本科毕业论文外文翻译

文献原文:

Chen C, Zheng X, Wang Y, et al. Capturing Semantic Correlation for Item Recommendation in Tagging Systems[C]//AAAI. 2016: 108-114.

利用标签的语义关联性进行物品推荐

摘要 标签系统的普及对于提升物品推荐的效果是一个很好的机会。虽然现有的方法对标签使用主题建模来挖掘物品的语义信息,但是他们忽略了一个重要的性质,标签是用户和物品之间连接的桥梁。因此,这些方法不能处理无共同评分项的数据稀疏性问题(DS-WO-CRI),从而限制了它们的推荐性能。为了解决这个问题,我们提出了一种新型的基于标签和评分的协同过滤推荐模型,首先使用主题建模依次挖掘每个用户和每个物品的语义信息,然后将这些语义信息纳入矩阵分解,同时捕获标签和评分在用户和物品间的桥接特性。因此,我们的模型捕获了用户和物品间的语义关联,极大的提高了推荐性能,尤其是在 DS-WO-CRI 的情况下。在两个流行的真实数据集上的实验表明,我们提出的模型在准确率和召回率上显著优于传统的协同过滤方法、最先进的基于社交关系的协同过滤和基于主题模型的的协同过滤方法,它是解决 DS-WO-CRI 问题的有效方法。

1 引言

近些年来,诸如 Delicious(社交书签),Last.fm(社交音乐),Flickr(照片分享)和 YouTube(视频分享)这些标记系统为用户提供了高效的方式来组织、管理、共享并搜索各种项目。例如,一个人在 Last.fm 听 Lady Gaga 的音乐时,他可以将她标记为"流行的"和"女歌手"。这些连同评分行为一起出现的标签很有价值,强烈建议使用这样的信息来提供个性化推荐 [21] 。

标签系统的普及促进了推荐系统的发展,尤其是标签系统中的协同过滤方法。目前为止,标签系统中主要有两种类型的协同过滤:标签推荐 [19] 的目的是为物品推荐合适的标签,另一种是基于标签的物品推荐 [22, 23],它关注利用标签和评分等信息为目标用户推荐相似的用户或物品。

目前,一个研究的趋势是在协同过滤中使用主题模型来处理标签信息

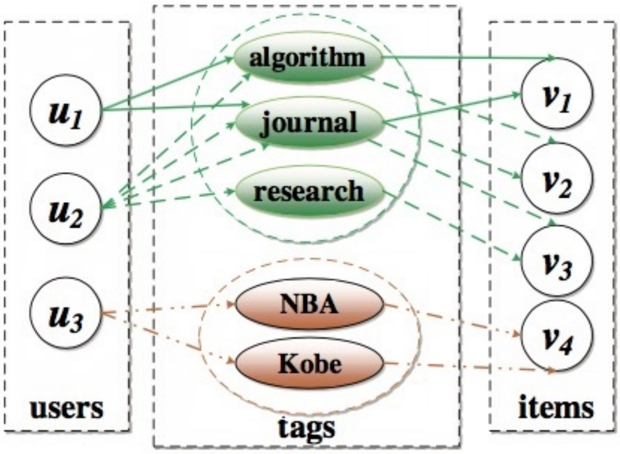


图 1.1　一个标签系统的例子。每个标签都对应了一个评分，为简洁起见省略了评分。

[1, 18, 15, 19, 6]。例如，Wang 和 Blei [18] 提出了一个协同主题回归（CTR）模型,可用于基于标签的物品推荐。Chen 等人 [6] 提出的将 CTR 与社交矩阵分解 [13] 结合的推荐方法获得了更好的预测效果。然而，这些方法只是将标签与用户和物品相关联，而忽略了标签重要性质，标签连接了用户和物品，这与评分的作用是一样的。但是标签可以反应用户和物品间的语义关联性，而评分没有这样的能力。

当一个用户为一些物品赋予了标签，那这些标签就反映了用户对物品的偏好。一个标签越频繁被一个用户使用，越可能表示这个用户喜欢这个标签所指代一类的物品 [21]。类似的，如果一个标签被越多的用户赋予一个物品，那么很可能这个物品匹配这个标签。因此，标签同时包含用户和物品的语义信息，而不仅仅是单独的用户或物品。

一个针对性的例子　图 1.1 描述了一个标签系统的例子，这个系统包含三个用户（u_1, u_2, u_3 ），四个物品（v_1, v_2, v_3, v_4 ）和五个标签（“algorithm”, “journal”, “research” , “NBA”, “Kobe”）。

这个例子中，用户 u_1 标注了物品 v_1，用户 u_2 标注了物品 v_2 和 v_3，因此，用户 u_1 和用户 u_2 之间没有共同评分项。我们定义这种情况为 * 无共同评分项的稀疏性问题（DS-WO-CRI），DS-WO-CRI 是标准的数据稀疏性问题（在所有的用户物品对中只有很少比例的已知项）的一个典型子集。在 DS-WO-CRI 的情况下，已有的协同

过滤方法,例如,PMF 和 CTR 模型,都无法将物品 v_1 推荐给用户 u_2 ,因为这些方法无法捕获它们之间的任何联系。然而,一个好的推荐系统应该能够将物品 v_2 和 v_3 推荐给用户 u_1 并且将物品 v_1 推荐给用户 u_2,因为在这个例子中,用户 u_1 和 u_2 很可能是算法方面的研究者,而物品 v_1, v_2 和 v_3 可能跟算法有关。

现有的研究表明,用户对一个物品的评分或标注等动作就可以表明用户喜好该物品,而不需考虑评分的等级 [11, 12]。换句话说,用户通过标记和评分隐式的表达了他的偏好 [11]。因此,一个用户和他所标注和评分的物品趋向于具有相似的潜在特征,我们在本文中将其定义为隐式偏好(implicit preference)。在上面的例子中,用户 u_2 将物品 v_3 标注为"journal"和"research",这在表明用户 u_2 可能是一名研究者同时,也指出物品 v_3 可能是一篇学术期刊或其他相关的东西。因此,在语义上,用户 u_2 和物品 v_3 的潜在特征应该具有某种程度的相似。

然而,现有的方法无法捕获用户和物品之间的语义关系,因此它们的推荐性能被局限了,尤其是在 DS-WO-CRI 的情况下。

我们的提议 为了解决上面提到的 DS-WO-CRI 问题,我们在这篇文章中提出了一种新型的协同过滤系统。我们首先利用主题模型依次为每一个用户和每一个物品挖掘标签的语义信息,然后将这些语义信息纳入矩阵分解,同时捕获标签和评分在用户和物品间的桥接特性(即,隐式偏好)。因此,我们的模型可以捕获用户和项目间的语义关联,并将具有相似语义信息的物品推荐给用户,即便是在 DS-WO-CRI 的情况下。

贡献 我们的工作主要有以下贡献:(1)我们首先指出标签的重要特性,即它们做为桥梁将用户和物品连接起来,概述了用户和物品之间的语义联系,然后我们说明了利用此特性可以帮助解决 DS-WO-CRI 问题。据我们所知,这是首个针对这个问题的研究。(2)我们提出了一种新型的基于标签和评分的协同过滤模型,它可以捕获用户和物品之间的语义联系,因此可以大大地提升推荐性能,尤其是在 DS-WO-CRI 的情况下。我们还提出了基于坐标上升的参数学习方法。据我们所知,这项研究是文献中对捕获用户和物品语义关联的首次尝试。(3)在两个流行的真实世界的数据

集上的实验表明,我们提出的模型在精确率和召回率方面都显著优于最先进的方法。实验还表明,我们的模型是一个解决 DS-WO-CRI 问题的有效方法。

2 相关工作

在本节中,我们将分三组来回顾已有的物品推荐方法,其中包括:传统的协同过滤方法、基于主题模型的协同过滤方法、以及基于社交关系的协同过滤方法。

基于已有的研究 [17],传统的协同过滤方法仅利用用户物品的评分进行推荐,主要分为两种类型:基于内存的协同过滤 [8] 和基于模型的协同过滤 [12, 11, 22, 20],这两种方法都可以用于标签系统的推荐。

传统的协同过滤方法无法借助文本内容的信息,因此,一些混合模型被提出来,它们结合基于内容的方法和协同过滤方法进行推荐 [14]。但是,这些方法简单的将内容表示为词向量的形式,因而无法发掘它们的语义信息。为了利用内容所提供的语义信息,研究者利用主题模型提高推荐效果,Agarwal 等人提出了 fLDA 模型 [1],该方法通过将 LDA 中学习到的先验信息加入物品向量,结合了 RLFM[1] 和隐式狄利克雷分布(LDA)。RLFM 和 fLDA 都能纳入额外的原信息,例如用户年龄和物品类别,然而,这些附加元特征信息不在本文的范围之内。稍后的,Wang 等人 [18] 将概率矩阵分解 [16] 与 LDA [4] 相结合,提出了协同主题回归模型。在 [18] 中证明了在相似的情况下 CTR 的性能要优于 fLDA,因为 fLDA 基本忽略了其他用户的评分。

此外,用户之间和物品之间的社会信息对于提升推荐的性能是有效的 [7]。首先,用户的社会信息被纳入常规的协同过滤模型 [10],例如,Ma 等人提出 Soreg 来约束具有联系的用户的潜在因子之间的差异性。之后,相邻用户和物品的社会信息被引入了基于主题模型的协同过滤中以进一步改善推荐性能,例如,Purushotham、Liu 和 Kuo [15] 以及 Chen 等人 [6] 提出了两种模型(CTR-SMF 和 CTR-SMF2),将用户社交关系网络纳入 CTR,以进一步提高项目推荐性能。Wang、Chen 和 Li [6] 提出了一个将物品的社会关系引入 CTR 的模型,以提高标签系统中的标签推荐性能。

3 提出的模型——TRCF

在本节中,我们提出一个新的基于标签和评分的协同过滤方法(TRCF)。我们首先正式确定基于标签的物品推荐问题并定义一些符号。然后,我们提出 TRCF ,这是一个分层的贝叶斯模型。最后,我们提出了基于坐标上升的参数学习方法。

3.1 初步定义

假定，我们有一个用户的集合 $U = \{u_1, \ldots, u_I\}$，这些用户用一组标签 $T = \{t_1, \ldots, t_N\}$ 标记了的一组物品 $V = \{v_1, \ldots, v_J\}$，以及评分的集合 $R = \{r_1, \ldots, r_O\}$，其中，I 、J、N 和 O 依次代表了用户、物品、标签和评分的数目。每一个用户-物品-标签-评分（U-I-T-R）的可观察数据是一个四元组 □u_i, v_j, T_{ij}, R_{ij}□，其中 $u_i \in U$，$v_j \in V$，T_{ij} 是用户 u_i 给予物品 v_j 的标签集合，并且 $T_{ij} \in T$ ，R_{ij} 是用户 u_i 给予物品 v_j 的评分，评分基于用户对该物品的喜好程度并在同时标注它。然而，用户-物品（U-I）的评分集合 R 一般是整数集合，例如，*MovieLens* 使用 [1,5] 范围内的评分。$U \in R^{K*I}$ 表示潜在的用户特征矩阵，其中列向量 U_i 表示属于用户 u_i 的 K 维潜在特征向量。$V \in T^{K*J}$ 表示潜在的物品特征矩阵，其中列向量 V_j 表示属于物品 v_i 的 K 维潜在特征向量。

对于基于标签和评分的物品推荐中，给定现有的四元组 U-I-T-R ，我们的目标是预测用户 u_i 对物品 v_j 的未知的评分。

3.2 基于标签和评分的协同过滤

TRCF 是一个新型的分层的贝叶斯模型，图 3.1 展示了它的图模型，其中 N_u 和 N_v 依次表示用户 u_i 和物品 v_j 的标签数目。我们首先将每个用户和物品的标签依次分组，然后使用隐式狄利克雷分布依次对每个用户和每个物品的标签集合进行语义挖掘（图 3.1 中以红色绘制）。最后将这些语义信息纳入矩阵分解用于分解评分信息（图 3.1 中以紫色绘制）以及捕获标签和评分所提供的隐式偏好（图 3.1 中以蓝色绘制）。

TRCF 同时在用户和物品两个方面上执行 LDA ，因此可以同时捕获用户和物品的语义信息，而不仅仅像现有的工作那样只捕获物品的语义信息。另外，在 TRCF 中，如果一个用户和一个物品通过标签或评分相关联，那么他们的隐式特征会在某些程度上比较相似，这被称为隐式偏好。相比之下，现有的基于主题建模的 CF 方法，例如，CTR、CTR-SMF 和 CTR-SMF2，都是假设用户和物品是独立的，并且忽略标签和评分在用户和项目之间的桥接作用。因此，TRCF 可以捕获用户和物品之间的语义关联，并且能够处理 DS-WO-CRI 问题。假定每个用户和每个物品都有 K 个主题，TRCF 的执行过程如下：

1. **挖掘用户标签的语义信息。** 对于每个用户 u_i ：

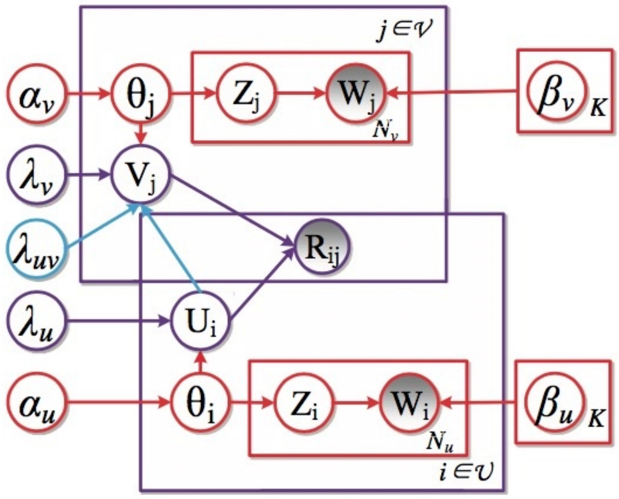


图 3.1　TRCF 的图模型。LDA 部分以红色绘制，隐式偏好部分以蓝色绘制，PMF 部分以紫色绘制。

(a) 选取主题分布 $\theta_i \sim Dirichlet(\alpha_u)$ ；

(b) 选取用户的潜在向量 $U_i \sim \mathcal{N}(\theta_i, \lambda_u^{-1} I_k)$ ；

(c) 对于用户 u_i 的每一个标签 w_{in_u} ：

i. 选取主题 $z_{in_u} \sim Mult(\theta_i)$ ；

ii. 选取标签 $w_{in_u} \sim Mult(\beta_{z_{in_u}})$ ；

2. **挖掘物品标签的语义信息，并且采集用户和物品间的隐式偏好。**对于每一个物品 v_j ：

(a) 选取主题分布 $\theta_j \sim Dirichlet(\alpha_v)$ ；

(b) 选取物品的潜在向量 $V_j \sim \mathcal{N}(\theta_j, \lambda_v^{-1} I_k) \times \prod_i I_{ij}^R \mathcal{N}(U_i, \lambda_{uv}^{-1} I_K)$ ；

(c) 对于物品 v_j 的每一个标签 w_{jn_v} ：

i. 选取主题 $z_{jn_v} \sim Mult(\theta_j)$ ；

ii. 选取标签 $w_{jn_v} \sim Mult(\beta_{z_{jn_v}})$ ；

3. **获取评分。**对于每一个用户-物品对 (i, j) ：

$$R_{ij} \sim (U_i^T V_j, c_{ij}^{-1}).$$

在上面的生成过程中,$\mathcal{N} \sim (x|\mu, \sigma^2)$ 是期望为 μ 方差为 σ^2 的高斯分布,I_K 是一个 K 行 K 列的单位矩阵。I_{ij}^R 是一个指示函数,如果用户 u_i 为物品 v_j 打分了,它的值为 1,否则为 0。C 是评分置信度矩阵,其中的项 c_{ij} 表示评分的置信度。更多细节请参考 [18]。

参数 λ_u 平衡了用户语义信息提供的标签和评分信息对模型性能的影响。类似地,参数 λ_v 平衡了由物品语义信息提供的标签和评分信息对推荐性能的影响。参数 λ_{uv} 平衡隐式偏好对模型性能的贡献,即通过评级和标签链接的用户和项目之前潜在特征相似度的程度。

观察到的评分的条件分布可以被形式化为:

$$p(R|U,V,C) = \prod_i \prod_j \mathcal{N}(R_{ij}|U_i^T V_j, c_{ij}).$$

用户和物品的潜在向量 U_i 和 V_j 生成的方式与 CTR 相似,它们可以被形式化为:

$$\begin{aligned} &p(U|\lambda_u) \sim \prod_i \mathcal{N}(\theta_i, \lambda_u^{-1} I_K), \\ &p(V|U, \lambda_v, \lambda_{uv}) \sim \prod_j \mathcal{N}(\theta_j, \lambda_v^{-1} I_K) \times \prod_i I_{ij}^R \mathcal{N}(U_i, \lambda_{uv}^{-1} I_K). \end{aligned} \tag{3.1}$$

给定了 U-I-T-R 信息,通过使用贝叶斯推理,我们可以得到 TRCF 的潜在特征向量的后验概率的如下式子:

$$p(U,V|R,C,\lambda_u,\lambda_v,\lambda_{uv}) \propto p(R|U,V,C)p(U|\lambda_u)p(V|U,\lambda_v,\lambda_{uv}). \tag{1}$$

3.3 TRCF 的参数学习

给定主体参数 β_u 和 β_v ,直接计算 $U_i, V_j, \theta_i, \theta_j$ 的完整后验是困难的。我们使用坐标上升的方法来学习最大后验概率。使等式 (1) 中具有固定超参数的两个潜在特征的后验最大化等价于,在给定 λ_u, λ_v 的条件下,使如下的 $U, V, \theta_{1:I}, \theta_{1:J}$ 的对数似

然函数最大：

$$
\begin{aligned}
L = & -\frac{\lambda_u}{2}\sum_i (U_i-\theta_i)^T(U_i-\theta_i) \\
& -\frac{\lambda_v}{2}\sum_j (V_j-\theta_j)^T(V_j-\theta_j) \\
& -\sum_{ij}\frac{c_{ij}}{2}(R_{ij}-U_i^TV_j)^2 \\
& +\sum_i\sum_{n_u} log\left(\sum_k \theta_{ik}\beta_{k,w_{in_u}}\right) \\
& +\sum_j\sum_{n_v} log\left(\sum_k \theta_{jk}\beta_{k,w_{jn_v}}\right) \\
& - -\frac{\lambda_{uv}}{2}I_{ij}^R\sum_{ij}(U_i-V_j)^T(U_i-V_j).
\end{aligned} \tag{2}
$$

我们省略一个常数并设置狄利克雷先验 $\alpha_u=\alpha_v=1$ 。这个函数可以通过使用坐标上升来优化，也就是说，我们固定 β_u 和 β_v ，并迭代优化 MF 变量 U_i, V_j 和主题分布 θ_i, θ_j 。具体来说，我们首先根据 θ_i, θ_j 当前的估计值更新 U_i 和 V_j ，我们计算等式 (2) 中 L 在 U_i□V_j 上的导数，并且将它设置为 0 ：

$$
\frac{\partial L}{\partial U_i}=0, \frac{\partial L}{\partial V_j}=0. \tag{3}
$$

解决上述的公式得到参数更新的式子：

$$
\begin{aligned}
U &\leftarrow \left(VC_iV^T+\lambda_u I_K+\lambda_{uv}\sum_i I_{ij}^R I_K\right)^{-1}(VC_iR_i+\lambda_u\theta_i+\lambda_{uv}\sum_j I_{ij}^R V_j), \\
V &\leftarrow \left(UC_jU^T+\lambda_v I_K+\lambda_{uv}\sum_i I_{ij}^R I_K\right)^{-1}(UC_jR_j+\lambda_v\theta_j+\lambda_{uv}\sum_i I_{ij}^R U_i).
\end{aligned} \tag{4}
$$

其中对于每一个用户 u_i ，C_i 是一个对角矩阵，它的对角元素是 $c_{ij}, j=1,\ldots,J$ ，并且 $R_i = R_{ij}{}_{j=1}^{J}$ 。对于物品 v_j ，C_j 和 R_j 的定义是相似的。

公式 (4) 显示了参数 λ_u、λ_v 和 λ_{uv} 是如何影响用户和物品的潜在特征的。越大的 λ_u 会导致用户的潜在特征越依赖于用户标签，而不是评分信息。类似的，较大的 λ_v 表示物品的潜在特征来自物品标签的比例更大，而不是评分信息。此外，更大的 λ_{uv} 意味着更强的约束，即通过标签和评分链接的用户和项目应当具有类似的潜在特征，即隐式偏好。从公式 (4) 可以看出，概率矩阵分解（PMF）和协同主题回归（CTR）都是 TRCF 的特殊形式。

接下来,给定当前的 MF 变量 U_i, V_j ,我们更新主题分布参数 θ_i 和 θ_j 。对于 θ_i ,我们先定义 $q(z_{in_u} = k) = \phi_{in_uk}$,然后分离包含 θ_i 的用户并应用 Jensen 不等式:

$$\begin{aligned}
L(\theta_i) \geq & -\frac{\lambda_u}{2}(U_i - \theta_i)^T(U_i - \theta_i) \\
& + \sum_{n_u}\sum_{k}\phi_{in_uk}(log\theta_{ik}\beta_{k,w_{in_u}} - log\phi_{in_uk}) \\
& = L(\theta_i, \phi_i).
\end{aligned} \tag{3.2}$$

其中,$\phi_i = \phi_{in_uk}{}_{n_u=1,k=1}^{N_u\times K}$ 。显然 $L(\theta_i, \phi_i)$ 是 $L(\theta_i)$ 的严格下界,因此我们可以使用映射梯度 [3] 来优化 θ_i 。最优的 ϕ_{in_uk} 正比于 $\theta_{ik}\beta_{k,w_{in_u}}$ 。对于 θ_j ,更新的规则是相似的。

对于 β_u ,我们像 LDA 那样为主题执行 M 步的更新。

$$\beta_{kw_i} \propto \sum_{i}\sum_{n_u}\phi_{in_uk}1[w_{in_u} = w].$$

对于 β_v ,它的更新策略是相似的。当参数 $U^*, V^*, {\theta_{1:I}}^*, {\theta_{1:J}}^*, {\beta_u}^*, {\beta_v}^*$ 的最优值学习完成后,我们的模型就可以进行评分预测了:

$${R_{ij}}^* \approx (U_i^*)^T V_j^*.$$

4 实验和分析

在本节中,我们介绍了对两个流行的现实世界数据集进行的实验,目的是回答如下的问题:(1)我们的模型相比现有的最先进的方法有什么改进? (2)我们的方法如何处理 DS-WO-CRI 问题? (3)参数 λ_u ,λ_v 和 λ_{uv} 如何影响 TRCF 的性能?

4.1 数据集

我们在实验中使用了两个现实世界的数据集: hetrec2011-delicious-2k (Delicious) 和 hetrec2011-lastfm-2k (Lastfm)[5]。这两个数据集已广泛用于标签系统的实验 [2],它们在表 4.1 中描述。

对于每个数据集,如果用户已将该项目设为书签(或收听),则我们认为该项目的用户评分为 1 ,否则,该项目的用户评分为 0 。

在我们的实验中,我们将每个数据集分为三个部分——训练数据集(80%),留出验证数据集(10%)和测试数据集(10%)。我们在训练数据集上训练我们的模型,在验证数据集上获得最佳参数,并在测试数据集上评估我们的模型。

表 4.1 数据集描述

数据集	用户	物品	标签	用户-标签-物品	评分
Delicious	1867	69226	53388	437593	104799
LastFm	1892	17632	11946	186479	92834

4.2 比较与评估

如相关工作中所述,存在许多种推荐方法,例如,基于存储的方法和混合方法。这里,我们将所提出的 TRCF 与以下三种现有的方法进行比较,即常规的协同过滤方法,基于社会关系的 CF 方法和基于主题建模的方法:

SVD ++ [11] 是一种经典的协同过滤方法,该方法仅使用 U-I 评分信息。

Soreg [13] 是基于社会关系的 CF 方法,其使用 U-I 评分和用户社交信息。

CTR [18] 是一种最先进的基于主题建模的 CF 方法,其使用类似于 TRCF 中使用的四元组 U-I-T-R 的信息。

CTR-SMF [15] 结合了用户社交矩阵分解和 CTR 。它包含除了在 TRCF 中使用的四元组 U-I-T-R 之外的附加用户社交信息。

CTR-SMF2 [6] 改进了 CTR-SMF ,它还将用户社交信息引入了 TRCF 中的 U-I-T-R 四元组。

精确率和召回率已被广泛用作评估推荐效果的指标 [9]。因此,我们使用 Precision 和 Recall 来评估推荐性能,并且计算召回率的方式也用于 CTR,CTR-SMF 和 CTR-SMF2。对于每个用户,Precision 和 Recall 定义如下:

$$\begin{aligned} Precision@M &= \frac{\#\text{Top M 中用户喜欢的物品}}{M}, \\ Recall@M &= \frac{\#\text{Top M 中用户喜欢的物品}}{\#\text{用户喜欢的所有物品}}, \end{aligned} \tag{4.1}$$

其中 M 是推荐列表中的物品数。我们计算测试数据集中所有项的精确率和召回率的平均值作为最终结果。

4.3 性能对比和分析

在对比不同模型时,我们使用在 CTR-SMF [15] 中设置的 SVD++、CTR 和 CTR-SMF 的最佳参数,它们使用相同的数据集。对于 Soreg、CTR-SMF2 和我们的模型,我们使用网格搜索来获得最佳参数。

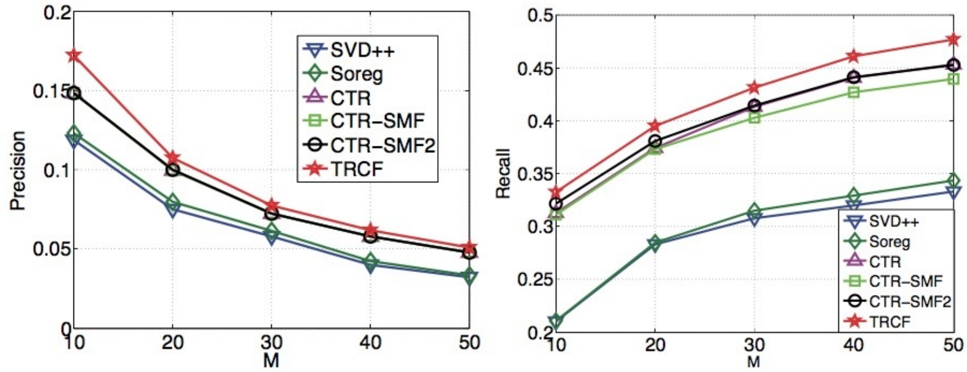


图 4.1　在 Delicious 数据集下设置不同的 M 值,精确率和召回率的对比,以及每个方法的最优参数

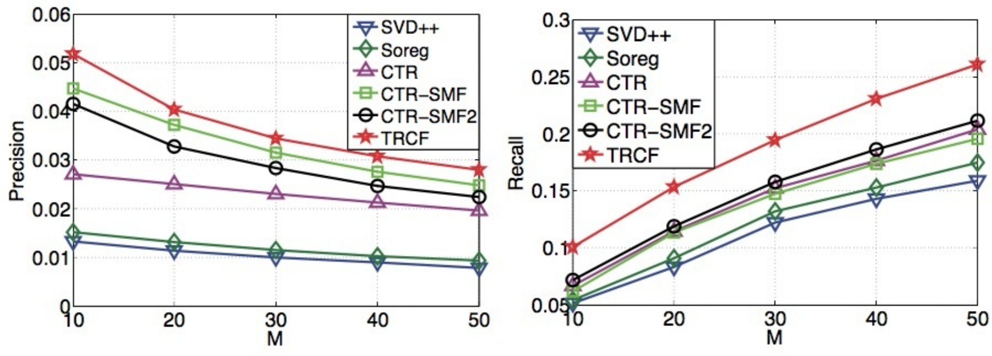


图 4.2　在 LastFm 数据集下设置不同的 M 值,精确率和召回率的对比,以及每个方法的最优参数

结果： 图 4.1 和图 4.2 显示了各个推荐方法在 Delicious 数据集和 Lastfm 数据集上的总体性能，其中我们设置 $M = 10, 20, 30, 40, 50$ 并将每个方法的参数固定为最佳值。结果表明，传统的 CF 方法（SVD ++）和基于社交关系的 CF 方法（即 Soreg）具有相似的性能。三个基于主题建模的 CF 方法（即 CTR、CTR-SMF 和 CTR-SMF2）具有类似的性能，并且显著优于 SVD++ 和 Soreg，这表明了标签信息在推荐中的重要性。

我们提出的 TRCF 方法在这两个数据集上显著优于 SVD++、Soreg、CTR、CTR-SMF 和 CTR-SMF2。具体在平均值上，在 Delicious 数据集上，TRCF 在精确度方面将 SVD++、Soreg、CTR、CTR-SMF 和 CTR-SMF2 提升了 46.75％、39.74％、8.62％、8.78％和 8.65％，并且在召回率方面分别提高了 44.99％、42.40％、5.23％、7.27％和 4.21％。在 Lastfm 数据集上，TRCF 在精确度方面将 SVD++、Soreg、CTR、CTR-SMF 和 CTR-SMF2 提升了 259.80％、210.27％、57.96％、11.64％和 23.85％，并且在召回率方面分别提高了 73.03％、60.39％、34.18％、39.04％和 28.12％。

分析和总结： 以上的对比表明了我们提出模型的有效性，它捕获了用户和物品之间的语义关联。实验结果表明，尽管用户的社交信息可以改善推荐性能，但是改善推荐性能的更有效的方式是考虑用户和物品的语义相关性。

4.4 DS-WO-CRI 实验

所有四种基于主题建模的 CF 方法（包括 TRCF）都可以通过采集物品的语义信息来提高推荐性能。为了研究它们处理 DS-WO-CRI 问题的能力，我们进行以下实验。我们首先依据 DS-WO-CRI 的程度将 LastFm 数据集分为四个子数据集，每个数据集在表 2 中描述了。DS-WO-CRI 的程度定义如下：

$$x\% = \frac{\#\text{没有共同评分物品的用户数目}}{\#\text{总的用户数目}}.$$

然后我们对每个子数据集进行对照实验。图 4.3 显示，我们的模型在不同的 DS-WO-CRI 程度下都达到最佳性能。我们的模型相对于其他三个主题建模方法的在 LC1、LC2、LC3、LC4 上精确率的平均改进分别为 49.67％、75.81％、345.77％和 428.97％，召回率的改进分别为 33.94％、65.00％、383.63 ％和 458.00％。实验结果表明，DS-WO-CRI 的程度越严重，我们的模型相对其它模型的优势越明显。

总结： DS-WO-CRI 实验表明了我们提出的模型在处理该问题时的有效性：越

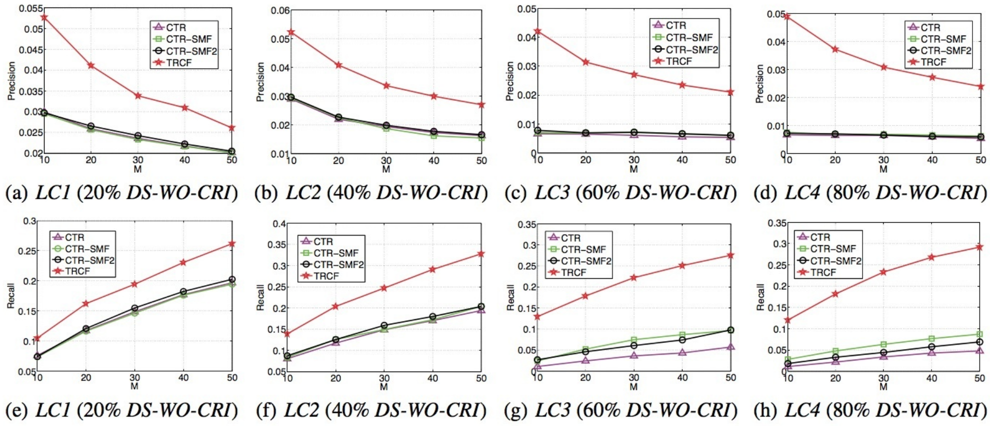


图 4.3 在每个 DS-WO-CRI 子数据集上设置每种方法的最优参数，精确率和召回率对比。

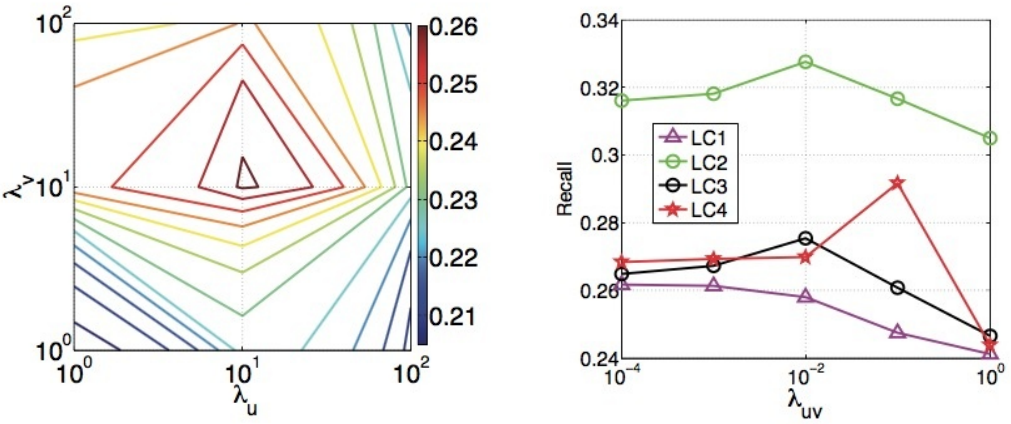
严重的 DS-WO-CRI 情况，我们的模型相比其它模型的优势越明显。这是由于我们的方法具有采集用户和物品间标签语义关联的能力。

4.5 参数影响

图 4.4(a) 显示了在 LastFm 数据集上固定公式 (2) 中的参数 λ_{uv} 时，参数 λ_u 和 λ_v 对 TRCF 性能的影响。我们可以看到，当 $\lambda_u = \lambda_v = 10$ 时，TRCF 达到最佳性能，这说明用户和物品语义信息都对模型性能有显着的贡献。图 4.4(b) 显示了在四个 DS-WO-CRI 子数据集上固定 λ_u 和 λ_v 为最佳值时，公式 (2) 中的参数 λ_{uv} 对 TRCF 性能的影响。我们可以看到，TRCF 的性能首先随着 λ_{uv} 的增大而上升，然后在某个阈值后开始下降。在 LC1、LC2、LC3、LC4 上最佳的 λ_{uv} 值分别是 0.0001、0.01、0.01 和 0.1。这个结果说明了 DS-WO-CRI 的程度越大，对应的 λ_{uv} 的最优值相应的越大。换句话来说，当 DS-WO-CRI 的问题越严重时，通过标签和评分（即隐式偏好）桥接用户和物品的特性越重要，这就解释了为什么我们的模型在可以在严重的 DS-WO-CRI 的情况下表现良好的原因。

5 总结

在本文中，我们首先介绍了在实际的标记系统中存在的 DS-WO-CRI 问题。然后，我们提出一个新型的基于标签和评分的 CF 模型来处理这个问题。该模型使用主题建模来分别挖掘用户和物品的标签的语义信息，并将语义信息结合到矩阵分解中以对评分信息进行因式分解，并捕获用户与物品之间的标签和评分的桥接特征。



(a) λ_u, λ_v 的影响　　(b) λ_{uv} 的影响

图 4.4　固定 M=50, $\lambda_u, \lambda_v, \lambda_{uv}$ 对 TRCF 召回率的影响。

据我们所知,这是文献中首次尝试引入 DS-WO-CRI 问题并提出一个处理它的模型。最后,对两个流行的数据集进行的实验表明,我们的模型在精确率和召回率方面显着优于目前最先进的方法,特别是在 DS-WO-CRI 情况下。

6　致谢

我们感谢匿名审稿人的有益建议。这项工作得到了中国国家自然科学基金（61379034）、国家重点技术研发计划（2014BAH28F05）、广东省科技计划（2013B040100004, 2013B040403002）和中国浙江省自然科学基金（LQ14F010006）的部分支持。

参 考 文 献

[1] Deepak Agarwal and Bee Chung Chen. flda:matrix factorization through latent dirichlet allocation. In International Conference on Web Search and Web Data Mining, WSDM 2010, New York, Ny, Usa, February, pages 91–100, 2010.

[2] Alejandro Bellogín, Iván Cantador, and Pablo Castells. A comparative study of heterogeneous item recommendations in social systems. Information Sciences An International Journal, 221(1):142–169, 2013.

[3] Dimitri P Bertsekas. Nonlinear programming. Athena scientific Belmont, 1999.

[4] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. Journal of Machine Learning Research, 3:993–1022, 2003.

[5] Ivan Cantador, Peter Brusilovsky, and Tsvi Kuflik. Second workshop on information heterogeneity and fusion in recommender systems (hetrec2011). In ACM Conference on Recommender Systems, Recsys 2011, Chicago, Il, Usa, October, pages 387–388, 2011.

[6] C. Chen, X. Zheng, Y. Wang, F. Hong, and Z. Lin. Context-aware collaborative topic regression with social matrix factorization for recommender systems. Palo Alto California Aaai Press, 3(3):239–242, 2014.

[7] Chaochao Chen, Jing Zeng, Xiaolin Zheng, and Deren Chen. Recommender system based on social trust relationships. In IEEE International Conference on E-Business Engineering, pages 32–37, 2013.

[8] Mukund Deshpande and George Karypis. Item-based top-n recommendation algorithms. ACM Transactions on Information Systems (TOIS), 22(1):143–177, 2004.

[9] Jonathan L Herlocker, Joseph A Konstan, Loren G Terveen, and John T Riedl. Evaluating collaborative filtering recommender systems. ACM Transactions on Information Systems (TOIS), 22(1):5–53, 2004.

[10] Mohsen Jamali and Martin Ester. A matrix factorization technique with trust propagation for recommendation in social networks. In ACM Conference on Recommender Systems, Recsys 2010, Barcelona, Spain, September, pages 135–142, 2010.

[11] Yehuda Koren. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 426–434, 2008.

[12] Yehuda Koren, Robert Bell, and Chris Volinsky. Matrix factorization techniques

for recommender systems. ieee, computer journal, 42(8), 30-37. 42(8):30–37, 2009.

[13] Hao Ma, Dengyong Zhou, Chao Liu, Michael R. Lyu, and Irwin King. Recommender systems with social regularization. In Forth International Conference on Web Search and Web Data Mining, WSDM 2011, Hong Kong, China, February, pages 287–296, 2011.

[14] Prem Melville, Raymod J. Mooney, and Ramadass Nagarajan. Content-boosted collaborative filtering for improved recommendations. In Eighteenth National Conference on Artificial Intelligence, pages 187–192, 2002.

[15] Sanjay Purushotham, Yan Liu, and C-C Jay Kuo. Collaborative topic regression with social matrix factorization for recommendation systems. arXiv preprint arXiv:1206.4684, 2012.

[16] R. Salakhutdinov and A. Mnih. Probabilistic matrix factorization. In International Conference on Machine Learning, pages 880–887, 2008.

[17] Yue Shi, Martha Larson, and Alan Hanjalic. Collaborative filtering beyond the user-item matrix: A survey of the state of the art and future challenges. ACM Computing Surveys (CSUR), 47(1):3, 2014.

[18] Chong Wang and David M. Blei. Collaborative topic modeling for recommending scientific articles. In ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, Ca, Usa, August, pages 448–456, 2011.

[19] Hao Wang, Binyi Chen, and Wu Jun Li. Collaborative topic regression with social regularization for tag recommendation. In International Joint Conference on Artificial Intelligence, pages 2719–2725, 2013.

[20] Jingwei Xu, Yuan Yao, Hanghang Tong, Xianping Tao, and Jian Lu. Ice-breaking: mitigating cold-start recommendation problem by rating comparison. In International Conference on Artificial Intelligence, pages 3981–3987, 2015.

[21] Nan Zheng and Qiudan Li. A recommender system based on tag and time information for social tagging systems. Expert Systems with Applications, 38(4):4575–4587, 2011.

[22] T. C Zhou, Hao Ma, I King, and M. R Lyu. Tagrec: Leveraging tagging wisdom for recommendation. In International Conference on Computational Science and Engineering, pages 194–199, 2009.

[23] Tom Chao Zhou, Hao Ma, Michael R. Lyu, and Irwin King. Userrec: A user recommendation framework in social tagging systems. In Twenty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2010, Atlanta, Georgia, Usa, July, pages 1486–1491, 2010.

Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16)

Capturing Semantic Correlation for Item Recommendation in Tagging Systems

Chaochao Chen, Xiaolin Zheng
College of Computer Science
Zhejiang University
{zjuccc,xlzheng}@zju.edu.cn

Yan Wang
Department of Computing
Macquarie University
yan.wang@mq.edu.au

Fuxing Hong, Deren Chen
College of Computer Science
Zhejiang University
{cstur4,drc}@zju.edu.cn

Abstract

The popularity of tagging systems provides a great opportunity to improve the performance of item recommendation. Although existing approaches use topic modeling to mine the semantic information of items by grouping the tags labelled for items, they overlook an important property that tags link users and items as a bridge. Thus these methods cannot deal with the *data sparsity without commonly rated items* (*DS-WO-CRI*) problem, limiting their recommendation performance. Towards solving this challenging problem, we propose a novel tag and rating based collaborative filtering (CF) model for item recommendation, which first uses topic modeling to mine the semantic information of tags for each user and for each item respectively, and then incorporates the semantic information into matrix factorization to factorize rating information and to capture the bridging feature of tags and ratings between users and items. As a result, our model captures the semantic correlation between users and items, and is able to greatly improve recommendation performance, especially in *DS-WO-CRI* situations. Experiments conducted on two popular real-world datasets demonstrate that our proposed model significantly outperforms the conventional CF approach, the state-of-the-art social relation based CF approach, and the state-of-the-art topic modeling based CF approaches in terms of both precision and recall, and it is an effective approach to the *DS-WO-CRI* problem.

Introduction

In recent years, tagging systems, such as *Delicious* (social bookmarking), *Last.fm* (social music), *Flickr* (photo sharing), and *YouTube* (video sharing), provide effective ways for users to organize, manage, share, and search various kinds of items (resources). For example, one may tag Lady Gaga with "pop" and "female vocalist" when he listens to her music on *Last.fm*. These valuable tags, which appear along with the tagging and rating behaviors, strongly suggest the need to use such information to provide personalized recommendation services (Zheng and Li, 2011).

The increasing popularity of tagging systems has promoted the development of recommender systems, especially collaborative filtering (CF) approaches, in tagging systems. So far, two main types of CF on tagging systems exist: *tag recommendation* (Wang, Chen, and Li, 2013; Fang et al., 2015), which aims to recommend appropriate tags for items, and *tag-based item recommendation* (Zhou et al., 2009; Xu et al., 2011; Zhou et al., 2010), which focuses on recommending similar users or items to the target user based on tag and other information (e.g., rating).

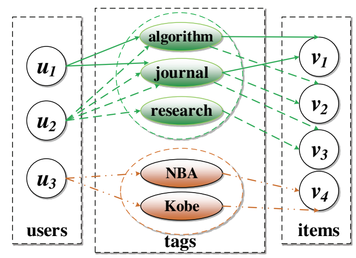


Figure 1: An example of tagging system. There is a rating behind each tagging behavior, and we omit the ratings for conciseness.

Currently, a trend in the literature is the use of topic modeling in CF to handle tag information (Agarwal and Chen, 2010; Wang and Blei, 2011; Purushotham, Liu, and Kuo, 2012; Wang, Chen, and Li, 2013; Chen et al., 2014). For example, Wang and Blei (2011) proposed a collaborative topic regression (CTR) model that can be used for tag-based item recommendation. Chen et al. (2014) proposed another item recommendation method that combined CTR with social matrix factorization (Ma et al., 2011) to make a better prediction. However, the existing approaches just associate tags with users or items, and overlook an important property of tags that tags link users and items as a bridge, as what ratings do. But tags can reflect the semantic correlation between users and items, which ratings cannot do.

When a user has tagged some items, these tags clearly represent the user's preference for the items. The more frequently a tag has been used by a user, the more likely this user is interested in the group of items that can be labeled by this tag (Zheng and Li, 2011). Similarly, the more frequently a tag has been given to an item by users, the more likely this item matches the tag. Thus, tags contain the semantic information of both users and items, not just one of them.

A motivating example. Figure 1 depicts an example of a tagging system, which consists of three users (u_1, u_2, and u_3), four items (v_1, v_2, v_3, and v_4), and five tags ("algorithm", "journal", "research", "NBA", and "Kobe").

In this example, user u_1 labelled item v_1, and user u_2 labelled items v_2 and v_3. Thus users u_1 and u_2 have no commonly rated items. We term this situation *data sparsity without commonly rated items* (*DS-WO-CRI*). DS-WO-CRI is a typical subset of the standard data sparsity problem (i.e., the known user-item actions are rare comparing with all the user-item pairs). In *DS-WO-CRI* situations, the existing CF approaches, e.g., PMF and CTR, cannot essentially recommend item v_1 to user u_2 because they cannot capture any relation between them. However, a good recommender system should recommend items v_2 and v_3 to user u_1 and recommend item v_1 to user u_2, because in this example, users u_1 and u_2 are probably researchers on algorithms, and items v_1, v_2, and v_3 are probably related to algorithms.

The existing studies have shown that a user's action on an item, e.g., tag and rate, has already indicated this user's interests in this item, regardless of how the user rated this item (Koren, 2008; Koren, Bell, and Volinsky, 2009). In other words, a user implicitly expresses his preferences by voicing his opinion through tagging and voting a (high or low) rating (Koren, 2008). Thus, a user and the items that he has tagged and rated tend to share similar latent features, and we term it *implicit preference* in this paper. In the above example, in particular, user u_2 gives item v_3 tags "journal" and "research" and this not only shows that user u_2 is likely to be a researcher, but also indicates that item v_3 is likely to be a research journal or something related to it. Thus, semantically, the latent features of user u_2 and item v_3 should be similar to some extent.

However, the existing approaches fail to capture the semantic correlation between users and items, and thus their recommendation performance is limited, especially in *DS-WO-CRI* situations.

Our proposal. To deal with the above mentioned *DS-WO-CRI* problem, in this paper, we propose a novel CF model. We first use topic modeling to mine the semantic information of tags for each user and for each item respectively, and then incorporate the semantic information into matrix factorization to factorize rating information and capture the bridging feature of tags and ratings between users and items (i.e., implicit preference). As a result, our model captures the semantic correlation between users and items, and can recommend an item to a user if they have similar semantic information, though they are in a *DS-WO-CRI* situation.

Contributions. The main contributions of our work are summarized as follows: (1) We first point out the important feature of tags, namely, they link users and items as a bridge, outlining the semantic correlation between users and items, and then we illustrate that utilizing this feature can help deal with the *DS-WO-CRI* problem. To the best of our knowledge, this is the first study in the literature to identify this problem; (2) We propose a novel tag and rating based CF model, which can capture the semantic correlation between users and items and thus can greatly improve the recommendation performance, especially in *DS-WO-CRI* situations. We also propose our parameter learning method based on coordinate ascent algorithm. To the best of our knowledge, this study is the first attempt in the literature to capture the semantic correlation between users and items provided by tags in tag-based item recommendation; (3) Experiments conducted on two popular real-world datasets demonstrate that our proposed model significantly outperforms the state-of-the-art approaches in terms of both precision and recall. The experiments also demonstrate that our proposed model is an effective approach to the *DS-WO-CRI* problem.

Related Work

In this section, we review the existing item recommendation methods in tagging systems in three groups, including (1) the conventional CF approaches, (2) the topic modeling based CF approaches, and (3) the social relation based CF approaches.

Based on the existing research (Shi, Larson, and Hanjalic, 2014), the conventional CF approaches, which only use user-item rating information to make recommendations, are in two major categories: the memory-based CF (Deshpande and Karypis, 2004) and model-based CF (Koren, Bell, and Volinsky, 2009; Koren, 2008; Zhou et al., 2009; Xu et al., 2015), both of which can be used to make recommendations in tagging systems.

The conventional CF approaches, e.g., TagRec (Zhou et al., 2009), cannot capture content (e.g., tag) information. Thus, some hybrid approaches were proposed to combine content-based approach and CF to do item recommendation (Melville, Mooney, and Nagarajan, 2002) and (Basilico and Hofmann, 2004). However, these methods take content simply as a vector of words, and thus cannot mine their semantic information. To take advantage of semantic information provided by content (e.g., tag), researchers use topic modeling to improve recommendation performance. Agarwal et al., proposed fLDA (Agarwal and Chen, 2010), which combines RLFM (Agarwal and Chen, 2009) with latent Dirichlet allocation (LDA) by assigning item factors through a richer prior learnt from LDA. Both RLFM and fLDA incorporate additional covariates that are obtained from additional meta-feature information, e.g., user age and item category, which, however, are out of the scope of this paper. Later on, Wang et al. (Wang and Blei, 2011) proposed CTR to combine probabilistic matrix factorization (PMF) (Mnih and Salakhutdinov, 2007) with LDA (Blei, Ng, and Jordan, 2003) to make recommendations. It has been proven in (Wang and Blei, 2011) that CTR performs better than fLDA in a similar setting, since fLDA largely ignores the other users' ratings.

Moreover, social information between users and between items is considered valuable to improve recommendation performance (Chen et al., 2013). First, user social information is incorporated into conventional CF models (Jamali and Ester, 2010; Ma et al., 2011). For example, Ma et al. (2011) proposed Soreg to constrain the difference between the user latent factors of connected users. Second, neighbor user or item social information is incorporated into topic modeling based CF models (e.g., CTR) to further improve recommendation performance. For example, Purushotham, Liu, and Kuo (2012) and Chen et al. (2014) proposed two models (i.e., CTR-SMF and CTR-SMF2) to incorporate user social network into CTR to further improve item recommendation performance. Wang, Chen, and Li (2013) proposed a model to incorporate item social relationship into CTR to further improve tag recommendation performance in social tagging systems.

However, all the above approaches not only overlook the semantic information between both users and items embedded in tags, but also neglect the bridging feature of tags and ratings between users and items (i.e., implicit preference). Therefore, they cannot capture the semantic correlation between users and items, and suffer from *DS-WO-CRI* problem. To overcome these shortcomings, in this paper, we propose a novel tag and rating based CF model, which can capture the semantic correlation between users and items. Hence, our model can help deal with the *DS-WO-CRI* problem and improve recommendation performance.

The Proposed Model–TRCF

In this section, we present a novel tag and rating based CF (TRCF) model. We first formalize the tag-based item recommendation problem and define notations. Then, we present TRCF, which is a hierarchical Bayesian model. Finally, we propose our parameter learning method based on coordinate ascent algorithm.

Preliminaries

Assume that we have a set of users $\mathbb{U} = \{u_1, ..., u_I\}$, who have labelled a set of items $\mathbb{V} = \{v_1, ..., v_J\}$ with a set of tags $\mathbb{T} = \{t_1, ..., t_N\}$ and a set of ratings $\mathbb{R} = \{R_1, ..., R_O\}$, where *I*, *J*, *N*, and *O* denote the numbers of users, items, tags, and ratings, respectively. Each *user-item-tag-rating* (U-I-T-R) observe data is a 4-tuple (u_i, v_j, T_{ij}, R_{ij}), where $u_i \in \mathbb{U}$, $v_j \in \mathbb{V}$, T_{ij} is a set of tags that user u_i gives to item v_j, and $T_{ij} \subseteq \mathbb{T}$. R_{ij} is the rating that user u_i gives to item v_j based on the extent to which he likes the item and tags it at the same time; however, the user-item (U-I) rating set $\mathbb{R}$ is typically of integers, e.g., in the range [1, 5] in *MovieLens*. Let $U \in \mathbb{R}^{K \times I}$ denote the latent user feature matrices, where the column vector U_i represents the *K*-dimensional user-specific latent feature vector of user u_i. Let $V \in \mathbb{R}^{K \times J}$ denote the latent item feature matrices, where the column vector V_j represents the *K*-dimensional item-specific latent feature vector of item v_j.

For tag and rating based item recommendation, given the existing U-I-T-R 4-tuples, our goal is to predict the unknown rating from a user u_i to an item v_j.

Tag and Rating based Collaborative Filtering

TRCF is a novel hierarchical Bayesian model, and its graphical model is shown in Figure 2, where N_u and N_v denote the number of tags for user u_i and for item v_j, respectively. TRCF first groups the tags for each user and for each item respectively, and then it uses latent Dirichlet allocation (LDA) to mine the semantic information of tags for each user and each item respectively (plotted in red in Figure 2). Finally it incorporates these semantic information into matrix factorization to factorize rating information (plotted in purple in Figure 2) and capture the implicit preference provided by tags and ratings (plotted in blue in Figure 2).

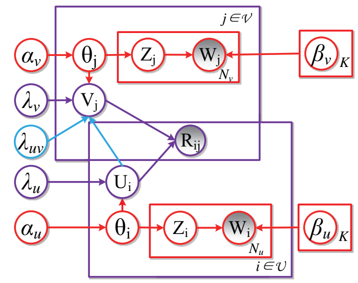


Figure 2: Graphical model of TRCF. The LDA part is plotted in red, the implicit preference part is shown in blue, and the PMF part is plotted in purple.

TRCF performs LDA on both the user side and the item side, and thus can capture the semantic information for both users and items, not just items as the existing works do. In addition, in TRCF, when a user and an item are linked by tags and ratings, their latent features are similar to each other to some extent, which is referred to as implicit preference. In contrast, the existing topic modeling based CF approaches, e.g., CTR, CTR-SFM, and CTR-SMF2, assume users and items are independent, and neglect the bridge feature of tags and ratings between users and items. Thus, TRCF can capture the semantic correlation between users and items, and is able to deal with the *DS-WO-CRI* problem. Assuming there are *K* topics for both users and items, the generative process of TRCF works as follows:

1. **Mining semantic information of tags for users**. For each user u_i,
   (a) Draw topic proportions $\theta_i \sim Dirichlet(\alpha_u)$;
   (b) Draw user latent vector as $U_i \sim \mathcal{N}(\theta_i, \lambda_u^{-1} I_K)$;
   (c) For each tag w_{in_u} of user u_i,
      i. Draw topic assignment $z_{in_u} \sim Mult(\theta_i)$;
      ii. Draw tag $w_{in_u} \sim Mult(\beta_{z_{in_u}})$;
2. **Mining semantic information of tags for items, and capturing implicit preference between users and items**. For each item v_j,
   (a) Draw topic proportions $\theta_j \sim Dirichlet(\alpha_v)$;
   (b) Draw item latent vector as $V_j \sim \mathcal{N}(\theta_j, \lambda_v^{-1} I_K) \times \prod_i I_{ij}^R \mathcal{N}(U_i, \lambda_{uv}^{-1} I_K)$;
   (c) For each tag w_{jn_v} of user v_j,
      i. Draw topic assignment $z_{jn_v} \sim Mult(\theta_j)$;
      ii. Draw tag $w_{jn_v} \sim Mult(\beta_{z_{jn_v}})$;
3. **Drawing the rating**. For each user-item pair (*i*, *j*),

$$R_{ij} \sim \mathcal{N}({U_i}^T V_j, c_{ij}^{-1}).$$

In the above generative process, $\mathcal{N}(x|\mu, \sigma^2)$ is a Gaussian distribution with a mean μ and a variance σ^2, and I_K is an identity matrix with K rows and K columns. I_{ij}^R is an indicator function the value of which equal to 1 if user u_i

rated item v_j, 0 otherwise. C is a rating confidence matrix with element c_{ij} denotes the rating confidence. Please refer to (Wang and Blei, 2011) for more details.

The parameter λ_u balances the contribution of user semantic information provided tags and rating information to the model performance. Similarly, the parameter λ_v balances the contribution of item semantic information provided by tags and rating information to the recommendation performance. The parameter λ_{uv} balances the contribution of implicit preference on model performance, i.e., the degree of the latent feature similarity of a user and an item linked by a rating and tags.

The conditional distribution of the observed ratings can be formalized as

$$p\left(R|U,V,C\right) = \prod_i \prod_j \mathcal{N}(R_{ij}|U_i^T V_j, c_{ij}).$$

The user and latent vectors U_i and V_j are generated in a similar way to CTR, which can be formalized as

$$p\left(U|\lambda_u\right) \sim \prod_i \mathcal{N}(\theta_i, \lambda_u^{-1} I_K),$$

$$p\left(V|U,\lambda_v,\lambda_{uv}\right) \sim \prod_j \mathcal{N}(\theta_j, \lambda_v^{-1} I_K) \times \prod_i I_{ij}^R \mathcal{N}(U_i, \lambda_{uv}^{-1} I_K).$$

Given the U-I-T-R information, by using Bayesian inference, we can obtain the following equation for the posterior probability of latent feature vectors of TRCF:

$$\begin{aligned} &p\left(U,V|R,C,\lambda_u,\lambda_v,\lambda_{uv}\right) \\ &\propto p\left(R|U,V,C\right) p\left(U|\lambda_u\right) p\left(V|U,\lambda_v,\lambda_{uv}\right). \end{aligned} \tag{1}$$

Parameter Learning of TRCF

Given topic parameters β_u and β_v, computing the full posterior of U_i, V_j, θ_i, and θ_j directly is intractable. We use coordinate ascent algorithm to learn the maximum a posteriori estimates. Maximizing the posterior over the two latent features with fixed hyper-parameters in Equation (1) is equivalent to maximizing the following complete log likelihood of *U*, *V*, $\theta_{1:I}$, $\theta_{1:J}$ and *R*, given λ_u and λ_v:

$$\begin{aligned} L = &-\frac{\lambda_u}{2}\sum_i \left(U_i-\theta_i\right)^T\left(U_i-\theta_i\right) \\ &-\frac{\lambda_v}{2}\sum_j \left(V_j-\theta_j\right)^T\left(V_j-\theta_j\right) \\ &-\sum_{ij}\frac{c_{ij}}{2}\left(R_{ij}-{U_i}^T V_j\right)^2 \\ &+\sum_i\sum_{n_u}\log\left(\sum_k \theta_{ik}\beta_{k,w_{in_u}}\right) \\ &+\sum_j\sum_{n_v}\log\left(\sum_k \theta_{jk}\beta_{k,w_{jn_v}}\right) \\ &-\frac{\lambda_{uv}}{2} I_{ij}^R \sum_{ij}\left(U_i-V_j\right)^T(U_i-V_j). \end{aligned} \tag{2}$$

We omit a constant and set the Dirichlet priors $\alpha_u = \alpha_v = 1$. This function can be optimized by using coordinate ascent. That is, we fix β_u and β_v, and iteratively optimize the MF variables U_i, V_j and the topic proportions θ_i and θ_j. Specifically, we first update U_i and V_j, given the current estimate of θ_i, θ_j. We take the gradient of L in Equation (2) with respect to U_i and V_j, and set it to zero,

$$\frac{\partial L}{\partial U_i} = 0, \frac{\partial L}{\partial V_j} = 0. \tag{3}$$

Solving the above equations will lead to the following update equation,

$$\begin{aligned} U_i &\leftarrow \left(V C_i V^T + \lambda_u I_K + \lambda_{uv}\sum_j I_{ij}^R I_K\right)^{-1} \\ &\quad (V C_i R_i + \lambda_u \theta_i + \lambda_{uv}\sum_j I_{ij}^R V_j), \\ V_j &\leftarrow \left(U C_j U^T + \lambda_v I_K + \lambda_{uv}\sum_i I_{ij}^R I_K\right)^{-1} \\ &\quad (U C_j R_j + \lambda_v \theta_j + \lambda_{uv}\sum_i I_{ij}^R U_i), \end{aligned} \tag{4}$$

where C_i is a diagonal matrix with $c_{ij}, j = 1, ..., J$ as its diagonal elements, and $R_i = R_{ij}{}_{j=1}^J$ for user u_i. For item v_j, C_j and R_j are similarly defined.

Equation (4) shows how parameters λ_u, λ_v, and λ_{uv} affect the user latent feature and the item latent feature. A bigger λ_u corresponds to a bigger proportion of the user latent feature from the user tags rather than the rating information. Similarly, a bigger λ_v indicates a bigger proportion of the item latent feature from the item tags, rather than the rating information. Also, a bigger λ_{uv} means a stronger constraint that the paired user and item linked by tags and ratings should have a similar latent feature, i.e., implicit preference. From Equation (4), we can see that probabilistic matrix factorization (PMF) and collaborative topic regression (CTR) are all special cases of TRCF.

Then, we update the topic proportions θ_i and θ_j given the current MF variables U_i and V_j. For θ_i, we first define $q\left(z_{in_u} = k\right) = \Phi_{in_u k}$, and then separate the users that contain θ_i and apply Jensen's inequality,

$$\begin{aligned} L(\theta_i) \geq &-\frac{\lambda_u}{2}\left(U_i-\theta_i\right)^T\left(U_i-\theta_i\right) \\ &+\sum_{n_u}\sum_k \Phi_{in_u k}(log\theta_{ik}\beta_{k,w_{in_u}} - log\Phi_{in_u k}) \\ &= L(\theta_i, \Phi_i). \end{aligned}$$

Here, $\Phi_i = \Phi_{in_u k}{}_{n_u=1,k=1}^{N_u \times K}$. Obviously $L(\theta_i, \Phi_i)$ is a tight lower bound of $L(\theta_i)$, and we can use projection gradient (Bertsekas, 1995) to optimize θ_i. The optimal $\Phi_{in_u k}$ is $\Phi_{in_u k} \propto \theta_{ik}\beta_{k,w_{in_u}}$. For θ_j, it is similarly updated.

As for β_u, we update the same M-step for topics as in LDA (Blei, Ng, and Jordan, 2003),

$$\beta_{kw_i} \propto \sum_i \sum_{n_u} \Phi_{in_u k} 1[w_{in_u} = w].$$

For β_v [1], it is similarly updated. After the optimal param-

[1]When using TF-CTR, a useful tactic is to fuse the user tags and item tags as the input of LDA, which ensures that users and items have the same semantic information in each element of *K*, that is, to make $\beta_u = \beta_v$.

latex表格

Dataset	users	items	tags	user-tags-items	ratings
Delicious	1,867	69,226	53,388	437,593	104,799
Lastfm	1,892	17,632	11,946	186,479	92,834

Table 1: Dataset description

eters U^*, V^*, ${\theta_{1:I}}^*$, ${\theta_{1:J}}^*$, ${\beta_u}^*$, and ${\beta_v}^*$ have been learned, our model can predict ratings:

$$ {R_{ij}}^* \approx (U_i^*)^T V_j^*. $$

Experiments and Analysis

In this section, we introduce the experiments conducted on two popular real-world datasets, which aim to answer the following questions: (1) How does our model perform comparing the state-of-the-art approaches? (2) How does our approach deal with the *DS-WO-CRI* problem? (3) How do parameters λ_u, λ_v, and λ_{uv} affect the performance of TRCF?

Datasets

We use two real-world datasets in our experiments: hetrec2011-delicious-2k (*Delicious*) and hetrec2011-lastfm-2k (*Lastfm*) (Cantador, Brusilovsky, and Kuflik, 2011). Both datasets have been widely used to conduct experiments in tagging systems (Bellogín, Cantador, and Castells, 2013), and they are described in Table 1.

For each of the two datasets, we consider a user rating for an item as '1' if the user has bookmarked (or listened) the item; otherwise, the user rating for the item is '0'.

In our experiments, we split each dataset into three parts — a training dataset (80%), a held-out validation dataset (10%), and a test dataset (10%). We train our model on the training dataset, obtain the optimal parameters on the validation dataset, and evaluate our model on the test dataset.

Comparison and Evaluation

As stated in related works, there are many kinds of recommendation approaches, e.g., memory-based approach and hybrid approach. Here, we compare the proposed TRCF with the following three kinds of state-of-the-art approaches, i.e., the conventional CF approach, the social relation based CF approach, and the topic modeling based approach:

SVD++ (Koren, 2008) is a classic conventional CF approach that only uses U-I rating information.

Soreg (Ma et al., 2011) is a state-of-the-art social relation based CF approach, which uses U-I rating and user social information.

CTR (Wang and Blei, 2011) is a state-of-the-art topic modeling based CF approach, which uses U-I-T-R information similar to the 4-tuple used in TRCF.

CTR-SMF (Purushotham, Liu, and Kuo, 2012) combines user social matrix factorization with CTR. It incorporates additional user social information additional to the U-I-T-R 4-tuple used in TRCF.

CTR-SMF2 (Chen et al., 2014) improves CTR-SMF, and it also incorporates user social information additional to the U-I-T-R 4-tuple used in TRCF.

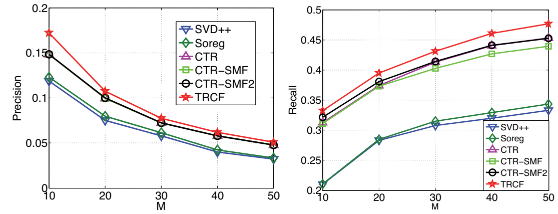


Figure 3: *Precision* and *Recall* comparison with different *M* and the best parameters of each method on *Delicious*.

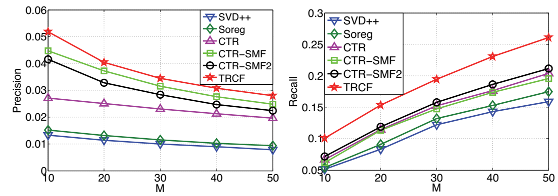


Figure 4: *Precision* and *Recall* comparison with different *M* and the best parameters of each method on *Lastfm*.

Precision and *Recall* have been widely used as the metrics to evaluate recommendation performance (Herlocker et al., 2004). Thus, we use both *Precision* and *Recall* to evaluate the recommendation performance, and note that the same way of computing recall is also used in CTR, CTR-SMF, and CTR-SMF2. For each user, *Precision* and *Recall* are defined as follows:

$$ Precision@M = \frac{\#\, items\, the\, user\, likes\, in\, Top\ M}{M}, $$

$$ Recall@M = \frac{\#\, items\, the\, user\, likes\, in\, Top\ M}{\#\, total\, items\, the\, user\, likes}, $$

where *M* is the number of returned items. We compute the average of all the items' precision and recall in the test dataset as the final result.

Performance Comparison and Analysis

During the comparison, we have used the best parameters for SVD++, CTR, and CTR-SMF that are set in CTR-SMF (Purushotham, Liu, and Kuo, 2012), which uses the same datasets. For each of Soreg, CTR-SMF2, and our model, we have used grid search to obtain the best parameters.

Results: Figures 3 and 4 show the overall performance for each recommendation approach on *Delicious* dataset and *Lastfm* dataset, in which we set *M* = 10, 20, 30, 40, 50 and fix the parameters of each approach to the best values. The results show that the conventional CF approach (i.e., SVD++) and the social relation based CF approach (i.e., Soreg) have the similar performance. The three topic modeling based CF approaches (i.e., CTR, CTR-SMF, and CTR-SMF2) significantly outperform SVD++ and Soreg, and also have similar performance, which shows the importance of tag information in recommendations.

Our proposed method, TRCF, significantly outperforms each of SVD++, Soreg, CTR, CTR-SMF, and CTR-SMF2

sub-dataset name	*LC1*	*LC2*	*LC3*	*LC4*
DS-WO-CRI degree	20%	40%	60%	80%
users	1,837	1,709	1,718	1,706
items	11,584	8,018	7,650	7,431

Table 2: Statistics of each *DS-WO-CRI* sub-dataset

on the two datasets in terms of different *M*. Specifically, on average, TRCF improves SVD++, Soreg, CTR, CTR-SMF, and CTR-SMF2 by 46.75%, 39.74%, 8.62%, 8.78%, and 8.65%, in terms of precision, and by 44.99%, 42.40%, 5.23%, 7.27%, and 4.21%, in terms of recall, on the *Delicious* dataset. On average, TRCF improves SVD++, Soreg, CTR, CTR-SMF, and CTR-SMF2 by 259.80%, 210.27%, 57.96%, 11.64%, and 23.85%, in terms of precision, and by 73.03%, 60.39%, 34.18%, 39.04%, and 28.12%, in terms of recall, on the *Lastfm* dataset.

Analysis and summary: The comparison demonstrates the effectiveness of our proposed method which captures the semantic correlation between users and items. Experimental results also indicate that though user social information (e.g., adopted in Soreg, CTR-SMF, and CTR-SMF2) can improve recommendation performance, considering user and item semantic correlation is a more effective way to improve item recommendation performance.

DS-WO-CRI Experiments

All four topic modeling based CF approaches, including TRCF, can improve recommendation performance by capturing the semantic information of items. To study their capability of handing the *DS-WO-CRI* problem, we conduct the following experiments.

We first randomly filter the original *Lastfm* datasets into four sub-datasets based on the degree of *DS-WO-CRI*, and each sub-dataset is described in Table . The degree of *DS-WO-CRI* is defined as follows:

$$x\% = \frac{\#\, users\, without\, commonly\, rated\, items}{\#\, total\, users}.$$

We then conduct comparison experiments on each of the sub-dataset. Figure 5 shows that our model always achieves the best performance under different *DS-WO-CRI* degrees. The average improvements of our model over other three topic modeling based approaches on *LC1*, *LC2*, *LC3*, *LC4* are 49.67%, 75.81%, 345.77%, and 428.97% respectively, in terms of precision, and are 33.94%, 65.00%, 383.63%, and 458.00% respectively, in term of recall. The experimental results show that a greater degree of *DS-WO-CRI* corresponds to a higher improvement of our model against other models on each of the two datasets.

Summary: The *DS-WO-CRI* experiments demonstrate the effectiveness of our method in dealing with the *DS-WO-CRI* problem: a greater degree of *DS-WO-CRI* corresponds to a higher improvement of our model against other models. This is due to the ability of our method to capture the semantic correlation between users and items provided by tags.

Parameter Effect Analysis

Figure 6(a) shows the effect of λ_u and λ_v in Eq.(2) on the performance of TRCF by fixing $\lambda_{uv} = 0$ on *Lastfm*. We can see that TRCF achieves the best performance when $\lambda_u = \lambda_v = 10$, which means that both user and item semantic information contribute significantly to model performance. Figure 6(b) shows how the performance of TRCF is affected by parameter λ_{uv} in Eq.(2) on each *DS-WO-CRI* sub-dataset with the best λ_u and λ_v. As we can see, the performance of TRCF first increases with λ_{uv} and then starts to decrease at a certain threshold. The best value of λ_{uv} on *LC1*, *LC2*, *LC3*, and *LC4* is 0.0001, 0.01, 0,01, and 0.1, respectively. These results demonstrate that a greater degree of *DS-WO-CRI* corresponds to a greater λ_{uv}. In other words, the feature of bridging users and items by tags and ratings (i.e., the implicit preference) is more important when the *DS-WO-CRI* problem is severer, which explains why our model performs well in severe *DS-WO-CRI* situations.

Conclusions

In this paper, we first introduce the *DS-WO-CRI* problem that exists in item recommendation in real tagging systems. Then, we present a novel tag and rating based CF model to deal with this problem. The proposed model uses topic modeling to mine the semantic information of tags for users and items respectively, and incorporates the semantic information into matrix factorization to factorize the rating information and to capture the bridging feature of tags and ratings between users and items. To the best of our knowledge, this is the first attempt in the literature to introduce the *DS-WO-CRI* problem and propose a model to deal with it. Finally, the experiments conducted on two well-known datasets have demonstrated that our model significantly outperforms the state-of-the-art approaches in terms of both precision and recall, especially in *DS-WO-CRI* situations.

Acknowledgments

We thank the anonymous reviewers for their helpful suggestions. This work was supported in part by the National Natural Science Foundation of China (No. 61379034), the National Key Technology R&D Program (No. 2014BAH28F05), the Guangdong Province Science and Technology Program (No. 2013B040100004 and No. 2013B040403002), and the Zhejiang Provincial Natural Science Foundation of China (LQ14F010006).

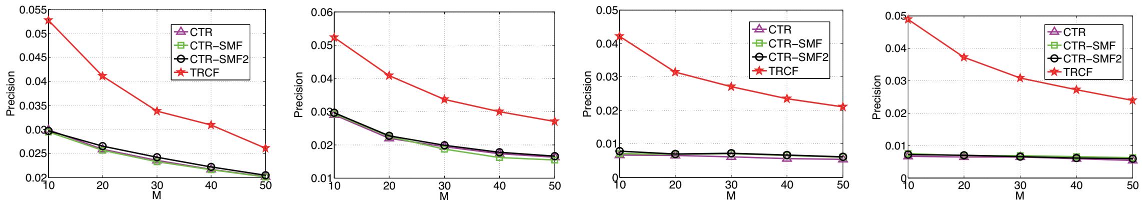
References

Agarwal, D., and Chen, B.-C. 2009. Regression-based latent factor models. In *SIGKDD*, 19–28.

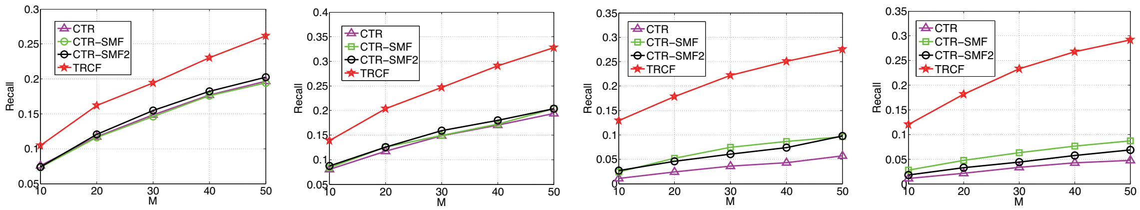
Agarwal, D., and Chen, B.-C. 2010. flda: matrix factorization through latent dirichlet allocation. In *WSDM*, 91–100.

Basilico, J., and Hofmann, T. 2004. Unifying collaborative and content-based filtering. In *ICML*, 65–72.

Bellogín, A.; Cantador, I.; and Castells, P. 2013. A comparative study of heterogeneous item recommendations in social systems. *Information Sciences* 221:142–169.

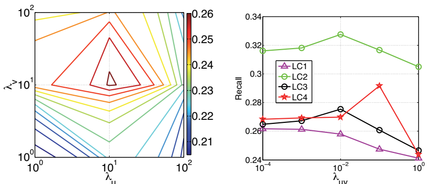


(a) *LC1* (20% *DS-WO-CRI*) (b) *LC2* (40% *DS-WO-CRI*) (c) *LC3* (60% *DS-WO-CRI*) (d) *LC4* (80% *DS-WO-CRI*)



(e) *LC1* (20% *DS-WO-CRI*) (f) *LC2* (40% *DS-WO-CRI*) (g) *LC3* (60% *DS-WO-CRI*) (h) *LC4* (80% *DS-WO-CRI*)

Figure 5: *Precision* and *Recall* comparison on each *DS-WO-CRI* sub-dataset with the best parameters of each method.



(a) Effect of λ_u and λ_v (b) Effect of λ_{uv}

Figure 6: Effect of λ_u, λ_v, and λ_{uv} on the recall performance of TRCF by fixing *M*=50.

Bertsekas, D. P. 1995. Nonlinear programming. *Athena Scientific*.

Blei, D. M.; Ng, A. Y.; and Jordan, M. I. 2003. Latent dirichlet allocation. *JMLR* 3:993–1022.

Cantador, I.; Brusilovsky, P.; and Kuflik, T. 2011. Second workshop on information heterogeneity and fusion in recommender systems (hetrec2011). In *RecSys*, 387–388.

Chen, C.; Zeng, J.; Zheng, X.; and Chen, D. 2013. Recommender system based on social trust relationships. In *ICEBE*, 32–37.

Chen, C.; Zheng, X.; Wang, Y.; Hong, F.; and Lin, Z. 2014. Context-aware collaborative topic regression with social matrix factorization for recommender systems. In *AAAI*. 9–15.

Deshpande, M., and Karypis, G. 2004. Item-based top-n recommendation algorithms. *TOIS* 22(1):143–177.

Fang, X.; Pan, R.; Cao, G.; He, X.; and Dai, W. 2015. Personalized tag recommendation through nonlinear tensor factorization using gaussian kernel. In *AAAI*, 439–445.

Herlocker, J. L.; Konstan, J. A.; Terveen, L. G.; and Riedl, J. T. 2004. Evaluating collaborative filtering recommender systems. *TOIS* 22(1):5–53.

Jamali, M., and Ester, M. 2010. A matrix factorization technique with trust propagation for recommendation in social networks. In *RecSys*, 135–142.

Koren, Y.; Bell, R.; and Volinsky, C. 2009. Matrix factorization techniques for recommender systems. *Computer* 42(8):30–37.

Koren, Y. 2008. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *SIGKDD*, 426–434.

Ma, H.; Zhou, D.; Liu, C.; Lyu, M. R.; and King, I. 2011. Recommender systems with social regularization. In *WSDM*, 287–296.

Melville, P.; Mooney, R. J.; and Nagarajan, R. 2002. Content-boosted collaborative filtering for improved recommendations. In *AAAI*, 187–192.

Mnih, A., and Salakhutdinov, R. 2007. Probabilistic matrix factorization. In *NIPS*, 1257–1264.

Purushotham, S.; Liu, Y.; and Kuo, C.-c. J. 2012. Collaborative topic regression with social matrix factorization for recommendation systems. In *ICML*, 759–766.

Shi, Y.; Larson, M.; and Hanjalic, A. 2014. Collaborative filtering beyond the user-item matrix: A survey of the state of the art and future challenges. *CSUR* 47(1):3.

Wang, C., and Blei, D. M. 2011. Collaborative topic modeling for recommending scientific articles. In *SIGKDD*, 448–456.

Wang, H.; Chen, B.; and Li, W.-J. 2013. Collaborative topic regression with social regularization for tag recommendation. In *IJCAI*, 2719–2725.

Xu, G.; Gu, Y.; Dolog, P.; Zhang, Y.; and Kitsuregawa, M. 2011. Semrec: a semantic enhancement framework for tag based recommendation. In *AAAI*. 1267–1272.

Xu, J.; Yao, Y.; Tong, H.; Tao, X.; and Lu, J. 2015. Ice-breaking: Mitigating cold-start recommendation problem by rating comparison. In *IJCAI*, 3981–3987.

Zheng, N., and Li, Q. 2011. A recommender system based on tag and time information for social tagging systems. *ESWA* 38(4):4575–4587.

Zhou, T. C.; Ma, H.; King, I.; and Lyu, M. R. 2009. Tagrec: Leveraging tagging wisdom for recommendation. In *CSE*, volume 4, 194–199.

Zhou, T. C.; Ma, H.; Lyu, M. R.; and King, I. 2010. Userrec: A user recommendation framework in social tagging systems. In *AAAI*. 1486–1491.