

基于景点标签的协同过滤推荐

史一帆^{1*}, 文益民^{1,2}, 蔡国永^{1,2}, 缪裕青^{1,2}

(1. 桂林电子科技大学 计算机科学与工程学院, 广西 桂林 541004; 2. 广西可信软件重点实验室(桂林电子科技大学), 广西 桂林 541004)

(* 通信作者电子邮箱 ymwen2004@aliyun.com)

摘 要:针对基于用户社会关系的协同过滤推荐算法有时无法给出目标用户对目标物品的评分的情况,以及基于物品的协同过滤推荐算法中存在的用户对不同类型物品的评分可能不具有可比性的问题,提出了两个基于物品标签的协同过滤推荐算法。这两个算法在计算物品相似度时引入了物品的类型标签信息。在景点评分数据上的实验结果表明:相比基于用户社会关系的协同过滤推荐算法,基于用户社会关系和物品标签的协同过滤推荐算法的准确率和覆盖率提升最高达 10% 和 4%;相比基于物品的协同过滤推荐算法,基于物品和物品标签的协同过滤推荐算法的准确率提升达 15%。这说明景点类型标签信息的引入能使得景点的相似度计算更准确。

关键词:社会关系; 标签; 协同过滤; 旅游推荐

中图分类号: TP391.4; TP18 **文献标志码:** A

Collaborative filtering recommendation based on tags of scenic spots

SHI Yifan^{1*}, WEN Yimin^{1,2}, CAI Guoyong^{1,2}, MIAO Yuqing^{1,2}

(1. School of Computer Science and Engineering, Guilin University of Electronic Technology, Guilin Guangxi 541004, China;

2. Guangxi Key Laboratory of Trusted Software (Guilin University of Electronic Technology), Guilin Guangxi 541004, China)

Abstract: In user-based collaborative filtering recommendation based on social relations, sometimes the ratings for the target items can not be predicted. What's more, in traditional item-based collaborative filtering, there are still some items which are not in the same class with the target item and not suitable to be references for predicting ratings. To handle these problems, two new algorithms of collaborative filtering recommendation were proposed, in which the tags of scenic spot's type were introduced to compute the similarity between two scenic spots. The experimental results on the data set of scenic spots' ratings show that, compared with the user-based collaborative filtering recommendation algorithms based on social relations, the algorithm based on the social relation and tag can increase the accuracy and the coverage by 10% and 4% respectively, and compared with the item-based collaborative filtering recommendation algorithms, the collaborative filtering recommendation algorithm based on item and tag can increase the accuracy by 15%, it also shows that introducing the tags of scenic spot's type can make the computation of the similarity between two scenic spots more accurate.

Key words: social relation; tag; collaborative filtering; travel recommendation

0 引言

协同过滤推荐算法是推荐系统中最基本的算法,它分为基于物品的协同过滤推荐与基于用户的协同过滤推荐。协同过滤推荐算法的基本思想是:根据用户对目标物品的评分计算其他物品与目标物品之间的相似度或其他用户与目标用户之间的相似度,然后根据相似度计算目标用户对目标物品的评分。

在旅游推荐领域,由于景点数量众多,两个用户之间很可能不存在共同给予评分的景点。因此,在基于用户的协同过滤推荐中,用户之间是否有共同评分的景点会影响用户间相似度的计算。为更有效地寻找目标用户的邻近用户,用户之间的信任度^[1-2],或是用户之间的信任度与相似度的结合

值^[3-4]被用来替代传统协同过滤中的相似度。这大大提高了邻近用户选取的有效性。但是,当存在有以下情形时,基于用户社会关系的协同过滤推荐算法无法给出目标用户对目标景点的评分:1) 其他用户与目标用户之间的信任度或相似度都小于设定的阈值;2) 在考虑了正相关相似度与负相关相似度的情况下,邻近用户的相似度之和为 0。

在基于物品的协同过滤推荐算法中,用户对不同景点评分的准确性或者可比性会影响景点之间相似度的计算。比如:对于不同类型的景点,用户给出的评分可能不具有可比性。假定“西湖”与“漓江”是同属于“自然风景”类型的景点,而“长城”是属于“古迹”类型的景点。若某用户对“西湖”的评分为 5,对“漓江”和“长城”的评分均为 4。从评分可以得知该用户喜欢“西湖”胜过“漓江”,但是很难说该用户喜欢

收稿日期:2014-06-09;修回日期:2014-06-26。 基金项目:国家自然科学基金资助项目(61363029);广西区科学研究与技术开发项目(桂科攻 14124005-2-1);广西可信软件重点实验室项目(KX201311)。

作者简介:史一帆(1988-),女,湖南郴州人,硕士,主要研究方向:机器学习、数据挖掘、旅游信息智能处理;文益民(1969-),男,湖南益阳人,教授,博士,CCF 高级会员,主要研究方向:机器学习、数据挖掘、推荐系统、智慧旅游;蔡国永(1971-),男,广西河池人,教授,博士,CCF 高级会员,主要研究方向:社交媒体分析、可信计算、软件工程;缪裕青(1966-),女,浙江台州人,副教授,博士,主要研究方向:数据挖掘、云计算、并行计算。

“西湖”一定胜过“长城”。

在一些旅游网站上,景点会被标注上标签。根据维基百科的定义,标签(Tag)是一种无层次化结构,是用来描述物品语义的关键词。根据给物品标注标签的人的不同,标签应用一般分为两种^[5]:一种是专家给物品标注标签;另一种是用户生成的(User Generated Content, UGC)标签,即让普通用户给物品标注标签。UGC标签体现了用户的旅游兴趣和旅游产品的语义。但是,相对于UGC标签而言,由专家标注的标签的噪声较小,语义更准确。因此,如何有效利用专家给景点标注的类型标签来提高协同过滤推荐算法的准确率和覆盖率是一个非常有价值的研究问题。

本文针对上述问题,将物品的类型标签信息引入协同过滤推荐算法以提高其覆盖率或准确率。

1 相关工作

在基于物品的协同过滤推荐^[6]中,利用目标用户曾经评分过且与目标物品之间相似度超过某个阈值的物品来计算目标用户对目标物品的评分。在基于用户的协同过滤推荐^[7]中,利用与目标用户相似且同时对一些物品都有评分的其他用户对目标物品的评分来计算目标用户对目标物品的评分。在引入社会关系后的基于用户的协同过滤推荐算法中,Massa等^[1-2]将用户之间的信任度替代传统协同过滤推荐中的用户相似度来寻找邻近用户。实验结果表明该算法比传统的协同过滤推荐方法预测效果更好。在传统的协同过滤推荐方法中,由于数据稀疏,用户共同评分的物品数很少,用户相似度为0的情况比较多,而该方法中的信任度则不然,它通过用户间社会关系构成的社会网络的最大传播路径以及用户间的最短路径计算可得。只要用户与目标用户之间存在一条路径,便可将其作为目标用户的邻近用户。结果表明,该方法的预测效果比传统的协同过滤推荐要好;然而,该算法却忽略了信任度为0,即用户之间不存在路径,但是相似度却很大的潜在用户。Yu等^[3]结合利用相似度和信任度的加权取值(文中称混合相似度)替代传统协同过滤推荐中的相似度作预测。O'donovan等^[4]将与目标用户之间的信任度超出阈值且相似度大于0的用户作为邻近用户集,并利用信任度和相似度的乘积与其二者之和的比值来替代传统协同过滤推荐中的相似度去预测未知评分。虽然文献[3]与文献[4]都综合考虑了用户间的相似度与信任度,但当不存在用户对目标物品有评分时,未知评分仍然也都得不到预测。Victor等^[8]将与目标用户之间的信任度大于0或相似度大于0的用户都作为其邻近用户集。若信任度大于0(即用户之间存在路径),则优先利用信任度替代传统协同过滤推荐中的相似度预测评分。它考虑了Massa等未考虑的潜在用户,但也存在与文献[3]与文献[4]类似的问题。

普通用户对物品标注的标签信息已经被广泛用于协同过滤推荐。Nakamoto等^[9]根据用户对物品共同标记的标签的情况,计算用户之间的相似度及用户偏好,并利用协同过滤推荐技术进行推荐。Zhao等^[10]根据用户对物品标注的标签信

息,首先利用WordNet计算标签的语义相似度,从而得到用户之间的相似度,然后再进行协同过滤推荐。

2 基于景点标签的协同过滤推荐

本章首先提出了基于物品及物品标签的协同过滤推荐算法,然后针对基于用户社会关系的协同过滤推荐算法^[1-4]有时不能给出对目标物品评分的情况,提出了基于用户社会关系和物品标签的协同过滤推荐算法。

2.1 基于物品及物品标签的协同过滤推荐

该算法在基于物品的协同过滤推荐中引入专家给物品标注的物品类型标签。利用标签信息使得景点相似度的计算更准确。

2.1.1 物品相似度的计算

1) 构建用户-景点评分矩阵。

利用用户、景点、以及用户对景点的评分信息构建评分矩阵 R :

$$R_{p,q} = \begin{bmatrix} r_{11} & r_{21} & \cdots & r_{1q} \\ r_{21} & r_{22} & \cdots & r_{2q} \\ \vdots & \vdots & & \vdots \\ r_{p1} & r_{p2} & \cdots & r_{pq} \end{bmatrix} \quad (1)$$

其中: p 为用户数, q 为景点数, r_{ij} 为用户 i 对景点 j 的评分,矩阵的第 i 行表示用户 i 的评分向量。若用户对景点未评过,则评分被赋值为0。

2) 计算景点之间的相似度。

①在目标用户曾经评过分的景点中,找到与目标景点标注有共同类型标签的景点集 I 。例如:若目标景点的标签为“古迹”“建筑”,则在用户曾经评过分的景点中找出所有被标注有“古迹”或“建筑”标签的景点。

②利用式(2)计算目标景点与 I 中的景点之间的相似度,找出与目标景点相似度大于0的相似景点集 I' :

$$\text{sim}(s, t) = \frac{\sum_{u \in U} (r_{us} - \bar{r}_s)(r_{ut} - \bar{r}_t)}{\sqrt{\sum_{u \in U} (r_{us} - \bar{r}_s)^2} \sqrt{\sum_{u \in U} (r_{ut} - \bar{r}_t)^2}} \quad (2)$$

其中: s 为目标景点, $t \in I$, U 为对景点 s 和 t 都有评分的用户集合, $u \in U$, r_{us} 为用户 u 对景点 s 的评分, \bar{r}_s 为用户集 U 中的用户对景点 s 的平均评分。

2.1.2 评分预测

1)若 I' 存在,利用式(3)计算目标用户 u 对目标景点 s 的评分:

$$\text{pred}(u, s) = \frac{\sum_{i \in I'} \text{sim}(s, i) \cdot r_{ui}}{\sum_{i \in I'} \text{sim}(s, i)} \quad (3)$$

2)若 I' 不存在,则不作评分预测。

2.2 基于用户社会关系和物品标签的协同过滤推荐

本节提出的基于用户社会关系和物品标签的协同过滤推荐算法是针对基于用户社会关系的协同过滤推荐算法^[8]有时不能计算目标用户对目标景点的评分时的改进。

2.2.1 邻近用户的选取

1)构建用户-景点评分矩阵(与2.1.1节中的方法一致)。

2) 计算用户之间的相似度。

根据上述评分矩阵,利用式(4)计算目标用户与其他用户之间的相似度:

$$\text{sim}(u, v) = \frac{\sum_{s \in S} (r_{us} - \bar{r}_u)(r_{vs} - \bar{r}_v)}{\sqrt{\sum_{s \in S} (r_{us} - \bar{r}_u)^2} \sqrt{\sum_{s \in S} (r_{vs} - \bar{r}_v)^2}} \quad (4)$$

其中: S 为用户 u 和 v 共同评过分的景点集合 $s \in S$; r_{us} 为用户 u 对景点 s 的评分 \bar{r}_u 为用户 u 的对其已评分景点的平均评分。

3) 构建社会网络。

利用用户之间的关注与被关注的社会关系信息,构建社会网络。

4) 计算用户之间的信任度。

利用构建的社会网络的最大传播路径与用户之间的最短路径,计算用户之间的信任度,公式如下:

$$\text{trust}(u, v) = \frac{d_{\max} - n + 1}{d_{\max}}; \quad d_{\max} = \left\lceil \frac{\ln(m)}{\ln(k)} \right\rceil \quad (5)$$

其中: m 为社会网络中用户的总数 k 代表用户的平均度数。如果用户 A 不能到达用户 B ,则 B 对 A 就没有信任价值,但若 A 能在最大传播距离之内到达 B ,则意味着 B 对 A 具有一定的信任度。

5) 寻找目标用户的邻近用户集 N 。

将与目标用户相似度大于 0 或者信任度大于 0,且对目标景点有评分的用户作为目标用户的邻近用户集 N 。

2.2.2 评分预测

1) 若 N 存在,利用以下预测公式计算评分^[8]:

$$\text{pred}(u, s) = \bar{r}_u + \frac{\sum_{v \in M} \text{sim}(u, v)(r_{vs} - \bar{r}_v) + \sum_{v' \in Z} \text{trust}(u, v')(r_{v's} - \bar{r}_{v'})}{\sum_{v \in M} \text{sim}(u, v) + \sum_{v' \in Z} \text{trust}(u, v')} \quad (6)$$

其中: u 为目标用户; s 为目标景点; M 为邻近用户集 N 中相似度大于 0 且信任度为 0 的用户集; v 为 M 中的一个用户; Z 为邻近用户集 N 中信任度大于 0 的用户集; v' 为 Z 中的一个用户; r_{vs} 为用户 v 对景点 s 的评分。

2) 若 N 不存在,则按照如下预测评分。

① 在用户曾经评过分的景点中,找到包含目标景点标签的景点集 T 。

② 若 T 存在:

a) 利用皮尔森相关系数计算目标景点与 T 中的景点之间的相似度,找出与目标景点相似度大于 0 的景点集 T' 。景点间的相似度计算如式(2)。

b) 若 T' 存在,利用基于物品及物品标签的协同过滤推荐算法计算目标用户 u 对目标景点 s 的评分,计算公式如式(3)。

③ 若 T 不存在或 T' 不存在:

a) 从所有景点中找到标有目标景点标签的景点集合 T^* 。

b) 若 T^* 不存在,则不对其进行评分预测。

c) 若 T^* 存在,则利用式(2)计算目标景点与 T^* 中的景

点之间的相似度,找出与目标景点相似度最大且大于 0 的景点 s' 。若 s' 不存在,则不对其进行评分预测。若 s' 存在,则再次利用基于用户社会关系及景点标签的协同过滤推荐算法预测目标用户对景点 s' 的评分,并将此评分作为目标用户对目标景点的评分。

3 实验方案设计

3.1 实验数据集

本文采用爬虫技术获取某旅游服务网站^[11]网页数据,通过内容析取得到需要的实验数据。获取的数据结构如图 1 所示,其中:圆圈表示用户,正方形表示景点,长方形表示标记景点类型的标签;用户之间的关注与被关注的社会关系构成了一个社会网络,它可用一个有向图表示。若存在一条边由用户 C 指向用户 A ,则表示 C 关注了 A ;图中带权的边表示用户对相应的景点的评分,边上的权值即为评分值;若景点与标签之间存在一条边,则表示该景点被标记了该标签。

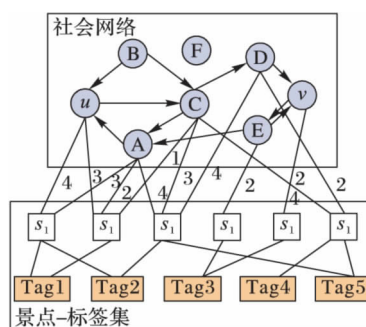


图1 实验数据结构

本实验获取的数据,包括中国五个城市(桂林、上海、厦门、杭州和香港)的 998 个景点、对这些景点有评分的 9202 个用户、用户对这些景点的 22 502 条评分记录、132 个景点类型标签、以及由 15 375 个用户和 26 931 条边构成的一个社会网络。其中 9 202 个用户也在这个社会网络中,由他们之间形成的边数为 10 471 条。

其中获取的景点标签信息如图 2 所示。

1	2888#地貌#河流#江河峡谷瀑布
2	5925#湖泊#河流#江河峡谷瀑布
3	14892#地貌#度假村#乡村风光#徒步
4	18147#洞穴#地貌#山岳/山岭
5	2881#公园
6	2886#山岳/山岭#古迹

图2 标签信息

如图 2 所示,一条记录代表一个景点的标签信息,其中的每个元素用“#”分隔,如“A#B#C#D”,即代表景点 A 的标签有 B、C、D 三个。如图中第一条记录表示:ID 为 2 888 的景点的标签有“地貌”、“河流”与“江河峡谷瀑布”。而与其标注有共同标签的景点有 5 925、14 892 与 18 147。

3.2 数据预处理

为了方便实验,对数据进行以下处理:

1) 从原始数据中提取本实验所需的数据,并加以整合。

2) 删除重复评分记录,选择用户对景点的最近评分作为该用户对该景点的评分。

3) 删除只有 1~2 条评分记录的用户。

4) 删除无用户评分的景点。

经过数据预处理后, 最终的实验数据集包括 988 个景点、1015 个用户、用户对景点的 12339 条评分记录, 以及由 15375 个用户和 26931 条边构成的一个社会网络, 之前的 1015 个用户也包含其中。数据的稀疏程度为 1.23%。

3.3 评测指标

根据常用的推荐系统性能评测指标以及大多数社会推荐系统研究论文^[1-2, 4, 8, 12]中利用的评测指标, 本文采用以下 5 个指标对算法进行性能评估。

1) 均方根误差 (Root-Mean-Square Error, RMSE): 衡量预测评分与真实评分之间的偏差。其计算公式如下:

$$RMSE = \sqrt{\frac{\sum_{(u_i, s_j) \in E} (r_{ij} - \bar{r}_{ij})^2}{|E|}} \quad (7)$$

2) 平均绝对误差 (Mean-Absolute Error, MAE):

$$MAE = \frac{1}{|E|} \sum_{(u_i, s_j) \in E} |r_{ij} - \bar{r}_{ij}| \quad (8)$$

其中: 为了更准确地反映预测出的评分的准确性, 将 E 定义为算法能够给出预测评分的样本集, $|E|$ 为该样本集的大小; (u_i, s_j) 表示一个目标用户与目标景点对; r_{ij} 为用户 u_i 对景点 s_j 的真实评分; \bar{r}_{ij} 为用户 u_i 对景点 s_j 的预测评分。

3) 覆盖率 (coverage)^[13]。能够计算预测评分的样本数占总样本数的比例。在本文中, 若未知评分能够预测, 则说明预测评分不为 0, 若不能预测则将预测评分设为 0, 假设预测评分不为 0 的样本条数为 k , 总样本条数为 M , 那么其计算公式如下:

$$coverage = k/M \quad (9)$$

4) 准确率 (precision)。本文利用文献[14]中将均方根误差映射为准确率的方法, 将其计算方法描述如下:

$$precision = 1 - (RMSE/4) \quad (10)$$

5) F-度量 (F-Measure)^[14]。因有时准确率和覆盖率二者会相互矛盾, 因此本文还采用了 F-度量这一个综合的评价指标, 其计算公式如下:

$$FMeasure = \frac{2 \times precision \times coverage}{precision + coverage} \quad (11)$$

3.4 对比实验设计

本文采取大多数社会推荐系统研究中用到的留一交叉验证方法 (leave-one-out cross validation)^[1-2, 14]对算法进行实验验证, 它循环利用整个数据集中的一条数据作为测试集, 其他数据作为训练集, 综合每次循环的预测结果计算各评测指标。同时, 为验证该方法的有效性, 还采用了本文相关工作部分提到的其中三种算法与提出的方法进行实验对比。本实验中的算法都采用最大传播路径方法计算信任度, 采用皮尔森相关系数度量用户或景点间的相似度。根据实验数据集中的社会网络信息, 利用式(5)计算得: 最大传播路径长度的值为 18。对比实验用到的算法如下:

1) 基于物品的协同过滤 (Item-based Collaborative Filtering, Item CF)。本文实现的关于文献[1]中的基于物品

的协同过滤推荐算法。其中相似度的度量方法采用皮尔森相关相似度。

2) 基于物品及物品标签的协同过滤 (Item and Tag-based Collaborative Filtering, ItemTag CF)。本文提出基于物品与物品标签的协同过滤推荐算法。

3) 基于用户的协同过滤 (user-based Collaborative Filtering, CF Pearson)。本文实现的关于文献[1]中的基于用户的协同过滤推荐算法。

4) MoleTrust (Trust-aware Recommendation)。本文实现的关于文献[2]提出的基于信任度意识的协同过滤推荐社会推荐算法。

5) EnsembleTrust (Ensembled Trust based Collaborative Filtering)。本文实现的关于文献[8]中的基于信任度整合的协同过滤推荐社会推荐算法。

6) Social RT (Social Recommendation based on the Tags of Scenic Spots)。本文基于文献[8]中算法提出的基于用户社会关系与物品标签的协同过滤推荐算法。

由于本实验采集的数据信息有限, 无法按照文献[3]与[4]中的信任度的计算实现信任度的计算, 因此本文未将它们列入对比实验。

4 实验结果与分析

最终的实验结果如表 1 所示。由于本实验数据获取的是来自几个比较热门的城市景点, 在数据预处理中删除了无用户评分的景点以及对所有景点只有 1~2 个评分记录的用户, 因此所有算法得到的覆盖率值相对来说都比较高。

表 1 各算法在各评测指标上的结果

方法	均方根误差	平均绝对误差	覆盖率	准确率	F-度量
Item CF	2.0124	1.5204	0.9974	0.4969	0.6633
ItemTag CF	1.7218	1.3678	0.7720	0.5695	0.6554
CF Pearson	1.2446	0.8204	0.9539	0.6889	0.8000
MoleTrust	1.1085	0.7466	0.9646	0.7229	0.8264
EnsembleTrust	0.9334	0.6609	0.9864	0.7666	0.8627
Social RT	0.8200	0.6085	0.9992	0.7950	0.8855

通过比较 Item CF 和 ItemTag CF 这两个算法可知: 使用专家标注的景点类型标签信息有效地降低了算法的预测评分误差。这说明利用景点类型标签能更合理、准确地对目标用户对目标物品的评分做出预测。这是因为利用标签过滤景点后, 利用与目标景点有共同标签的其他景点的评分作为参考, 将使得预测评分更加客观、合理。当然, 由于标签的使用, 减少了可用于评分参考的景点数目。因此, 相对于 Item CF 算法, ItemTag CF 算法的预测覆盖率也降低了。

通过比较 CF Pearson、MoleTrust 和 EnsembleTrust 三个算法可知: 利用用户之间的社会关系, 的确有效地提高了各项评测指标。这是因为利用用户之间的社会关系使得邻近用户集的选取更加合理。这再一次验证了文献[2]和文献[8]中的研究结论。

根据表 1 中的实验结果可知: 本文提出的算法 Social RT 在各方面的性能指标值都比其他对比算法都要高。可见, 景点标签信息的引入, 使得基于用户社会关系的协同过滤推荐算法不仅有效地提高了对旅游景点评分预测的覆盖率, 还减少了预测评分的误差。表 2 列出了各个算法不能给出预测评分的具体分析。

表 2 不能预测的原因分析表

方法	不能预测数目	原因
Item CF	32	3 条用户曾经评过分的景点与目标景点间的相似度与评分的乘积之和为 0; 29 条用户曾经评过分的景点与目标景点之间的相似度之和为 0
ItemTag CF	2 810	目标用户曾经评过分的景点中, 不存在标记有目标景点标签的景点
CF Pearson	568	419 条无相似度大于 0 的邻近用户; 149 条无对测试景点有评分的用户
MoleTrust	437	288 条无信任度大于 0 的邻近用户; 149 条无对测试景点有评分的用户
EnsembleTrust	168	19 条无信任度或相似度大于 0 的邻近用户; 149 条无用户对测试景点有评分
Social RT	10	3 条无相似景点; 7 条对相似景点有评分的用户中不存在信任度或相似度大于 0 的邻近用户

以基于用户的协同过滤(CF Pearson) 推荐算法为例: 在 12 339 条数据记录中, 存在着 568 条无法给出预测评分的记录, 其中 149 条是因为邻近用户对测试景点没有评分。另外 419 条中的目标用户不存在邻近用户。本文提出的算法 Social RT 虽然也存在着 10 条记录不能预测, 其中有 3 条是因为找不到与目标景点最相似的其他景点, 7 条是因为对相似景点有评分的用户中, 不存在与目标用户间的信任度或相似度大于 0 的邻近用户。因此, 本文提出的算法较好地改善了其他算法不能给出预测评分的状况, 并能在提高覆盖率的同时, 提高预测准确率, 减少预测评分误差。

5 结语

本文针对基于用户社会关系的协同过滤推荐算法有时无法给出目标用户对目标景点的评分, 以及在基于物品的协同过滤推荐算法中, 用户对不同类型物品评分的准确性或者可比性会影响物品之间相似度的计算这两个问题, 提出了基于物品及物品标签的协同过滤推荐算法和基于用户社会关系和物品标签的协同过滤推荐算法。通过从携程网上采集关于中国五个城市(桂林、上海、厦门、杭州和香港) 的 988 个景点、1 015 个用户对这些景点的评分信息及 15 375 个用户之间的社会关系及携程网给景点标注的类型标签信息, 设计了对比实验, 并进行了留一交叉验证。实验结果表明: 景点类型标签信息的引入使得景点的相似度计算更准确, 有效地提高了基于用户社会关系的协同过滤推荐算法的准确率和覆盖率, 以及基于物品的协同过滤推荐算法的准确率。

未来的工作包括: 1) 结合文献[3-4]中信任度的计算方法, 获取相应的旅游数据实现这些文献中的算法, 并根据旅游数据的特点对信任度的计算方法进行改进; 2) 获取并分析其他更多类型的旅游数据, 如游记、旅游照片等, 提取更多的信息以更准确地计算用户之间的相似度。

参考文献:

[1] MASSA P, AVESANI P. Trust-aware collaborative filtering for recommender systems [C]// Proceedings of the 2004 Conference on Cooperative Information Systems. Berlin: Springer, 2004: 492-508.

[2] MASSA P, AVESANI P. Trust-aware recommender systems [C] // Proceedings of the 2007 ACM Conference on Recommender Systems. New York: ACM Press, 2007: 17-24.

[3] YU X, WANG Z. A enhanced trust model based on social network and online behavior analysis for recommendation [C]// Proceedings of the 2010 International Conference on Computational Intelligence and Software Engineering. Piscataway: IEEE Press, 2010: 1-4.

[4] O'DONOVAN J, SMYTH B. Trust in recommender systems [C]// Proceedings of the 10th International Conference on Intelligent User Interfaces. New York: ACM Press, 2005: 167-174.

[5] XIANG L. Practice of recommendation system [M]. Beijing: Posts & Telecom Press, 2012. (项亮. 推荐系统实践[M]. 北京: 人民邮电出版社, 2012.)

[6] SARWAR B, KARYPIS G, RIEDL J. Item-based collaborative filtering recommendation algorithms [C]// Proceedings of World Wide Web Conference Series. New York: ACM Press, 2001: 285-295.

[7] GOLDBERG D, NICHOLS D, OKI B M, et al. Using collaborative filtering to weave an information tapestry [J]. Communications of the ACM, 1992, 35(12): 61-70.

[8] VICTOR P, CORNELIS C, COCK M D, et al. A comparative analysis of trust-enhanced recommenders for controversial items [C]// Proceedings of the 3rd International Conference on Weblogs and Social Media. Palo Alto: AAAI Press, 2009: 342-345.

[9] NAKAMOTO R, NAKAJIMA S, MIYAZAKI J, et al. Tag-based contextual collaborative filtering [J]. IAENG International Journal of Computer Science, 2007, 34(2): 214-219.

[10] ZHAO S, DU N, NAUERZ A, et al. Improved recommendation based on collaborative tagging behaviors [C]// Proceedings of the 13th International Conference on Intelligent User Interfaces. New York: ACM Press, 2008: 413-416.

[11] Ctrip. Ctrip.com [EB/OL]. [2014-02-12]. <http://www.ctrip.com>.

[12] TANG J, HU X, LIU H. Social recommendation: a review [J]. Social Network Analysis and Mining, 2013, 3(4): 1113-1133.

[13] VICTOR P, COCK M D, CORNELIS C. Trust and recommendations [M]// KANTOR P B, ROKACH L, RICCI F, et al. Recommender Systems Handbook. Berlin: Springer, 2011: 645-675.

[14] JAMALI M, ESTER M. Trustwalker: a random walk model for combining trust-based and item-based recommendation [C]// Proceedings of 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2009: 397-406.