

基于网络表示学习的个性化商品推荐

李宇琦 陈维政 闫宏飞 李晓明

(北京大学计算机科学与技术系 北京 100871)

摘要 近些年来,互联网不断普及,其应用场景也在不断增加.电子商务是互联网普及、成熟的一大重要产物.这种新型的商业模式,便利了大众的生活,同时也创造了巨大的利润.对于电子商务而言,推荐系统是其中最关键的组成部分.推荐系统可以针对不同的用户,推荐其感兴趣的商品.好的推荐系统无论是对于用户体验还是公司盈利而言,都有着非常正面的作用.近几年间,网络表示学习受到关注,出现了一些利用网络表示学习的推荐算法研究.将网络表示学习应用于推荐系统中乃至商品推荐中,可以有效地利用近期网络表示学习研究成果.该文提出了一种利用网络表示学习进行个性化商品推荐的方法 PGE(Product Graph Embedding).首先,作者通过历史购买记录获取商品的顺序信息,从而构建商品网络.基于商品网络和网络表示学习算法,商品可以被映射至低维向量空间中.一旦作者获取了商品的低维向量表示,动态的用户偏好便可以基于用户购买过的商品记录及商品的时序线性计算得出,并和商品映射到相同的低维向量空间中.由此,商品和用户的相关性可利用商品和用户的低维向量相似度进行评估.作者在京东数据集上进行实验.实验表明,作者的算法在个性化商品推荐方面相较于最好的基准方法在 $P@10$ 上提升了 10% 以上,这显示出了作者算法的优越性.

关键词 网络表示学习;商品推荐;动态的用户偏好

中图法分类号 TP18 **DOI号** 10.11897/SP.J.1016.2019.01767

Learning Graph-Based Embedding for Personalized Product Recommendation

LI Yu-Qi CHEN Wei-Zheng YAN Hong-Fei LI Xiao-Ming

(Department of Computer Science and Technology, Peking University, Beijing 100871)

Abstract In recent years, the Internet has continued to be growing and progressing, its application scenarios are constantly increasing and developing at the same time. For example, e-commerce is an important product with the popularity of the Internet. This new business model has not only facilitated the lives of the public, but also created very huge profits. For e-commerce, the recommendation system is one of the most critical component which can impact it in many facets. Generally speaking, the recommendation system can recommend products to a large number of different users according to their various interest. A good recommendation system would have a very positive effect on both the user experience and the profitability of e-commerce's company. In the meantime, network representation learning has become a research hotspot and received a lot of attention by many researchers in recent years. There are a lot of different practical application scene utilizing network embedding. In addition, some research on recommendation algorithms using network representation learning method has emerged. The methods that apply network representation learning to recommendation systems and more specifically even to product recommendations, can be beneficial by leveraging the recent excellent network representation learning

收稿日期:2017-08-14;在线出版日期:2018-11-15. 本课题得到“九七三”重点基础研究发展规划项目基金(2014CB340400)、国家自然科学基金(U1536201,61472013)资助. 李宇琦,硕士研究生,中国计算机学会(CCF)会员,主要研究方向为机器学习. E-mail: liyuqi@pku.edu.cn. 陈维政,博士研究生,主要研究方向为机器学习. 闫宏飞(通信作者),博士,副教授,主要研究方向为信息检索. E-mail: yhf1029@gmail.com. 李晓明,博士,教授,主要研究领域为搜索引擎、网络数据挖掘、社会网络分析和慕课.

research results effectively. In this paper, we propose a novel product graph embedding (PGE) model. In this model, we investigate personalized product recommendation by leveraging the network representation learning technique. At first, our proposed model captures the sequential influences of products by transforming the historical adoption records into a product graph. After we construct the product graph, products can then be transformed into a low dimensional vector by utilizing the network embedding model on the product graph. Up to now, we have gotten the low dimensional vector representation of all the products in the product graph and we hope to get the low dimensional vector representation of users for further exploration. Thus, once products are projected into the latent low dimensional space, we present a novel method to compute user's latest preferences based on the embedding of sequential products that the user purchased, which projects users into the same latent space as products. By doing this way, relatedness between a product and a user can be easily measured by the similarity between the two low dimensional embedding vector, one embedding vector that represent the product and the other embedding vector which represent the user's preferences. At last, we conduct experiments on purchase records crawled from JINGDONG which is one of the largest online shopping site in China. Compared with the best baseline, our method achieves gain of 10% in the measure of $P@10$, which shows the superiority of our proposed framework for personalized product recommendation.

Keywords network embedding; product recommendation; dynamic user preference

1 引 言

随着互联网的普及,电子商务这种新型的商业模式也在飞速发展.推荐系统对于电子商务和在线购物领域具有重要的意义.所谓推荐系统,是指针对特定的用户甄别出其感兴趣的产品集合,从而挖掘用户的潜在需求,帮助用户进行决策^[1].推荐系统的应用领域非常广泛,其中包括电影^[2-3]、音乐^[4-5]、书籍^[6]和电子商务^[7-9]等.本文主要针对电子商务的个性化商品推荐领域.值得注意的是,本文所阐述的推荐系统算法具有一般性,也可以适用于其他的推荐领域.

目前,很多论文对推荐系统进行了深入的研究,提出了很多推荐算法.推荐系统一般会收集有用的信息,例如用户自身的特征,以获得用户对于产品的偏好.根据提取特征的不同,推荐算法主要分为:基于人口统计学的推荐算法、基于内容的推荐算法和协同过滤等^[10].基于人口统计学的推荐算法^[11-13]主要是利用用户自身的人口统计学特征信息(年龄、国籍、性别等)来进行推荐.这种推荐算法假定如果两个用户有一些相同的个人属性,则这两个用户会有相近的偏好.基于内容的推荐算法^[11,14]则是依据产品的内容(文本、图片、声音等)构造推荐算法.协同

过滤算法^[15-17]是最普遍、应用最广泛的推荐算法之一.基于用户的协同过滤技术的主要思想是收集用户的历史数据,并以此获取和当前用户兴趣最相近的用户集合,从而利用用户集合中最感兴趣的产品向用户进行推荐.这些已有的推荐算法虽然在一些应用场景下有着较好的效果,然而随着数据规模的不断加大,现有的推荐算法具有很大的局限性,很难应用于大数据.例如,常见的矩阵分解算法需要利用特征分解,其时间消耗很大且通常需要近似计算^[18].

近几年来,网络表示学习^[19-20]成为一个新的热点,也出现了一些利用网络表示学习的推荐算法研究.Xie等人将网络表示学习 LINE 算法^[20]应用于地点推荐领域^[21-22].Zhao等人文献^[8]中通过构建商品和用户的异构网络进行商品推荐.然而,用户的偏好会随着时间变化,其方法难以高效的表示用户的动态变化.

当前主要的网络表示学习算法比如 LINE 算法,只能用于学习同构网络中节点的低维向量表示,主要应用于社会网络、词语共现网络等网络中,并利用获取的节点的低维向量表示进行节点分类等任务.而在商品推荐领域,我们的数据格式是用户对商品的购买记录. NERM 算法^[8]的思想是将用户-商品交互网络当作同构网络,然后利用 LINE 算法同

时学习用户和商品的低维向量表示, 这样没有考虑时间因素对用户购物偏好的影响. 而本文则针对商品推荐的时间相关特性, 将表示学习算法分为两步. 我们的算法首先构建商品网络并利用网络表示学习获取商品的低维向量表示. 随后, 基于用户的历史购买记录和商品低维向量表示, 我们采用了一种高效的获取用户动态的偏好的算法, 以此获取用户低维向量表示. 随后针对每个用户, 我们利用用户和商品的低维向量表示通过 KNN(k -Nearest Neighbors) 算法得到有序的推荐商品列表, 并以此针对用户进行个性化推荐. 通过实验, 我们发现我们的算法在商品推荐下在各项评测指标中都优于比较的算法, 由此可以证明我们算法的有效性.

本文第 2 节介绍相关工作; 第 3 节给出问题描述和本文方法的概览; 第 4 节详细介绍基于网络表示学习的个性化商品推荐模型; 第 5 节描述使用的数据集以及模型在数据集上的实验效果; 第 6 节总结本文的内容.

2 相关工作

2.1 表示学习在推荐系统中的应用

随着近些年来互联网信息的增长, 推荐系统在很多在线网站和移动应用程序中成为重要工具, 对提升销售和服务起着至关重要的作用^[23]. 在日常生活中, 人们会参考别人口头推荐的话、新闻报道、旅行指南等进行选择和规划. 推荐系统增强实现这个自然过程, 以此帮助人们筛选可用的书籍、文章、网页、电影、餐馆、笑话、杂货等等, 以此提供最有趣和最有价值的信息给用户^[24].

最近几年, 学术界有很多基于表示学习进行推荐的研究工作, 工业界也有许多公司引入表示学习来进一步提升其推荐效果. 一些学者在研究工作中, 利用循环神经网络的表示学习方法来刻画用户画像^[25-26]. Covington 等人提出了一种基于深度神经网络的表示学习方法进行 YouTube 的推荐, 主要分为生成候选集模块和精排模块^[27]. 由此可见, 表示学习无论是在研究领域还是工业推荐应用中, 都有至关重要的位置.

多层感知机是一种简单而有效的模型, 目前也有很多针对基于多层感知机的表示学习的推荐系统的研究. 其中, NCF 模型^[28]输入用户和商品的边信息, 并利用多层感知机获取用户和物品之间的非线性关系, 最终预测两者的相似度从而进行推荐.

Wide&Deep 学习模型^[29]初始用于 Google Play 的 App 推荐, 它结合了一般的线性模型(wild)和多层感知机模型(deep), 实现了推荐的多样性.

基于 Autoencoder 的推荐系统在近些年来也有很多研究成果. AutoRec^[30]就是一种基于 Autoencoder 的推荐算法, 这种算法将用户或者物品的向量作为输入, 而后经过编码层和解码层, 在输出层重构相应的向量. 主要包括基于物品的 AutoRec 和基于用户的 AutoRec 两种.

卷积神经网络可以捕捉到局部信息, 在处理图片、文本等信息领域具有比较好的效果, 有很多学者在针对基于卷积神经网络的推荐系统进行研究. Gong 等人针对微博中的标签推荐提出了一种基于注意力机制(Attention)的卷积神经网络模型^[31]. 模型的输入是词向量, 主要包括一个全局的通道(global channel)和局部注意力机制的通道(local attention channel). 其中全局通道由卷积层和 max-pooling 层组成, 局部注意力机制通道主要包括一个注意力层以此收集有信息有价值的单词.

基于循环神经网络的推荐系统非常适合处理推荐系统中时序相关的特征. RRN 模型^[28]假定用户偏好和电影属性都是随着时间变化的, 有着季节时令性等性质. 鉴于以上的假定, 模型利用 LSTM 网络对用户和电影属性进行建模, 从而进行排序打分.

综上所述, 表示学习在推荐系统中的应用非常广泛, 针对其进行的研究也层出不穷.

2.2 网络表示学习

在现实世界中, 网络是一种非常重要的数据组织形式. 例如, 大量用户之间通过关注或者好友关系组成的在线社交网络如微博、微信等、城市之间通过道路连接组成的交通网络、论文之间通过引用关系组成的学术论文网络. 应用机器学习和数据挖掘技术对这些网络进行分析的首要步骤是对网络进行表示. 经典的方法是把网络用邻接矩阵来表示, 每个节点都对应一个高维稀疏向量. 这种表示的缺点在于需要消耗大量的运行时间和计算空间. 网络表示学习把网络中每一个节点映射成一个低维稠密实数向量, 不仅可以降低时间和空间上的计算开销, 而且可以提升节点向量在各种网络分析任务中的性能, 比如节点分类、链接预测、社区发现和网络重建.

传统的网络表示学习算法主要分为两类. 第一类是谱方法, 通过计算特定矩阵的特征值向量或者奇异值向量得到数据的低维表示. 例如, 可以将网络

的邻接矩阵直接作为 PCA 或者 SVD 算法的输入,从而得到节点的低维表示,但这种表示在数据挖掘任务中的性能通常很不理想^[31]. 另外几种代表性的谱方法则对网络中节点的性质做了一些假设, LLE (Locally Linear Embedding) 算法假设每一个节点的向量都可以由其近邻居节点的向量的线性加权组合构造得到^[32], Laplacian Eigenmaps 算法的思想则是希望相邻的节点在低维向量空间中对应的位置尽可能地接近^[33]. 第二类是基于最优化的方法,通过最小化或者最大化一个以节点向量作为参数的目标函数得到节点低维表示. 例如, LSHM 算法的目标函数可以同时学习节点的向量表示和节点标签的线性分类函数^[34], CSDK 算法^[35]通过网络上的信息扩散问题建模为低维空间上的信息扩散问题学习节点的低维向量表示. 以上两类传统方法的算法复杂度通常是网络节点数量的平方甚至立方量级,所以并不适用于大数据时代的大规模复杂信息网络分析任务.

近年来,深度神经网络算法在计算机视觉、自然语言处理等领域得到了成功应用,更多的研究者也开始关注把深度学习技术应用到网络数据分析任务. DeepWalk 算法^[19]通过在网络上进行随机游走得到节点访问序列,然后把节点序列作为 skip-gram 神经网络^[36]的输入从而得到节点向量. node2vec^[37]是 DeepWalk 的改进算法,通过平衡网络随机游走中的深度优先搜索和宽度优先搜索从而更好的建模网络结构信息. DeepWalk 和 node2vec 使用的都是浅层神经网络,在刻画网络结构上有先天的局限. SDNE^[38]算法使用多层的自编码器获取深层次网络结构信息,在节点分类等任务上取得了更好的效果.

网络表示学习的一个重要扩展是在考虑网络结构信息的基础之上,进一步融合网络中的其他信息,如文本、标签、时间等. TADW^[39]算法扩展了矩阵分解形式的 DeepWalk 算法,在学习网络表示的时候综合考虑了节点的文本属性信息. PNE^[40]算法通过同时学习网络中节点和节点标签的向量表示,可以进一步提高网络节点表示在分类任务中的性能. DANE^[41]算法考虑了网络随时间的动态演化,可以学习动态的节点表示.

现实世界中,特定领域的数据往往可以用合适的异构网络来建模表示,异构网络由多种类型的节点和连接关系组成. 例如电子商务网络^[42]有商品、评论、标签等节点,论文社交网络^[43]有作者、论文等

节点,单细胞生物数据网络^[44]有细胞、基因等节点. 已有的同构网络表示学习算法不适合直接应用于异构网络. metapath2vec^[45]算法提出了在异构网络上通过元路径指导随机游走策略,可以把多种类型的节点映射到同一空间. 参考文献^[41-43]分别给出了特定领域的异构网络表示学习算法.

综上所述,网络表示学习是一个迅猛发展的数据分析技术,但在个性化商品推荐这个问题上的应用研究还处于刚起步的不完善阶段.

3 问题描述

为了方便阅读,本章节主要介绍一些模型中用到的结构和定义. 首先定义购买记录 and 用户概要.

定义 1. 购买记录. 一个用户 u 在时间点 t 购买商品 p 这种行为就是一次购买记录,可以用三元组 $\langle u, p, t \rangle$ 表示.

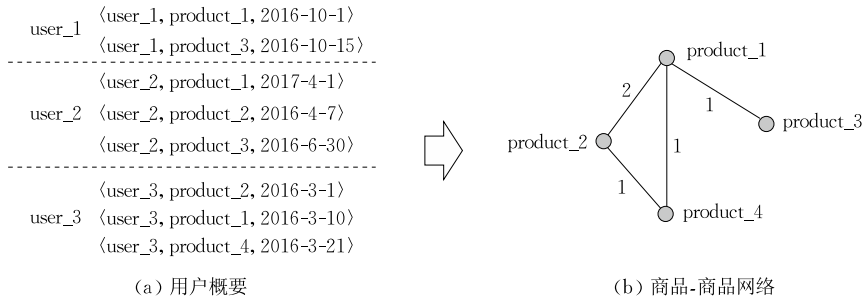
定义 2. 用户概要. 对于用户 u 而言,用户概要 D_u 可以表示成一个用户 u 的购买记录 $\langle u, p, t \rangle$ 的列表,且这个列表按照购买时间 t 排序. 数据集 D 是所有用户的用户概要的集合, $D = \{D_u : u \in U\}$.

当给定所有的购买记录集合时,我们可以对于每个用户都构造出当前用户的用户概要 D_u . 一旦我们获得了所有的用户概要 D ,根据相应的购买记录和规则可以统计出商品对之间的共现次数,从而构建出商品-商品网络. 接下来我们会给出商品共和商品-商品网络的定义.

定义 3. 商品共现. 给定一个时间区间 ΔT ,对于用户 u 的用户概要 D_u 列表中的每一对购买记录 $\{\langle u, p_i, t_i \rangle, \langle u, p_j, t_j \rangle\}$ 而言,如果 $|t_i - t_j| < \Delta T$,那么这两种商品共现.

定义 4. 商品-商品网络. 商品-商品网络主要是抽取商品对之间的共现信息构建获得. 可以将商品-商品网络表示成 $G = (V, E)$ 的加权无向图形式,其中 V 表示网络的节点信息,每个节点代表着一个商品; E 表示商品和商品之间的无向边,代表商品之间共现信息. 节点 v_i 和节点 v_j 之间边的权重 $w_{i,j}$ 是考虑所有的用户概要集合 D 后商品 i 和商品 j 之间共现次数.

根据以上的定义,已经可以根据历史的购买记录获得用户概要集合 D ,而后基于用户概要集合 D 获取商品-商品网络. 图 1 就是一个从用户概要生成商品-商品网络的简单示例. 在示例中,设置时间区间 ΔT 为 30 天. 在构建商品-商品网络的过程中,分

图 1 从用户概要生成商品-商品网络的示例($\Delta T=30$ 天)

析用户概要中每个用户在时间区间内 ΔT 的共同购买记录,从而获得商品-商品网络中边的权重. 举例来说,用户 2 和用户 3 均在 30 天内同时购买了商品 1 和商品 2,因此在商品-商品网络中,商品 1 和商品 2 对应节点的边的权重为 2. 而对于商品 1 和商品 3 而言,虽然用户 1 和用户 2 都购买过商品 1 和商品 3,但是用户 2 购买商品 1 和商品 3 的时间差超过了 30 天,因此,商品 1 和商品 3 之间的共现次数仅为 1 次,对应商品-商品网络中的权重为 1.

一旦获得了商品-商品网络,可以利用网络表示学习的相关算法获取商品的低维向量表示. 假定用户的偏好是时间相关的,也即用户的偏好是动态的,会随着时间而产生变化. 本文利用用户 u 的历史购买记录和商品的低维向量表示可以获得用户在时间点 t 的偏好.

最终,计算商品和用户的低维向量表示的相关性以实现商品推荐任务,因为用户偏好是时间相关的,本文对应的商品推荐任务也是时间相关的,定义如下.

定义 5. 时间相关的个性化商品推荐任务. 给定一个查询请求 $q=(u, t)$, 表示在时间点 t 的用户 u , 时间相关的个性化商品推荐任务的目标是返回一个有序的商品列表推荐给用户 u , 尽可能地符合用户 u 在时间 t 的兴趣偏好.

4 PGE 模型

本节具体介绍本文提出的基于网络表示学习的 PGE(Product Graph Embedding) 个性化商品推荐模型. 本章节的第一部分主要介绍如何利用商品-商品网络信息将商品映射到低维向量; 第二部分主要介绍如何表示动态的用户偏好; 第三部分则是介绍利用商品和用户的低维向量表示进行 top- N 的个性化商品推荐.

4.1 商品网络表示学习

当给定了商品-商品网络之后,需要尽可能多地利用商品网络信息来获得商品的低维向量表示. 如果仅利用相邻节点之间的信息的话,将会丢失很多网络结构的信息. 而一旦利用过于深层的网络节点信息,又将会引入过多的噪音,因此,利用节点的邻边作为当前节点的上下文信息,并假定如果两个节点共享更多的相邻节点,也即两个节点的上下文信息更接近,那么这两个节点会比较相似^[20]. 基于这种假定,可以更多的利用网络全局信息且不容易引入过多的噪音.

根据这种假定可以给出本文的商品网络表示学习模型的形式化表示. 首先,对于每个节点 v_i 而言,其共有两个表示形式:节点本身和作为其他节点的上下文(context). 对于节点 v_i ,共有两个 d 维的低维向量表示:

$$\mathbf{p}_i, \mathbf{p}'_i \in \mathbb{R}^d \quad (1)$$

其中, \mathbf{p}_i 表示节点作为本身的低维向量表示, \mathbf{p}'_i 表示节点作为上下文时的低维向量表示.

首先,定义对于每一个边 $\langle i, j \rangle$, 由节点 v_i 生成上下文 v_j 的条件概率,利用 softmax 函数获得:

$$P(v_j | v_i) = \frac{\exp(\mathbf{p}'_j{}^T \cdot \mathbf{p}_i)}{\sum_{k=1}^{|V|} \exp(\mathbf{p}'_k{}^T \cdot \mathbf{p}_i)} \quad (2)$$

其中, $|V|$ 是节点的个数.

已知了条件概率 $P(v_j | v_i)$ 的含义,可以获得给定节点 v_i 在全部节点集合上生成上下文的条件分布 $P(\cdot | v_i)$. 正如上文曾经提及,假定如果两个节点的上下文分布比较近似则这两个节点是比较相似的. 由此,希望条件分布 $P(\cdot | v_i)$ 尽可能地接近其经验分布 $\hat{P}(\cdot | v_i)$. 经验分布可定义为

$$\hat{P}(v_j | v_i) = \frac{\omega_{i,j}}{d_i} = \frac{\omega_{i,j}}{\sum_{k \in N(i)} \omega_{i,k}} \quad (3)$$

其中, $\omega_{i,j}$ 是边 $\langle i, j \rangle$ 的权重, d_i 是节点 v_i 的度,

$N(i)$ 是和节点 v_i 相邻的节点.

KL 散度是一种常见的用于衡量两个概率分布差异的函数. KL 散度越大, 则两个概率分布差异越大. 因此, 采用最小化这两种分布的 KL 散度的方式, 以保证条件分布 $P(\cdot | v_i)$ 和其经验分布 $\hat{P}(\cdot | v_i)$ 尽可能接近. 因此, 我们需要最小化如下的目标函数:

$$O = \sum_{i \in V} \lambda_i D_{KL}(P(\cdot | v_i) \| \hat{P}(\cdot | v_i)) \quad (4)$$

其中, λ_i 用于衡量节点 v_i 在整个网络中的重要程度, 在本文用节点 v_i 的度 d_i 表示.

可以消掉一些常数项以对目标函数进行化简, 最终可以获得如下的目标函数:

$$O = - \sum_{(i,j) \in E} w_{i,j} \log P(v_j | v_i) \quad (5)$$

当获得了上述的目标函数后, 需要对目标函数进行优化. 然而, 因为目标函数要遍历网络中所有的节点, 对目标函数的优化的时间复杂度很高. 为了解决时间消耗过大的问题, 引入了负采样的方法^[20]. 具体而言, 负采样是对每个边 $\langle i, j \rangle$ 都根据噪音分布采样几个负边, 以此降低时间消耗. 在本文中对应的目标函数进行负采样后可以转化为

$$O = - \sum_{(i,j) \in E} w_{i,j} \left\{ \log \sigma(\mathbf{p}_j^T \cdot \mathbf{p}_i) + \sum_{k=1}^K \mathbb{E}_{v_n \sim P_n(v)} [\log \sigma(-\mathbf{p}_n^T \cdot \mathbf{p}_i)] \right\} \quad (6)$$

其中, $\sigma(\cdot)$ 是 sigmoid 函数, 可以表示成 $\sigma(x) = 1/(1 + \exp(-x))$ 的形式, K 是每条边负采样的数目, $P_n(v)$ 是噪音的分布函数. 依据文献^[36]的设定, 本文设置 $P_n(v) \propto (d_v)^{0.75}$. \mathbb{E} 表示数学期望. 这个公式的含义是, 每次采样商品网络中的一条边 $\langle i, j \rangle$ 作为正样本, 然后从噪音分布 $P_n(v)$ 中采样 K 个商品节点 n , 组成 K 个负样本 $\langle i, n \rangle$. 我们使用正样本和负样本对目标函数进行优化.

为了优化负采样后的目标函数, 针对每一条边使用异步随机梯度下降 (ASGD) 的优化算法. 在异步随机梯度下降算法中, 每一轮的迭代都会采样少量几个 (min-batch) 网络中的边, 算出梯度并更新模型参数. 如果边 $\langle i, j \rangle$ 在优化算法中被采样到, 则需要计算节点 v_i 本身的低维向量表示 \mathbf{p}_i 对目标函数的导数, 计算可得导数如下:

$$\frac{\partial O}{\partial \mathbf{p}_i} = w_{i,j} \cdot \frac{\partial \log P(v_j | v_i)}{\partial \mathbf{p}_i} \quad (7)$$

由上述的导数可知, 导数是边 $\langle i, j \rangle$ 的权重 $w_{i,j}$ 和条件概率导数的乘积, 这样会导致导数受边权重

影响很大. 对不同边而言, 如果权重 w 相差很多, 方差很大时, 则导数相差可能很大, 导致优化过程产生问题. 举个简单的例子, 如果在一个商品交易平台中, 有些商品之间共现过几千乃至几万次, 而有些商品之间只共现过一两次. 那么, 在商品-商品网络中, 有些边的权重会是几千乃至几万的数量级, 而另一些边权重则是个位数. 这样在优化目标函数的过程中, 对不同的边进行采样并计算导数, 会获得数据规模差距很大的导数, 这样难以选择一个合适的学习速率. 为了解决这种问题, 本文在优化的过程中利用加权随机采样的方式, 根据边权重的大小获得随机采样概率的大小. 在实际的实现中, 使用别名表 (alias table) 的方式以 $O(1)$ 的时间复杂度可以获得加权随机采样得到的采样结果.

设定 d 是商品低维向量表示的维度, K 是对于每一条边而言负采样的数目, $|E|$ 是边的数目. 因为加权随机采样的时间复杂度为 $O(1)$, 因此在优化算法的每一轮迭代的时间复杂度为 $O(dK)$. 而一般情况下, 迭代的轮数是 $O(E)$ 的. 综上所述, 整个优化算法的时间复杂度为 $O(dK|E|)$, 这表明优化算法的时间复杂度和商品网络中的边数目 $|E|$ 成线性关系, 而与节点数目 $|V|$ 是无关的. 由此, 这可以被推广到商品数量很多的情况.

4.2 动态的用户偏好表示

正如在本文的第 2 节所提及的, 假定用户的偏好是时间相关的, 也即用户的偏好是动态的, 会随着时间而产生变化. 很多传统模型都需要消耗大量的时间以实现用户偏好的动态性, 因为大多数模型都需要随着时间的变化重新训练更新模型才可以得到不同的用户偏好表示. 为了降低时间复杂度提高效率, 本文使用一种利用用户 u 的历史购买记录和商品的低维向量表示获得用户在时间点 t 偏好的算法, 这种算法的时间复杂度和用户曾经购买过的商品数目成线性相关, 这样可以进行个性化的商品推荐.

利用一个 d 维的向量表示用户的动态的偏好:

$$\mathbf{u}_{i,t} \in \mathbb{R}^d \quad (8)$$

$\mathbf{u}_{i,t}$ 表示用户 u_i 在时间 t 的偏好对应的低维向量表示. 用户的向量表示和商品的向量表示都是 d 维, 两者在相同的向量空间中.

假定用户的偏好可以在其购买过的商品中展现, 且距离当前时间点越近的购买记录对用户偏好的影响越大. 基于上述的假设, 为了计算 $\mathbf{u}_{i,t}$ 的取值, 我们首先需要获取用户 u_i 在时间点 t 前购买的商品

集合. 一旦获得了用于表示用户动态的偏好的商品集合, 将会利用时间衰减函数, 以此来保证在不同时间点购买的商品对用户偏好影响不同, 且距离时间 t 越近购买的商品对用户的偏好影响越大.

在具体的实现中, 可以注意到用户概要 D_u 是一个按照用户 u 购买时间 t 排序的购买记录列表. 因此, 可以形式化的表示时间相关的动态的用户偏好如下所示:

$$\mathbf{u}_{i,t} = \sum_{\langle (u_i, p_j, t_j) \in D_{u_i} \rangle \cap (t_j < t)} \Gamma(t-t_j) \cdot \mathbf{p}_j \quad (9)$$

其中, \mathbf{p}_j 是商品 j 本身的低维向量表示, $\Gamma(\cdot)$ 是时间衰减函数. 时间衰减函数 $\Gamma(t-t_j)$ 是一个单调递减的函数, 也即随着时间差 $t-t_j$ 的增加, 函数 $\Gamma(t-t_j)$ 的值将会减小. 这样保证了时间差越小, 购买商品的时间距离当前时间越近, 则商品 j 的权重越大, 对用户的动态的偏好的影响也就越大.

鉴于用户偏好的低维向量表示是根据商品的低维向量表示和时间衰减函数两部分构成, 因此时间衰减函数的定义很重要. 在本文的模型中, 因为预测的是用户未来将要购买的商品, 所以, $t > t_j$. 为保证时间衰减函数的单调性, 我们利用截断的高斯分布密度函数作为时间衰减函数:

$$\Gamma(t-t_j) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{(t-t_j)^2}{2\sigma^2}\right) \quad (10)$$

4.3 个性化商品推荐

之前的章节已经介绍了如何在商品-商品网络中学习商品的低维向量表示以及利用历史购买记录和时间衰减函数构建动态的用户偏好的低维向量表示的方法. 本节主要介绍如何利用商品和用户的低维向量表示来针对用户进行个性化商品推荐.

根据在“问题描述”章节中对“时间相关的商品推荐任务”的定义, 给定一个在时间点 t 的用户 u , 推荐任务是推荐给这个用户一个有序的商品列表. 为了完成这个任务, 本文的方法是对于用户 u 未购买过的商品集合中的商品均计算出一个分数, 这个分数值是衡量商品和用户 u 之间的相似度. 然后, 我们选择前 k 个和用户最相关的商品推荐给当前的用户. 定义商品和用户之前的相关性分数如下所示:

$$\text{score}(u_i, p_j) = \mathbf{u}_{i,t} \cdot \mathbf{p}_j \quad (11)$$

其中, $\mathbf{u}_{i,t}$ 是用户 u_i 在时间 t 的低维向量表示, \mathbf{p}_j 是商品 j 的低维向量表示.

在我们的模型中, 我们使用 L2 正则化的方式

使得商品和用户的低维向量长度均为 1, 即

$$|\mathbf{u}_{i,t}| = 1, |\mathbf{p}_j| = 1 \quad (12)$$

因此, 商品和用户之前的相关性分数等价于余弦相似度.

知晓了商品和用户的相似度计算方法后, 利用 KNN 的方法即可获得 top- k 的商品, 本文是利用了库 Flann^[46] 进行 KNN 检索.

5 实验

本节主要介绍实验, 包括实验的数据集、评测指标、实验结果以及结果分析等多个方面的内容.

5.1 实验数据集

本文使用含有 1.3 亿个购买记录的京东数据集^[9]进行实验. 首先, 需要对这个数据进行预处理. 预处理的过程主要是依据文献[9]中的过滤流程对数据集进行过滤. 过滤的流程是先将购买记录按照用户进行分组, 从而获得对于每个用户自己的购买记录; 而后, 将少于 150 次购买记录的用户和少于 80 次购买记录的商品都删掉; 完成上述的过滤步骤后, 可以构造出一个用户-商品的矩阵, 如果某个商品曾被用户购买过, 则把对应的矩阵元素设置为 1; 接下来, 迭代的进行下述操作: 删除购买记录少于 50 次的用户集合和商品记录一直到无法删除则迭代结束. 由此, 获得了过滤后的数据集, 过滤后的用户-商品矩阵的稀疏率是 0.9519%.

为了在实验中模拟真实的线上商品推荐, 将数据集按照用户购买记录的时间的先后顺序切分为训练集和测试集. 具体而言, 对于每个用户, 将其购买记录按照购买时间排序, 而后取 80% 的购买记录作为训练集, 后 20% 作为测试集. 综上所述, 京东数据集的概要统计如表 1 所示.

表 1 京东数据集的摘要统计

数据集	用户数目	商品数目	购买记录
整个集合	37064	20235	7138933
训练集	37064	20235	5696269
测试集	37064	20235	1442664

5.2 实验方法、参数设置和评测指标

在实验中, 本文和几种基准方法 (baseline) 进行了比较, 主要包括以下几种:

(1) BPR 算法^[47]. BPR (Bayesian Personalized Ranking) 是一种基于贝叶斯后验优化的个性化排序算法. 这种算法利用了用户对物品的反馈情况, 优

化目标是成对的组合,假定用户购买过的产品的权重会大于用户未购买的产品。

(2) DeepWalk 算法^[19]. DeepWalk 是最近提出的一种基于 word2vec^[36]的网络表示学习方法. 这种算法首先在网络中随机游走获取一些随机的路径,然后再利用 word2vec 算法获取网络中节点的低维向量表示.

(3) NERM 算法^[8]. NERM 算法是基于 LINE^[13]的 first-order 的网络表示学习方法来进行商品推荐和商品标签推荐的方法. 对于商品推荐而言,算法的主要思想是先将用户和商品表示成二部图的形式,然后再利用 LINE 的 first-order 算法获得用户和商品的低维向量表示,从而进行推荐.

BPR 算法是目前推荐系统中效果比较好的推荐算法,而 DeepWalk 和 NERM 都是最近比较热门的基于网络表示学习的方法,我们以这三种作为比较方法. DeepWalk 和 NERM 构建的用户-商品网络相同. 为了使得实验公平,将这三种算法和本文的算法的向量维数均设置为 500. 其他的参数设置为:对 BPR 算法,我们在训练过程中共迭代 50 次(每次迭代为遍历所有的训练集进行训练);对于 NERM 算法,负采样的数目设置为 8;对于本文的 PGE 方法,我们设置 ΔT 为 30 天, σ 为 10,负采样数目为 8.

本文的评测方法主要包括准确率($P@k$)、召回率($R@k$)和平均精度均值($MAP@k$),本文 k 的选取为 1,10,各项评测指标的定义如下:

$$P@k = \frac{\text{correct}@k}{k} \quad (13)$$

$$R@k = \frac{\text{correct}@k}{\text{total_purchased}} \quad (14)$$

$$AP_j@k = \sum_{i=1}^k p_j(i)/m_j, \quad (15)$$

$$MAP@k = \sum_{j=1}^N AP_j@k/N$$

其中, m_j 是用户 j 的购买商品总个数, N 是用户的总数.

鉴于商品集合很大,对每个用户如果考虑所有的商品集合很耗时. 因而,我们利用文献[48]中的方法重构测试集:对于测试集中的每个用户 u_i 购买的每个商品 p_j ,都随机抽样 50 个未被该用户购买的商品作为负例,从而和原来购买过的集合构成一个整体的商品集合组成新的测试集进行测试.

5.3 实验结果和分析

根据上面章节的设定,可以获得实验的结果如

表 2 所示,加粗表示最好的结果.

表 2 商品推荐的实验结果比较

评测指标	BPR	DeepWalk	NERM	PGE
$P@1$	0.2845	0.2863	0.3595	0.6322
$R@1$	0.0073	0.0074	0.0092	0.0162
$MAP@1$	0.0076	0.0081	0.0102	0.0174
$P@10$	0.2114	0.2126	0.2572	0.3647
$R@10$	0.0543	0.0546	0.0661	0.0936
$MAP@10$	0.0424	0.0440	0.0542	0.0893

我们主要对比了 BPR、DeepWalk、NERM、PGE 这四种算法的准确率($P@k$)、召回率($R@k$)和平均精度均值($MAP@k$),其中 $k=1,10$.

从实验的结果中,可以看出:首先,利用网络表示学习的商品推荐算法 DeepWalk、NERM 和本文 PGE 算法均高于 BPR 算法,由此可见网络表示学习应用于商品推荐系统的有效性;其次,本文的 PGE 算法的各项评测指标都要远高于其他的利用网络表示学习的推荐算法 DeepWalk 和 NERM,由此可见本文 PGE 算法的优越性.

本文 PGE 算法的主要优势在于:首先,利用了网络表示学习的方法,考虑商品网络的全局信息,这样可以更好更多地使用数据集的信息,减小有用信息的丢失;其次,我们利用动态的用户偏好表示的方式,可以更贴合地表示用户的偏好. 通俗而言,如果某个商品与用户向量的相关度越高,则我们更有可能向用户推荐这个商品,因为这个商品和用户之前购买过的产品在向量空间中更相似,用户会倾向于购买与其买过的商品类似的商品. 通过把商品网络中的商品都映射到同一个向量空间,我们可以更好地衡量商品之间的相似度.

5.4 参数敏感度分析

本文调整参数对实验结果有着很重要的影响,因而本节主要对参数敏感度进行分析. 在 PGE 算法中,对结果影响比较大的参数是向量的维度 d 和构建商品网络的时间区间 ΔT ,因此本节针对向量维度 d 和时间区间 ΔT 对实验结果的影响进行探讨. 同时,也对 σ 和负采样数目对实验结果的影响进行了讨论. 其中不同维度的实验结果如图 2 所示,不同时间区间 ΔT 结果如图 3 所示, σ 的实验结果如图 4 所示,负采样数目如图 5 所示.

对于向量维度 d ,我们主要对比了其 100,300,500,700,1000 这 5 个不同维度的准确率($P@k$)、召回率($R@k$)和平均精度均值($MAP@k$),其中 $k=1,10$,结果如图 2 所示.

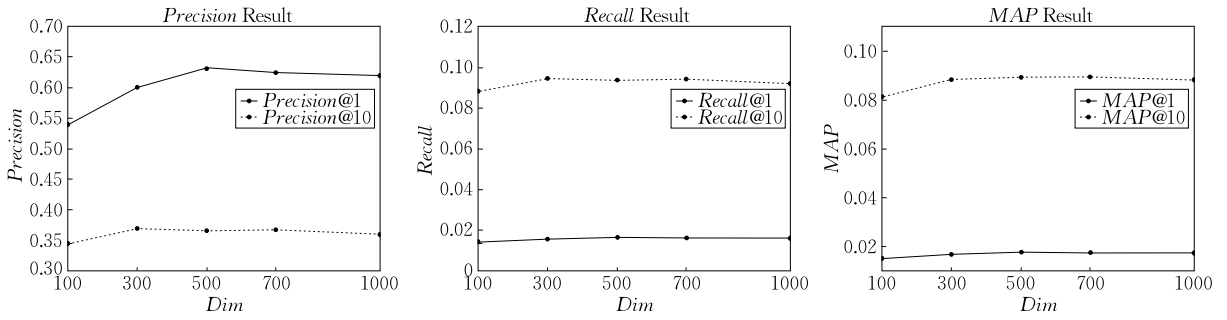


图 2 向量维度 d 的影响

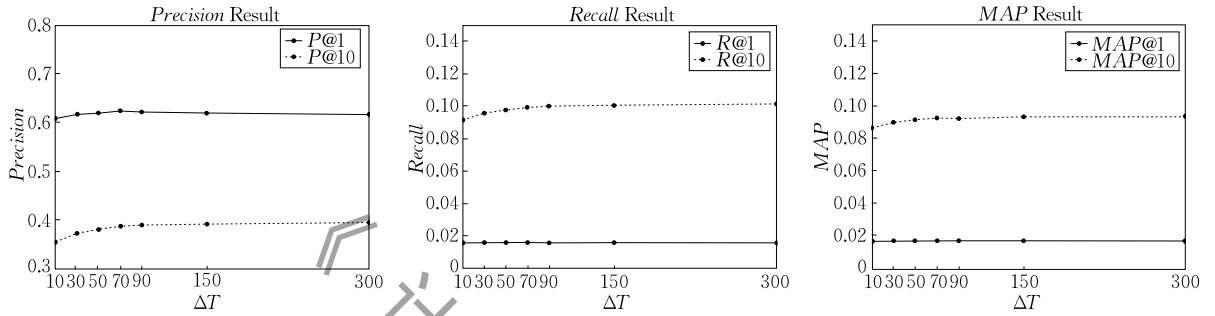


图 3 时间区间 ΔT 的影响

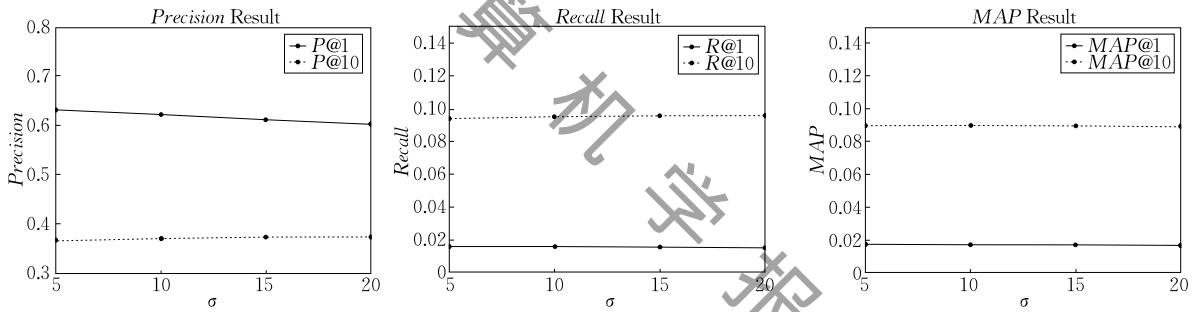


图 4 σ 的影响

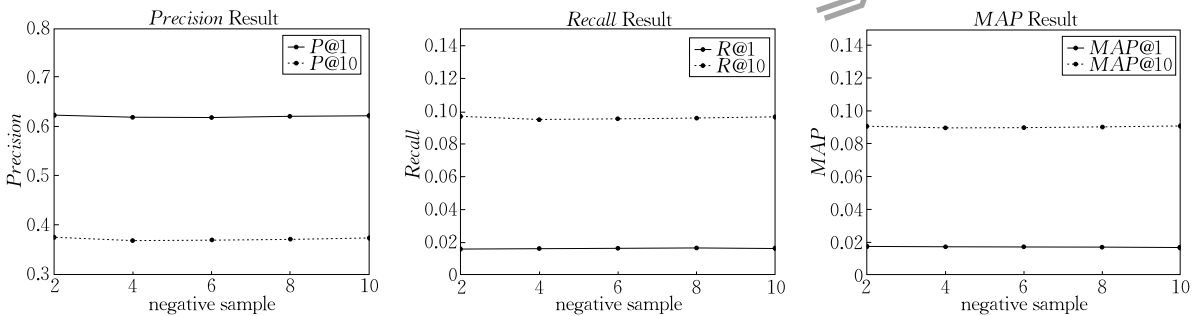


图 5 负采样数目的影响

从图 2 的实验结果可以看出,随着维度的增加,各项评测指标:准确率($P@k$)、召回率($R@k$)和平均精度均值($MAP@k$),均呈现先上升后逐渐平稳并略微下降的趋势.最初,随着维度的增加,更大的维度的灵活性更高,能够更好的表达信息;然而当后期维度已经增加到了了一定的阈值,维度所能表达的信息已经达到上限,这时再增加维度则效果不会更

理想,且时间和空间的消耗会更大.因此,本文选用维度 d 为 500 的值,以权衡效果(performance)和效率(efficiency).

对于构建商品网络的时间区间 ΔT ,我们主要对比了 ΔT 为 10 天、30 天、50 天、70 天、90 天、150 天和 500 天的这 7 个不同维度的准确率($P@k$)、召回率($R@k$)和平均精度均值($MAP@k$),其中 $k=1$,

10, 结果如图 3 所示.

从图 3 实验结果可以看出, 随着时间区间 ΔT 的增加, 各项评测指标: 准确率($P@k$)、召回率($R@k$)和平均精度均值($MAP@k$), 均呈现先上升后逐渐平稳的趋势. 可见起初时间区间 ΔT 越大, 获取的有效信息越多; 随后随着时间区间 ΔT 的增加, 噪音和有效信息都在增加, 因而性能并没有进一步提升.

与此同时, 我们统计了不同的时间区间 ΔT 所构造出的商品网络边的数目, 以比较不同的时间区间 ΔT 对于时间效率的影响, 统计结果如图 6 所示.

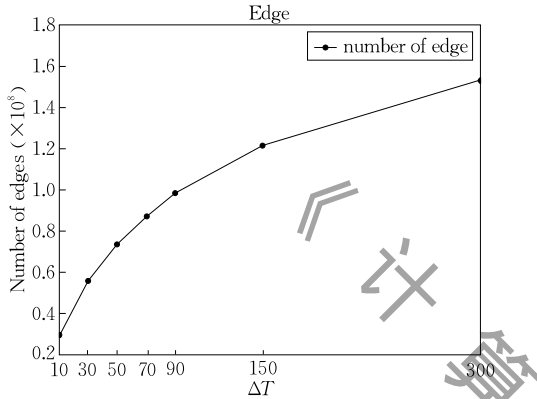


图 6 不同时间区间 ΔT 对应的商品网络边数对比

从图 6 的统计结果可以看出, 商品网络边的数目随着时间区间 ΔT 的增加持续呈现一个比较明显的上升. 在 3.1 节中曾阐述过, 优化算法的时间复杂度和商品网络中的边数目 $|E|$ 成线性关系. 因此, 随着时间区间 ΔT 的持续增加, 时间效率有显著的降低, 而对各项评测指标的实验结果则没有改善. 基于以上的分析, 我们在本文中选择不时间区间 ΔT 的取值为 30 天.

由图 4 和图 5 的实验结果可知, σ 和负采样数目对实验结果的影响并不大, 在本文中我们设置 σ 为 10, 负采样数目为 8.

6 总 结

本文在商品推荐方面, 提出了一种基于网络表示学习的个性化商品推荐算法. 本文的 PGE 算法首先利用训练集的用户购买记录信息构建商品网络, 并基于网络表示学习的方法获取商品的低维向量表示; 随后, 用商品的低维向量表示和时间衰减函数获取动态的用户偏好的低维向量表示; 最终, 利用商品和用户的低维向量表示进行商品推荐的方法. 实验中, 发现本文所述方法在各项评测指标上均有显著提高, 其中在准确率 $P@10$ 中相较于其他效果最好

的基准方法提升了 10%, 由此验证了本文算法的有效性.

参 考 文 献

- [1] Deshpande M, Karypis G. Item-based top- N recommendation algorithms. *ACM Transactions on Information Systems*, 2004, 22(1): 143-177
- [2] Carrer-Neto W, Hernández-Alcaraz M L, Valencia-García R, et al. Social knowledge-based recommender system. Application to the movies domain. *Expert Systems with Applications*, 2012, 39(12): 10990-11000
- [3] Winoto P, Tang T Y. The role of user mood in movie recommendations. *Expert Systems with Applications*, 2010, 37(8): 6086-6092
- [4] Lee S K, Cho Y H, Kim S H. Collaborative filtering with ordinal scale-based implicit ratings for mobile music recommendations. *Information Sciences*, 2010, 180(11): 2142-2155
- [5] Park H S, Yoo J O, Cho S B. A context-aware music recommendation system using fuzzy Bayesian networks with utility theory//*Proceedings of the International Conference on Fuzzy Systems and Knowledge Discovery*. Berlin, Germany, 2006: 970-979
- [6] Núñez-Valdéz E R, Lovelle J M C, Martínez O S, et al. Implicit feedback techniques on recommender systems applied to electronic books. *Computers in Human Behavior*, 2012, 28(4): 1186-1193
- [7] Huang Z, Zeng D, Chen H. A comparison of collaborative-filtering recommendation algorithms for e-commerce. *IEEE Intelligent Systems*, 2007, 22(5): 68-78
- [8] Zhao W X, Huang J, Wen J R. Learning distributed representations for recommender systems with a network embedding approach//*Proceedings of the Asia Information Retrieval Societies Conference*. Beijing, China, 2016: 224-236
- [9] Wang J, Zhao W X, He Y, et al. Leveraging product adopter information from online reviews for product recommendation//*Proceedings of the International AAAI Conference on Web and Social Media*. Oxford, UK, 2015: 464-472
- [10] Bobadilla J, Ortega F, Hernando A, et al. Recommender systems survey. *Knowledge-Based Systems*, 2013, 46(1): 109-132
- [11] Pazzani M J. A framework for collaborative, content-based and demographic filtering. *Artificial Intelligence Review*, 1999, 13(5-6): 393-408
- [12] Krulwich B. Lifestyle finder: Intelligent user profiling using large-scale demographic data. *AI Magazine*, 1997, 18(2): 37
- [13] Vozalis M G, Margaritis K G. Using SVD and demographic data for the enhancement of generalized collaborative filtering. *Information Sciences*, 2007, 177(15): 3017-3037

- [14] Balabanović M, Shoham Y. Fab; Content-based, collaborative recommendation. *Communications of the ACM*, 1997, 40(3): 66-72
- [15] Resnick P, et al. GroupLens: An open architecture for collaborative filtering of netnews//*Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work*. Chapel Hill, USA, 1994: 175-186
- [16] Linden G, Smith B, York J. Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet Computing*, 2003, 7(1): 76-80
- [17] Sarwar B, et al. Item-based collaborative filtering recommendation algorithms//*Proceedings of the 10th International Conference on World Wide Web*. Hong Kong, China, 2001: 285-295
- [18] Rendle S, Schmidt-Thieme L. Pairwise interaction tensor factorization for personalized tag recommendation//*Proceedings of the 3rd ACM International Conference on Web Search and Data Mining*. New York, USA, 2010: 81-90
- [19] Perozzi B, Al-Rfou R, Skiena S. DeepWalk, Online learning of social representations//*Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, USA, 2014: 701-710
- [20] Tang Jian, et al. Line: Large-scale information network embedding//*Proceedings of the 24th International Conference on World Wide Web*. Florence, Italy, 2015: 1067-1077
- [21] Xie Min, et al. Graph-based metric embedding for next poi recommendation//*Proceedings of the 27th Workshop on Information Systems and Economics*. Shanghai, China, 2016: 207-222
- [22] Xie Min, et al. Learning graph-based POI embedding for location-based recommendation//*Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*. Indianapolis, USA, 2016: 15-24
- [23] Zhang Shuai, Yao Lina, Sun Aixin. Deep learning based recommender system: A survey and new perspectives. *arXiv preprint arXiv:1707.07435*, 2017
- [24] Su Xiaoyuan, Khoshgoftaar T M. A survey of collaborative filtering techniques. *Advances in Artificial Intelligence*, 2009, 2009(12): 4
- [25] Wu Chao-Yuan, et al. Recurrent recommender networks//*Proceedings of the 10th ACM International Conference on Web Search and Data Mining*. Cambridge, UK, 2017: 495-503
- [26] Yu Feng, et al. A dynamic recurrent model for next basket recommendation//*Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*. Pisa, Italy, 2016: 729-732
- [27] Covington P, Adams J, Sargin E. Deep neural networks for YouTube recommendations//*Proceedings of the 10th ACM Conference on Recommender Systems*. Boston, USA, 2016: 191-198
- [28] He Xiangnan, et al. Neural collaborative filtering//*Proceedings of the 26th International Conference on World Wide Web*. Perth, Australia, 2017: 173-182
- [29] Cheng Heng-Tze, et al. Wide & deep learning for recommender systems//*Proceedings of the 1st Workshop on Deep Learning for Recommender Systems*. Boston, USA, 2016: 7-10
- [30] Sedhain S, et al. AutoRec: Autoencoders meet collaborative filtering//*Proceedings of the 24th International Conference on World Wide Web*. Florence, Italy, 2015: 111-112
- [31] Le T M V, Lauw H W. Probabilistic latent document network embedding//*Proceedings of the 2014 IEEE International Conference on Data Mining*. Shenzhen, China, 2014: 270-279
- [32] Hyvärinen A, Oja E. Independent component analysis: Algorithms and applications. *Neural Networks*, 2000, 13(4-5): 411-430
- [33] Belkin M, Niyogi P. Laplacian eigenmaps and spectral techniques for embedding and clustering//*Proceedings of the Advances in Neural Information Processing Systems*. Vancouver, UK, 2002: 585-591
- [34] Jacob Y, Denoyer L, Gallinari P. Learning latent representations of nodes for classifying in heterogeneous social networks//*Proceedings of the 7th ACM International Conference on Web Search and Data Mining*. New York, USA, 2014: 373-382
- [35] Bourigault S, Lagnier C, Lamprier S, et al. Learning social network embeddings for predicting information diffusion//*Proceedings of the 7th ACM International Conference on Web search and Data Mining*. New York, USA, 2014: 393-402
- [36] Mikolov T, et al. Distributed representations of words and phrases and their compositionality//*Proceedings of the Advances in Neural Information Processing Systems*. Stateline, Nevada, 2013: 3111-3119
- [37] Grover A, Leskovec J. node2vec: Scalable feature learning for networks//*Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco, USA, 2016: 855-864
- [38] Wang D, Cui P, Zhu W. Structural deep network embedding //*Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco, USA, 2016: 1225-1234
- [39] Yang C, Liu Z, Zhao D, et al. Network representation learning with rich text information//*Proceedings of the 24th International Joint Conference on Artificial Intelligence*. Buenos Aires, Argentina, 2015: 2111-2117
- [40] Chen W, Mao X, Li X, et al. PNE: Label embedding enhanced network embedding//*Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Jeju, South Korea, 2017: 547-560
- [41] Li J, Dani H, Hu X, et al. Attributed network embedding for learning in a dynamic environment//*Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. Singapore, 2017: 387-396

- [42] Chen W, Liu C, Yin J, et al. Mining E-commercial data: A text-rich heterogeneous network embedding approach// Proceedings of the 2017 International Joint Conference on Neural Networks. Anchorage, Alaska, 2017: 1403-1410
- [43] Chen T, Sun Y. Task-guided and path-augmented heterogeneous network embedding for author identification// Proceedings of the 10th ACM International Conference on Web Search and Data Mining. Cambridge, UK, 2017: 295-304
- [44] Li X, Chen W, Chen Y, et al. Network embedding-based representation learning for single cell RNA-seq data. *Nucleic Acids Research*, 2017, 45(19): e166-e166
- [45] Dong Y, Chawla N V, Swami A. metapath2vec: Scalable representation learning for heterogeneous networks// Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Halifax, Canada, 2017: 135-144
- [46] Muja M, Lowe D G. Fast approximate nearest neighbors with automatic algorithm configuration// Proceedings of the International Conference on Computer Vision Theory and Applications. Lisboa, Portugal, 2009: 331-340
- [47] Rendle S, Freudenthaler C, Gantner Z, et al. BPR: Bayesian personalized ranking from implicit feedback// Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence. Montreal, Canada, 2009: 452-461
- [48] Zhao W X, Wang J, He Y, et al. Mining product adopter information from online reviews for improving product recommendation. *ACM Transactions on Knowledge Discovery from Data*, 2016, 10(3): 29



LI Yu-Qi, M.S. candidate. Her research interest is machine learning.

CHEN Wei-Zheng, Ph.D. candidate. His research interest is machine learning.

YAN Hong-Fei, Ph.D., associate professor. His research interest is information retrieval.

LI Xiao-Ming, Ph.D., professor. His research interests include search engine, Web mining, social network analysis and MOOCs.

Background

In this paper, we investigate the personalized product recommendation problem, which is crucial to improve the user experience in the online e-commerce websites. During the past few years, there is a surge of research on utilizing the network embedding technique to enhance the performance of product recommendation systems. The key idea of these network embedding models is to map all the products and all the users to a common low-dimensional real vector space in which the user preference is preserved. The major disadvantage of these existing models is that the time factor is not considered, which may harm the accuracy of user preference modeling.

Our paper propose a novel model named PGE to tackle

the aforementioned weakness. First, PGE can learn similarity preserved product embedding by using a sliding time window to transform the historical adoption records into a product co-occurrence graph. Second, the user's latest preference is computed as a weighted linear combination of the embeddings of the products that the user purchased. Extensive experiments carried on a real large scale dataset show the superiority of PGE over several existing state-of-the-art methods.

This work is supported by the National Basic Research Program of China (973 Program) (Grant No. 2014CB340400), the National Natural Science Foundation (Grant No. U1536201, Grant No. 61472013).