doi:10.11959/j.issn.1000-436x.2015130

基于一跳信任模型的协同过滤推荐算法

王兴茂，张兴明，邬江兴
（国家数字交换系统工程技术研究中心，河南 郑州 450002）

摘 要：基于社会信任网络的协同过滤推荐算法存在节点之间多下一跳带来的复杂路径选择和信任弱传递问题。针对这 2 个问题，给出基于项目的一跳信任模型，该模型通过用户对项目信任度的计算，定义用户的直接和间接社会信任属性，然后一步跳转计算用户之间的直接和间接信任距离，进而计算用户之间的信任度。基于此模型设计推荐算法，同时分析了信任度与传统相似度的理论关系并二维拟合。仿真实验表明，该算法提高了推荐准确度(约 0.02 MAE)，降低了训练时间(约 50%)。
关键词：推荐算法；一跳信任模型；信任距离；信任度
中图分类号：TP393 **文献标识码**：A

Collaborative filtering recommendation algorithm based on one-jump trust model

WANG Xing-mao, ZHANG Xing-ming, WU Jiang-xing
(National Digital Switching System Engineering and Technological R&D Center, Zhengzhou 450002, China)

Abstract: A collaborative filtering recommendation algorithm based on the trust network of social brings two problems that the choice of complex paths between nodes and the weak transfering of trust. Toward to these two problems, a one-jump trust model based on items was put forward, the model calculated the trust between users and items, defined the consumer's trust attribute vector of social and calculated the direct and indirect distance one-jump by items, and then calculated the trust between users. A collaborative filtering algorithm(OneJ-TCF) is degined based on the model, moreover analysed and reorganized the relation between trust and similarity. The experiments show that this algorithm improves the degree of accuracy(reducing about 0.02 MAE), and saves about 50% training time at the same time.
Key words: recommendation algorithm; one-jump trust model ; trust distance; trust

1 引言

推荐系统能够根据用户的偏好进行推荐，这种能力缓和了“信息爆炸”加重的传统广告式的“广播骚扰”，已经成为学术研究的一个热点[1~3]。协同过滤是推荐系统中应用最广泛的推荐算法[4]，但随着互联网的爆炸式扩张，传统的协同过滤推荐系统普遍存在数据稀疏性的问题。将社会网络中人与人之间的信任关系应用到推荐系统中[5]，能够有效地缓解数据稀疏性问题。

Polo Massa 等[6]最早开始对协同过滤推荐系统中的信任问题进行研究，“撬开”了推荐系统中信任研究的大门。Avesani 等[7]基于社会信任网络，采用一定长度的路径值来计算目标用户和其他用户之间的信任值；Yuan 等[8]将社会中的朋友关系、用户参加的群组信息及用户选择项目的信息联合构建 3 种节点类型的信任网络图，采用基于图的随机游走方法产生推荐结果；Jebrin 等[9]采用用户之间的信任信息和用户对项目的评分信息来对用户进行推荐；Ma 等[10]提出将社会信任作为一个推荐的约束，

收稿日期：2014-06-04；修回日期：2014-11-25
基金项目：国家重点基础研究发展计划(“973”计划)基金资助项目(2012CB315901)；国家高技术研究发展计划(“863”计划)基金资助项目(2011AA01A103)
Foundation Items: The National Basic Research Program of China (973 Program) (2012CB315901); The National High Technology Research and Development Program of China (863 Program)(2011AA01A01)

使用概率因子分析的方式来结合用户与其所信任的用户间的信任关系来进行推荐。通过对这些文献的研究和分析，这种基于社会网络中信任机制的推荐算法存在共同的2个问题：一是当中间用户节点繁多时，存在复杂路径的选择问题；二是信任具有弱传递性，经过多次传递会带来准确度降低的问题。究其根本原因是计算模型[11]的问题，如图1所示，图中目标用户节点A和节点B之间存在很多中间节点，两节点在建立信任度的时候就需要进行复杂的中间路径选取，同时考虑到信任具有弱传递性，在多次传递后建立的信任很可能会存在信任度计算偏差过大的问题。

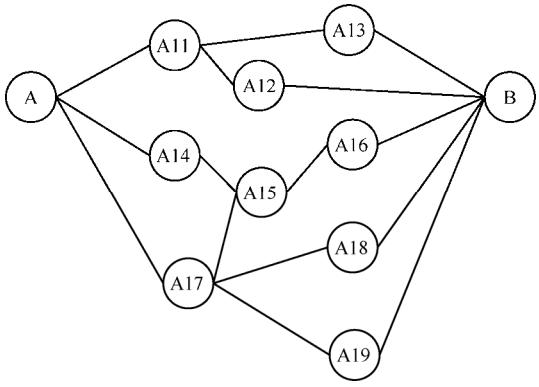

图1 基于社会信任网络节点多跳的信任模型

针对这2个问题，本文对模型进行改进，提出一种新的信任模型——基于项目的一跳信任模型，该模型没有采用传统的基于用户节点的多下一跳方式计算信任度，而是以系统中项目整体为中间节点，一步跳转计算用户之间信任度。本文在此模型的基础上提出基于一跳信任模型的协同过滤推荐算法(OneJ-TCF)，本算法主要分为3步：首先通过用户对项目信任度的计算获取用户的社会信任属性向量；然后通过直接社会信任属性和间接社会信任属性来计算他们之间的直接信任和间接信任距离；最后计算用户之间信任度并和用户之间相似度进行二维拟合。

在MovieLens 100k数据集上的仿真结果表明，OneJ-TCF明显比传统的基于信任的协同过滤推荐算法有更高的推荐准确度，推荐准确度指标MAE提升约0.02，而且训练时间降低了50%左右。

2 基于一跳信任模型的协同过滤推荐算法

2.1 一跳信任模型的设计及相关概念

针对社会网络节点多跳的信任模型存在的2个问题，本文提出一种新的信任模型，在此模型的基础上建立用户之间新的信任机制。为了更好地描述本文模型和便于后文计算说明的简洁和形象，先进行相关概念的说明。

2.1.1 相关概念

信任是目标节点A对节点B能够按照其预想完成任务的行为和能力的综合期望值，分为直接信任和间接信任。具体相关定义如下。

1) 用户对项目的信任：信任是交易中建立的一个主观概念，本文采用用户对项目的评分建立用户对项目的信任，即用户A对某项目j的评分大小象征了一种信任程度，用$tr_{ij} \in [0,1]$表示信任程度(i为用户A的索引号)，简称信任度，tr_{ij}越大，表明用户对该项目越信任。

2) 用户对项目的直接信任：如果用户A对项目j存在直接评分，则用户A对该项目存在直接信任，用$trD^{A\to j}$表示。

3) 用户对项目的间接信任：如果用户A对项目j不存在直接评分，则用户A对该项目存在间接信任，用$trI^{A\to j}$表示。

4) 用户社会信任属性向量：系统中每个项目象征了某种特殊的属性，可以看作用户的一个社会属性，即用户对一个项目越信任，表示用户的这个社会属性越强，由系统中n个项目可以构建出用户A的n维社会信任属性向量$(tr_{i1},tr_{i2},tr_{i3},\cdots,tr_{in})$，$i$为用户A的索引号。

5) 直接信任属性：如果用户A的社会信任属性向量中的某一属性tr_{ik}(i为A的索引号，k为用户A的第k维属性)是由用户对项目直接信任计算而来，则称这一属性为直接信任属性。

6) 间接信任属性：如果用户A的社会信任属性向量中的某一属性tr_{ik}是由用户对项目间接信任计算而来，则称这一属性为间接信任属性。

7) 用户之间的信任距离：用户A和用户B的社会信任属性向量分别看作多维空间上的一个点，计算出用户A和B这2个点之间的加权欧几里得距离为2个用户之间的信任距离$D_{A\to B}$。

8) 直接信任距离：2个用户A和B由直接信任属性计算出的距离为直接信任距离。

9) 间接信任距离：2个用户A和B由间接信任属性计算出的距离为间接信任距离。

10) 信任矩阵：系统中m个用户对n个项目信任度的矩阵，文中用TR表示。

2.1.2　模型的设计

基于上述概念说明，本文提出的基于项目的一跳信任模型如图 2 所示。在该模型中用户 A 和 B 通过系统中项目建立各自的社会信任属性向量，然后通过社会信任属性向量中的 k 个共同的直接信任属性和$(n-k)$个非共同信任属性来计算它们之间的直接信任距离和间接信任距离。直接信任距离表征了用户之间直接信任，间接信任距离表征了用户之间的间接信任，通过两者的加权计算用户 A 和用户 B 的信任度。该模型可以形象地看作用户 A 和用户 B 通过系统中项目一跳建立用户之间的信任度。

2.2　算法的相关计算

本节在上节信任模型的基础上提出了基于一跳信任模型的协同过滤推荐算法，主要计算步骤为用户对项目信任度的计算、用户之间信任度计算以及信任度和传统的相似度二维拟合。

2.2.1　用户对项目的信任度计算

设用户对项目的评分矩阵为 $\boldsymbol{R}$，是一个 m 个用户对 n 个项目的评分矩阵，如式(1)所示。

$$\boldsymbol{R}=\begin{pmatrix} r_{11} & \cdots & r_{1n} \\ \vdots & \ddots & \vdots \\ r_{m1} & \cdots & r_{mn} \end{pmatrix} \tag{1}$$

矩阵元素 r_{ij} 表示用户 i 对项目 j 的评分，不经特殊说明，下文中的 m 都表示系统中用户数，n 都表示系统中项目数。

本文引入信任模型中的模糊集合理论[12]来定义并计算用户对项目的信任度，为了便于表达，首先引入隶属度函数的定义。

隶属度函数定义：设 U 是论域，称映射 $\mu_A: U \to [0,1], x \mapsto u_A(x) \in [0,1]$ 确定了一个 u 上的模糊子集 A，映射 μ_A 称为 A 的隶属函数，$\mu_A(x)$ 成为 x 对 A 的隶属度。

这样，把论域定为用户对项目的信任集，信任等级可以表述为 U 上的多个模糊子集，本文模糊子集划分为：“完全不信任”子集、“一般信任”子集，“信任”子集、“非常信任”子集、“完全信任”子集。

用户对项目的信任度的计算：一般用户对项目的评分区间为[0,5](可以根据实际情况按比例调整)，设用户的评分为 r，信任度 $tr=\frac{r}{5}$，$tr\in[0,1]$，tr 的值越大表明用户对该项目的信任度越强[13]。$0\le tr<0.2$ 对应“完全不信任”模糊子集、$0.2\le tr<0.4$ 对应“一般信任”子集、$0.4\le tr<0.6$ 对应“信任”子集、$0.6\le tr<0.8$ 对应“非常信任”子集、$0.8\le r\le 1$ 对应“完全信任”子集。

上述信任度的计算可以看作用户对项目直接信任度的计算，但用户项目评分矩阵很稀疏，所以单一的直接信任不能满足计算需求，用户之间的间接信任尤为重要，下面给出用户对项目直接信任度计算和间接信任度计算过程。

用户对项目直接信任度的计算如下。

如果用户 A 对项目 j 存在直接评分为 r_{ij} (i 为用户 A 的索引号)，则用户 A 对该项目存在直接信任，直接信任度的计算如式(2)所示。

$$trD^{\mathrm{A}\to j}=\frac{r_{ij}}{5} \tag{2}$$

用户对项目间接信任度的计算如下。

间接信任的计算是基于这样一个事实，如果用

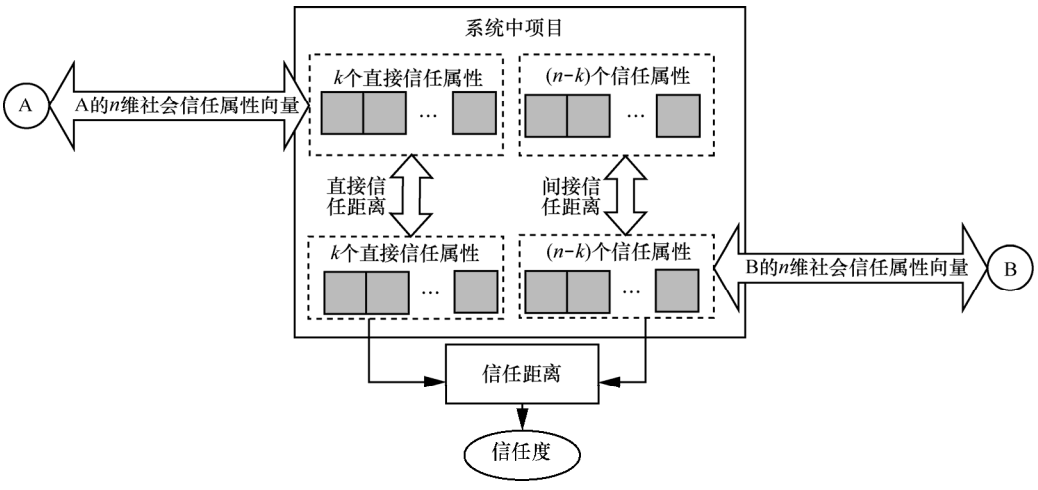

图 2　基于项目的一跳信任模型

户 A 对项目 j 信任存在直接信任, 若用户 A 未评分的项目 k 与项目 j 很相似，那么用户 A 对项目 k 存在间接信任[14]。

设 $i_1,i_2,i_3,\cdots,i_k$ 为用户 A 评价过的 k 个项目，A 对项目 $i_1,i_2,i_3,\cdots,i_k$ 的信任度为 $trD_{i_1},trD_{i_2},trD_{i_3},\cdots,trD_{i_k}$，这是用户对项目的直接信任。设 j 为用户 A 未直接评价过的项目，项目 j，$j\notin\{i_1,i_2,i_3,\cdots,i_k\}$ 与项目 i_u 之间的相似度为 $Isim_{i_u j}$，则用户 A 对项目 j 的间接信任度计算如式(3)所示。

$$trI^{\mathrm{A}\to j}=\frac{\sum_{u=1}^{k}trD_{i_u}Isim_{i_u j}}{\sum_{u=1}^{k}Isim_{i_u j}} \tag{3}$$

$$Isim_{i_u j}=\cos(\vec{\boldsymbol{u}},\vec{\boldsymbol{v}})=\frac{\vec{\boldsymbol{u}}\times\vec{\boldsymbol{v}}}{\left\|\vec{\boldsymbol{u}}\right\|\times\left\|\vec{\boldsymbol{v}}\right\|} \tag{4}$$

其中，$\vec{\boldsymbol{u}}$ 为项目 i_u 的用户评分向量，$\vec{\boldsymbol{v}}$ 为项目 j 的用户评分向量。通过间接信任和直接信任的计算，可以得到用户对项目的信任矩阵 $\boldsymbol{TR}$，如式(5)所示。在这个矩阵中，如果用户 A 对项目 j 存在直接信任，则 $tr_{ij}=trD^{\mathrm{A}\to j}$，否则 $tr_{ij}=trI^{\mathrm{A}\to j}$，$i$ 为用户 A 的索引号，i=1,2,3,⋯,m。

$$\boldsymbol{TR}=\begin{pmatrix}tr_{11}&\cdots&tr_{1n}\\ \vdots&\ddots&\vdots\\ tr_{m1}&\cdots&tr_{mn}\end{pmatrix} \tag{5}$$

2.2.2 用户之间的信任度计算

由信任矩阵 $\boldsymbol{TR}$ 可以得出用户 A 的社会信任属性向量为 $(tr_{i1},tr_{i2},tr_{i3},\cdots,tr_{in})$，用户 B 的社会信任属性向量为 $(tr_{j1},tr_{j2},tr_{j3},\cdots,tr_{jn})$，$i$ 和 j 分别为用户 A 和 B 的索引号。将用户的社会信任属性向量看成 n 维空间的一个点，为了计算时突出直接信任距离和间接信任距离的权重，本文采用 2 个点之间的加权欧几里德距离来计算用户之间信任距离。

计算目标用户 A 与用户 B 的信任距离的规则。

规则 1 如果当第 k 维社会信任属性分量 tr_{ik} 与 tr_{jk}，$k\in(1,2,3,\cdots,n)$，分别为用户 A 和用户 B 对项目 k 的直接信任度时，那么这个分量之间的距离为直接信任距离。其他情况分量之间存在间接信任距离，设直接信任距离的权重为 p，间接信任距离的权重为 q。本文计算时令 p=0.6，q=0.4，经验证这样设置会带来最佳的效果。

规则 2 如果目标用户 A 的社会信任属性分量 $tr_{ik}=0$，$k\in(1,2,3,\cdots,n)$，那么 2 个用户之间的第 k 维分量信任距离为零。

设目标用户 A 和用户 B 之间存在 k 个直接信任属性，$n-k$ 个间接信任属性，则目标用户 A 与用户 B 的信任距离为 $D_{\mathrm{A}\to\mathrm{B}}$，计算如式(6)所示。

$$D_{\mathrm{A}\to\mathrm{B}}=\sqrt{\sum_{i=1}^{i=k}p(trD^{\mathrm{A}\to j_i}-trD^{\mathrm{B}\to j_i})^2+\sum_{i=k+1}^{i=n}q(trI^{\mathrm{A}\to j_i}-trI^{\mathrm{B}\to j_i})^2} \tag{6}$$

其中，$j_1,j_2,j_3,\cdots,j_k$ 为目标用户 A 和用户 B 之间的直接信任属性分量的索引号，$j_{k+1},j_{k+2},j_{k+3},\cdots,j_n$ 为目标用户 A 和用户 B 之间的间接信任属性分量的索引号，即式中 $\sum_{i=1}^{i=k}p(trD^{\mathrm{A}\to j_i}-trD^{\mathrm{B}\to j_i})^2$ 为直接信任距离计算，象征了用户之间的直接信任度。$\sum_{i=k+1}^{i=n}q(trI^{\mathrm{A}\to j_i}-trI^{\mathrm{B}\to j_i})^2$ 为间接信任距离计算，象征了用户之间的间接信任度。通过以上信任距离公式计算可知 $D_{\mathrm{A}\to\mathrm{B}}\neq D_{\mathrm{B}\to\mathrm{A}}$，满足现实中的信任不对称性。

从上述计算可以看到信任距离越大，用户与目标用户之间的信任度越小，这是一个反相关的关系。本文为了将信任度归一化到[0,1]区间内，对信任距离进行处理，得到 2 个用户之间的信任度为 $Tr^{\mathrm{A}\to\mathrm{B}}$，如式(7)所示。

$$Tr^{\mathrm{A}\to\mathrm{B}}=1/\left(1+D_{\mathrm{A}\to\mathrm{B}}\right) \tag{7}$$

2.2.3 信任度与相似度整合

通过前两小节的计算，可以看出本文的信任度是基于用户之间的信任距离，侧面地反映了 2 个用户之间具体评分差异，而没有考虑用户之间的总体评分趋势差异，为了弥补这一缺点，本文引入用户之间的总体评分趋势相似度与本文计算得出的信任度进行互补，用户总体评分趋势采用余弦相似性公式计算，计算如式(8)所示。

$$sim^{\mathrm{A}\to\mathrm{B}}=\cos(\vec{\boldsymbol{u}},\vec{\boldsymbol{v}})=\frac{\vec{\boldsymbol{u}}\times\vec{\boldsymbol{v}}}{\left\|\vec{\boldsymbol{u}}\right\|\times\left\|\vec{\boldsymbol{v}}\right\|} \tag{8}$$

其中，$\vec{\boldsymbol{u}}$ 和 $\vec{\boldsymbol{v}}$ 分别为用户 A 和用户 B 的评分向量。为了使信任度和相似度拟合更加精确，本文在拟合时，摒弃了以往的线性拟合的方式，而是将用户 A 和用户 B 的相似度与信任度联合看作二维平面的一个点 $(sim^{\mathrm{A}\to\mathrm{B}},Tr^{\mathrm{A}\to\mathrm{B}})$，改变了原有的单一直线上点的拟合，扩展一个信任维度，在二维的

平面上计算权重，这种计算方式会能够把原来降低维度后缺省的“个体评分”差异体现出来。这个点离原点的加权距离代表了用户的最终权重，计算如式(9)所示。

$$weight^{A\to B}=\sqrt{\alpha(sim^{A\to B})^2+\beta(Tr^{A\to B})^2}\ ,\alpha+\beta=1 \quad (9)$$

2.3　算法的流程

综合前面对信任度和相似度的计算，本小节给出该算法对目标用户 A 预测评分和进行推荐的流程。

输入：用户对项目的评分矩阵 ***R***。

Begin

step1　用户相似度计算。

根据余弦相似性计算公式计算目标用户 A 和其他用户 B 之间相似度 $sim^{A\to B}$。

step2　用户之间信任度的计算。

通过直接信任 $trD^{A\to j_i}$ 和间接信任 $trI^{A\to j_i}$ 计算用户对项目 j，$j\in\{1,2,3,\cdots,n\}$ 的信任度。

根据式(6)计算目标用户 A 和其他用户 B 之间的信任距离 $D_{A\to B}$。

通过公式 $Tr^{A\to B}=1/\left(1+D_{A\to B}\right)$ 计算目标用户 A 与其他用户 B 的信任度。

step3　最终权重，信任度与相似度整合。

通过权重公式 $weight^{A\to B}=\sqrt{\alpha(sim^{A\to B})^2+\beta(Tr^{A\to B})^2}$，$\alpha+\beta=1$ 计算用户之间的最终权重。

step4　产生推荐集

1) 挑选出权重最高的 k 个用户，生成一个集合，为邻居集。

2) 通过邻居用户对项目 j 的评分来预测目标用户 A 对为评分项目 j 的评分，预测评分的经典公式[14]如式(10)所示。

$$P_{A,j}=\overline{R_A}+\frac{\sum_{i=1}^{k}(R_i(j)-\overline{R_i})weight^{A\to i}}{\sum_{i=1}^{n}weight^{A\to i}} \quad (10)$$

其中，$P_{A,i}$ 表示用户 A 对项目 i 的预测评分，$weight^{A\to i}$ 表示用户 A 和用户 i 之间的权重，k 是用户 A 的近邻用户组中用户的个数，$\overline{R_j}$ 表示用户 A 项目评分的均值。

从评分列表中挑选出评分高的项目推荐给用户。

End

3　算法的性能分析

本文算法的主要目的是在数据稀疏情况下能够提高准确度，但一个算法的复杂度也是衡量这个算法的重要指标，所以本节进行准确度分析和算法复杂度分析。

3.1　准确度分析

本文的信任度计算的实质是基于用户之间的信任距离，侧面反映了 2 个用户之间的具体评分差异，而且在最终权重计算时又引入了用户之间的余弦相似性来弥补信任度计算时没有考虑到的用户之间的总体评分趋势差异。由于多维空间无法画出原图，本文只在二维空间进行图文说明，这两者如图 3 所示，用户 A 和用户 B 相似度表示向量 ***OA*** 和 ***OB*** 夹角 θ 的大小，信任度形象地表征了目标点 A 与向量 ***OB*** 上点 B 的距离 L，通过两者的共同“定位”能够更准确地找出与目标 A 接近的点 B，在寻找邻居时会更准确。

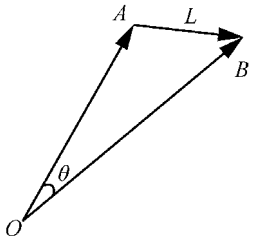

图 3　用户 A 和用户 B 之间相似度和信任度关系

为了进一步提高准确度，本文在相似度与信任度拟合时，摒弃了以往原有的单一直线上点的拟合，扩展一个信任维度，在二维的平面上计算权重，这种计算方式会更能够把原来降低维度后缺省的“个体评分”差异体现出来，所以在权重计算会更准确，使邻居用户的选择更加合理。

例如向量(1,2)与向量(2,4)，用传统的推荐算法中余弦相似度计算得到 2 个向量的相似度为 1，然而并没有考虑向量中每一维度的具体值，即用户的具体爱好程度。很显然这 2 个向量表示的用户对同一项目爱好的程度是不同的，相似度为 1 并不够精确。通过本文信任度计算得两者信任度为 0.309，最终权重为 0.680 3，更能符合实际用户之间的兴趣情况，即本文的算法从数学描述说明和推理的角度上，分析得出推荐准确度会有所提高。

3.2　复杂度分析

随着科技的不断发展，空间存储对算法的影响逐渐弱化。所以本节主要对时间复杂度进行分析和比较。本文的时间开销主要是项目相似度、用户对项目信任度、用户之间信任度计算。

1) 项目相似度计算

$Isim_{aj,i}$：项目之间相似度 //执行 $n(n+1)$次

2) 用户对项目的信任度计算

tr_{ij}：用户项目信任度，储存，构成用户的社会信任属性向量 //执行 mn 次

3) 用户之间信任度计算及最终权重计算

$D_{i\to j}$：用户之间的信任距离 //执行 $m(m+1)$次

Tr_{ij}：信任度 //执行 $m(m+1)$次

sim_{ij}：相似度 //执行 $m(m+1)$次

$weight_{ab}=\alpha sim_{ij}+\beta Tr_{ij}$ //执行 $m(m+1)$次

4) 为用户进行推荐 //执行 k(k 为邻居数)次

算法执行总次数

$$f(n)=4m^2+n^2+nm+5m+2n+k+1$$

计算得出时间复杂度为

$$T(n)=T(m^2)+T(n^2)+T(mn)+T(m)+T(n)+T(1)$$

$$T(n)=O(m^2)+O(n^2)+O(mn)$$

本文算法复杂度为 $O(m^2)+O(n^2)+O(mn)$，但在实际的系统中，项目数 n 相对于用户数 m 来说是比较固定的，而且成熟的推荐系统中项目数 n 要远小于用户数 m，所以 $T(n)\approx O(m^2)$。而基于多下一跳信任模型的协同过滤推荐算法的复杂度约为 $O(m^3)$，即本文的复杂度比节点多下一跳推荐算法的复杂度降低了一个级别，和传统的基于用户的协同过滤推荐算法的复杂度是一个量级的。

4 仿真实验及结果分析

4.1 仿真实验环境设置和评价指标

本实验是在基于 java 的 Eclipse 开发环境下进行的。为了验证本文算法的有效性，实验中采用 Grouplens 提供的 MovieLens 100k 数据集，该数据集为 943 个用户对 1 682 部电影的评分，评分范围为 1~5，稀疏度为 94.3%。通过对数据集的随机 2/8 分割，80%为训练数据，20%为测试数据，进行本文算法的仿真实验。

为了验证本算法的性能，本文采用准确度指标、半衰期效用指标[15]和时间指标。

1) 准确度指标

准确度采用平均绝对误差 *MAE* 来度量

$$MAE=\frac{1}{N}\sum_{i=1}^{N}|p_i-r_i| \tag{11}$$

其中，p_i 为算法预测的评分，r_i 为测试数据中的时间评分，N 为测试集中项目数。*MAE* 越小，推荐精度越高。

2) 半衰期效用指标

$$HLU=\sum_{i\in V(u,K)}\frac{\max(r_{ui}-d,0)}{2^{(l_{ui}-1)/(h-1)}} \tag{12}$$

其中，r_{ui} 为用户对列表中排名为 i 的项目的打分，$V(u,K)$ 为推荐列表，l_{ui} 为推荐列表中项目 i 的排名，h 为用户关注度为 50%的位置。半衰期效用指标度量的是推荐系统对一个用户推荐项目的排序准确性。

3) 时间指标

主要考虑算法的训练时间，因为推荐过程中训练占主导地位。

4.2 实验结果和分析

仿真实验首先对本文算法 OneJ-TCF 的权值 α 进行验证和选取，然后比较以下 3 种推荐算法的准确度和复杂度。

1) 传统基于用户的协同过滤推荐算法(相似度计算采用 pearson 相关系数)CF；

2) 基于传统信任模型(节点多下一跳)的协同过滤推荐算法 TCF；

3) 基于一跳信任模型的协同过滤推荐算法 OneJ-TCF。

4.2.1 权值的验证及选取

在实验中，针对本文相似度和信任度的权值 α 和 β，$\alpha+\beta=1$，在取值时根据实验结果和最小二乘法拟合，调整 α 的值，使效果达到最好。在仿真实验中令 α=0,0.2,0.4,0.5,0.6,0.7，调整邻居数 N 比较准确率指标 *MAE*。

如图 4 所示，针对不同的邻居数，通过取不同的 α 值，发现当 α<0.4 时，所有的 *MAE* 有下降趋势，当 α>0.5 时，*MAE* 又有上升趋势。经过多次实验，得出 α=0.42 时本文算法 OneJ-TCF 实验效果最好，所以以下实验 OneJ-TCF 都取 α=0.42。

4.2.2 准确度的仿真实验分析

图 5 展示了 3 种算法的 *MAE*，OneJ-TCF 的 *MAE* 要好于 CF 和 TCF。这与第 3 节的准确度分析相符合。

在邻居数为 5 时，3 种算法的 *MAE* 接近，分析原因是当邻居少时，其中的邻居用户与目标用户之间的相似度计算都较为精确，OneJ-TCF 虽然能够提高计算准确度，但效果相对不明显。

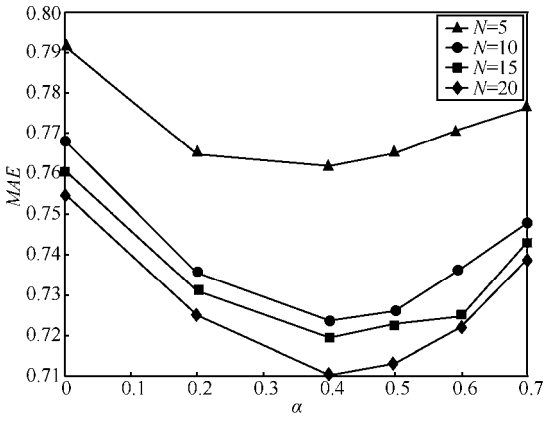

图 4　α 与准确度 *MAE* 的关系

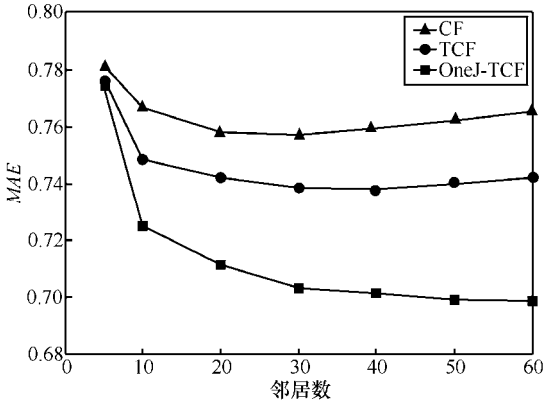

图 5　推荐准确度 *MAE* 对比

OneJ-TCF 随着邻居数的增多准确率逐渐提高，当邻居数大于 40 时，逐渐变平稳。而 CF 和 TCF 在邻居数为 30 左右时效果最好，当邻居数再增加时准确率会降低。分析这种现象，是因为 OneJ-TCF 在计算信任度时采用了基于项目一跳信任模型, 该模型在计算信任距离时对间接信任距离采用项目相似传递进行填充。这种计算方式在数据稀少时仍会取得较好的效果。所以当邻居数增加时尤其当邻居数大于 30 时, OneJ-TCF 的准确度仍然会有所提高，而 CF 和 TCF 的邻居列表中靠后的邻居会由于数据稀疏造成相似度计算不够精确, 所以准确度明显下降。

为了更好地度量本文算法的推荐质量，本文采用了排序准确度的度量指标 *HLU*。图 6 所示为 3 种算法的 *HLU*，由图中可以看出 OneJ-TCF 的 *HLU* 要高于 CF 和 TCF, 尤其当邻居数增大时, OneJ-TCF 的 *HLU* 指标明显优于其他 2 种算法。表明 OneJ-TCF 在对用户推荐时的推荐列表中用户喜欢的商品排序更加合理，推荐成功的可能性更大，这也侧面反映了该算法的推荐准确率。

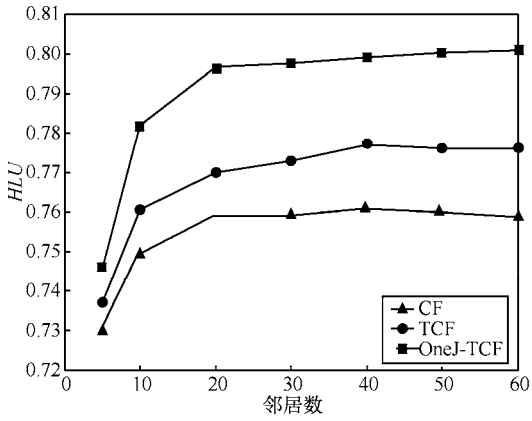

图 6　排序准确度 *HLU* 对比

4.2.3　复杂度的仿真实验分析

如图 7 给出 3 种算法随邻居数变化时训练时间的对比。可以看到 OneJ-TCF 的训练时间约为 CF 的 1.7 倍，OneJ-TCF 在信任度计算时确实会产生额外开销，但不符合第 3 节中的复杂度分析, 即 OneJ-TCF 和 CF 复杂度接近。分析产生这种训练时间差异过大现象的根本原因是本数据集中项目数是约为用户数的 2 倍，导致了 OneJ-TCF 比 CF 训练时间升高了约 70%，但实际推荐系统中用户数要多于项目数，OneJ-TCF 的训练时间会接近 CF 的训练时间。在都引入用户之间信任度的同时，OneJ-TCF 的训练时间要比 TCF 降低了 50%左右，这是因为在计算时降低路径选择的复杂度，降低了计算开销。

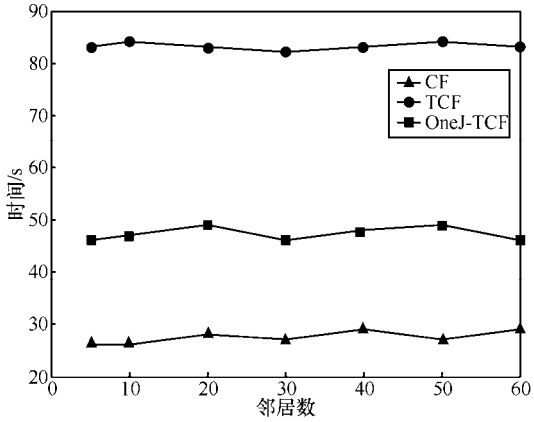

图 7　3 种算法训练时间对比(项目数约为用户数 2 倍)

为了验证实际中项目数低于用户数时 OneJ-TCF 与 CF 的时间复杂度，本文将数据集中项目数缩减至 800 个进行仿真，2 个算法的训练时间如图 8 所示。OneJ-TCF 只比 CF 训练时间提升了 1.5%左右，可以看到 2 个算法的时间复杂度接近，这也验证 3.2 节的时间复杂度分析。

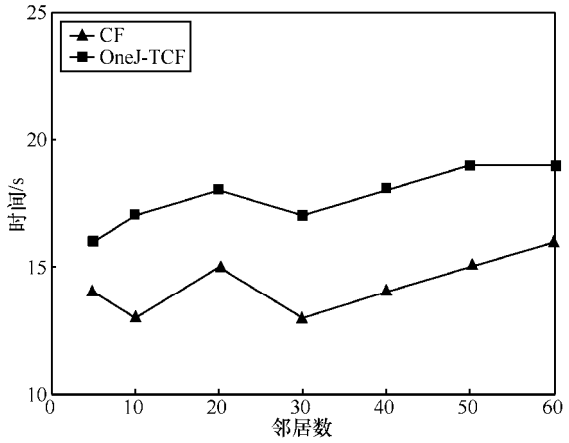

图 8 OneJ-TCF 与 CF 在项目数小于用户数时的训练时间对比

5 结束语

本文提出了基于项目的一跳信任模型，该模型通过用户对系统中项目信任度的计算，定义了用户的社会信任属性向量，以这种属性向量为桥梁一步跳转建立用户之间的信任度，一定程度上解决了传统的基于中间用户节点多下一跳的信任模型存在的复杂的路径选取和信任弱传递带来的计算准确度降低的问题。在此模型的基础上，本文又提出基于一跳信任模型的协同过滤推荐算法，该算法通过对能体现个体评分差异的信任度与表征总体评分趋势的余弦相似度进行二维拟合，使用户之间的权重计算更加合理。实验证明，本文算法 OneJ-TCF 较 CF 显著的提高了准确度(约 0.05 MAE)，比 TCF 准确度指标 MAE 提高约 0.02，而且时间复杂度较 TCF 大大降低(约 50%), 和 CF 的时间复杂度接近，验证了第 3 节的理论性能分析。本文下一步研究工作将探索直接信任、间接信任和相似度的多维直接拟合问题，以进一步提高推荐准确度。

参考文献：

[1] 荣辉桂，火生旭，胡春华. 基于用户相似度的协同过滤推荐算法[J]. 通信学报, 2014, 35(2):16-24.
RONG H G, HUO S X, HU C H. User similarity-based collaborative filtering recommendation algorithm[J]. Journal on Communications, 2014, 35(2):16-24.

[2] 李英壮，高拓，李先毅. 基于云计算的视频推荐系统的设计[J].通信学报,2013,34(Z2):138-140.
LI Y Z, GAO T, LI X Y. Design of video recommender system based on cloud computing[J]. Journal on Communications, 2013, 34(Z2): 138-140.

[3] 丁欣，马严，吴军. 适用于校园网的视频推荐系统的设计与实现[J]. 通信学报, 2013, 34(Z2):175-179.
DING X, MA Y, WU J. Design and implementation of a video recommendation system in campus network[J]. Journal on Communications, 2013,34(Z2):175-179.

[4] YOU W, YE S S. A survey of collaborative filtering algorithm applied in E-commerce recommender system[J]. Computer Technology and Development, 2006, 16(9):70-72.

[5] GOLBECK J, HENDLER J. Inferring trust relationships in Web based social networks[J]. ACM Transactions on Internet Technology, 2006,6(4):497-529.

[6] POLO M, BOBBY B. Using trust in recommender system: an experimental analysis[A].Proceedings of iTrust2004 International Conference[C]. 2004.

[7] AVESANI P, MASSA P, TIELLA R. A trust-enhanced recommender system application: molesking[A]. Proceedings of the 2005 ACM Symposium on Applied Computing[C]. Santa, 2005. 1589-1593.

[8] YUAN Q, ZHAO S W, LI C, *et al*. Augmenting collaborative recommender by fusing explicit social relationships[A]. ACM Conference on Recommender System Workshop on Recommender Systems and the Social Web[C]. New York, 2009.49-56.

[9] JEBRIN A S, WILLIAMS M A. Credibility-aware web-based social network recommender: follow the leader[A]. Proceedings of the 2nd ACM RecSys'10 Workshop on Recommender System and the Social Web[C]. Barcelona, 2010.

[10] MA H，KING I，LYU M R. Learning to recommend with social trust ensemble[A]. Proceedings of the 32rid International ACM SIGIR Conference on Research and Development Information Retrieval[C]. New York, USA, 2009. 203-210.

[11] 田俊峰，鲁玉臻，李宁. 基于推荐的信任链管理模型[J].通信学报,2011, 32(10):1-9.
TIAN J F, LU Y Z, LI N. Trust chain management model based on recommendation[J]. Journal on Communications, 2011,32(10):1-9.

[12] 唐文，陈钟. 基于模糊集合理论的主观信任管理模型研究[J]. 软件学报, 2003, 14(8):1401-1408.
TANG W, CHEN Z. Research of subjective trust management model based on the fuzzy set theory[J]. Journal on Software, 2003, 14(8): 1401-1408.

[13] 王新洲，史文中，王树良. 模糊空间信息处理[M]. 武汉：武汉大学出版社, 2003.
WANG X Z, SHI W Z, WANG S L. Fuzzy Spatial Information Processing[M]. Wuhan: Wuhan University Press, 2003.

[14] ZIEGLER C N, LAUSEN G. Analyzing correlation between trust and user similarity in online communities[A]. The Second International Conference on Trust Management[C]. Oxford, 2004.251-265.

[15] 朱郁筱，吕琳媛. 推荐系统评价指标综述[J]. 电子科技大学学报, 2012, 41(2): 163-175.
ZHU Y X, LV L Y. Evaluation metrics for recommender systems[J]. Journal of University of Electronic Science and Technology of China, 2012, 41(2): 163-175.

作者简介：

王兴茂（1989-），男，辽宁营口人，国家数字交换系统工程技术研究中心硕士生，主要研究方向为数据挖掘、用户行为分析、推荐算法。

张兴明（1963-），男，河南新乡人，国家数字交换系统工程技术研究中心教授，主要研究方向为通信与信息系统、宽带信息网络等。

邬江兴（1953-），男，浙江嘉兴人，中国工程院院士，国家数字交换系统工程技术研究中心教授，主要研究方向为通信与信息系统、计算机网络、拟态安全等。