推荐中, N 的值设置为{5,10},其中召回率 Recall 和 MRR 均为 $N=10$ 时的实验结果。本文实验结果如表 2 所示。

表 2 选取不同路径条数的实验结果

数据集	评价指标	路径条数		
		5	10	15
Movie Lens-1M	Precision@ 5	0.2782	0.2836	0.2793
	Precision@ 10	0.2324	0.2569	0.2439
	Recall	0.2238	0.2385	0.2301
	MRR	0.8559	0.8756	0.8622
Yelp	Precision@ 5	0.0846	0.0875	0.0859
	Precision@ 10	0.0795	0.0802	0.0798
	Recall	0.0743	0.0762	0.0749
	MRR	0.3972	0.4027	0.3994

由表 2 实验结果可以看出,Yelp 数据集的预测精确率、召回率和 MRR 均比 MovieLens-1M 数据集低,这是由于 Yelp 相比 MovieLens-1M 数据集更稀疏,因此其预测精确率更低。除此之外,随着路径条数的增多,在 2 个数据集上的预测精确率也在随之增高,这是由于多条路径信息的引入丰富了实体的语义信息,对于某一用户而言,可更加全面地学习到多条路径的特征。但当路径条数为 15 时,引入过多的路径条数,会使无关路径信息增加,影响预测的精确率。从本文实验结果中可观察到,当路径条数为 10 时,2 个数据集的精确率和召回率到达了最优值,因此本文选取 10 作为最终路径条数。

3.2.2 路径长度选取

本文路径长度在{3,5,7} 中进行选取,并且 N 的值同样设置为{5,10},召回率 Recall 和 MRR 均为 $N=10$ 时的实验结果,路径条数设置为 10。实验结果如表 3 所示。

表 3 选取不同路径长度的实验结果

数据集	评价指标	路径长度		
		3	5	7
Movie Lens-1M	Precision@ 5	0.2841	0.3062	0.2896
	Precision@ 10	0.2327	0.2601	0.2401
	Recall	0.2321	0.2326	0.2399
	MRR	0.8628	0.8771	0.8669
Yelp	Precision@ 5	0.0946	0.0982	0.0962
	Precision@ 10	0.0924	0.0947	0.0934
	Recall	0.0892	0.0901	0.0893
	MRR	0.3841	0.4064	0.3981

由表 3 实验数据可以看出,随着路径长度的增加,在 2 个不同数据集上的预测精确度和召回率都有所增加。这是由于随着路径长度的增加,由该路径学习得到的特征也在增加,因此可以融合多种实体特征从而增强预测的精确性。但当路径长度为 7 时,2 个数据集的预测精确性和召回率开始下降,这是由于选取较长的路径,会引入过多的远距离结点信息,导致引入更多的无关结点,从而影响预测的精确率。因此,本文选取 5 作为最终路径长度。

3.2.3 不同模型对比结果

本节将本文提出的 KG-BiLSTM 算法与目前主流推荐算法相比较,其中用于对比实验的推荐算法包括:

1) 基于嵌入的协同知识库推荐算法 CKE。该算法在数据集层以有向图的形式构建知识图谱,在嵌入层对知识图谱利用 TransR 算法进行知识嵌入,得到低维向量,协同过滤层整合图谱、文本、图片等各类型向量信息,结合协同过滤算法进行推荐结果预测。

2) 基于知识图谱嵌入的推荐算法 MKR。该算法主要是利用知识图谱嵌入技术,结合传统的矩阵分解算法以及深度神经网络,来进行模型预测。

3) 基于协同深度学习的推荐算法 CDL。该算法抽取实体中的特征信息,通过计算不同物品、用户之间的相似度来进行物品推荐。

4) 深度知识感知推荐算法 DKN。该算法利用 TransE 对实体、单词和上下文信息进行编码,生成特征向量,将卷积神经网络与注意力机制相融合进行推荐。

5) 基于用户偏好传播的推荐算法 RippleNet。这是基于知识图谱路径的推荐算法,将用户对物品的交互信息按层进行扩散,结合多层扩散结果实现推荐结果预测。

本文针对各个算法的预测精确率、召回率和 MRR 指标进行实验,得出在 MovieLens-1M 数据集上关于 Precision@ 5 和 Precision@ 10 的实验结果如表 4 所示。在 Yelp 数据集上关于 Precision@ 5 和 Precision@ 10 的实验结果如表 5 所示。

表 4 MovieLens-1M 数据集上的不同模型精确率对比结果

算法模型	Precision@ 5	Precision@ 10
CKE	0.2623	0.2388
MKR	0.2598	0.2305
CDL	0.2613	0.2392
DKN	0.2693	0.2401
RippleNet	0.2739	0.2476
KG-BiLSTM	0.2836	0.2569

表 5 Yelp 数据集上的不同模型精确率对比结果

算法模型	Precision@ 5	Precision@ 10
CKE	0.0728	0.0693
MKR	0.0654	0.0634
CDL	0.0681	0.0603
DKN	0.0742	0.0723
RippleNet	0.0834	0.0771
KG-BiLSTM	0.0875	0.0802

由多种模型对比实验得出,本文提出的 KG-BiLSTM 算法和 RippleNet 算法的预测精确率均比其他基于知识图谱嵌入技术算法的精确率更高,这是因为基于知识图谱路径的推荐算法,不仅局限于学习知识的嵌入向量,更重要的是引入了路径信息的概念,将路径信息与知识的嵌入相结合,利用模型搭建,学习到整条路径的向量表示,由此来更好地融合多特征信息,因此预测精确率高于其他几种算法。并且本文由于考虑到较远结点特征信息无法全面有效地融入到向量种的问题,因此引入了双向循环神经网络,对整条路径的特征进行前向和后向的双向特征提取,除此之外,加入了一层注意力机制,更好地将前端结点特征融合到路径特征向量之中,因此本文提出的算法预测精确率较 RippleNet 算法更高,当 $N = 10$ 时,2 个数据集的精确率较 RippleNet 分别高出 3.76% 和 4.02%。

不同算法模型在召回率 Recall 和 MRR 指标上的实验对比结果如图 5、图 6 和表 6 所示。

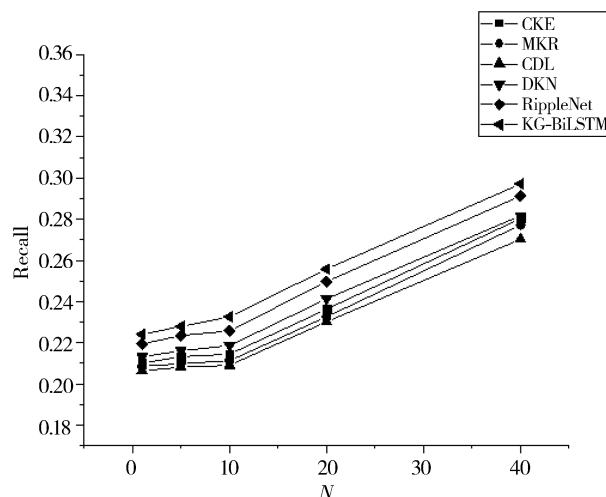


图 5 MovieLens-1M 数据集上不同模型的 Recall 对比结果

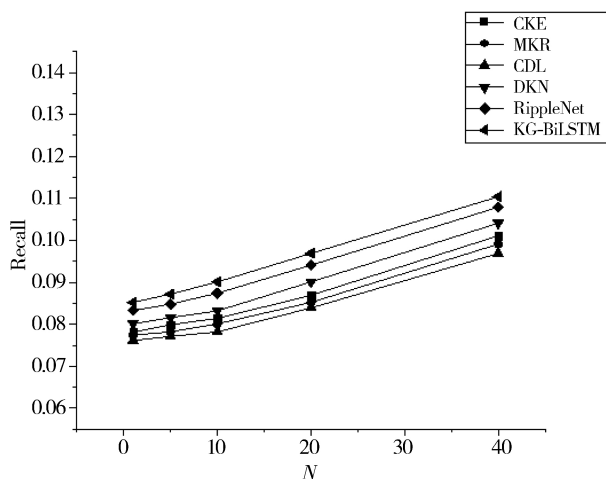


图 6 Yelp 数据集上不同模型的 Recall 对比结果

表 6 2 种数据集上不同模型的 MRR@10 对比结果

算法模型	MRR@10	
	MovieLens-1M	Yelp
CKE	0.8519	0.3929
MKR	0.8508	0.3917
CDL	0.8503	0.3909
DKN	0.8531	0.3947
RippleNet	0.8549	0.3961
KG-BiLSTM	0.8771	0.4054

根据图 5、图 6 和表 6 实验对比结果可以看出, 本文提出的 KG-BiLSTM 和 RippleNet 算法在 2 种数据集上的召回率 Recall 和 MRR 指标方面, 整体上高于其他 4 种对比算法, 这是由于引入了多种路径信息, 而不仅是考虑对实体进行嵌入向量的训练。除此之外, 本文提出的算法在 2 种数据集上的召回率和 MRR 都优于其他算法, 这是由于本文提出的算法, 将知识图谱嵌入向量和路径提取相结合, 并考虑较远路径结点的问题, 从而保证实体信息的全面获取, 从而为用户提供更加精准的 Top-N 推荐。当 $N=10$ 时, 2 个数据集在召回率上较主流算法 RippleNet 提高了 3.06% 和 3.09%。在 MRR 上提高了 2.60% 和 2.35%。证实了本文算法的有效性。

3.3 可解释性分析

本文由于引入了知识图谱路径的概念, 而知识图谱天生具有可解释性, 因此可以为用户推荐的 Top-N 列表提供一定的可解释性。由于推荐算法的可解释性提高, 可更加有效地帮助用户理解推荐行为, 从而使用户更加信任该推荐, 增加用户接受被推荐物品的概率。因此提高推荐算法的可解释性对于推荐质量而言十分重要。本节以一个具体的推荐实例, 来详细阐述推荐的可解释性。具体实例如图 7 所示。

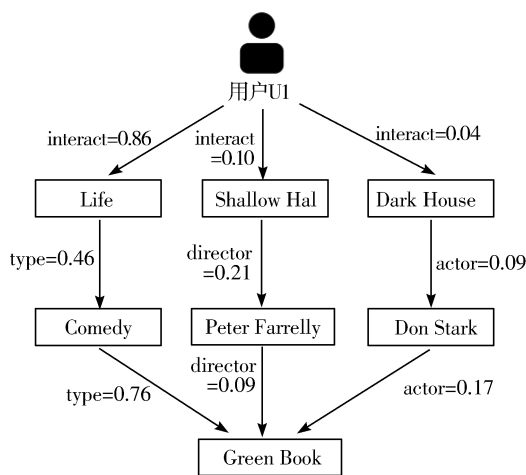


图 7 算法可解释性分析图

图 7 展示了对于 MovieLens-1M 数据集来说, 存在用户 U1 对 3 部电影 Life、Shallow Hal、Dark House 感兴趣, 且存在 3 条不同的路径, 分别根据电影类型、电影导演、电影演员来进行推荐, 根据图中注意力系数可知, 路径 $U1 \xrightarrow{\text{interact}} \text{Life} \xrightarrow{\text{type}} \text{Comedy} \xrightarrow{\text{type}} \text{Green Book}$ 中系数权重最高, 因此可推测出用户对喜剧类型的电影最为感兴趣, 因此向该用户推荐了具有相同喜剧类型的电影 Green Book。

除此之外, 由于用户对其他拥有不同类型、导演、演员的电影同样感兴趣, 因此该模型可从不同的角度来分析推断出用户 U1 的具体偏好, 并生成相应的推荐列表。并且, 对于推荐列表中的每一部预测电影来说, 都是由多条路径的不同属性融合得出的推荐结果, 而不仅由单一的路径生成。因此这意味着, 对于每个预测结果来说, 用户-物品的交互信息会受到不同路径不同属性信息的影响, 本文模型可有效对多路径的重要性进行区分, 从而有效地提高推荐的精确率, 并为推荐结果提供合理解释。

3.4 冷启动分析

本文选取用户评分小于等于 4 的用户-项目对作为冷启动数据集, 进行实验分析。实验参数与 3.2 节实验参数设置相同, 该算法与其他主流算法在 2 种冷启动数据集上的算法精确率 Precision、召回率 Recall 和 MRR 实验结果如表 7 和表 8 所示。

表 7 MovieLens-1M 数据集上不同模型的实验对比结果

算法模型	Precision@5	Precision@10	Recall	MRR
CKE	0.0859	0.0902	0.0832	0.3329
MKR	0.0837	0.0893	0.0822	0.3304
CDL	0.0801	0.0886	0.0805	0.3291
DKN	0.0926	0.0903	0.0908	0.3468
RippleNet	0.0983	0.0926	0.0933	0.3496
KG-BiLSTM	0.1163	0.1003	0.1129	0.3517

表 8 Yelp 数据集上不同模型的实验对比结果

算法模型	Precision@5	Precision@10	Recall	MRR
CKE	0.0271	0.0269	0.0269	0.1831
MKR	0.0269	0.0261	0.0258	0.1826
CDL	0.0254	0.0250	0.0233	0.1776
DKN	0.0283	0.0276	0.0276	0.1843
RippleNet	0.0321	0.0319	0.0298	0.1864
KG-BiLSTM	0.0326	0.0312	0.0313	0.1968

由表 7 和表 8 实验结果可以看出,本文提出的算法在 MovieLens-1M 冷启动数据集上的实验结果优于其他主流模型,而在 Yelp 冷启动数据集上的 Precision@10 实验结果略低于 RippleNet 模型,存在的原因是由于 Yelp 冷启动数据集的知识图谱较为稀疏,无法有效地获取到全面的知识图谱路径信息,因此相比主流模型精确率略低。

4 结束语

本文针对推荐算法中存在的可解释性和冷启动等问题提出了 KG-BiLSTM 算法,该算法将知识图谱实体嵌入和路径提取相结合,并充分考虑路径中各个结点的特征信息,针对用户已交互的物品,根据其各结点的类型,学习不同路径的特征,并最终进行向量融合,实现 Top-N 预测。该模型在 2 个真实数据集上进行了实验分析,其结果均优于目前主流算法。在未来,还将考虑融入用户结点的属性信息,从而进一步提高算法的预测精确率。

参考文献:

- [1] SHAPIRA B. Recommender Systems Handbook [M]. Springer, 2011.
- [2] JANNACH D, ZANKER M, Felfernig A, et al. Recommender Systems: An Introduction [M]. Cambridge University Press, 2010.
- [3] XIA F, LIU H F, LEE I, et al. Scientific article recommendation: Exploiting common author relations and historical preferences [J]. IEEE Transactions on Big Data, 2016,2(2): 101-112.
- [4] 杨武,唐瑞,卢玲. 基于内容的推荐与协同过滤融合的新闻推荐方法 [J]. 计算机应用, 2016,36(2): 414-418.
- [5] LINDEN G, SMITH B, YORK J. Item-to-item collaborative filtering [J]. Internet Computing, 2003,7(1): 76-80.
- [6] 李玉省. 个性化推荐系统关键技术研究 [D]. 北京: 北京邮电大学, 2016.
- [7] PRIYANKA R D, VIJENDRA S. Systematic evaluation of social recommendation systems: Challenges and future [J]. International Journal of Advanced Computer Science and Applications, 2016,7(4), DOI: 10.14569/IJACSA.2016.070420.
- [8] 徐增林,盛泳潘,贺丽荣,等. 知识图谱技术综述 [J]. 电子科技大学学报, 2016,45(4): 589-606.
- [9] KETHAVARAPU U P K, SARASWATHI S. Concept based

dynamic ontology creation for job recommendation system [J]. Procedia Computer Science, 2016,85:915-921.

- [10] 漆桂林,高桓,吴天星. 知识图谱研究进展 [J]. 情报工程, 2017,3(1): 4-25.
- [11] ZHANG F Z, YUAN N J, LIAN D F, et al. Collaborative knowledge base embedding for recommender systems [C] // Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016: 353-362.
- [12] WANG H, WANG N Y, YEUNG D Y. Collaborative deep learning for recommender systems [C] // Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2015: 1235-1244.
- [13] WANG H W, ZHANG F Z, XIE X, et al. DKN: Deep knowledge-aware network for news recommendation [J]. Machine Learning, 2018, arXiv: 1801.08284.
- [14] QIAN F L, ZHAO S, TANG J, et al. SoRS: Social recommendation using global rating reputation and local rating similarity [J]. Physica A Statistical Mechanics & Its Applications, 2016,461: 61-72.
- [15] SHI C, ZHANG Z Q, LUO P, et al. Semantic path based personalized recommendation on weighted heterogeneous information networks [C] // Proceedings of the 24th ACM International Conference on Information and Knowledge Management. 2015: 453-462.
- [16] WANG H W, ZHANG F Z, WANG J L, et al. RippleNet: Propagating user preferences on the knowledge graph for recommender system [C] // Proceedings of the 27th ACM International Conference on Information and Knowledge Management. 2018: 417-426.
- [17] 唐晓波,魏巍. 基于本体的推荐系统研究综述 [J]. 图书馆学研究, 2016(18): 7-12.
- [18] SHI C, HU B B, ZHAO W X, et al. Heterogeneous information network embedding for recommendation [J]. IEEE Transactions on Knowledge and Data Engineering, 2019,31(2): 357-370.
- [19] FENG Y, QIANG L, SHU W, et al. A dynamic recurrent model for next basket recommendation [C] // International ACM SIGIR Conference on Research & Development in Information Retrieval. 2016: 729-732.
- [20] ZHU Y, ZHU J X, HOU J, et al. A brand-level ranking system with the customized attention-GRU model [C] // Proceedings of the 27th International Joint Conference on Artificial Intelligence. 2018: 3947-3953.
- [21] SUN Y Z, HAN J W, YAN X F, et al. PathSim: Meta path-based Top-K similarity search in heterogeneous information networks [J]. Proceedings of the VLDB Endowment, 2011,4(11): 992-1003.
- [22] YANG X L, GOBEAWAN L, YEO S Y, et al. Automatic segmentation of left ventricular myocardium by deep convolutional and de-convolutional neural networks [C] // 2016 Computing in Cardiology Conference. 2016: 81-84.
- [23] 朱郁筱,吕琳媛. 推荐系统评价指标综述 [J]. 电子科技大学学报, 2012,41(2): 163-175.
- [24] JAIN A, SINGH P, DHAR J. Multi-objective item evaluation for diverse as well as novel item recommendations [J]. Expert Systems with Applications, 2019,139(4), DOI: 10.1016/j.eswa.2019.112857.