

基于时序逆影响的随机游走推荐算法*

肖春景^{1,2a}, 夏克文^{1†}, 乔永卫^{2b}

(1. 河北工业大学 电子信息工程学院, 天津 300300; 2. 中国民航大学 a. 计算机科学与技术学院; b. 工程技术训练中心, 天津 300300)

摘要: 相似性计算是协同过滤推荐的关键步骤。针对传统相似性计算认为相似关系是对等的且没有考虑消费顺序和时间间隔的问题,提出了基于时序逆影响的随机游走推荐算法。首先,基于用户时序关联图提出一种新的称为时序逆影响的相似性度量,利用随机游走得到了目标用户近邻集合;其次,利用随机游走在项目时序关联图上进一步改进推荐的多样性和覆盖率。它不但认为用户间相似是不对称的,考虑了用户消费项目的顺序和时间间隔,获得了用户全局的直接和间接近邻,而且考虑了项目间的时序逆影响。通过在真实数据集上的大量实验结果表明,与其他随机游走方法相比,其不但能提高推荐性能、缓解数据稀疏,而且通过提高多样性和覆盖率解决了过拟合的问题。

关键词: 相似性计算; 随机游走; 时序信息; 时序关联图; 协同过滤

中图分类号: TP391.1; TP301.6

文献标志码: A

文章编号: 1001-3695(2018)08-2304-04

doi:10.3969/j.issn.1001-3695.2018.08.016

Temporal inverse influence based recommendation method by using random walk

Xiao Chunjing^{1,2a}, Xia Kewen^{1†}, Qiao Yongwei^{2b}

(1. School of Electronic & Information Engineering, Hebei University of Technology, Tianjin 300300, China; 2. a. School of Computer Science & Technology, b. Engineering & Technical Training Center, Civil Aviation University of China, Tianjin 300300, China)

Abstract: Similarity computation is a very critical step in traditional collaborative filtering (CF) recommendation. However, traditional similarity always thought that the relationships are symmetric and rarely considered the order of common consumed items and the time interval of them. To get better recommendation performance, this paper proposed a random walk based recommendation method based on temporal inverse influence. It firstly proposed a new similarity measurement called temporal inverse influence on user temporal correlation graph and selected the similar users who most influenced on the target user using random walk with restart. Secondly, it used the item temporal correlation graph to further improve the diversity and coverage of CF recommendation results. It not only thinks that the influence is asymmetric, takes the order of common consumed items of users and the time interval of them into account to find direct and indirect neighbors from a global perspective, but also considers the inverse influence between items to improve the recommendation results. The experimental results on a real dataset show that this proposed method can not only achieve better recommendation performance, but also alleviate data sparsity and tackle the over fitting problem by getting better coverage and diversity compared to several random-walk-based methods.

Key words: similarity computation; random walk (RW); temporal information; temporal relation graph; collaborative filtering

0 引言

随着网上可用信息的增多,推荐系统已经成为帮助用户选择感兴趣信息的重要工具。协同过滤已经成为最成功的推荐算法。相似性计算是协同过滤非常关键的过程,它通过相似用户或项目的反馈信息来预测目标用户偏好,认为相似用户会对相似项目有相同的偏好^[1]。用户间有足够多的共同评分项目时它是十分有效的,但是数据往往是十分稀疏的。例如,用户1只对项目1和2有评分,用户2只对项目3和4有评分,用户1和2间的相似性为0。但是如果从全局的角度去衡量他们的相似关系,可能并非如此。因为可以通过对四个项目都有评分的用户3间接地挖掘用户1和2间的关系。

近年来,为改进协同过滤中局部相似性算法,基于图的方法通过中间用户/项目计算间接用户/项目的相似性来缓解数据稀疏问题。随机游走(random walk, RW)方法广泛应用于从全局的角度挖掘用户/项目间的关系。这些传统的基于随机行走的推荐方法根据项目/用户相似性建立转移概率,并利用随

机游走完成项目推荐^[2-5]。Fouss等人^[2]提出了利用马尔可夫链随机游走计算有向和无向图节点间相似度的新视角。Zhang等人^[3]在三元交互图上利用随机游走在项目相似图上扩展用户偏好来处理数据稀疏。在文献[4]中建立用户、用户组和项目间的二部图,并利用随机游走预测用户组和项目间的关联。Feng等人^[5]提出一个结合主题模型的用户项目二分图,并应用随机游走度量用户和项目集用户间的关联。单晓菲等人^[6]以选择性随机游走计算用户间的相似性,完成项目推荐。这些方法不仅可利用节点间的直接关系,也可利用间接关系。然而只利用了用户×项目评分矩阵,并没有考虑其他上下文信息(如位置、社交关系、时间)。在文献[7,8]中考虑用户签到行为利用随机游走完成地点推荐。在文献[9]中构建用户社交关系利用随机游走进行项目推荐。在文献[10,11]中基于微博的关注关系和多用户社交行为建立用户兴趣模型,没有利用随机游走完成项目和微博粉丝的推荐。Xia等人^[12]提出的MVCWalker通过合作者、最后合作时间和次数定义链接的重要性来进行学术网络上的推荐。由于用户偏好的不断变化,时

收稿日期: 2017-04-21; **修回日期:** 2017-06-08 **基金项目:** 国家自然科学基金资助项目(U1533104);河北省自然科学基金资助项目(E2016202341);天津市自然科学基金资助项目(14JCZDJC32500);中央高校基本科研业务费专项资金资助项目(ZXH2012P009)

作者简介: 肖春景(1978-),女,河北唐山人,讲师,博士研究生,主要研究方向为推荐系统、数据挖掘;夏克文(1965-),男(通信作者),湖南武冈人,教授,博士,主要研究方向为智能信息处理、数据挖掘(kwxia@hebut.edu.cn);乔永卫(1976-),男,山西祁县人,讲师,主要研究方向为民航信息智能处理、机务维修。

间成为重要的上下文信息。研究人员把时间信息作为一种新的节点或集成时间信息到相似性度量,然后用随机行走完成推荐^[13-15]。在文献[13]中将上下文信息作为网络中的节点,并利用随机游走进行项目推荐。在文献[14]中根据显式反馈信息构建了一个多层上下文(如时间、地点)关系图,并应用随机游走进行上下文感知推荐。Hu等人^[15]将时间信息应用到相似性度量和网络服务质量预测。Yu等人^[16]利用重启随机游走(random walk with restart, RWR)进行权值随时间不断变化的动态网络的增量计算。Xiang等人^[17]在基于会话的时序图上同时建立用户长期和短期偏好,并利用随机游走进行项目推荐。Nie等人^[18]提出一种改进的随机游走算法以提高个性化和多样性。这些方法通过游走后的节点稳定分布得到节点的重要性排序,并认为这种影响是对称的。然而实际中这种影响往往是不对称的。例如在电影推荐中,如果用户B总是在用户A观看了某部电影后也去观看此电影,那么用户A对用户B的影响则较大。文献[19]研究了在社交网络中的社会逆影响的问题,但没有考虑时间信息。Liu等人^[20]提出一种基于用户时序行为计算用户间权值的方法。孙光福等人^[21]构建了一个项目的时序消费网络,利用概率矩阵分解为用户推荐项目,但只考虑了时间先后关系,没有考虑时间间隔的影响。

为了克服上述问题,本文提出基于时序逆影响的随机游走推荐算法来缓解数据稀疏和提高多样性。首先,提出了一种基于用户时序关联图的新的称为时序逆影响的相似性度量,它不仅考虑了消费时间顺序,而且考虑了时间间隔影响;其次,利用随机游走模型从全局找到对目标用户影响最大的直接和间接的近邻集合;最后通过构建一个项目时序关联图进一步提高推荐多样性和覆盖率。实验结果表明,它不仅能提高推荐性能、缓解数据稀疏,而且提高了多样性和覆盖率。

1 基于时序逆影响的近邻选择

1.1 基于用户时序关联图的时序逆影响

基于随机游走的方法将用户和项目作为图的节点,边的权值通常采用传统的相似性方法,如余弦相似度,计算得到且认为节点间的权值是对称的,如用户 u_i 和 u_j 的边权重是相等的 $\omega_{ij} = \omega_{ji}$ 。但是在推荐过程中,消费项目的时序信息隐含着用户间的影响,并且这种影响是不相等的。例如,用户 u_i 和 u_j 共同看过很多电影,并且总是在 u_i 看过某部电影之后 u_j 也观看了它。因此 u_i 对 u_j 的影响要远大于 u_j 对 u_i 的影响,并且时间间隔越小影响越强。用户间影响在挖掘用户偏好时起着重要的作用。在确定用户之间的影响时不仅需要考察他们是否有共同行为,而且要考虑这些行为发生顺序和时间间隔。为了挖掘用户间的潜在影响,建立用户的时序关联,如图1所示。

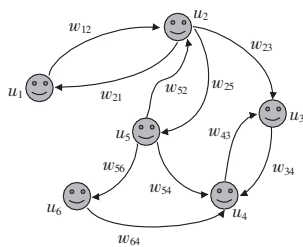


图1 用户时序关联图

这里 $G = \{U, E\}$, U 是用户集合, E 是边集合。 ω_{ij} 表示边权值。 u_i 对 u_j 的影响强度称为时序逆影响^[10], 定义为

$$\omega_{ij} = \frac{|P_{u_i \rightarrow u_j}|}{|I_{u_i} \cap I_{u_j}|} \sum_{k \in P_{u_i \rightarrow u_j}} \exp\left(\frac{t_{u_i}^k - t_{u_j}^k}{t_f^k - t_l^k}\right) \quad (1)$$

其中: $P_{u_i \rightarrow u_j}$ 表示 u_i 在 u_j 之前消费项目集合; $|P_{u_i \rightarrow u_j}|$ 表示集合 $P_{u_i \rightarrow u_j}$ 中的项目数; $|I_{u_i} \cap I_{u_j}|$ 表示用户 u_i 和 u_j 共同消费项目数; $t_{u_i}^k, t_{u_j}^k$ 分别代表用户 u_i 和 u_j 消费项目 k 的时间; t_f^k, t_l^k 表示所有用户消费项目 k 的最早和最晚时间。式(1)表示如果 u_i

和 u_j 消费了许多共同项目且 u_i 总是在 u_j 之前消费这些项目, 那么 u_i 对 u_j 有较强的影响, 且消费时间间隔越短, 影响越强。将时序逆影响作为用户间相似性的度量, 不仅考虑了他们间的共同行为, 而且考虑了行为的顺序和时间间隔。

1.2 基于时序逆影响利用随机游走选择近邻

带重启的随机游走(RWR)从源节点开始, 每一步都以 α 概率回到源点或以 $1 - \alpha$ 的概率随机跳转到邻居节点。当游走步数趋于无穷大时, 每个节点的概率分布趋向一个常数。因此, 基于 RWR 的方法能够返回源节点对图中其他节点重要性的排序。在图1中, 希望得到对源节点重要的节点, 而不是源节点对其他节点的重要性。可将活跃用户最重要的 top- k 用户作为其近邻集合。

为了得到目标用户 u_i 的近邻集合, 从每个用户节点 j 出发, 带重启的随机游走可回到节点 j 或以转移概率 $\omega(k/j)$ 随机游走到另一个用户节点 k 并最终停止在活跃用户 u_i 的节点 i 。通过这种逆向游走后每个节点得到的概率分布 $R(t)$ 为这些节点对活跃用户 u_i 节点的影响力大小, 定义如式(2)所示^[3]。

$$R(t) = \alpha \cdot R(t-1) \cdot P + (1 - \alpha) \cdot v \quad (2)$$

其中: P 为转移概率矩阵, 由用户间的时序影响矩阵定义, 表示为式(3); α 是重启概率, 介于 $[0, 1]$; 仅有 $i = j$ 时行向量 $v = 1$, 否则为 0。

$$P = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1M} \\ w_{21} & w_{22} & \cdots & w_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ w_{M1} & w_{M2} & \cdots & w_{MM} \end{bmatrix} \quad (3)$$

其中: w_{ij} 为式(1)定义的用户 u_i 对 u_j 的时序逆影响。基于时序逆影响采用随机游走进行近邻选择, 不但从全局角度挖掘直接和间接近邻以缓解数据稀疏性, 而且由于在相似性计算过程中考虑了时间上下文信息, 使得能够挖掘到与目标用户真正相似的近邻。

2 项目推荐

2.1 基于传统协同过滤的项目推荐

在1.2节中得到了所有用户对目标用户的影响排序, 选择一组对其影响最大的用户作为近邻集合。通过两个传统的 CF 模型, 累积加权总和和平均及偏差预测进行项目推荐。具体评分预测公式如式(4)(5)所示^[22]。

$$\hat{r}_{u_i}^k = \frac{\sum_{u_j \in N(u_i)} w_{ji} r_{uj}^k}{\sum_{u_j \in N(u_i)} w_{ji}} \quad (4)$$

$$\hat{r}_{u_i}^k = \bar{r}_{u_i} + \frac{\sum_{u_j \in N(u_i)} w_{ji} (r_{uj}^k - \bar{r}_{uj})}{\sum_{u_j \in N(u_i)} w_{ji}} \quad (5)$$

其中: $N(u_i)$ 表示目标用户 u_i 近邻集合; w_{ji} 表示 u_j 对 u_i 的时序逆影响; r_{uj}^k 表示 u_j 对项目 k 的评分; $\bar{r}_{u_i}, \bar{r}_{u_j}$ 分别是 u_i, u_j 的平均评分。

传统相似性可以根据消费历史发现用户间显示和局部的相似偏好, 而时序逆影响可发现他们潜在的和全局的相似偏好。因此结合传统的相似性与时序逆影响, 可以更好地利用局部和全局的、显示的和潜在的兴趣来计算用户之间的相似性, 以提高推荐性能。混合相似性被定义为

$$\text{sim}_{\text{hybrid}}(u_i, u_j) = \text{sim}_{\text{temporal}}(u_i, u_j) \text{sim}_{\text{traditional}}(u_i, u_j) \quad (6)$$

其中: $\text{sim}_{\text{temporal}}(u_i, u_j)$ 表示用户 u_j 和 u_i 间的时序逆影响; $\text{sim}_{\text{traditional}}(u_i, u_j)$ 为用户间的传统相似性, 如余弦相似性、修正的余弦相似性或皮尔森的相似性。因此, 优化后的评分预测公式如式(7)所示。

$$\hat{r}_{u_i}^k = \bar{r}_{u_i} + \frac{\sum_{u_j \in N(u_i)} \text{sim}_{\text{hybrid}}(u_i, u_j) (r_{uj}^k - \bar{r}_{uj})}{\sum_{u_j \in N(u_i)} \text{sim}_{\text{hybrid}}(u_i, u_j)} \quad (7)$$

2.2 基于项目关联图的项目推荐

用户 u_i 消费项目 I_m 和 I_k 的顺序隐含着项目之间的时序

影响,而这种影响也是不对称的。如果存在着很多用户总是先消费了 I_m 后也消费了 I_k ,却很少有用户消费 I_k 后消费 I_m ,则 I_m 对 I_k 的时序影响大于 I_k 对 I_m 的影响,且用户数越多,相隔时间越短,这种影响越强。因此类似于用户时序关联图构建方式,建立项目的时序关联图,并根据项目的时序影响定义边的权值,如式(8)所示。

$$\phi_{km} = \frac{|P_{I_k \rightarrow I_m}|}{|U_{I_k} \cap U_{I_m}|} \sum_{u_i \in U_{I_k \rightarrow I_m}} \exp\left(\frac{t_{u_i}^{I_k} - t_{u_i}^{I_m}}{t_{u_i}^{I_k} - t_{u_i}^{I_i}}\right) \quad (8)$$

其中: $P_{I_k \rightarrow I_m}$ 和 $|P_{I_k \rightarrow I_m}|$ 分别表示消费项目 I_m 后消费 I_k 的用户集合及用户数目; U_{I_k} 、 U_{I_m} 分别为消费过项目 I_m 和 I_k 的用户集合; $|U_{I_k} \cap U_{I_m}|$ 为共同消费项目 I_m 和 I_k 的用户集合; $t_{u_i}^{I_k}$ 、 $t_{u_i}^{I_m}$ 表示用户 u_i 消费项目 I_m 和 I_k 的时间; $t_{u_i}^{I_i}$ 、 $t_{u_i}^{I_i}$ 代表用户 u_i 消费第一个项目和最后一个项目的时间。

类似于用户时序关联图上游走过程,从一个源节点 $i \in I(u_i)$ 出发,每步或以 β 的概率回到初始节点,或以 $1 - \beta$ 的概率跳转到相邻的节点 j 。最后每个节点的稳定状态的概率分布为所有项目对目标项目的时序影响大小。

在 1.2 和 2.1 节中利用用户间的时序逆影响选择近邻,完成基于用户的协同过滤推荐。在推荐过程中只利用了近邻的评分。项目时序关联图反映了项目间的时序逆影响,目标用户已消费项目对待预测项目的时序影响反映了目标用户可能喜欢此项目的概率。因此结合项目间的时序影响,能进一步修正用户 u_i 的项目偏好以提高推荐的性能,如式(9)所示。

$$P_{u_i}^k = \sum_{j \in S(u_i)} \phi_{jk} \varphi_{jk} \quad (9)$$

其中: $\hat{r}_{u_i}^k$ 为 2.1 节评分预测结果; φ_{jk} 为项目 j 对项目 k 的时序逆影响; $S(u_i)$ 为用户 u_i 已消费过的项目集合。

基于项目时序关联图的推荐,不仅可以推荐那些近邻用户的高评分项目,而且可以推荐目标用户已消费项目在时序关联图中影响程度较高的项目,也是目标用户最可能喜欢的项目。这些被推荐的项目不但可以反映评分特征,而且反映了时序信息,提高了推荐精度和多样性。

3 实验和结果

3.1 数据集

本文使用 MovieLens 数据集 (<http://www.grouplens.org/node/12>) 来进行算法评估。选取每个至少评估过 20 个以上的项目的用户,至少被 5 个以上的用户评分的项目,它包含 14 111 个用户对 9 180 个项目的 224 452 评分,评分以 0.5 为间隔,范围是 0~5 分。每条数据包含用户和项目的 ID、评分和项目被观看的时间。

3.2 度量

3.2.1 精度度量

使用平均绝对误差(mean absolute error, MAE)来评估评分预测,使用精度(P)、召回率(R)和 F 评分(F)来评估 top- n 排序预测。

3.2.2 覆盖率

覆盖率描述一个推荐系统对物品长尾的发掘能力。覆盖率有不同的定义方法,简单地定义为推荐系统能够推荐出来的物品占总物品集合的比例。假设系统的用户集合为 U ,推荐系统给每个用户推荐一个长度为 N 的物品列表 $R(u)$,那么推荐系统的覆盖率如式(10)所示^[23]。

$$\text{coverage} = \frac{|U \cup_{u \in U} R(u)|}{|I|} \quad (10)$$

3.2.3 多样性

多样性描述了推荐列表中物品两两之间的不相似性。假设 $s[i, j] \in [0, 1]$ 定义了物品 i 和 j 间的相似性,那么用户 u 的推荐列表 $R(u)$ 的多样性定义^[23]为

$$\text{diversity}(R(u)) = 1 - \frac{\sum_{i, j \in R(u), i \neq j} s(i, j)}{\frac{1}{2}|R(u)|(|R(u)| - 1)} \quad (11)$$

而推荐系统的整体多样性可以定义为所有用户推荐列表多样性的平均值,如式(12)所示^[23]。

$$\text{diversity} = \frac{1}{|U|} \sum_{u \in U} \text{diversity}(R(u)) \quad (12)$$

3.3 与传统相似性比较

比较了时序逆影响相似性度量和其他传统的相似性度量,包括余弦相似性、余弦调整相似性和皮尔森相似性,在累计求和预测和平均值预测上的性能结果如图 2、3 所示。

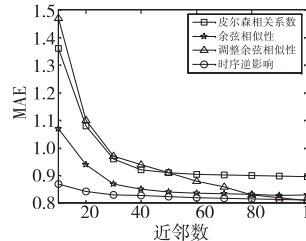


图2 累计求和预测

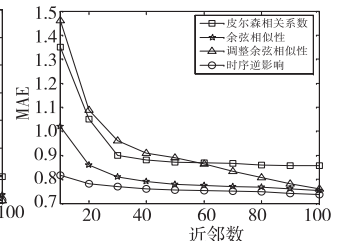


图3 均值预测

从图 2 和 3 中可以看出,随着近邻用户的增加,四种相似性度量的性能得到提高且变化趋势一致。余弦相似性优于调整余弦相似性和皮尔森相似性,调整余弦相似性的性能与皮尔森相似性的性能接近。本文提出的时序逆影响优于其他所有方法。具体来说,从图 2 中看到,当近邻数较小时调整余弦相似性与皮尔森相似性的性能是相同的。这主要是因为它们只考虑目标用户和近邻用户共同观看的项目。当邻居的数目增加到 60,皮尔森相似性的性能优于调整余弦相似性。这主要是因为随着邻居数量的增加,皮尔森相似性不仅反映了全局影响,而且消除了邻居用户评分的影响,但是由于共同项目数的减少,调整余弦相似性的性能降低。余弦相似性的性能优于以上两种度量,主要是因为它考虑了目标用户和近邻在项目空间上的全局相似性。本文提出的时序相似性获得了最佳的性能,收敛速度快,受近邻用户数据的影响最小,因为它不仅考虑了共同项目的评分,而且考虑了时序影响挖掘全局的近邻用户。但在图 3 中,当邻居大于 60 时,调整余弦相似性的性能明显优于皮尔森相似性。因为调整余弦相似性消除了随着近邻数据的增加共同项目减少后平均评分的影响来更好地描述用户间的相似关系。

3.4 混合相似性

传统的相似性反映了用户之间的局部相似性,而时间逆影响是从全局的角度度量用户间的相似。混合相似性结合了他们的优势。本文比较混合相似性、余弦相似性和时序逆影响利用平均值评分预测的性能结果,如图 4 所示。

从图 4 中可以看出,混合相似性的性能接近时间逆影响,优于余弦相似性。当邻居的数目变大时,混合相似性的性能优于时间逆影响的性能。这是因为当邻居集较小时,全局性能具有绝对优势,但随着邻居的增加,局部性能则更加突出。

3.5 与其他方法的比较

为了验证方法的有效性,将本文方法(temporal similarity random walk with restart, TSRWR)与其他推荐算法进行了对比:a)基于用户的协同过滤(user-based collaborative filtering, UCF)^[1],基于该方法,使用余弦相似性和平均偏差评分预测;b)带重启随机游走(random walk with restart, RWR)^[2],利用节点间传统相似性作为节点间转移概率的基于马尔可夫链的随机游走模型;c)时间依赖的网络服务推荐(time-aware Web service recommendation, TWSRec)^[15],结合时间信息计算节点相似度,并利用个性化随机游走进行服务推荐;d)短长期喜好融合模型(short term and long term random walk, SLTRW)^[17],基于会话时序图的同时建模用户长期和短期偏好,并利用随机游走进行项目推荐。

对于每种方法,选择 P 、 R 和 F 为标准来衡量 top- n 推荐的性能,设置 n 为 10、20 和 30。为所有用户计算平均精度、召回率和 F 得分,结果如表 1 所示。

表1 与其他方法 P 、 R 、 F 性能对比

方法	top@10			top@20			top@30		
	P	R	F	P	R	F	P	R	F
TSRWR	0.297	0.020	0.037	0.464	0.077	0.132	0.512	0.108	0.178
UCF	0.247	0.016	0.031	0.409	0.054	0.096	0.482	0.096	0.161
RWR	0.287	0.019	0.035	0.434	0.058	0.102	0.499	0.101	0.166
TWSRec	0.292	0.019	0.036	0.443	0.057	0.102	0.506	0.100	0.167
SLTRW	0.291	0.021	0.036	0.454	0.061	0.108	0.509	0.104	0.173

从表1可知,基于随机行走的方法在精度、召回率和 F 值上都优于 UCF,说明利用随机游走可更好地捕获邻居。这是因为基于随机行走的方法找到的邻居不仅考虑用户间的直接关系,也从全局的角度找到用户间的间接关系,真正捕获用户间的实际相似关系,但 UCF 只能依靠局部两两相似度,算法可用信息少。本文方法性能优于 RWR、TWSRec 和 SLTRW,简单来说是因为本文方法不仅考虑了共同的评分,还考虑观看时间的先后关系和时间间隔,而 TWSRec 和 SLTRW 简单地将时间信息集成到相似性度量或构建了包含时间信息的隐式反馈图,没有考虑先后关系方向和时间间隔, RWR 只构建基于节点间相似性的网络,没有考虑时间信息。

与其他方法 MAE 预测结果对比如图5所示。MAE 的结果与精度、召回率和 F 值有相似的趋势。MAE 基于随机行走的方法优于 UCF,本文方法优于 RWR、TWSRec 和 SLTRW。考虑时间信息的 TWSRec 和 SLTRW 优于没有考虑时间信息的 RWR。

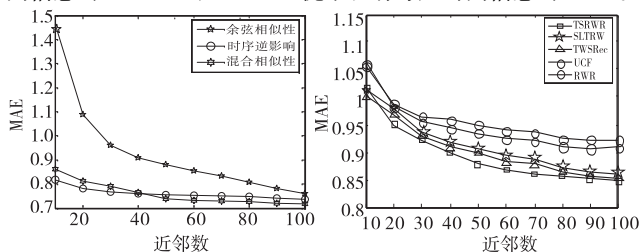


图4 混合相似性预测

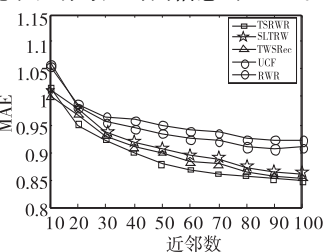


图5 与其他方法MAE性能对比

此外,在表2中给出推荐覆盖率和多样性的结果,与精度、召回率和 F 有相似的趋势。基于随机行走的方法覆盖率和多样性优于 UCF。这是因为基于随机游走的方法使用全局信息,减轻了数据稀疏性和过拟合问题。本文方法进一步优于 RWR、TWSRec 和 SLTRW,因为它集成了将局部信息融合到全局信息。

表2 与其他方法覆盖率和多样性的对比

方法	top@10		top@20		top@30	
	覆盖率	多样性	覆盖率	多样性	覆盖率	多样性
TSRWR	0.510 9	0.620 4	0.631 3	0.620 9	0.705 7	0.661 2
UCF	0.246 1	0.562 5	0.441 1	0.594 7	0.564 2	0.618 4
RWR	0.489 3	0.604 0	0.624 2	0.619 7	0.690 8	0.635 1
TWSRec	0.476 5	0.610 2	0.621 6	0.609 8	0.681 9	0.640 3
SLTRW	0.491 5	0.608 6	0.628 7	0.612 6	0.687 3	0.651 7

4 结束语

本文提出基于时序逆影响的随机游走推荐算法。首先提出基于用户时序关联图的时序逆影响相似性度量,它考虑了消费项目的先后关系和时间间隔,更好地利用了时序信息。利用随机游走不但为目标用户从局部挖掘直接近邻,而且从全局找到间接近邻,真正捕获了用户间的实际相似性以缓解数据稀疏性。利用项目时序关联图进一步提高了协同过滤推荐覆盖率和多样性。实验结果表明,该方法不仅得到了更好的推荐性能,缓解了数据稀疏性,而且解决了过度拟合问题,得到更好的覆盖和多样性。

参考文献:

[1] Zhang Lisheng, Deng Xiaoliang, Lei Dajiang. Collaborative filtering recommendation algorithm based on user interest characteristics and

item category [J]. Journal of Computational Information Systems, 2013, 19(15): 5973-5986.

- [2] Fouss F, Pirotte A, Renders J M, et al. Random-walk computation of similarities between nodes of a graph with application to collaborative recommendation [J]. IEEE Trans on Knowledge and Engineering, 2007, 19(3): 355-369.
- [3] Zhang Zhu, Zeng D D, Abbasi A, et al. A random walk model for item recommendation in social tagging systems [J]. ACM Trans on Management Information Systems, 2013, 4(2): 8-24.
- [4] Feng Shanshan, Cao Jian. Improving group recommendations via detecting comprehensive correlative information [J]. Multimedia Tools and Applications, 2017, 76(1): 1355-1377.
- [5] Feng Weisi, Jing Chenyang, Li Li. Recommender system based on random walk with topic model [C]//Proc of the 6th International Advanced Computing Conference. Piscataway, NJ: IEEE Press, 2016: 727-732.
- [6] 单晓菲, 米传民, 马静. 基于选择性随机游走的协同过滤推荐算法研究 [C]//第十六届中国管理科学学术年会论文集. 2014: 73-78.
- [7] Ying J J C, Kuo Wenning, Tseng V S, et al. Mining user check-in behavior with a random walk for urban point-of-interest recommendations [J]. ACM Trans on Intelligent Systems and Technology, 2014, 5(3): 40-66.
- [8] Xu Guangdong, Fu Bin, Gu Yanhui. Point-of-interest recommendations via a supervised random walk algorithm [J]. IEEE Intelligent Systems, 2016, 31(1): 15-23.
- [9] Bagci H, Karagoz P. Random walk based context-aware activity recommendation for location based social networks [C]//Proc of IEEE International Conference on Data Science and Advanced Analytics. Piscataway, NJ: IEEE Press, 2015.
- [10] 刘梦娟, 王巍, 李杨曦, 等. AttentionRank+: 一种基于关注关系与多用户行为的图推荐算法 [J]. 计算机学报, 2017, 40(3): 634-647.
- [11] 曹云忠, 邵培基, 李良强. 基于信任随机游走模型的微博粉丝推荐 [J]. 系统管理学报, 2017, 26(1): 117-123.
- [12] Xia Feng, Chen Zhen, Wang Wei, et al. MVCWalker: random walk-based most valuable collaborators recommendation exploiting academic factors [J]. IEEE Trans on Emerging Topics in Computing, 2014, 2(3): 364-375.
- [13] Wu Hao, Liu Xiaoxin, Pei Yijian, et al. Enhancing context-aware recommendation via a unified graph model [C]//Proc of International Conference on Identification, Information and Knowledge in the Internet of Things. Piscataway, NJ: IEEE Press, 2014: 76-79.
- [14] Yao Weilong, He Jing, Huang Guangyan, et al. A graph-based model for context-aware recommendation using implicit feedback data [J]. World Wide Web, 2015, 18(5): 1351-1371.
- [15] Hu Yan, Peng Qimin, Hu Xiaohui. A time-aware and data sparsity tolerant approach for Web service recommendation [C]//Proc of the 21st IEEE International Conference on Web Services. Piscataway, NJ: IEEE Press, 2014: 330-340.
- [16] Yu Weiren, Lin Xuemin. IRWR: incremental random walk with restart [C]//Proc of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press, 2013: 1017-1020.
- [17] Xiang Liang, Yuan Quan, Zhao Shiwan, et al. Temporal recommendation on graphs via long-and short-term preference fusion [C]//Proc of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2010: 723-732.
- [18] Nie Dacheng, Fu Yan, Zhou Junlin, et al. A personalized recommendation algorithm via biased random walk [C]//Proc of the 11th International Joint Conference on Computer Science and Software Engineering. Washington DC: IEEE Computer Society, 2014: 292-296.
- [19] Jin Zhaoyan, Wu Quanyuan, Shi Dianxi, et al. Random walk based inverse influence research in online social networks [C]//Proc of IEEE International Conference on High Performance Computing and Communication & IEEE International Conference on Embedded and Ubiquitous Computing. Washington DC: IEEE Computer Society, 2013: 2206-2213.
- [20] Liu N N, He Luheng, Zhao Min. Social temporal collaborative ranking for context aware movie recommendation [J]. ACM Trans on Intelligent Systems and Technology, 2013, 4(1): 15-26.
- [21] 孙光福, 吴乐, 刘洪, 等. 基于时序行为的协同过滤推荐算法 [J]. 软件学报, 2013, 24(11): 2721-2733.
- [22] Yin Fengjiang, Wang Zhenwen, Tan Wentang, et al. Sparsity-tolerated algorithm with missing value recovering in user-based collaborative filtering recommendation [J]. Journal of Information & Computational Science, 2013, 10(15): 4939-4948.
- [23] 项亮. 推荐系统实践 [M]. 北京: 人民邮电出版社, 2012: 23-34.