

doi:10.11835/j.issn.1000-582X.2020.237

# 基于用户关系网络表征学习的服务推荐方法

杨宇凌<sup>a</sup>, 王澳蓉<sup>a</sup>, 吴浩<sup>a</sup>, 董琳<sup>a</sup>, 何鹏<sup>a,b</sup>

(湖北大学 a. 计算机与信息工程学院; b. 应用数学湖北省重点实验室, 武汉 430062)

**摘要:** 网络嵌入学习是深度学习的一个热门分支, 它将网络节点映射到一个拓展的低维向量空间。针对用户共用标签网络和社交网络, 利用表征学习方法得到用户标签标注关系和社交关系的向量表征, 并提出一种新的服务推荐方法。该方法利用用户的向量表征得到相似用户集, 由最终得到的用户特征信息返回 Top-k 个相似用户, 并根据相似用户的偏好情况向目标用户推荐合适的服务。为验证方法的可行性, 在公开数据集 Delicious 和 Last.FM 上进行了实验, 结果表明: 相比 4 种基准方法, 文中方法准确率可提升 13%, 召回率提升 18.6%, F-measure 值可提升 13.1%; 在学习用户表征向量时, 用户之间共用标签关系与社交关系同样重要; 推荐过程中, 为目标用户返回的相似用户值在 [25, 30] 区间更为适宜。

**关键词:** 表征学习; 网络嵌入; 标签共用网络; 服务推荐

**中图分类号:** TP391.3

**文献标志码:** A

**文章编号:** 1000-582X(2020)07-051-12

## Service recommendation based on user network representation learning

YANG Yuling<sup>a</sup>, WANG Aorong<sup>a</sup>, WU Hao<sup>a</sup>, DONG Lin<sup>a</sup>, HE Peng<sup>a,b</sup>

(a. School of Computer Science and Information Engineering; b. Hubei Key Laboratory of Applied Mathematics, Hubei University, Wuhan 430062, P. R. China)

**Abstract:** Network embedding has been a popular branch of deep learning, which represents the network information by mapping network nodes to an extended low-dimensional space. According to user co-tag network and social network, we employed network representation learning to extract the representation vector of tagging relationship and social relationship of users respectively, and proposed a novel service recommendation method, using user's representation vector learned to calculate the similar user set and recommend appropriate services to the target users according to the preferences of top-k similar users. To investigate the feasibility of our approach, experiments were carried out on two open data sets, Delicious and Last.FM. The results show that our method outperforms the four benchmarks, with an average improvement of 13% in precision, 18.6% in recall and 13.1% in F-measure. It is also found that when

**收稿日期:** 2019-12-10

**基金项目:** 国家重点研发计划(2018YFB1003801); 国家自然科学基金项目(61902114); 湖北省教育厅青年人才项目(Q20171008); 应用数学湖北省重点实验室开放基金(HBAM201901)。

Supported by National Key Research and Development Program of China(2018YFB1003801), National Natural Science Foundation of China(61902114), Hubei Province Education Department Youth Talent Project(Q20171008) and Hubei Provincial Key Laboratory of Applied Mathematics(HBAM201901).

**作者简介:** 杨宇凌(1999—), 女, 主要研究领域为服务计算、深度学习。

**通讯作者:** 何鹏, 男, 副教授, 主要从事软件工程和复杂网络研究, (E-mail) penghe@hubu.edu.cn。

learning user representation, the co-tag relationship between users is as important as the social relationship. Meanwhile, during the process of collaborative recommendation, the number of similar users returned for a target user is suitable in the range from 25 to 30.

**Keywords:** representation learning; network embedding; co-tag network; service recommendation

随着各种资源信息的爆炸式增长<sup>[1]</sup>,用户需要花费大量的时间去搜索他们感兴趣的服务。准确地向用户推荐所需服务是一项迫切的需求。标签能提高共享、发现和检索的效率<sup>[2]</sup>,服务的标签能直接体现服务本身的特性,用户标注的标签既是用户兴趣偏好的体现,也能间接地反映用户之间的关系(如标签共用关系)。因此,社会化标签被广泛地应用于个性化的服务推荐<sup>[3]</sup>。

随着社交网络平台的出现,用户之间的交互更加便捷,用户的社交关系变得更加丰富。相同喜好的用户会聚集成社区,一个用户又会同时参与多个社区<sup>[4]</sup>,这些信息有助于更好地挖掘用户对服务的偏好和用户之间影响。因此,基于社交网络的个性化服务推荐也被许多研究者所关注。例如,Deng 等<sup>[5]</sup>提出了一种基于社交网络的增强信任的服务推荐方法。

众所周知,分析和处理大规模的网络是费时费力的。近年来,网络表征学习(network representation learning)作为一种新的方便有效的学习方法,被广泛关注,成为深度学习和数据挖掘领域中一个快速发展的方向<sup>[6]</sup>。有学者成功地将网络表征学习方法运用于服务推荐,发现这样可以有效地获取用户与事项之间复杂的非线性关系,克服传统推荐模型的一些不足,达到提高推荐质量的效果<sup>[7]</sup>。但是,目前还少有研究者利用网络表征学习方法(network embedding technique)从用户社交网络 and 用户标签共用网络中提取用户的偏好特征。

因此,笔者根据用户的标签标注关系和社交关系,构建相应的用户信息网络,并引入典型的网络表征学习方法 Node2vec<sup>[8]</sup>,尝试从不同用户信息网络中学习用户的特征偏好,并最后用于服务的协同推荐。1)引入网络表征学习方法提取用户的隐性特征向量,取代传统的 TF-IDF(term frequency-inverse document frequency)加权方法。2)综合考虑用户标签共用关系和社交关系,提出一种新的服务推荐模型,并在 2 个公开数据集上进行了实证分析。3)针对 2 个数据集进行相关的参数训练,使得模型的推荐质量达到最优效果。内容的组织框架为:第二部分归纳了目前已取得的一些相关成果;第三部分详细描述了文中服务推荐模型;第四部分陈述了实验数据和实验结果分析。

## 1 相关工作

推荐系统是当下一种重要的信息过滤手段,其中,协同过滤(collaborative filtering)因其高效、准确和可扩展性被广泛应用于大规模的推荐引擎中<sup>[9]</sup>。协同过滤按照推荐类型可划分为基于模型(model-based)、基于记忆(memory-based)和基于上下文(context-based)的推荐 3 大类<sup>[10-13]</sup>。

### 1.1 基于标签的服务推荐

Shi 等<sup>[14]</sup>为了预测用户对全新的媒体服务的评分,利用基于主题向量去类比全新的媒体服务与以往的媒体服务的相似程度,并根据得到的相似关系对用户进行推荐。该方法提高了推荐效率,有效地解决了冷启动的问题;Zhang 等<sup>[15]</sup>针对冷启动和简化问题,提出一种基于矩阵分解的混合标签多样化推荐算法,根据用户对服务的行为数据挖掘出用户与标签的隐性关系,该方法不仅能保证推荐的准确率,也提高了标签多样性;Qiang 等<sup>[16]</sup>提出一种基于多标签传播的算法去预测出一个多维的社区结构,利用多维的社区结构向用户推荐服务,推荐精度和效率都有所提高。

与此同时,TF-IDF 作为一种用于信息检索与数据挖掘的常用加权技术,常被应用于基于标签的服务推荐中,用以评估用户间的相似度,比如,Ba 等<sup>[17]</sup>利用 TF-IDF 方法计算目标用户与邻居之间的相似度,而不用考虑事项标签的详细评级,该方法可以提高推荐系统的精度、实时性和可扩展性;Otsuka 等<sup>[18]</sup>提出 TF-

IDF 变体排名方案,依据相似用户的标签信息帮助用户找到可能感兴趣的新标签,并且在 Twitter 提供的公开数据集上证明了该方法的有效性。

### 1.2 基于社交关系的服务推荐

Jiang 等<sup>[19]</sup>提出一种基于社会关系和时间影响的个性化服务推荐,综合考虑了用户熟悉度和偏好相似度,该方法不依赖于用户的历史数据,可以满足为新用户和不活跃用户进行服务推荐的需求;Yuan 等<sup>[20]</sup>将用户之间的好友关系和组员关系 2 种不同类型的社交关系通过因子分解过程融合到协同推荐系统中,综合考虑 2 种关系能够有效改善推荐的质量和效率;Meo 等<sup>[21]</sup>提出一种基于矩阵分解和社交关系组合的协同服务推荐,该方法能为用户提供更准确的建议和排名;Rafailidis 等<sup>[22]</sup>发现选择亲密的社交朋友可以提高推荐的准确性,提出基于亲密度比较的社交关系模型 SCR,该模型的性能优于目前最先进的 CR 模型;Liu 等<sup>[23]</sup>也提出一种结合用户社交关系和个性化标签的服务推荐模型,用图模型模拟协作用户的社会标注和他们之间的社交关系,在公开数据集上证明模型的有效性;Chen 等<sup>[24]</sup>提出了变异 SimRank 算法,通过提取用户社交关系的显式信息计算偏好数据,提高原模型的推荐质量。

### 1.3 基于表征学习的服务推荐

网络表征学习旨在将网络中的节点表示成低维、实值、稠密的向量形式,使得到的向量可以在向量空间中具有表示和推理的能力,可以更加灵活地应用于不同的数据挖掘和推荐任务中,引起了越来越多的关注。网络表征学习方法也成功应用到服务推荐领域,例如,Tran 等<sup>[25]</sup>在人脸识别领域,提出表征学习生成对抗网络 DR-GAN(disentangled representation learning-generative adversarial network),评价结果表明,改良后的生成对抗网络优于目前的技术水平;Tang 等<sup>[26]</sup>提出一种基于卷积神经网络的联合表征学习模型,将用户和事件投影到同一个潜在空间来捕捉特征信息,该模型的预测精度 AUC 值提高了 6%;Kong 等<sup>[27]</sup>利用表征学习得到引用网络论文的特征向量,开发了科学的论文推荐系统 VOPRec(vector representation learning of papers with text information and structural identity for recommendation),能满足用户进行科学论文推荐的需求;Zhuang 等<sup>[28]</sup>提出新的表征学习框架,通过自动编码器学习用户和项目间的隐性表示,经比较发现该学习模型能解决矩阵分解存在的缺陷,并且表示的多样性和可扩展性都优于现有的矩阵分解方法;Wu 等<sup>[29]</sup>提出新的上下文感知用户表征学习模型 CARL(context-aware user-item representation learning)进行评分预测,通过因子分解机器,建模高阶特征交互,用于评级预测。Han 等<sup>[30]</sup>提出了一种基于异质信息网络嵌入和卷积神经网络混合的表征学习方法,通过表征学习从用户与项目交互结构中获取用户和项目的特征表示;Wu 等<sup>[31]</sup>提出一种利用知识图谱的表征学习方法,将现有的语义数据表征到低维语义空间,提高了协同过滤推荐在语义层面的效果;Shi 等<sup>[32]</sup>提出基于元路径的随机游走策略来生成有意义的节点序列进行网络嵌入,并将其扩展到 MF 模型中,证明了该模型能有效地解决冷启动问题。

## 2 研究方法

文中方法的研究框架如图 1 所示,主要包括 3 部分:1)构建用户信息网络模型:根据用户在服务上标注的标签信息,构建用户-标签异质信息网络,再通过投影方式得到用户标签共用网络;另外,根据用户之间的 contact 和 follow 关系,构建用户社交网络。2)网络表征学习:对于得到的 2 类用户信息网络,分别利用网络表征学习方法通过网络嵌入(network embedding)学习用户的特征向量;再通过线性组合方式将得到的 2 类表征进行融合,得到最终的用户偏好特征向量。3)基于用户的协同推荐:指定一个目标用户,利用最终得到的用户特征信息,为其返回 Top-k 个相似用户,根据相似用户的偏好情况向目标用户推荐合适的服务。

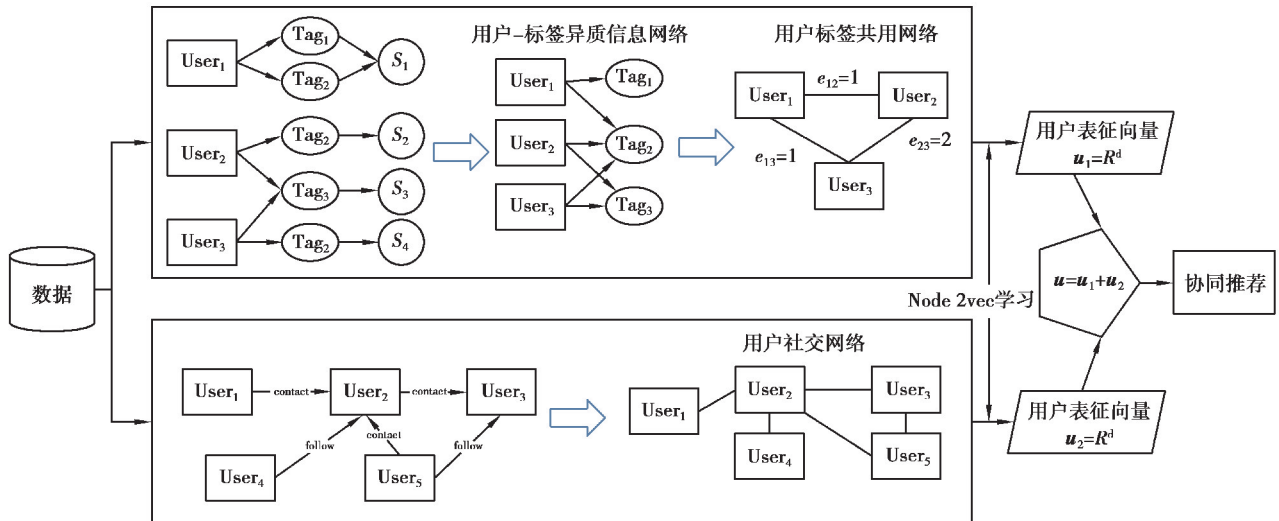


图 1 文中方法的主要框架

Fig. 1 The main framework of our approach

## 2.1 用户信息网络建模

用户-标签异质信息网络(user-tag heterogamous information network):异质信息网络简单的理解是指网络中节点或连边的种类在 2 种及以上的网络。在文中,根据用户对服务进行标注时使用的标签信息,可构建用户-标签异质信息网络  $N_{\text{UTHIN}} = \{U \cup T, E\}$ ,其中,  $U = \{u_1, u_2, \dots, u_m\}$  为用户集,  $T = \{t_1, t_2, \dots, t_n\}$  为标签集,  $E = \{e_{ij} | i=1, 2, \dots, n, j=1, 2, \dots, m\}$  为连边集,当存在一条连边  $e_{ij} = w$  时,代表用户  $u_i$  在服务  $w$  上标注了  $t_j$  标签。

用户标签共用网络(user co-tag network):为了提取出用户之间的关系,对用户-标签异质信息网络  $N_{\text{UTHIN}}$  进行投影,可得到用户标签共用网络  $N_{\text{UCTN}} = \{U', E'\}$ ,其中,  $U' \subseteq U$ ,如果用户  $u_i$  和  $u_j$  使用过相同标签,则视他们之间有一条连边,表现为  $e'_{ij} = w_{ij}$ ,  $w_{ij}$  为 2 个用户共有的标签数。

用户社交网络(user social network):现在流行很多公共服务平台,比如,Delicious 和 Last.FM。通过这些平台用户可以给服务标注标签,也可以跟其他用户交流。根据平台上提供的用户交互数据,假如,一个用户  $u_i$  关注(contact)或 follow 了另一个用户  $u_j$ ,则视 2 个用户之间有一次社交关联。因此,用户之间的社交关系网络  $N_{\text{Socialnet}} = (U, E)$ ,  $U = \{u_i\}_{i=1, \dots, m}$  为用户集,  $E = \{e_{ij} | i, j=1, 2, \dots, m\}$  为连边集,若用户  $u_i$  和  $u_j$  之间有交互关系,则  $e_{ij} = 1$ , 否则为 0。考虑到虚拟社交平台中交互频繁的用户可能正是因为彼此之间不熟悉,导致沟通效率不高;反而,2 个彼此熟悉的人在某个问题上可能只需通过一次简单的交互就能达到很好的沟通效果。因此,把用户之间的社交网络只作为一个无向非加权的网络处理。

## 2.2 网络嵌入方法

将信息网络映射到一个低维向量空间的方法,已在机器学习领域得到有效验证,网络表征学习方法能够有效地学习信息网络中节点之间的局部和整体结构信息。在服务推荐过程中,针对用户标签共用网络与社交网络,引入网络嵌入方法处理 2 类用户信息网络,将网络中每个用户节点  $u$  转化为一个  $d$  维的向量  $\mathbf{u}$ ,来表征每个用户的信息,将得到的 2 类用户表征向量通过组合得到用户最终表征向量。

首先,利用随机游走的方法将网络转换成一个节点序列集合。一条节点序列可类比为为一个句子,节点序列集合则类比为一段文本。然后,借鉴 Skip-gram 模型的学习方法,利用节点序列集合学习出每个节点的特征。

### 2.2.1 随机游走策略

2 种常规游走策略:深度优先遍历和广度优先遍历。深度优先遍历能保留网络的结构信息,而广度优先遍历能保留网络的邻域信息。借鉴 Node2vec 方法引入参数  $p$  和  $q$  优化随机游走策略。在网络  $\text{Net} = \{V, E\}$  中,随机选择节点  $u$  为源节点,令  $N(u)$  为随机游走产生的节点序列,  $n_i$  表示节点序列中第  $i$  个节点,

选择  $n_i$  的条件概率计算公式如下:

$$p(n_i = x \mid n_{i-1} = v, n_{i-2} = t) = \begin{cases} \frac{\beta_{pq}(t, x) \omega_{vx}}{Z} & (v, x) \in E, \\ 0 & (v, x) \notin E, \end{cases} \quad (1)$$

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

如式(1)所示,在边权重  $\omega_{vx}$  前加了一个偏重  $\beta_{pq}(t, x)$ ,  $Z$  是一个常数。在式(2)中,  $d_{tx}$  表示节点  $t$  和  $x$  之间的网络局部结构,当节点  $t$  和  $x$  之间有连边时,  $d_{tx} = 1$ ;当节点  $t$  和  $x$  重合时,  $d_{tx} = 0$ ;当节点  $t$  和  $x$  之间没有连边时,  $d_{tx} = 2$ 。因此, Node2vec 的随机游走策略可以根据不同的网络局部结构动态计算,选择不同方向的条件概率,从而动态选择遍历的方向。

### 2.2.2 表征学习

对于每个源节点  $v \in V$ ,  $N(v) \subset V$  是从节点  $v$  出发,通过随机游走得到的节点序列集合。由节点  $v$  得到节点序列  $N(v)$  的条件概率计算公式为

$$p(N(v) \mid v) = \prod_{i=1}^{|N(v)|} p(v_i \mid v). \quad (3)$$

其中,将  $p(v_i \mid v)$  模型化为一个由各节点表征向量组成的 softmax 单元,为

$$p(v_i \mid \mathbf{v}) = \frac{\exp(\mathbf{v}T\mathbf{v}_i)}{\sum_{k=1}^{|V|} \exp(\mathbf{v}T\mathbf{v}_k)}. \quad (4)$$

式中,  $\mathbf{v} \in R^d$  为节点  $v$  的  $d$  维表征向量,  $|V|$  表示网络的节点个数。考虑到直接对目标函数  $f_1$  进行训练,资源消耗较大,文中引入负采样方法进行优化。其思想是利用负采样代替全局参数进行训练,从而减少大量的参数训练带来的资源消耗。其中,  $N_{\text{NEG}}(v_i)$  表示以  $v_i$  为中心的负采样节点集。

$$P(z \mid v_i) = \begin{cases} q(\mathbf{v}_i T \mathbf{v}_z) & z \in N(v_i) \\ 1 - q(\mathbf{v}_i T \mathbf{v}_z) & z \in N_{\text{NEG}}(v_i) \end{cases} \quad q(x) = \frac{1}{1 + e^{-x}}. \quad (5)$$

将网络中所有节点的条件概率累加,就得到目标函数:

$$f_1 = \sum_{v \in V} \sum_{i=1}^{|N(v)|} p(v_i \mid v). \quad (6)$$

通过学习训练找到合适的向量  $\{\mathbf{v}_i\}_{i=1, \dots, |V|}$ , 使式(5)最大化,采用随机梯度下降(stochastic gradient descent)方法进行优化。

### 2.3 基于用户的协同推荐

对于每个节点  $v_i$ , 令  $\mathbf{v}_i^1$  为从用户标签共用网络中学习得到的用户表征向量,  $\mathbf{v}_i^2$  为从用户社交网络中学习得到的表征向量。节点  $v_i$  最终的表征向量  $\mathbf{v}_i$  表示为  $\mathbf{v}_i^1$  与  $\mathbf{v}_i^2$  的线性组合,即:

$$\mathbf{v}_i = \alpha \mathbf{v}_i^1 + (1 - \alpha) \mathbf{v}_i^2. \quad (7)$$

基于所得到的用户表征向量,文中通过向量的余弦相似度来获取目标用户的相似用户集  $S(u)$ , 取代传统计算用户标签集相似度或者用户社交网络中共同好友个数的方式。

$$\text{Sim}(u_i, u_j) = \frac{\mathbf{u}_i \mathbf{u}_j}{\|\mathbf{u}_i\| \times \|\mathbf{u}_j\|}. \quad (8)$$

对于目标用户  $u$  是否选择服务  $s$ , 根据其相似用户集中用户的选择情况进行判断。对于服务  $s$ , 用户  $u$  选择它的概率为

$$p(u, s) = \sum_{n \in S(u) \cap N(s)} \text{sim}(u, n) \cdot r_{ns}. \quad (9)$$

式中,  $r_{ns}$  表示用户  $u$  的相似用户  $v$  对服务  $s$  的喜欢程度, 例如评分大小。  $N(s)$  为已经选择了服务  $s$  的用户集。最后可返回用户  $u$  对一系列未曾关注过的服务感兴趣的概率列表, 并将靠前的 Top-k 个服务进行优先推荐。

### 3 实验分析

#### 3.1 实验数据

实验数据来源于 Grouplens 平台。Delicious 数据集中包括用户 1 867 个,服务(书签)104 799 个和标签 53 388 个。Last.FM 数据集包括用户 1 892 个,服务(音乐)17 632 个和标签 11 946 个。数据集中,有用户对服务标注标签的行为数据,也有用户之间的交互记录。用户信息网络建模后,所得网络信息统计如表 1 所示。为提高实验数据的质量,过滤掉了关注服务个数少于 5 个的用户。

表 1 实验数据

Table 1 Experimental data

名称	用户	服务	标签	网络	节点数	连边数
Last.FM	1 892	17 632	11 946	用户标签共用网络	1 207	4 483
				用户社交网络	1 132	2 357
Delicious	1 867	104 799	53 388	用户标签共用网络	1 194	7 621
				用户社交网络	1 037	4 889

#### 3.2 评价指标

用于评价推荐结果的方法有很多,文中采用常用准确率(precision)、召回率(recall)和综合评价指标 F-Measure 3 个指标。对于目标用户  $u_i$ ,令  $s_i$  表示该用户真实选择的服务集合,令  $s_i^{\text{TopK}}$  表示向该用户推荐的 Top-K 个服务。对于测试用户集  $|U_{\text{test}}|$ ,指标的定义如下:

$$P_{\text{Precision@K}} = \frac{1}{|U_{\text{test}}|} \sum_1^{|U_{\text{test}}|} \frac{|s_i \cap s_i^{\text{TopK}}|}{|s_i^{\text{TopK}}|}, \quad (10)$$

$$R_{\text{Recall@K}} = \frac{1}{|U_{\text{test}}|} \sum_1^{|U_{\text{test}}|} \frac{|s_i \cap s_i^{\text{TopK}}|}{|s_i|}, \quad (11)$$

$$F\text{-}M_{\text{Measure@K}} = \frac{2 \times P_{\text{Precision@K}} \times R_{\text{Recall@K}}}{P_{\text{Precision@K}} + R_{\text{Recall@K}}}. \quad (12)$$

为了验证文中方法的有效性,选取 UPCC (user-based pearson correlation coefficient)<sup>[33]</sup>、WSRec (web service recommender)<sup>[34]</sup>、SoRec (social recommendation)<sup>[35]</sup> 和 Truser (trustworthy user based service recommendation)<sup>[36]</sup>, 4 种已有方法作为实验的基准方法。UPCC 是一种基于用户对服务的偏好找到相似用户,将相似用户喜欢的服务推荐给当前用户的方法;WSRec 是一种基于用户和事项的协同过滤(混合)推荐方法;SoRec 是一种基于社交网络和评分矩阵融合的推荐方法;Truser 是一种基于可信用户的两阶段聚类服务推荐的最新方法。文中基于关系网络表征学习的服务推荐方法为 RLSR (represent learning-based service recommendation)。

#### 3.3 实验结果

##### 3.3.1 推荐结果评价

为验证 RLSR 方法的有效性,随机从数据集中抽取一定比例  $\gamma$  (1%~10%) 的用户进行测试。在网络表征学习阶段,设置用户向量维度为默认值 ( $d=128$ ),游走步长  $l=50$ ,每个节点游走次数  $r=10$ ,偏重参数  $p=q=1$ ,参数设置参照文献[8]中的训练经验值。在基于用户协同过滤时,设置表征向量组合参数  $\alpha=0.5$ ,相似用户集  $|S(u)|=20$ 。当 Top-k=10 时,在 2 个数据集上的召回率(Rec@k)和准确率(Pre@k)结果如表 2 所示。

实验结果显示,在 2 组数据集的实验中,无论  $\gamma$  比例如何变化,RLSR 方法的准确率和召回率始终优于 4 种基准方法,即推荐质量得到提高。尤其是与常规的推荐方法 UPCC 和 WSRec 相比,改进更为明显。对于 SoRec 和 Truser 这 2 种较新的方法,在 Delicious 数据集上,RLSR 方法的平均准确率达 0.635,提升幅度分别为 12.9% 和 5.7%;平均召回率达 0.593,提升幅度更大,分别为 19.2% 和 7%。而在 Last.FM 数据集上,RLSR 方法的平均

准确率达 0.572,提升幅度分别为 13%和 11.7%;平均召回率达 0.534,提升幅度分别为 18.9%和 13.2%。

表 2 各方法在 2 个数据集上的推荐结果

Table 2 Recommended results of different methods on two datasets

数据集	指标	方法	比例 $\gamma$										平均值
			1	2	3	4	5	6	7	8	9	10	
Delicious	Rec@10	UPCC	0.34	0.42	0.39	0.42	0.31	0.41	0.37	0.33	0.37	0.33	0.369
		WSRec	0.51	0.47	0.49	0.43	0.41	0.51	0.43	0.41	0.47	0.49	0.462
		SoRec	0.51	0.51	0.53	0.51	0.57	0.57	0.54	0.50	0.51	0.50	0.525
		Truser	0.51	0.57	0.53	0.59	0.57	0.53	0.59	0.59	0.59	0.54	0.531
		RLSR	0.61	0.59	0.57	0.60	0.60	0.58	0.59	0.61	0.61	0.57	0.593
	Pre@10	UPCC	0.39	0.45	0.41	0.47	0.39	0.42	0.40	0.40	0.41	0.39	0.413
		WSRec	0.51	0.49	0.50	0.52	0.47	0.57	0.43	0.44	0.50	0.54	0.497
		SoRec	0.54	0.57	0.53	0.57	0.42	0.51	0.57	0.53	0.56	0.53	0.534
		Truser	0.53	0.57	0.57	0.57	0.51	0.51	0.51	0.50	0.54	0.50	0.561
		RLSR	0.65	0.63	0.65	0.61	0.63	0.63	0.63	0.61	0.64	0.67	0.635
Last.FM	Rec@10	UPCC	0.31	0.38	0.35	0.38	0.28	0.37	0.33	0.30	0.33	0.30	0.332
		WSRec	0.46	0.42	0.44	0.39	0.37	0.46	0.39	0.37	0.42	0.44	0.416
		SoRec	0.46	0.46	0.48	0.46	0.51	0.51	0.49	0.45	0.46	0.45	0.473
		Truser	0.46	0.53	0.53	0.48	0.49	0.53	0.53	0.51	0.48	0.51	0.505
		RLSR	0.55	0.53	0.51	0.54	0.54	0.52	0.53	0.51	0.55	0.51	0.534
	Pre@10	UPCC	0.35	0.41	0.37	0.42	0.35	0.38	0.36	0.36	0.37	0.35	0.372
		WSRec	0.46	0.44	0.45	0.47	0.42	0.51	0.39	0.40	0.45	0.49	0.447
		SoRec	0.49	0.51	0.48	0.51	0.38	0.46	0.51	0.48	0.51	0.48	0.480
		Truser	0.48	0.51	0.51	0.51	0.46	0.46	0.46	0.45	0.49	0.45	0.534
		RLSR	0.59	0.57	0.59	0.55	0.57	0.57	0.57	0.55	0.58	0.60	0.572

为了进一步比较各方法综合推荐效果,图 2 给出了各方法推荐结果的 F-measure 值。相比 2 个较新的基准方法,在 2 个数据集上,RLSR 方法  $F_1$  值的提升幅度为 13.1%和 14.1%,11.7%和 8.5%,能进一步提高服务推荐的质量。网络表征学习方法可以得到局部和全局的结构信息,笔者的实验也证明,在服务推荐中,2 种用户网络通过表征学习方法,有利于更好地找出用户的相似用户。

### 3.3.2 参数影响分析

1)参数  $l, r, p$  和  $q$  的影响分析。

对于随机游走步长  $l$ ,令其他参数为默认值,分析随步行长度  $l$  从 10~100 的变化,推荐结果的 F-Measure

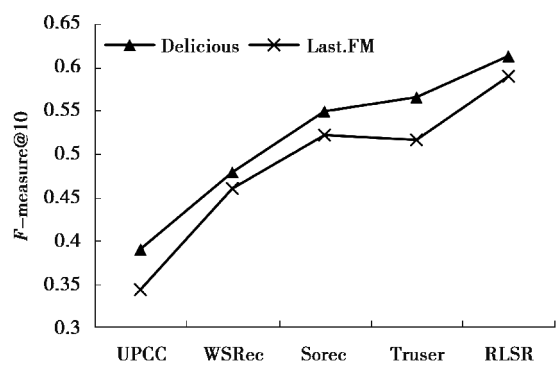


图 2 各方法在 2 个数据集上 F-measure 结果对比  
Fig. 2 Comparison of F-measure results of each method on two datasets

值,如图 3 所示。结果在 2 组数据集上都呈现出上升的趋势,而且当  $l$  超过 50 时,曲线的上升幅度不断减小,最后表现为趋于稳定。例如,当  $l$  增加到 100 时,在 Delicious 数据集上的 F-Measure@10 值为最大值 0.63;相应的,Last.FM 数据集上的 F-Measure@10 值也达到最大值 0.59。对于 F-Measure 值提升的一种解释是,较长的随机游走步行可以更好地保留网络的全局结构信息,提高网络表征学习的质量。

对于每个节点游走次数  $r$ ,设置其他参数都采用默认值,分析  $r$  从 10~100 变化时的 F-Measure 值变化,如图 4 所示。在 2 组数据集上结果都呈现出一种上升的趋势,当  $l$  超过 80 后,曲线的上升幅度减小,最后趋于稳定。一种解释是,网络中每个节点的游走次数越多,得到的节点序列集更能代表节点在网络中的结构信息,提高网络表征学习的质量。还涉及 2 个重要参数,即控制游走策略的 2 个参数  $p$  和  $q$ 。为分析其影响,保持其他参数不变,令  $\log_2 p$  和  $\log_2 q$  分别从 -4~4 变化。如图 5 和图 6 所示,整体上  $p$  和  $q$  值越小,推荐结果的 F-Measure 值越大。其中,Delicious 数据集在  $\log_2 p = -4$  时的 F-Measure 为 0.68,在  $\log_2 q = -4$  时的 F-Measure 为 0.67。这表明网络表征学习的质量会随着  $p$  和  $q$  值的降低而提高。因为较小的  $p$  值鼓励深度游走,有利于获取全局结构信息,但较小的  $q$  值反过来又可以助力平衡游走深度,使得向外游走时又始终离起始节点不至于太远。因此,参数  $p$  和  $q$  不应该设置得太大。如果参数太小,对游走的方向和速度的控制将不明显,可考虑取值范围在  $[2^{-3}, 2^{-2}]$ 。

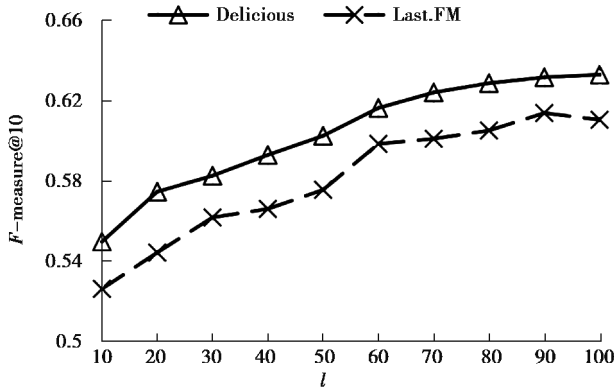


图 3 参数  $l$  对推荐结果的影响

Fig. 3 Impact of parameter  $l$  on recommended results

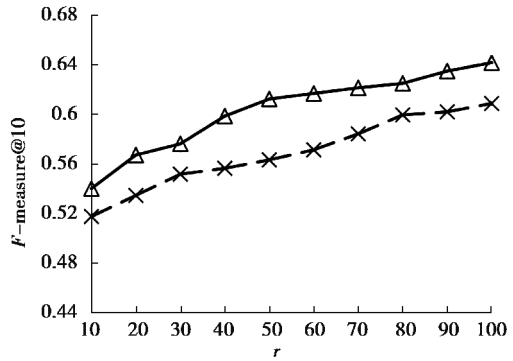


图 4 参数  $r$  对推荐结果的影响

Fig. 4 Impact of parameter  $r$  on recommended results

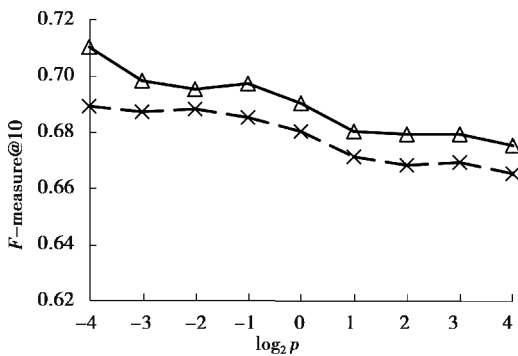


图 5 参数  $p$  对推荐结果的影响

Fig. 5 Impact of parameter  $p$  on recommended results

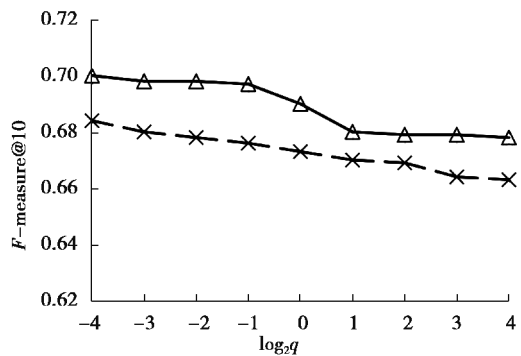


图 6 参数  $q$  对推荐结果的影响

Fig. 6 Impact of parameter  $q$  on recommended results

2) 参数  $|S(u)|$  和  $\alpha$  的影响分析。

在推荐过程中,返回给目标用户的相似用户集  $|S(u)|$  的大小也会对推荐性能产生一定影响。进一步分析参数  $|S(u)|$  的影响,设置  $\alpha = 0.5$ ,  $l = 50$ , 参数  $p = q = 1$ , 令  $|S(u)|$  从 5~50 变化,结果如图 7 所示。总体来说,随着  $|S(u)|$  值的增加,F-Measure 值先增加后减少。例如,当  $|S(u)|$  增加到 25 时,Delicious 数据集



上的 F-Measure@5 值为最大值 0.61。然而,对于 F-Measure@10 和 F-Measure@15,当  $|S(u)|$  值为 30 和 25 的时候为最大,分别为 0.62 和 0.64。同样,在 Last.FM 数据集上, $|S(u)|=30$  时结果最佳,分别为 0.55 和 0.59。此外,当  $|S(u)|=5$  时,3 种  $k$  值情况下的 F-Measure 值都为最低,说明返回给目标用户的相似用户集太少也会导致推荐质量下降。但是返回相似用户过多,同样不利于推荐效果。

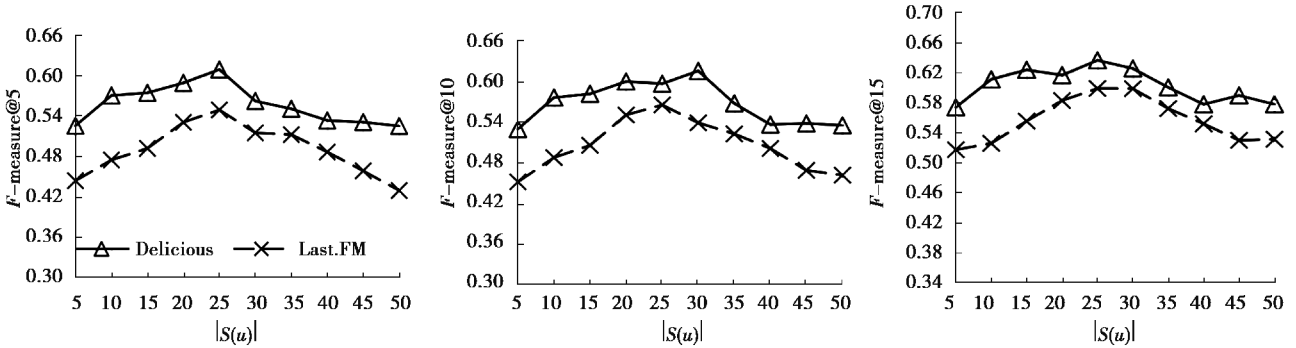


图 7 参数  $|S(u)|$  对推荐结果的影响

Fig. 7 Impact of parameter  $|S(u)|$  on recommended results

在融合 2 类用户表征过程中,权重系数  $\alpha$  用于衡量用户标签共用关系和社交关系的重要性。采取保持其他参数为默认值,令参数  $\alpha$  从 0~1 变化,且每次增加幅度为 0.1,结果如图 8 所示。整体上,从两类用户信息网络中学习到的表征向量对于服务推荐都重要,其中标签共用关系提供的信息相对更充沛。具体表现为:对于 Delicious 数据集,F-Measure@5、F-Measure@10 和 F-Measure@15 的最佳值分别为 0.59、0.61 和 0.64。对于 Last.FM 数据集,F-Measure@5、F-Measure@10 和 F-Measure@15 的最优值分别为 0.55、0.56 和 0.58。

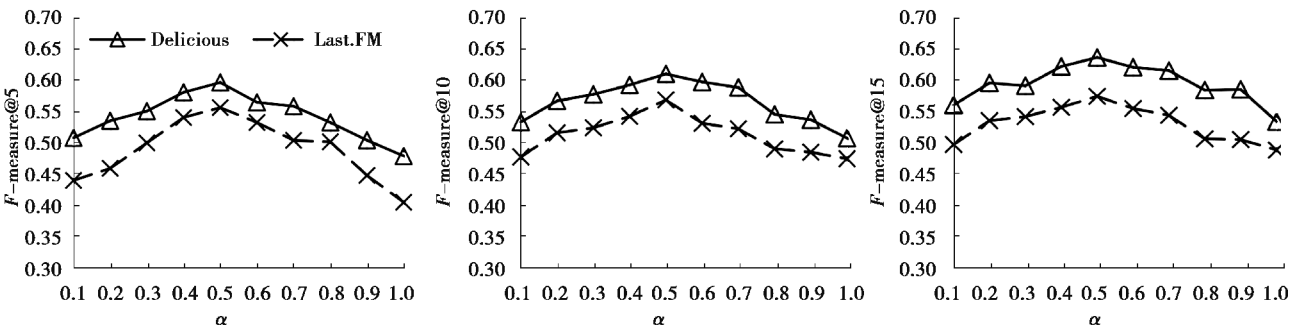


图 8 参数  $\alpha$  对推荐结果的影响

Fig. 8 Impact of parameter  $\alpha$  on recommended results

## 4 结 论

文中提出一种基于关系网络表征学习方法的服务推荐方法。构建了用户共用标签网络和社交网络,利用表征学习方法分别从这 2 个网络中提取用户表征向量,并将学习到的表征向量进行线性组合的方式集成为最终的用户表征。利用扩展的用户表征与协同过滤联合优化服务推荐。方法在 2 个公共数据集上的验证。结果表明,文中的 RLSR 方法优于 4 种基准方法,表现为准确率、召回率和 F-measure 值的平均提升幅度分别可达 13%、18.6% 和 13.1%。实验结果还表明,在融合用户表征向量时,用户的共用标签关系与社会关系同样重要;另外,在协同推荐时,为目标用户返回的相似用户不宜过少与过多,25 个较为适宜。

新的网络表征学习方法,如 LINE (large-scale information network embedding)<sup>[37]</sup>、SNE (social network embedding)<sup>[38]</sup>、SDNE(structural deep network embedding)<sup>[39]</sup>等,都对传统的网络表征学习进行了改进。在对用户信息网络进行表征学习时,并没有考虑用户节点本身的属性信息,例如,用户的活跃度、年龄、性别等。活跃度不同的用户对周边用户的影响程度往往有所不同,不同年龄段的用户对同一个服务的标签标注行为也会有所差异。因此,后期工作可进一步在网络表征学习过程中引入节点的属性信息,从而更精准地表征用户偏好。

#### 参考文献:

- [ 1 ] Tan S, Li Y, Sun H, et al. Interpreting the public sentiment variations on twitter[J]. IEEE Transactions on Knowledge and Data Engineering, 2014, 26(5): 1158-1170.
- [ 2 ] Rafailidis D, Daras P. The TFC model: tensor factorization and tag clustering for item recommendation in social tagging systems[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2013, 43(3): 673-688.
- [ 3 ] Fan M, Xiao Y, Zhou Q. Bringing the associative ability to social tag recommendation[C]. Workshop Proceedings of TextGraphs-7 on Graph-based Methods for Natural Language Processing. Stroudsburg, PA, USA: Association for Computational Linguistics, 2012:44-54.
- [ 4 ] Kohi A, Ebrahimi S J, Jalali M. Improving the accuracy and efficiency of tag recommendation system by applying hybrid methods[C/OL]. 2011 1st International eConference on Computer and Knowledge Engineering (ICCKE). New York, USA: IEEE, 2013(2013-01-17)[2019-09-25]. <https://ieeexplore.ieee.org/document/6413358>.
- [ 5 ] Deng S G, Huang L T, Xu G D. Social network-based service recommendation with trust enhancement[J]. Expert Systems With Applications, 2014, 41(18): 8075-8084.
- [ 6 ] Ignacio Fernández-Tobías, Blanco R. Memory-based recommendations of entities for web search users[C]. CIKM '16 Proceedings of the 25th ACM International on Conference on Information and Knowledge Management. New York, USA: ACM Press 2016:35-44.
- [ 7 ] Noh K J, Park Y, Moon K. Topic model-based recommendation system for media re-creation service[C/OL]. 2016 International Conference on Information and Communication Technology Convergence (ICTC). New York, USA: IEEE, 2016(2016-12-05)[2019-09-25]. <https://ieeexplore.ieee.org/document/7763452>.
- [ 8 ] Grover A, Leskovec J. Node2vec: Scalable feature learning for networks[C]. AcmSigkdd International Conference on Knowledge Discovery & Data Mining. New York, USA: ACM Press, 2016:855-866.
- [ 9 ] Yuan J B, Shalaby W, Korayem M, et al. Solving cold-start problem in large-scale recommendation engines: A deep learning approach[C/OL]. 2016 IEEE International Conference on Big Data (Big Data). New York, USA: IEEE, 2017 (2017-02-06)[2019-09-25]. <https://ieeexplore.ieee.org/document/7840810>.
- [10] Zhao Z, Lu H Q, Cai D, et al. User preference learning for online social recommendation[J]. IEEE Transactions on Knowledge and Data Engineering, 2016, 28(9): 2522-2534.
- [11] Younes, Idrissi E B E, Rachida, et al. Context-based services selection and recommendation through P-learning platform [C/OL]. 2010 9th International Conference on Information Technology Based Higher Education and Training (ITHET). New York, USA: IEEE, 2010(2010-06-07)[2019-09-25]. <https://doi.org/10.1109/ithet.2010.5480054>
- [12] Noh K J, Park Y, Moon K. Topic model-based recommendation system for media re-creation service[C/OL]. 2016 International Conference on Information and Communication Technology Convergence (ICTC). New York, USA: IEEE, 2016(2016-12-05)[2019-09-25]. <https://ieeexplore.ieee.org/document/7763452/>.
- [13] Ignacio Fernández-Tobías, Blanco R. Memory-based recommendations of entities for web search users[C]. CIKM '16 Proceedings of the 25th ACM International on Conference on Information and Knowledge Management. New York, USA: ACM Press, 2016:35-44.
- [14] Shi W S, Liu X M, Yu Q. Correlation-aware multi-label active learning for web service tag recommendation[C/OL]. 2017 IEEE International Conference on Web Services (ICWS). New York, USA: IEEE, 2017(2017-09-11)[2019-09-25]. <https://doi.org/10.1109/icws.2017.37>.

- [15] Zhang S S, Xia Y, Li X R. Diversified recommendation algorithm for hybrid label based on matrix factorization[C/OL]. 2018 IEEE 3rd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA). New York, USA: IEEE, 2018(2018-06-18)[2019-09-25]. <https://doi.org/10.1109/icccbda.2018.8386484>.
- [16] Qiang H, Yan G. A method of personalized recommendation based on multi-label propagation for overlapping community detection[C/OL]. 2012 3rd International Conference on System Science, Engineering Design and Manufacturing Informatization, New York, USA: IEEE, 2012 (2012-12-12) [2019-09-25]. <https://ieeexplore.ieee.org/document/6340748/>.
- [17] Ba Q, Li X, Bai Z. A Similarity calculating approach simulated from TF-IDF in collaborative filtering recommendation[C]. International Conference on Multimedia Information NETWORKING & Security. Washington, DC, USA: IEEE Computer Society, 2013:738-741.
- [18] Otsuka E, Wallace S A, Chiu D. Design and evaluation of a Twitter hashtag recommendation system[C]. Proceedings of the 18th International Database Engineering & Applications Symposium on-IDEAS '14. New York, USA: ACM Press, 2014:330-333.
- [19] Jiang Z G, Zhou A, Wang S G, et al. Personalized service recommendation for collaborative tagging systems with social relations and temporal influences[C/OL]. 2016 IEEE International Conference on Services Computing (SCC). New York, USA: IEEE, 2016(2016-09-01)[2019-09-25]. <https://doi.org/10.1145/2628194.2628238>.
- [20] Yuan Q, Chen L, Zhao S. Factorization vs. regularization: fusing heterogeneous social relationships in top-n recommendation[C]. ACM Conference on Recommender Systems. New York, USA: ACM Press, 2011:245-252.
- [21] de Meo P, Ferrara E, Fiumara G, et al. Improving recommendation quality by merging collaborative filtering and social relationships[C/OL]. 2011 11th International Conference on Intelligent Systems Design and Applications. New York, USA: IEEE, 2012(2012-01-03)[2019-09-25]. <https://doi.org/10.1109/isda.2011.6121719>.
- [22] Rafailidis D, Crestani F. Collaborative ranking with social relationships for top-N recommendations[C]. Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval -SIGIR '16. New York, USA: ACM Press, 2016:785-788.
- [23] Liu K P, Fang B X. Integrating social relations into personalized tag recommendation[C/OL]. 2010 Second International Conference on Intelligent Human-Machine Systems and Cybernetics. New York, USA: IEEE, 2010(2010-09-30)[2019-09-25]. <https://ieeexplore.ieee.org/document/5590829>.
- [24] Chen L, Nayak R, Xu Y. A recommendation method for online dating networks based on social relations and demographic information[C/OL]. 2011 International Conference on Advances in Social Networks Analysis and Mining. New York, USA: IEEE, 2011(2011-08-18)[2019-09-25]. <https://doi.org/10.1109/asonam.2011.66>.
- [25] Tran L, Yin X, Liu X M. Disentangled representation learning GAN for pose-invariant face recognition[C/OL]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York, USA: IEEE, 2017(2017-11-09) [2019-09-25]. <https://doi.org/10.1109/cvpr.2017.141>.
- [26] Tang L J, Liu E Y. Joint user-entity representation learning for event recommendation in social network[C/OL]. 2017 IEEE 33rd International Conference on Data Engineering (ICDE). New York, USA: IEEE, 2017(2017-05-18)[2019-09-25]. <https://ieeexplore.ieee.org/document/7929983>.
- [27] Kong X J, Mao M Y, Wang W, et al. VOPRec: vector representation learning of papers with text information and structural identity for recommendation[J]. IEEE Transactions on Emerging Topics in Computing, 2019: 1.
- [28] Zhuang F Z, Zhang Z Q, Qian M D, et al. Representation learning via dual-autoencoder for recommendation[J]. Neural Networks, 2017, 90: 83-89.
- [29] Wu L, Quan C, Li C, et al. A context-aware user-item representation learning for item recommendation[J]. ACM Transactions on Information Systems, 2019, 37(2):1-29.
- [30] Han X T, Shi C, Zheng L, et al. Representation learning with depth and breadth for recommendation using multi-view data[C]. Web and Big Data. Cham: Springer International Publishing, 2018: 181-188.
- [31] 吴玺煜, 陈启买, 刘海, 等. 基于知识图谱表示学习的协同过滤推荐算法[J]. 计算机工程, 2018, 44(2): 226-232, 263.

- WU Xiyu, CHEN Qimai, LIU Hai, et al. Collaborative filtering recommendation algorithm based on representation learning of knowledge graph[J]. *Computer Engineering*, 2018, 44(2): 226-232, 263..(in Chinese)
- [32] 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.
- [33] Shao L S, Zhang J, Wei Y, et al. Personalized QoS prediction for Web services via collaborative filtering[C/OL]. *IEEE International Conference on Web Services (ICWS 2007)*. New York, USA: IEEE, 2007(2007-07-30)[2019-09-25]. <https://doi.org/10.1109/icws.2007.140>.
- [34] Zheng Z B, Ma H, Lyu M R, et al. WSRec: A collaborative filtering based web service recommender system[C/OL]. *2009 IEEE International Conference on Web Services*. New York, USA: IEEE, 2009(2009-07-31)[2019-09-25]. <https://ieeexplore.ieee.org/document/5175854>.
- [35] Ma H, Yang H, Lyu M R, et al. SoRec: social recommendation using probabilistic matrix factorization[C]//*Proceedings of the 17th ACM conference on Information and knowledge management*. ACM, 2008: 931-940.
- [36] 何鹏, 吴浩, 曾诚, 等. Truser: 一种基于可信用户的服务推荐方法[J]. *计算机学报*, 2019, 42(4): 851-863.  
HE Peng, WU Hao, ZENG Cheng, et al. Truser: an approach to service recommendation based on trusted users[J]. *Chinese Journal of Computers*, 2019, 42(4): 851-863.(in Chinese)
- [37] Tang J, Qu M, Wang M, et al. LINE: Large-scale information network embedding[C]. *International World Wide Web Conferences Steering Committee*, 2015:1067-1077.
- [38] Liao L Z, He X N, Zhang H W, et al. Attributed social network embedding[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2018, 30(12): 2257-2270.
- [39] Wang D, Peng C, Zhu W. Structural deep network embedding[C]. *ACM Sigkdd International Conference on Knowledge Discovery & Data Mining*. New York, USA: ACM Press, 2016:1225-1234.

(编辑 陈移峰)