(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 111259133 A
(43)申请公布日 2020.06.09

(21)申请号 202010054209.1

(22)申请日 2020.01.17

(71)申请人 成都信息工程大学
地址 610225 四川省成都市双流区西南航空港经济开发区学府路1段24号
申请人 四川省金科成地理信息技术有限公司
成都探码科技有限公司

(72)发明人 乔少杰 韩楠 沈杰 宋学江 程维杰 魏军林 张小辉 丁超 肖月强 陈文林 李斌勇 张吉烈 张永清 何林波 元昌安 彭京 周凯 余华 范勇强 冉先进

(74)专利代理机构 成都正华专利代理事务所（普通合伙）51229
代理人 李蕊 陈选中

(51)Int.Cl.
G06F 16/335(2019.01)
G06F 16/36(2019.01)
G06F 16/9535(2019.01)

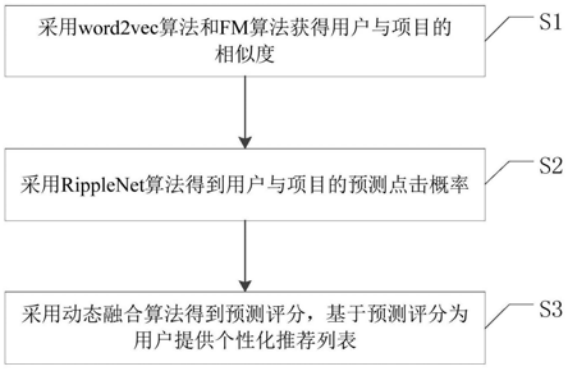
权利要求书2页 说明书10页 附图3页

(54)发明名称

一种融合多信息的个性化推荐方法

(57)摘要

本发明公开了一种融合多信息的个性化推荐方法，该方法包括采用word2vec算法和FM算法获得用户与项目的相似度，采用RippleNet算法得到用户与项目的预测点击概率，采用动态融合算法得到预测评分，基于预测评分为用户提供个性化推荐列表。本发明将知识图谱与评论内容作为多源数据，并使用不同算法对数据进行处理，并采用动态融合方法进行有效结合，为用户提供更精准的个性化推荐服务，能够实现更好的推荐效果，并且可以有效地解决数据稀疏带来的推荐准确性降低的问题。



CN 111259133 A

1.一种融合多信息的个性化推荐方法,其特征在于,包括以下步骤:

S1、获取用户-项目评论数据集,采用word2vec算法分别获取用户和项目的特征词向量,并采用FM算法获得用户与项目的相似度;

S2、根据用户的历史点击项目信息构建用户与项目的交互矩阵,并结合知识图谱,采用RippleNet算法得到用户与项目的预测点击概率;

S3、将步骤S1得到的用户与项目的相似度和步骤S2得到的用户与项目的预测点击概率采用动态融合算法进行动态融合得到预测评分,基于预测评分为用户提供个性化推荐列表。

2.如权利要求1所述的融合多信息的个性化推荐方法,其特征在于,所述步骤S1具体包括以下分步骤:

S1-1、获取数据库中所有用户-项目评论信息,采用word2vec算法将一个用户对所有项目的评论合成代表用户信息的文本数据,并将一个项目接受到的所有用户的评论集成为项目的文本数据;

S1-2、采用word2vec算法分别对步骤S1-1得到的用户信息的文本数据和项目的文本数据进行向量化处理,得到用户和项目的特征词向量;

S1-3、采用FM算法将步骤S1-2得到的用户和项目的特征词向量进行两两组合,并添加交叉项特征,得到用户与项目的相似度。

3.如权利要求2所述的融合多信息的个性化推荐方法,其特征在于,所述步骤S1-3中,FM算法的模型表示为:

$$\bar{y}_{u,v}(z_{uv}) = m_0 - m^T z_{uv} + \frac{1}{2} z_{uv}^T M z_{uv}, M_{j,c} = i_j^T i_c, j \neq c$$

其中,m_0表示全局偏差项,m为用户u与项目v的特征向量z_{uv}的系数向量,M为二阶交互的权重矩阵,$M_{j,c}$为M的第j行c列的值,i_j,i_c为与z_{uv}的特征维度j和c相关的i维隐向量。

4.如权利要求3所述的融合多信息的个性化推荐方法,其特征在于,所述步骤S1采用平方损失作为参数优化的目标函数,表示为:

$$J_{sqr} = \sum_{(u,v)\in O} (y_{u,v} - \bar{y}_{u,v})^2 - \lambda_\Theta \|\Theta\|^2$$

其中,O表示观察到的用户-项目评分对集合,$y_{u,v}$表示用户u与项目v的交互历史,Θ表示所有参数,λ_Θ表示L2正则化参数。

5.如权利要求4所述的融合多信息的个性化推荐方法,其特征在于,所述步骤S2具体包括以下分步骤:

S2-1、设定用户集和项目集分别为U=$\{u_1,u_2,...,u_m\}$和V=$\{v_1,v_2,...,v_n\}$,构建用户与项目的交互矩阵,表示为:

$Y_{uv}=\{y_{uv}|u\in U,v\in V\}$

其中,$y_{u,v}$表示用户u与项目v的交互历史,m表示用户数量,n表示项目数量;

S2-2、根据用户与项目的交互矩阵和包含关系-实体三元组的知识图谱,定义用户u的第k个关联实体为:

$$\varepsilon_u^k = \left\{ t | (h,r,t) \in G, h \in \varepsilon_u^{k-1} \right\}, k = 1,2,...H$$

其中，(h,r,t)表示知识图谱包含的关系-实体三元组，h表示头实体，r表示关系，t表示尾实体，H表示与原点项目所关联的最远位置；

定义用户u在知识图谱G上的第k跳波纹集为：

$$S_u^k=\left\{(h,r,t)|(h,r,t)\in G,h\in \varepsilon_u^{k-1}\right\};$$

S2-3、对于每一个项目v对应创建一个d维的嵌入向量v，将用户u的第1跳波纹集的每一个三元组(h_i,r_i,t_i)与v的相关系数为：

$$p_i=softmax(v^T\cdot R^i\cdot h^i)=\frac{\exp(v^T\cdot R^i\cdot h^i)}{\sum_{(h_i,r_i,t_i)}\exp(v^T\cdot R^i\cdot h^i)}$$

其中，R^i表示关系r_i的嵌入向量，h^i表示头实体h_i的嵌入向量；

S2-4、根据相关系数对用户u的第一跳波纹集的尾实体t_i计算加权和得到用户u对于项目v的一阶反响为：

$$o_u^1=\sum_{(h,r,t)\in S_u^1}p_it_i$$

根据用户u对于项目v的多阶反响，定义项目v的用户u的嵌入向量为：

$$u=\alpha_1o_u^1+\alpha_2o_u^2+...+\alpha_Ho_u^H$$

其中，α_i为正项混合参数；

S2-5、根据项目v的用户u的嵌入向量得到用户与项目的预测点击概率，表示为：

$$\hat{y}_{uv}(z_{KG})=\sigma(u^Tv)=\frac{1}{1+\exp(-u^Tv)}$$

其中，z_{KG}表示基于知识图谱数据的推荐结果。

6.如权利要求5所述的融合多信息的个性化推荐方法，其特征在于，所述步骤S2中RippleNet算法的损失函数表示为：

$\Gamma=\Sigma_{(u,v)\in Y}-y_{uv}\log\sigma(u^Tv)+(1-y_{uv})\log(1-\sigma(u^Tv))$。

7.如权利要求6所述的融合多信息的个性化推荐方法，其特征在于，所述步骤S3中，将用户与项目的相似度和用户与项目的预测点击概率采用动态融合算法进行动态融合得到预测评分，表示为：

$$\bar{y}_{u,v}=\alpha\bar{y}_{u,v}(z_{review})+(1-\alpha)\bar{y}_{u,v}(z_{KG})$$

其中，$\alpha=\frac{\bar{y}_{u,v}(z_{review})}{\bar{y}_{u,v}(z_{review})+\bar{y}_{u,v}(z_{KG})}$，$z_{review}$表示基于文本评论数据的推荐结果。

一种融合多信息的个性化推荐方法

技术领域

[0001] 本发明属于推荐系统技术领域,具体涉及一种融合多信息的个性化推荐方法。

背景技术

[0002] 随着诸如人工智能,云计算与大数据技术以及移动互联网等先进技术的迅速发展,各种信息数据的规模也呈现爆炸式的增长。在享受这些数据带来的便利的同时,需要处理数据量过大而导致的“信息过载”问题。推荐系统则正是解决“信息过载”问题的有效方法之一,可以根据用户以及项目(item)的相关属性找到用户的兴趣点,并且将用户感兴趣的项目以个性化目录的方式推荐给用户。

[0003] 目前,基于协同过滤的推荐系统考虑到用户与项目的历史交互,然后根据其潜在特征为用户提出推荐建议,已经取得一些效果。但是基于协同过滤的推荐系统通常会面临用户和商户历史交互数据的稀疏性问题及伴随的冷启动问题。为了解决这些局限性,研究人员将诸如用户/项目属性、社交网络、图像、背景等辅助信息纳入基于协同过滤的推荐系统。

[0004] 在各种辅助信息中,知识图谱(Knowledge Graph,简称KG)由于其具有效率极高的事实描述能力以及可解释项目之间的关联信息,得到研究人员的广泛关注。知识图谱是一种定向异构图,其中节点对应于实体,边对应于关系。研究人员已经提出了诸多知识图谱,例如:NELL,DBpedia,以及诸如Google Knowledge Graph和Microsoft Satori的商业知识图谱。这些知识图谱已成功应用于多个领域,例如知识图谱填充,人机问答,单词嵌入(word embedding)[10]和文本分类。

[0005] 深度学习是当前互联网和人工智能的一个研究热点。深度学习主要是通过将低层的属性特征生成高层语义抽象,自动挖掘出数据的分布式特征表示,解决了传统机器学习中需要人工设计特征的问题,在图像识别,机器翻译等诸多领域取得了较大进展。基于深度学习的推荐系统近期得到广泛关注,是将与用户和商品项目相关数据作为输入,通过深度学习模型获得具有相应属性特征的用户和项目的隐表示,再基于这类隐表示为用户推荐项目。

[0006] 知识图谱在各个领域中得到广泛应用,研究人员尝试利用知识图谱来改进推荐系统的性能。现有基于知识图谱的推荐系统分为如下两类:

[0007] (1)基于嵌入(embedding)的方法,此类方法使用知识图谱嵌入(knowledge graph embedding,KGE)算法预处理KG,并将学习到的实体嵌入到推荐系统框架中。基于嵌入的方法利用KG辅助推荐系统提升算法的灵活性,但是上述方法采用的KGE算法更适用于链路预测而不是推荐系统。

[0008] (2)基于路径的方法,此类方法探索KG中各实体之间的关联模式,作为推荐系统的额外辅助信息。基于路径的方法以更直观的方式使用KG,但是会严重依赖于人工设定的元路径,通用性无法保障,不同的应用场景需要设定不同的元路径。此外,实体和关系不在一个领域内的某些场景(例如,新闻推荐)中是无法人工设计元路径的。

[0009] 文献较早将图嵌入技术应用于推荐领域。将Movielens中电影与用户信息嵌入(embedding)到同一个向量空间，进而计算用户与电影之间的空间距离，生成推荐列表。Wang等人将医学知识图谱、疾病&患者二部图、疾病&药物二部图分别嵌入低维向量空间，为病患推荐更为安全的药物治疗方式。通过加权平均将知识图谱与二部图结合生成包含更加细粒度属性信息的患者和药物向量，最终生成对给定患者的药物top-k列表。

[0010] Ostuni等人融合KG路径中隐含的语义反馈信息，提出基于隐式语义反馈的路径算法SPrank。基于路径特征对数据集进行挖掘，以捕获项目之间的复杂关系。SPrank的主要思想是探索语义图中的路径，以便找到与用户感兴趣项目相关的项目。通过分析路径，提取基于路径的特征，利用随机森林与渐变增强回归树相结合的学习算法来生成推荐结果。

发明内容

[0011] 为了更加有效地融合多种数据信息，解决数据稀疏的问题，提高推荐系统的准确性，本发明提供了一种融合多信息的个性化推荐方法。

[0012] 为了达到上述发明目的，本发明采用的技术方案为：

[0013] 一种融合多信息的个性化推荐方法，包括以下步骤：

[0014] S1、获取用户-项目评论数据集，采用word2vec算法分别获取用户和项目的特征词向量，并采用FM算法获得用户与项目的相似度；

[0015] S2、根据用户的历史点击项目信息构建用户与项目的交互矩阵，并结合知识图谱，采用RippleNet算法得到用户与项目的预测点击概率；

[0016] S3、将步骤S1得到的用户与项目的相似度和步骤S2得到的用户与项目的预测点击概率采用动态融合算法进行动态融合得到预测评分，基于预测评分为用户提供个性化推荐列表。

[0017] 进一步地，所述步骤S1具体包括以下分步骤：

[0018] S1-1、获取数据库中所有用户-项目评论信息，采用word2vec算法将一个用户对所有项目的评论合成代表用户信息的文本数据，并将一个项目接受到的所有用户的评论集成为项目的文本数据；

[0019] S1-2、采用word2vec算法分别对步骤S1-1得到的用户信息的文本数据和项目的文本数据进行向量化处理，得到用户和项目的特征词向量；

[0020] S1-3、采用FM算法将步骤S1-2得到的用户和项目的特征词向量进行两两组合，并添加交叉项特征，得到用户与项目的相似度。

[0021] 进一步地，所述步骤S1-3中，FM算法的模型表示为：

[0022] $$\bar{y}_{u,v}(z_{uv}) = m_0 - m^T z_{uv} + \frac{1}{2} z_{uv}^T M z_{uv}, M_{j,c} = i_j^T i_c, j \neq c$$

[0023] 其中，m_0表示全局偏差项，m为用户u与项目v的特征向量z_{uv}的系数向量，M为二阶交互的权重矩阵，$M_{j,c}$为M的第j行c列的值，i_j，i_c为与z_{uv}的特征维度j和c相关的i维隐向量。

[0024] 进一步地，所述步骤S1采用平方损失作为参数优化的目标函数，表示为：

[0025] $$J_{sqr} = \sum_{(u,v)\in O} (y_{u,v} - \bar{y}_{u,v})^2 - \lambda_\Theta \|\Theta\|^2$$

[0026] 其中，O表示观察到的用户-项目评分对集合，$y_{u,v}$表示用户u与项目v的交互历史，

Θ表示所有参数，λ_{Θ}表示L2正则化参数。

[0027] 进一步地，所述步骤S2具体包括以下分步骤：

[0028] S2-1、设定用户集和项目集分别为$U=\{u_1,u_2,...,u_m\}$和$V=\{v_1,v_2,...,v_n\}$，构建用户与项目的交互矩阵，表示为：

[0029] $Y_{uv}=\{y_{uv}|u\in U,v\in V\}$

[0030] 其中，$y_{u,v}$表示用户u与项目v的交互历史，m表示用户数量，n表示项目数量；

[0031] S2-2、根据用户与项目的交互矩阵和包含关系-实体三元组的知识图谱，定义用户u的第k个关联实体为：

[0032] $$\varepsilon_u^k=\left\{t|(h,r,t)\in G,h\in\varepsilon_u^{k-1}\right\},k=1,2,...H$$

[0033] 其中，(h,r,t)表示知识图谱包含的关系-实体三元组，h表示头实体，r表示关系，t表示尾实体，H表示与原点项目所关联的最远位置；

[0034] 定义用户u在知识图谱G上的第k跳波纹集为：

[0035] $$S_u^k=\left\{(h,r,t)|(h,r,t)\in G,h\in\varepsilon_u^{k-1}\right\};$$

[0036] S2-3、对于每一个项目v对应创建一个d维的嵌入向量v，将用户u的第1跳波纹集的每一个三元组(h_i,r_i,t_i)与v的相关系数为：

[0037] $$p_i=softmax(v^T\cdot R^i\cdot h^i)=\frac{\exp(v^T\cdot R^i\cdot h^i)}{\sum_{(h_i,r_i,t_i)}\exp(v^T\cdot R^i\cdot h^i)}$$

[0038] 其中，R^i表示关系r_i的嵌入向量，h^i表示头实体h_i的嵌入向量；

[0039] S2-4、根据相关系数对用户u的第一跳波纹集的尾实体t_i计算加权和得到用户u对于项目v的多阶反响为：

[0040] $$o_u^1=\sum_{(h,r,t)\in S_u^1}p_it_i$$

[0041] 根据用户u对于项目v的多阶反响，定义项目v的用户u的嵌入向量为：

[0042] $$u=\alpha_1o_u^1+\alpha_2o_u^2+...+\alpha_Ho_u^H$$

[0043] 其中，α_i为正项混合参数；

[0044] S2-5、根据项目v的用户u的嵌入向量得到用户与项目的预测点击概率，表示为：

[0045] $$\hat{y}_{uv}(z_{KG})=\sigma(u^Tv)=\frac{1}{1+\exp(-u^Tv)}$$

[0046] 其中，z_{KG}表示基于知识图谱数据的推荐结果。

[0047] 进一步地，所述步骤S2中RippleNet算法的损失函数表示为：

[0048] $\Gamma=\sum_{(u,v)\in Y}-y_{uv}\log\sigma(u^Tv)+(1-y_{uv})\log(1-\sigma(u^Tv))$。

[0049] 进一步地，所述步骤S3中，将用户与项目的相似度和用户与项目的预测点击概率采用动态融合算法进行动态融合得到预测评分，表示为：

[0050] $$\bar{y}_{u,v}=\alpha\bar{y}_{u,v}(z_{review})+(1-\alpha)\bar{y}_{u,v}(z_{KG})$$

[0051] 其中，$\alpha=\frac{\bar{y}_{u,v}(z_{review})}{\bar{y}_{u,v}(z_{review})+\bar{y}_{u,v}(z_{KG})}$，$z_{review}$表示基于文本评论数据的推荐结果。

[0052] 本发明具有以下有益效果：本发明将知识图谱与评论内容作为多源数据，并使用不同算法对数据进行处理，并采用动态融合方法进行有效结合，为用户提供更精准的个性化推荐服务，能够实现更好的推荐效果，并且可以有效地解决数据稀疏带来的推荐准确性降低的问题。

附图说明

[0053] 图1是本发明的融合多信息的个性化推荐方法流程示意图；

[0054] 图2是本发明实施例中波纹集结构示意图；

[0055] 图3是本发明实施例中REME模型结构示意图；

[0056] 图4是本发明实施例中数据集AZ不同模型的召回率对比示意图；

[0057] 图5是本发明实施例中数据集SC不同模型的召回率对比示意图。

具体实施方式

[0058] 为了使本发明的目的、技术方案及优点更加清楚明白，以下结合附图及实施例，对本发明进行进一步详细说明。应当理解，此处所描述的具体实施例仅用以解释本发明，并不用于限定本发明。

[0059] 如图1所示，本发明实施例提供了一种融合多信息的个性化推荐方法，包括以下步骤S1至S3：

[0060] S1、获取用户-项目评论数据集，采用word2vec算法分别获取用户和项目的特征词向量，并采用FM算法获得用户与项目的相似度；

[0061] 在本实施例中，当前大多数社交媒体网站和电子商务系统允许用户发表文本评论。文本蕴含着丰富的信息，可以找到用户潜在的兴趣点，因此本发明将用户评论文本应用于推荐系统，从而提高推荐系统的准确性。

[0062] 本发明将基于深度学习的word2vec模型应用到推荐系统中，Word2vec是基于Skip-gram或CBOW(Continuous Bag-of-Words)的词嵌入模型。在没有词性标注的情况下，利用word2vec能够从原始语料学习到单词的向量表示，比较单词间的语义、句法的相似性。

[0063] 本发明利用word2vec对文本进行处理，将一个用户对所有商户的评论合成代表用户信息的文本数据，同理将一个商户接受到的所有用户的评论集成为商户的文本数据，提取出用户和项目的文本潜在特征，根据其特征进行匹配，最终进行合理推荐。

[0064] 上述步骤S1具体包括以下分步骤：

[0065] S1-1、获取数据库中所有用户-项目评论信息，采用word2vec算法将一个用户对所有项目的评论合成代表用户信息的文本数据，并将一个项目接受到的所有用户的评论集成为项目的文本数据；

[0066] S1-2、采用word2vec算法分别对步骤S1-1得到的用户信息的文本数据和项目的文本数据进行向量化处理，得到用户和项目的特征词向量；

[0067] Word2vec可以视为一个神经网络，其主要作用为将自然语言中的每个词通过一个三层神经网络训练成为一个词向量，从而很好地解决了传统词袋模型(bag-of-words,BOW)

无法表示文本上下文语义信息以及造成的维数灾难的问题，使得语义上相似的词具有相似的向量表示。

[0068] 本发明中word2vec算法采用CBOW预测模型和Hierachical softmax(huffman树，HS)训练模型，CBOW是根据已知的上下文单词来预测中心词的后验概率，模型结构如下：

[0069] 1)输入层，上下文词向量context(w)。

[0070] 2)投影层，将输入层的2c个context(w)词向量相加。

[0071] 3)输出层，输出中间词向量。

[0072] CBOW的训练函数为：

[0073] $\max\Phi=\sum_{w\in C}\log p(w|Context(w))$

[0074] S1-3、采用FM算法将步骤S1-2得到的用户和项目的特征词向量进行两两组合，并添加交叉项特征，得到用户与项目的相似度。

[0075] 本发明首先设置FM算法的输入：基于Word2vec构建出用户与项目的特征向量为：

[0076] $t_u=word2vec(T_u)$

[0077] $t_v=word2vec(T_v)$

[0078] 其中，T_u，T_v分别表示用户u与项目v的评论，t_u与t_v为相对应的用户和项目特征向量。

[0079] 将用户和项目的特征词向量进行两两组合，表示为：

[0080] $z_{uv}=t_u\odot t_v$

[0081] 其中，$\odot$表示向量点乘操作，z_{uv}是u与v之间的相关系数向量。

[0082] 本发明采用FM算法将用户和项目的特征词向量进行两两组合，并添加交叉项特征，从而显著提高模型的准确性。

[0083] FM算法的模型表示为：

[0084] $$\bar{y}_{u,v}(z_{uv})=m_0-m^Tz_{uv}+\frac{1}{2}z_{uv}^TMz_{uv},M_{j,c}=i_j^Ti_c,j\neq c$$

[0085] 其中，m_0表示全局偏差项，m为用户u与项目v的特征向量z_{uv}的系数向量，M为二阶交互的权重矩阵，$M_{j,c}$为M的第j行c列的值，i_j，i_c为与z_{uv}的特征维度j和c相关的i维隐向量。

[0086] 最后，采用平方损失作为参数优化的目标函数，表示为：

[0087] $$J_{sqr}=\sum_{(u,v)\in O}(y_{u,v}-\bar{y}_{u,v})^2-\lambda_\Theta\|\Theta\|^2$$

[0088] 其中，O表示观察到的用户-项目评分对集合，$y_{u,v}$表示用户u与项目v的交互历史，Θ表示所有参数，λ_Θ表示L2正则化参数，第二项$\lambda_\Theta||\Theta||^2$实现了防止模型过拟合。

[0089] S2、根据用户的历史点击项目信息构建用户与项目的交互矩阵，并结合知识图谱，采用RippleNet算法得到用户与项目的预测点击概率；

[0090] 在本实施例中，现有的RippleNet算法仅使用了用户的历史点击记录与结构化知识构成的知识图谱，没有考虑蕴含丰富知识的用户和项目评论数据，因此本发明利用word2vec提取出用户以及商户的隐特征，通过Factorization Machine(FM)算法处理隐特征，进而计算得到用户点击概率值；再通过加入一个动态参数，将RippleNet算法所得值与word2vec+FM所得值结合，最终获得点击率预测数值。

[0091] 上述步骤S2具体包括以下分步骤：

[0092] S2-1、设定用户集和项目集分别为U＝{u_1,u_2,...,u_m}和V＝{v_1,v_2,...,v_n}，构建用户与项目的交互矩阵，表示为：

[0093] Y_{uv}＝{y_{uv}|u∈U,v∈V}

[0094] 其中，$y_{u,v}$表示用户u与项目v的交互历史，m表示用户数量，n表示项目数量；y_{uv}的值为1和0，当取值为1时，表示用户u和项目v存在历史交互，即用户u曾经点击观看过项目v。

[0095] S2-2、根据用户与项目的交互矩阵和包含关系-实体三元组的知识图谱，定义用户u的第k个关联实体为：

[0096] $$\varepsilon_u^k=\left\{t\,|\,(h,r,t)\in G,h\in\varepsilon_u^{k-1}\right\},k=1,2,...H$$

[0097] 其中，(h,r,t)表示知识图谱包含的关系-实体三元组，h表示头实体，r表示尾实体，t表示关系；h∈E,r∈R,t∈E,E和R分别代表着知识图谱G中的实体集合和关系集合，H表示实验所设置的与原点项目所关联的最远位置。

[0098] 本发明中RippleNet算法的目标是在已有交互矩阵Y和知识图谱G的情况下，得到用户u和待定项目v的点击预测评分。即将用户u和项目v作为输入，输出用户u会点击项目v的概率。

[0099] 定义用户u在知识图谱G上的第k跳波纹集为：

[0100] $$S_u^k=\left\{(h,r,t)\,|\,(h,r,t)\in G,h\in\varepsilon_u^{k-1}\right\};$$

[0101] 其中，ε_u^0＝{v|y_{uv}＝1}表示用户u曾经点击过项目v，即用户u在G中的种子集。上标0表示种子结点。

[0102] “波纹”的含义包括：

[0103] 1)将用户的历史点击视为一个个水滴滴在知识图谱水面上形成了很多波纹，波纹的传播可以用于表示用户的潜在兴趣传播路径。

[0104] 2)用户的潜在兴趣程度随着k的增大而变小，即传播距离越远，与初始项目的相似性越小。

[0105] “波纹集”如图2所示：三角形代表的是用户最初点击的“种子集”，方块代表的是与种子集直接相连的第一跳波纹集(Hop1)，实心圆则代表的第二跳波纹集(Hop2)，第h跳波纹集皆以此类推。

[0106] S2-3、对于每一个项目v对应创建一个d维的嵌入向量v，项目嵌入向量为基于one-hot ID，分布，词袋等特征信息表示的一个项目。在已有用户u的Hop1波纹集S_u^1与项目嵌入向量v的条件下，将用户u的第1跳波纹集的每一个三元组(h_i,r_i,t_i)与v的相关系数为：

[0107] $$p_i=softmax(v^T\cdot R^i\cdot h^i)=\frac{\exp(v^T\cdot R^i\cdot h^i)}{\sum_{(h_i,r_i,t_i)}\exp(v^T\cdot R^i\cdot h^i)}$$

[0108] 其中，R^i表示关系r_i的嵌入向量，是一个d×d的矩阵；h^i表示头实体h_i的嵌入向量，是一个d维的向量；相关系数p_i表示项目v与头实体h_i在关系R_i上的相似程度。

[0109] S2-4、在获得相关系数p_i后，对于S_u^1的尾实体t_i计算加权和得到向量O_u^1：

[0110] $$O_u^1=\sum_{(h,r,t)\in S_u^1}p_it_i$$

[0111] 向量O_u^1表示基于用户u的历史交互对于项目v的1阶响应(Responding)，等价于用

项目v的特征表示用户u，而不是使用一个单独的特征向量来表示。同理，可以得到用户u对于v的2阶反响及多阶反响。

[0112] 根据用户u对于项目v的多阶反响，定义项目v的用户u的嵌入向量为：

[0113] $$u=\alpha_1 o_u^1+\alpha_2 o_u^2+\ldots+\alpha_H o_u^H$$

[0114] 其中，α_i为正项可训练混合参数，$\alpha_i>0$，且其和为1；

[0115] S2-5、根据项目v的用户u的嵌入向量得到用户与项目的预测点击概率，表示为：

[0116] $$\hat{y}_{uv}(z_{KG})=\sigma(u^T v)=\frac{1}{1+\exp(-u^T v)}$$

[0117] 其中，z_{KG}表示基于知识图谱数据的推荐结果。

[0118] 根据上式得到RippleNet算法的损失函数表示为：

[0119] $\Gamma=\sum_{(u,v)\in Y}-y_{uv}\log\sigma(u^T v)+(1-y_{uv})\log(1-\sigma(u^T v))$。

[0120] 其中，y_{uv}表示用户u与项目v的交互历史，值为1和0，当取值为1时，表示用户u和项目v存在历史交互，即用户u曾经点击并观看过项目v。定义的损失函数用于训练和调节参数。

[0121] S3、将步骤S1得到的用户与项目的相似度和步骤S2得到的用户与项目的预测点击概率采用动态融合算法进行动态融合得到预测评分，基于预测评分为用户提供个性化推荐列表。

[0122] 在本实施例中，上述RippleNet算法、word2vec+FM算法分别从评分和文本两个不同的信息源获得隐特征。为了使这两个隐特征的整合可以相互补充并产生更好的预测结果，加入了一个线性插值α，提出了一种动态融合推荐模型REME(RippleNet and word2vec fusion Model)，如图3所示，将步骤S1得到的用户与项目的相似度和步骤S2得到的用户与项目的预测点击概率进行动态融合得到预测评分，表示为：

[0123] $$\bar{y}_{u,v}=\alpha\bar{y}_{u,v}(z_{review})+(1-\alpha)\bar{y}_{u,v}(z_{KG})$$

[0124] 其中，$\alpha=\frac{\bar{y}_{u,v}(z_{review})}{\bar{y}_{u,v}(z_{review})+\bar{y}_{u,v}(z_{KG})}$，$z_{review}$表示基于文本评论数据的推荐结果。

[0125] 本发明采用随机梯度下降与反向传播的方法对上式的参数进行优化，具体过程为：

[0126] 首先为每一个用户统计出其波纹集与这个用户的所有评论的集合，使用word2vec算法将评论集合文件转化为对应的用户特征向量；

[0127] 在预设好的迭代次数T内，使用随机梯度下降算法与反向传播算法更新参数$\{\alpha_i, i=1,2,\ldots,H\}$；

[0128] 利用计算用户特征向量相同的操作，为每一个项目计算出对应的项目特征向量；

[0129] 在计算出所有用户-项目特征向量后，遍历出测试集的用户-项目对，计算出用户-项目相关系数向量z_{uv}；

[0130] 并在基于FM算法使用随机梯度下降算法与反向传播算法更新参数Θ；

[0131] 最终输出参数$\{\alpha_i, i=1,2,\ldots,H\}$与Θ。

[0132] 为了说明REME算法在提高了算法准确性的同时仍具有着较好的时间性能，本发明将分析REME算法的时间复杂性。

[0133] 首先创建用户特征向量:计算用户波纹集的时间复杂度为O(a×m),其中a为用户数量,是一个常数;word2vec算法的时间复杂度为O(alog(n))。综合上述步骤,创建用户特征向量的时间复杂度为O(a(m+log(n))),因为n的值远大于m,所以近似为O(log(n))。与创建用户特征向量相似,创建项目特征向量的时间复杂度为O(log(n))。计算用户特征与项目特征的交叉向量的时间复杂度为O($log^2(n)$),综上,REME的算法时间复杂度为O($log^2(n)$)。

[0134] 下面本发明采用具体实例对本发明与不同算法性能进行对比分析。

[0135] 在实验中使用通用的Yelp数据集进行推荐性能分析。本发明抽取Yelp数据集中亚利桑那州(AZ)与南卡洛来纳州(SC)两个不同地区的餐馆数据,包含用户的评论数据以及商家的属性数据集。在用户的评论数据中主要包含用户的评论,评分等信息。在实验中将用户评论视为签到一次。在商家的属性数据集中主要包含了商家的ID,名称,位置(地区,城市,经纬度等),餐馆种类以及标签等信息。实验利用的是Microsoft Satori来为Yelp商户建立知识图谱。

[0136] 对于筛选完后的两个地区的数据集,其统计信息如表1所示。

[0137] 表1数据集的各项统计信息

[0138]

	Yelp (AZ)	Yelp (SC)
用户数量	26688	10550
商户数量	5079	1162
签到数量	75239	21273

[0139] 从表1中可以发现AZ的用户数量是SC的两倍左右,然而商户数量是SC的五倍,因此带来了数据稀疏性的不同,从而导致最终的实验结果有着一定的差异。

[0140] 在Ripplenet中,将其波纹跳数H设置为2,根据实验结果证明大量的波纹跳数不会提高性能,相反会增加额外的计算开销。

[0141] 实验完整的参数设置为:商户和知识图谱的嵌入维度d=16,学习率η=0.02,正则化参数λ_1=10-7,λ_2=0.01,H=2。

[0142] 对于word2vec,设置其所得嵌入向量维度为也为d。通过验证数据集上的AUC曲线来确定超参数。

[0143] 为了取得更好的实验结果,对于每个数据集进行训练,训练、评估、测试集的比例为6:2:2。每个实验重复5次,取其平均值作为最终数据。

[0144] 本发明采用如下两个评价指标,评价算法的性能:

[0145] 1)对于点击率(CTR)预测,本文使用ACC(accuracy)与AUC来评估CTR预测的性能。

[0146] 2)对于top-k推荐,本文采用的是recall@k作为评价指标,recall@k的定义如公式:

[0147] $$recall@k = \frac{hit}{recall}$$

[0148] 其中,recall@k表示的是在top-k推荐列表中的召回率,即用户在推荐列表中点击

的概率。其中,hit表示的是测试集中的用户点击推荐列表中的餐馆的次数,recall代表的是测试集的签到总次数。

[0149] 在本发明中主要对比如下三个经典推荐算法:

[0150] 1)CKE:CKE主要将协同过滤与结构知识,文本知识和图像知识结合在一个统一的框架中进行推荐。

[0151] 2)DKN:DKN将实体嵌入和字嵌入视为多个通道,并将它们组合在CNN中以进行CTR预测。实验中使用商户标签作为DKN的文本输入。

[0152] 3)PMF:PMF主要是利用了用户的签到信息,将“用户-兴趣点”的签到矩阵分解为用户隐式因子矩阵和兴趣点隐式因子矩阵,利用这些隐式因子矩阵预测用户对于兴趣点的评分,进而为用户生成推荐列表。

[0153] 不同算法的top-k推荐与CTR预测的结果如图4和5所示,以及表2所示。

[0154] 表2点击率预测中AUC与Accuracy结果

[0155]

模型	Yelp(AZ)		Yelp(SC)	
	AUC	ACC	AUC	ACC
REME	**0.7823**	**0.8229**	**0.8024**	**0.8675**
Ripplenet	0.7045	0.7739	0.7538	0.8038
CKE	0.6698	0.7213	0.7031	0.7334
DKN	0.6187	0.6523	0.6321	0.6718
PMF	0.5897	0.6123	0.6048	0.6513

[0156]

[0157] 实验结果表明:

[0158] (1)在不同数据集上,SC的实验效果总是比AZ的好,这是因为两个地区的数据稀疏度有着差异,AZ的每个商户的平均流量比SC的更少。

[0159] (2)CKE在这里只使用了结构化知识,因此其效果相比RippleNet来说较差。RippleNet与其他模型相比结果较好,但是其仅考虑了知识图谱,没有有效利用评论文本信息等数据,因此推荐效果没有REME好。DKN的实验结果并不好,因为这里仅使用了标签信息,没有考虑其他的有效信息。

[0160] (3)无论在哪个数据集中PMF的推荐效果总是最差的,因为用户的签到数据是稀疏的。此外,PMF算法没有融合其他的内容数据信息。

[0161] (4)在两个数据集中,REME都取得了最好的推荐效果,其分别在AZ和SC两个数据集AUC上相比于其它基线提升了7.8%-19.3%和4.9%-20%,同时在recall@k的测试中也取得了最佳效果。

[0162] 本发明提出的REME模型相比于现有典型模型，在有效融合多种数据的情况下，明显提升了推荐效果，且在数据稀疏的情况下可以得到不错的推荐效果，表明REME模型可以有效解决数据稀疏对于推荐结果所带来的负面影响。

[0163] 本领域的普通技术人员将会意识到，这里所述的实施例是为了帮助读者理解本发明的原理，应被理解为本发明的保护范围并不局限于这样的特别陈述和实施例。本领域的普通技术人员可以根据本发明公开的这些技术启示做出各种不脱离本发明实质的其它各种具体变形和组合，这些变形和组合仍然在本发明的保护范围内。

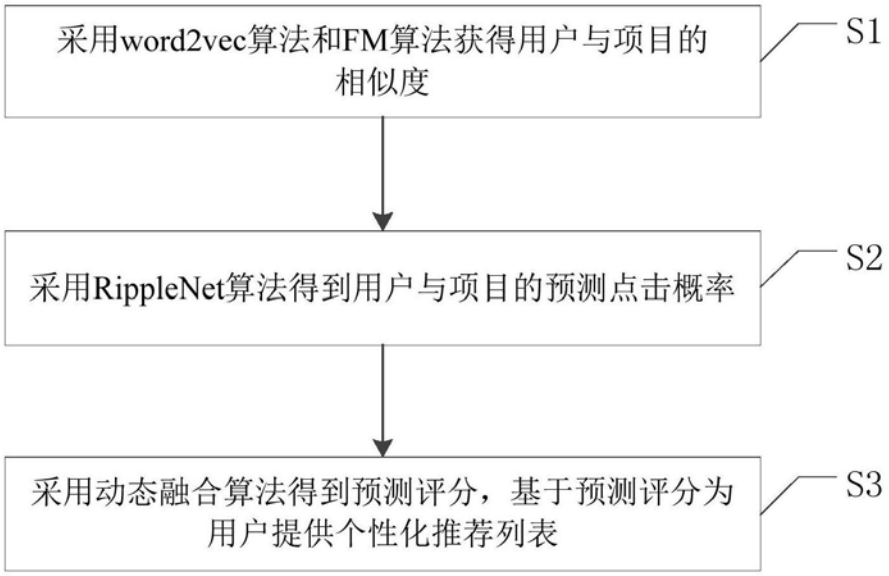

采用word2vec算法和FM算法获得用户与项目的相似度
S1
采用RippleNet算法得到用户与项目的预测点击概率
S2
采用动态融合算法得到预测评分，基于预测评分为用户提供个性化推荐列表
S3

图1

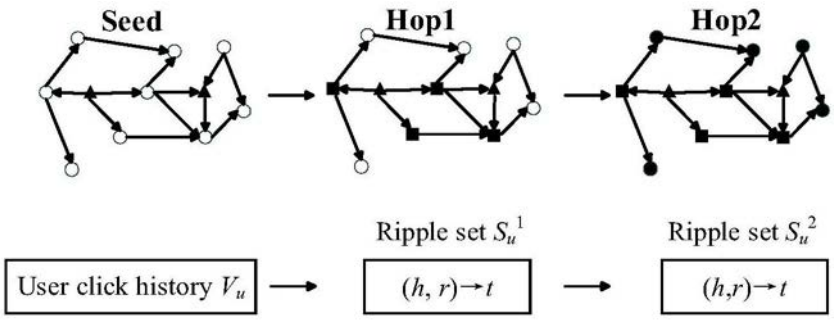

Seed
Hop1
Hop2
User click history V_u
Ripple set S_u^1
$(h, r)\rightarrow t$
Ripple set S_u^2
$(h,r)\rightarrow t$

图2

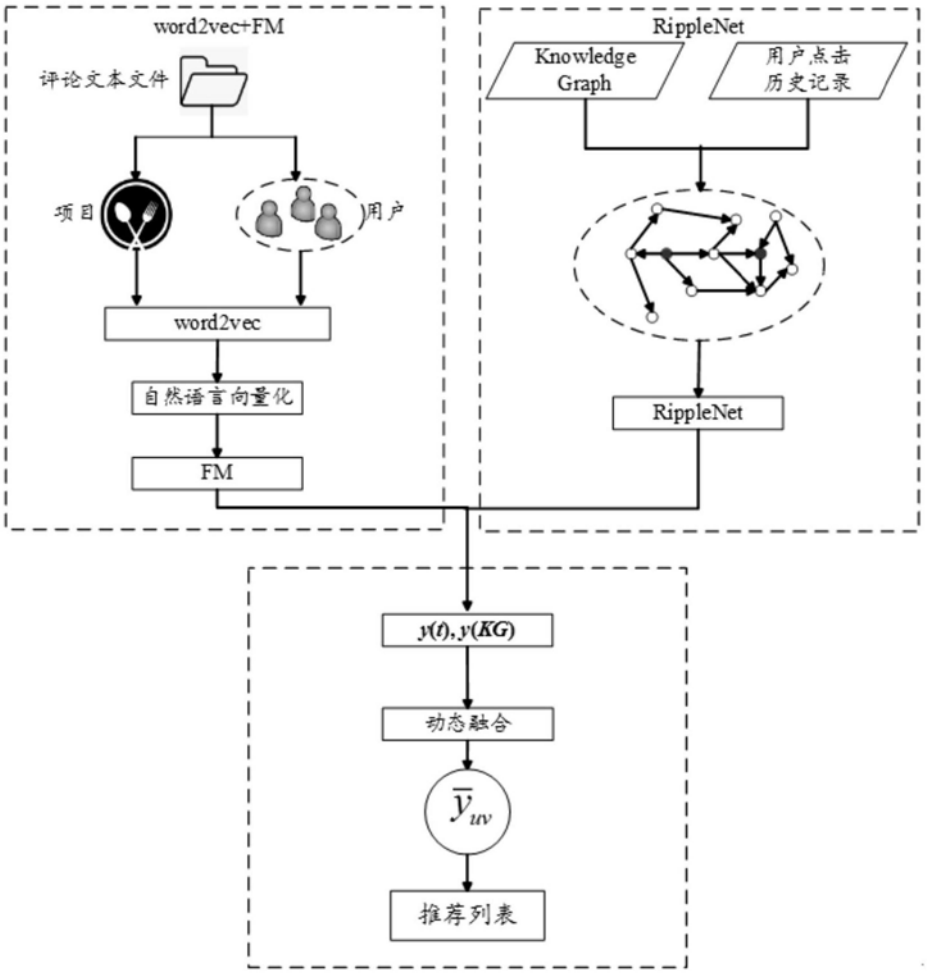

word2vec+FM
评论文本文件
项目
用户
word2vec
自然语言向量化
FM
RippleNet
Knowledge Graph
用户点击历史记录
RippleNet
y(t), y(KG)
动态融合
$\bar{y}_{uv}$
推荐列表

图3

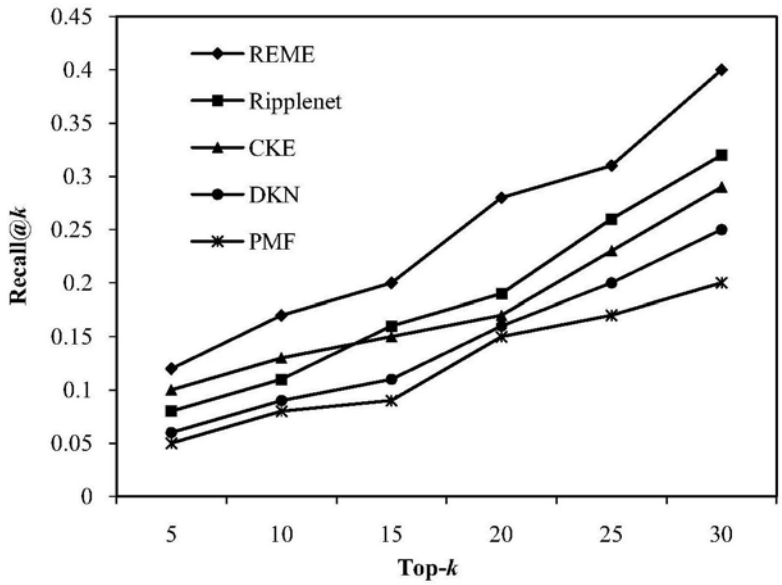

0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Recall@k
REME
Ripplenet
CKE
DKN
PMF
5
10
15
20
25
30
Top-k

图4

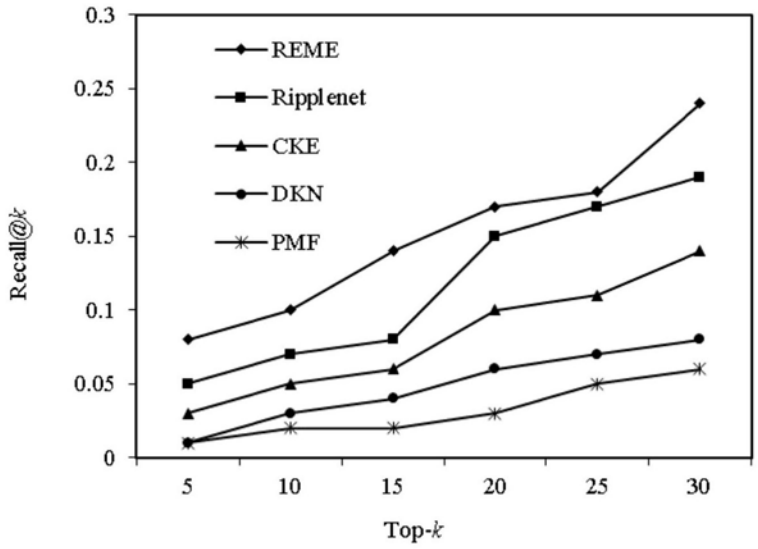

0.3
0.25
0.2
0.15
0.1
0.05
0
Recall@k
REME
Ripplenet
CKE
DKN
PMF
5
10
15
20
25
30
Top-k

图5