基于社交网络用户信任度的混合推荐算法研究

文俊浩 何 波 胡远鹏

(重庆大学软件学院 重庆 400030)

摘 要 为了解决当前社交网络中基于用户信任的Web服务推荐算法存在的覆盖率不足的问题,整合了当前有关直接信任、间接信任及群体信任度的研究思路,对相关的信任度计算方式进行了扩展研究。在此基础上,提出了一种新的混合信任度算法。实验结果表明,在召回率、用户信任度和用户争议度等指标上该混合推荐算法优于现有算法ModelTrust,证明了该算法具有覆盖率较高的特点,能解决由单一信任度算法数据稀疏性造成推荐结果不佳的问题。

关键词 社交网络,用户信任度,混合推荐算法,ModelTrust

中图法分类号 TP311 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2016.1.055

Hybrid Recommendation Algorithm Based on User's Trust in Social Networks

WEN Jun-hao HE Bo HU Yuan-peng

(College of Software Engineering, Chongqing University, Chongqing 400030, China)

Abstract In order to solve the problem of insufficient coverage of current social networking Web service recommendation algorithm based on user' trust, this paper incorporated the research ideas of direct trust, indirect trust and community trust, and did a lot of extended research on the calculation about trust. Based on these work, we proposed a new hybrid trust algorithm. We did a lot of experiments based on recall rate, user trust and user dispute. The experimental result shows that the hybrid algorithm has a good performance on coverage problem. Meanwhile, it can solve the problem that single trust algorithm's sparse data lead to poor recommendation result.

Keywords Social network, User trust, Hybrid recommendation algorithm, ModelTrust

推荐系统的面世有效解决了现代社会"信息爆炸"带给普通人群的信息选择难题。推荐系统目前主要分为3类:基于内容过滤的推荐、基于协同过滤的推荐和基于社会化过滤的推荐。得益于Facebook、微博等社交网络应用的流行,社会化过滤推荐系统是当前研究的热点领域。而信任度在社会化过滤中扮演着重要角色,Golbeck[1]针对信任值的传播提出了Tidal-Trust。Li Xiong等[2]针对平均值算法不能满足信任动态性提出了基于个人相似度的信任信息聚合方法PSM。

当前信任度算法研究主要基于直接信任度和间接信任度进行,文献[3]提出了一种基于IOWA算子的直接信任度算法,该算法通过对IOWA算子求解,得到不同的权重序列,从而计算出用户间的直接信任度,这种算法忽略了信任在传播过程中的衰减因素,同时存在数据稀疏性问题;在间接信任度研究中,文献[4]提出了一种基于信任链的信任度计算方法,而对于信任链的确定和计算是本文研究的一个重点。近年来,群体信任度研究在社交网络中得到广泛普及,而大多数群体信任算法存在复杂的迭代聚类过程,增加了推荐过程中资源和时间的开销。基于此等情况,本文进行了相关的算法扩展改进并提出了新的算法思路。

1 直接信任度研究

1.1 IOWA算子研究

文献[1]提出了一种基于IOWA算子的直接信任算法。该算法通过求解IOWA算子获得最优权重值,计算出信任结果。对IOWA算子的定义如下。

定义1 设 $f_w:R^m\rightarrow R$ 为 m 元函数,$W=(w_1,w_2,w_3,\cdots,w_n)^T$ 为与 f_w 相关的加权向量,其中 W 满足:$\sum w_i=1$,$w_i>0$,二维向量(v_i,a_i)满足:

$$f_w((v_1,a_1),(v_2,a_2),(v_3,a_3),\cdots,(v_n,a_n))=\sum_{i=1}^{m}w_ib_{ui} \quad (1)$$

且 b_{ui} 是 a_i 按照 v_i 进行排序后的序列中第 i 大的数,则称 f_w 为 m 维诱导有序加权平均算子,简称IOWA。

根据上述的定义,IOWA算子将按照$(v_1,v_2,v_3,\cdots,v_n)$排序后对应的$(a_1,a_2,a_3,\cdots,a_n)$进行加权相加。因此可以使用信任的时间作为诱导值对信任值进行排序。

根据文献[1]对IOWA算子的求解方法,令IOWA算子的"或度量"(orness)为:

$$orness(W)=\frac{1}{n-1}\sum_{i=1}^{n}(n-i)w_i \quad (2)$$

在IOWA算子的求解过程中同时涉及到"与度量"和"离

到稿日期:2014-11-03 返修日期:2015-03-21 本文受基于异构服务网络分析的Web服务推荐研究(61379158)资助。

文俊浩(1969-),男,博士,教授,主要研究方向为数据挖掘、面向服务的计算与面向对象的软件工程,E-mail:jhwen@cqu.edu.cn;何 波(1989-),男,硕士生,主要研究方向为Web服务推荐及算法;胡远鹏(1989-),男,硕士生,主要方向为数据挖掘、Web服务推荐。

散度”两个概念及其对应求解公式，但在本文的改进型研究中并未采用，故略过不提。

1.2 基于时间衰减的直接信任度模型

上述基于 IOWA 算子的直接信任计算情况忽视了信任在传播过程中因为时间流逝而导致的信任减弱问题，对此，本文在用户交互记录的过程中加入了时间因素，并重新给出信任算法定义。

定义二维向量(t_k, r_{ij})为用户 i,j 的一次交互记录，t_k 为此次交互时间，r_{ij} 为此次交互行为后 i 对 j 的评价。定义 p_{ij} 为节点 i 对 j 的直接信任度，向量组 $I(i,j)=((t_1,r_{ij1}),(t_2,r_{ij2}),(t_3,r_{ij3}),\cdots,(t_n,r_{ijn}))^{\mathrm{T}}$ 为 i,j 交互记录的合集，则直接信任度计算公式如下所示。

$$P_{ij}=IOWA(I(i,j)) \tag{3}$$

上式即是将 i 对 j 节点的评价 r_{ij} 按照对应的 t_k 以从大到小的顺序进行排序，在排序以后按照一定的 orness 计算出的权重向量$(w_1,w_2,w_3,\cdots,w_n)^{\mathrm{T}}$ 进行权重相加，得出 i 对 j 节点的直接信任度。

2 基于图模型的间接信任度研究

在社交网路中，个体用户之间的信任被映射为网络节点之间的询问。对于不相邻的节点，其信任值根据路径的不同而有所不同。对此，本文构建了基于图模型的信任链(信任路径)研究模型，通过描述信任链的算法步骤，从而探讨用户间接信任的计算方法，并最终给出了基于信任链的间接信任度计算公式。

本节划分了 4 种典型的路径传播模型来进行节点间信任计算的研究。首先，定义集合 $I=\{i_1,i_2,i_3,\cdots\}$为 A,E 节点间连通路径的集合，不同模型传递方式分类如下。

2.1 简单路径下的信任链

对$\forall i_k,i_v\in I$，$node(i_k)\cap node(i_v)=\{A,C\}$且 $length(i_k)=length(i_v)$，即 A、E 间的连通路径长度相同且各不相交，则集合 I 中所有的路径均为 $A\rightarrow E$ 的信任链，如图 1 所示。

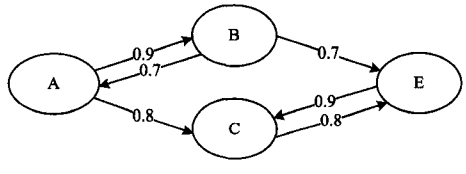

图 1 简单路径下的信任链

2.2 重复节点路径下的信任链

若$\exists i_k,i_v\in I$，$node(i_k)\cap node(i_v)\neq\{A,E\}$且$\forall i_k,i_v\in I$，$length(i_k)=length(i_v)$，即 A、E 间路径长度相同且存在相交节点，则此时需要引入信任链入口值的概念，用以在信任链的确定过程中剔除重复的中间路径，如图 2 所示。对于路径中的重复节点，其信任链入口值应该取其所在路径到节点路径中信任度的积的最大值，只有当路径中最小信任度大于等于入口值时，路径才能被选为信任链。

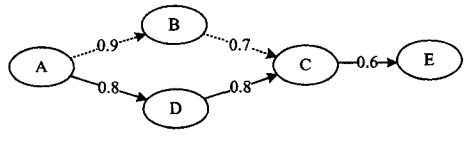

图 2 重复节点路径下的信任链

2.3 不同长度路径下的信任链

若$\forall i_k,i_v\in I$，使得 $node(i_k)\cap node(i_v)=\{A,E\}$且$\exists i_k$，$i_v\in I$，$length(i_k)\neq length(i_v)$，即 A、E 间存在不同长度的路径，此时需要确定信任链的最大长度。如图 3 所示，令 $L=\min\{length(i_k),i_k\in I\}$，$t_{A;l_{\min};E}$为最短路径上的信任度之积，则对于 $\forall i_k\in I$，i_k 为信任链的条件为 $t_{A;l_k;E}\geqslant t_{A;l_{\min};E}$ 且 $length(i_k)\leqslant L+\alpha$。$\alpha$ 为整数，根据实际情况进行约定设置。

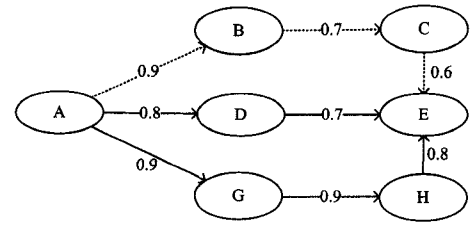

图 3 不同长度路径的信任链

2.4 复杂路径下的信任链

若$\exists i_k,i_v\in I$，使得 $node(i_k)\cap node(i_v)\neq\{A,E\}$且$\exists i_k$，$i_v\in I$，$length(i_k)\neq length(i_v)$，此时的复杂情况将面临前 3 种情况的集合，如图 4 所示。对于 $i_k,i_v\in I$，当遭遇重复节点时，根据情况的不同选择不同的路径：

(1)如果重复节点是中间节点，则选取信任度之积最大的路径。

(2)如果重复节点是目标节点，则选取最短路径和信任度之积大于最短路径且长度之差为 1 的路径。

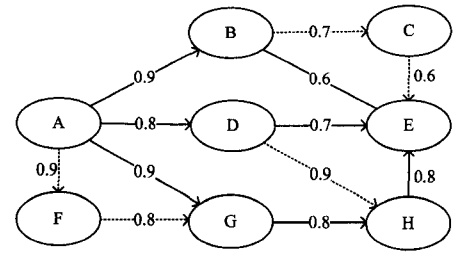

图 4 复杂路径下的信任链

2.5 信任链算法步骤总结

基于上述 4 种情况，本文研究得出任意两点之间的信任链只与连接目标节点的路径长度和信任度之积有关。设初始节点为 *Source*，目标节点为 *Sink*，则两点间的信任链计算步骤如下：

(1)建立队列 P，加入 *Source*，建立已访问节点集合 S，设定信任链最大长度 *Maxpath* 为 3。

(2)若 P 为空则跳转(6)。否则设定当前节点 C 为对头元素，遍历节点 C 的邻接节点并存入集合 N 中。若 N 包含 *Sink* 则进入(3)，否则进入(4)。

(3)设定 *Maxpath* 为 *Source* 到 C 的路径长度＋1，信任链入口值为通过该路径到 *Sink* 上直接信任度的积。将该路径存入结果集 R 中，进入(4)。

(4)若 *Source* 通过节点 C 连接集合 N 中元素的路径长度小于 *Maxpath*＋1，且路径上信任度最小值大于信任链入口值时，检查 N 中元素是否已被访问过。若 N 中任意节点 T 都不在 S 中，则将节点 T 存入队列 P 和集合 S 中，并记录 *Source* 到 T 的路径长度和路径中的信任度相乘的结果。

(5)若节点 T 在集合 S 中，则将当前节点对应的信任度之积、路径长度与队列 P 中相同节点对应的信任度之积、路

径长度进行比较，选取较短的路径，若路径长度相同，选取信任度最小值较大的路径。若两者都相同，则随机选取。跳转至(2)。

(6)若 P 为空，信任链的选择即为结果集 R 中的信任路径。

在此信任链选取算法基础上，本文给定新的间接信任度计算公式如下：

$$t_{ij}=\frac{\sum_{k\in N(i),t_{ik}\geqslant\max} t_{ik}t_{kj}}{\sum_{k\in N(i),t_{ik}\geqslant\max} t_{ik}} \tag{4}$$

其中，$N(i)$为 i 的邻域节点集合，max 为信任链的入口值的最小值，$Trust(i,j)$为 i 到 j 节点信任链的集合。

3 基于直接信任的群体信任度

3.1 基于 FOAF 的群体划分

在基于社交网络信任推荐的研究中，群体信任度算法将有助于解决“冷启动”以及覆盖率较低的问题。本文基于 FOAF(Friend of a Friend)进行了群体划分：确定群体的中心节点为被推荐的用户，以该用户为中心，用户路径长度小于一个给定值 n 的所有用户节点均为该用户的群体。

对于给定值 n 的确定，本文以公开数据集 Epinions 为例。数据集中包含 131828 个节点，841372 条边，也就是平均每个节点与 6.38 个节点相连，最长连接路径为 14。数据显示，90%的节点可以通过长度小于 4.9 的路径相连。通过计算研究，本文最终取平均值，即 n 值为 3。

本文给出的群体信任度公式如下：

$$s(i,G)=\frac{1}{\sum_{j\in V,i\neq j} s(j,G)}\sum_{j\in V,i\neq j} s(j,G)*e_{ji} \tag{5}$$

式中，$G=\langle V,E\rangle$为节点 i 所在的群体构成的子图，$s(i,G)$为节点 i 在群体 G 中的群体信任度，V 为 G 中的所有节点集合，e_{ji}为节点 j 对节点 i 的直接信任度。若节点 j 和 i 没有交互记录则 e_{ji}为 0。

3.2 算法求解过程

在群体信任的计算中，对 n 个用户的群体存在 n 个如同式(6)的计算公式，包含 n 个未知数，需要采取迭代的方式进行方程组求解。

$$\Delta=\sum_{i\in V}|\overline{s(i,G)}-s(i,G)| \tag{6}$$

当 Δ 小于一个特定值时迭代结束。最后一次计算的结果就可以近似地认为是用户节点的群体信任度。算法过程如下：

(1)对于群体节点所构成的子图 $G=\langle V,E\rangle$中，为每一个节点的群体信任度 $s(i,G)$设定一个值，使得 $s(i,G)\in[0,1]$。

(2)对于群体当中的每一个节点，根据式(6)计算每个节点迭代出的群体信任度$\overline{s(i,G)}$，后续的节点可以使用之前计算出的$\overline{s(i,G)}$以加快迭代速度。

(3)当 Δ 小于设定的近似范围 ε 时，则当前的$\overline{s(i,G)}$就是群体中节点的群体信任度。否则跳转到步骤(2)继续对$\overline{s(i,G)}$进行迭代。

4 混合信任度算法研究

4.1 单类信任算法比较分析

本文在前面 3 节中，分别介绍了关于直接信任、间接信任以及群体信任的相关算法研究。其中既有对前人的改进，也有自己的新观点。但总体而言，依然存在一些不足：

(1)实验表明，通过直接信任算法进行权重向量$(w_1,w_2,w_3,\cdots,w_n)^T$的计算时，由于 IOWA 因子的取值不同，会呈现较大的结果分歧，在引入时间概念后，这种数据计算量将更加庞大，难以操作。此外，采用直接信任的计算方式仍然存在数据稀疏性的问题。

(2)本文提出的基于图模型的间接信任算法是在网络节点信息传递的研究基础上，提出了关于信任链的路径计算方法，为非邻节点(陌生用户)之间的信任关系提供了一种合理量化标准。但考虑到信任在传播过程中的衰减因素，该方法只对信任链长度在一定范围以内的模型具备良好的推荐效果，超出范围后推荐精度将受较大削弱。

(3)提出群体信任度的研究，是为了解决前面两种算法都面临的“冷启动”问题，通过群体的相似性来加强信任关系。

4.2 混合推荐算法的提出

针对以上问题，本文在对以上所述各类算法的研究基础上，依据节点之间最短信任链长度取值大小，提出了一种高覆盖率的混合信任度算法。该算法的主要思想就是根据不同的信任群体，合理使用不同信任算法并进行加权平均，在克服单一算法缺点的同时，取得更加准确的信任计算结果值。该算法可用以计算任意两点间的信任度。其计算公式如下。

令 T_{ij}为节点 i 对节点 j 的信任度，则其计算公式如下：

$$T_{ij}=\begin{cases}\dfrac{\alpha*t_{ij}^{direct}+\beta*t_j^{group}}{\alpha+\beta}, & length(i,j)=1\\ \dfrac{\alpha*t_{ij}^{indirect}+\beta*t_j^{group}}{\alpha+\beta}, & length(i,j)>1\\ t_j^{group}, & lenght(i,j)=0\end{cases} \tag{7}$$

式中，t_{ij}^{direct}表示直接信任度，$t_{ij}^{indirect}$表示间接信任度，t_j^{group}表示群体信任度，它们对应的计算公式分别为式(3)－式(5)。$length(i,j)$为两节点间最短信任链的长度。$length(i,j)=0$表示用户 i 只被其他用户信任而自身并不信任其他任何用户的情况，可以理解为新用户的加入。

α 和 β 作为计算公式适应参数，需要根据实际情况进行调整。对于稀疏率较高的数据，应该将 β 的值设定为大于 α，以提高模型的覆盖率。对于稀疏率较低的数据，过高的群体信任权重将抵消掉局部信任的个体化信任的优势，降低模型的准确率，因此需要设定较低的权重参数。

5 混合算法实验对比分析

5.1 实验数据集

实验数据来源于用户评论网站 Epinions.com 的数据集，为了减少数据计算量，提高实验效率，本文对原有的数据集通过随机游走的方式选取了 2000 个节点，其详细信息如表 1 所列。用户之间的信任度在数据集中只有两种表现形式，相互信任为 1，互不信任为－1。

表 1 用于测试的数据集的详细信息

项目	数值
节点数	2000
边数	137522
平均度数	69.7
最大信任数	1736
最小信任数	0
信任数>500	29
信任数>100	381
信任数>10	751
信任数>5	839

5.2 实验对比算法及测试方法

实验中采用的对比算法 ModelTrust 由文献[7]提出，是一种针对时间效应而提出的应对局部信任算法，能够计算在一定的信任长度限制下连通的所有用户节点之间的信任度。实验采用文献[6]提出的 MAUE 评测方法，该方法通过计算每位用户所获得的推荐结果差距的平均值来评价算法优劣。

5.3 实验结果分析

(1)MAUE 及召回率结果分析

按照式(7)，将 α 和 β 均取值为 1，再利用以上的数据开始对每一个节点计算 MAUE 和召回率，计算结果如表 2 所列。

表 2 用户绝对平均误差和召回率实验结果

	ModleTrust (MAUE/召回率)	混合信任算法 (MAUE/召回率)
信任数≤5	0.8833 / 17.49%	0.6838 / 42.49%
5<信任数≤10	0.7113 / 46.64%	0.6415 / 77.70%
10<信任数≤100	0.3878 / 59.75%	0.3867 / 79.35%
100<信任数≤500	0.2062 / 60.83%	0.2361 / 89.42%
信任数>500	0.3526 / 79.31%	0.3599 / 96.55%

如表 2 所列，混合信任度算法在召回率上较 ModleTrust 有较大的提升，主要原因在于本文中采用的数据集数据半径为 4.9，而群体发现采用的 FOAF 设定长度为 3，因此大部分的用户节点都能够通过群体信任度获得计算结果，提高了算法的召回率。

(2)基于用户信任度数量的实验结果分析

实验结果如图 5 所示，纵坐标表示 MAUE，横坐标表示节点的出度。两种算法在信任数量较少的情况下都出现较大的 MAUE 值，其原因在于数据集中用户间的信任度只有信任与不信任两种，缺乏中间值用以表示信任的程度，在计算 MAUE 时得到的结果值会相对偏大。

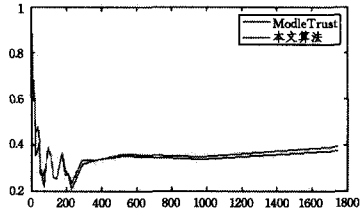

图 5 基于用户信任数的对比实验结果

对于信任数量较少的节点，由于可供参考的节点数量较少同样造成结果偏差较大，但本文算法采用了全局信任作为参考，因此平均后的结果相对 ModelTrust 算法较优。另一方面，对于信任数量较大的节点，全局信任算法并不考虑特点用户节点之间的信任关系，因此造成算法在这一部分效果并不理想。

(3)基于用户争议度的实验结果分析

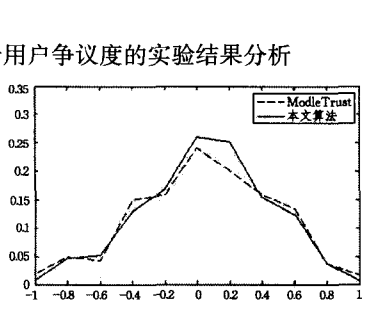

图 6 基于用户争议度的对比实验结果

图 6 中横坐标表示数据在用户群体中的争议程度(即正反评价差异性)，纵坐标表示算法推荐效率。从实验结果来看，本文算法在争议较小的用户群体当中表现效果较佳，但是在争议较大的用户当中效果不明显，原因在于本文算法中 α 和 β 的取值问题。目前，我们无法做到根据数据集进自动行调整。但根据经验，争议度较高的数据集应该增加直接和间接信任度对应的权重，相反，争议度较低而聚类系数较低的数据集适用于全局信任对应的权重较高的因子。

结束语 本文基于社交网络中用户信任传递模型的基础研究，整合了当前有关直接信任度、间接信任度和群体信任度的相关计算方法，对相关算法进行了扩展研究。在此基础上，提出了一种新的混合信任度算法。通过实验对比，验证了该算法的执行效果和可行性。

总结实验结果，该算法的研究还存在一些不足之处，例如加权因子的取值问题，科学合理的取值将对该算法的准确性产生重要影响，因此，这也将是本文下一步研究工作的重点。

参 考 文 献

[1] Golbeck J A. Computing and Applying Trust in Web-Based Social Networks[D]. University of Marylard at College Park, 2005

[2] Li X, Liu L. Peer-Trust: Supporting reputation-based trust in peer-to-peer communities[J]. IEEE Trans. on Data and Knowledge Engineering, Special Issue on Peer-to-Peer Based Data Management, 2004, 16(7): 843-857

[3] Sen S, Vig J, Riedl J. Tagommenders: connecting users to items through tags[C]// Proceedings of the 18th International Conference on World Wide Web(WWW '09). New York, NY, USA, 2009: 671-680

[4] Kurant M, Gjoka M, Butts C T, et al. Walking on a graph with a magnifying glass: stratified sampling via weighted random walks [C]// Proceedings of the ACM SIGMETRICS Joint International Conference on Measurement and Modeling of Computer systems. ACM, 2011: 281-292

[5] Massa P, Avesani P. Trust metrics in recommender systems [M]// Computing with social trust. Springer London, 2009: 259-285

[6] Liang Z Q, Shi W S. Analysis of recommendations on trust inference in open environment[J]. Journal of Performance Evaluation, 2008, 65(2): 99-128

[7] Massa P, Avesani P. Controversial users demand local trust metrics: An experimental study on epinions. com community[C]// Proceedings of the National Conference on artificial Intelligence. AAAI Press, 2005: 121-126

[8] Bao Jie, Cheng Jiu-jun. Group Trust Algorithm Based on Social Network[J]. Computer Science , 2012, 39(2): 38-41 (in Chinese)
鲍捷，程久军. 基于社交网络的群体信任算法[J]. 计算机科学，2012，39(2)：38-41

[9] Li Xiao-yong, Gui Xiao-lin. Cognitive Model of Dynamic Trust Forecasting[J]. Journal of Software, 2010, 21(1): 163-176 (in Chinese)
李小勇，桂小林. 动态信任预测的认知模型[J]. 软件学报，2010，21(1)：163-176

[10] Zhang Feng, Wang Jian, Zhao Yan-fei, et al. Trust Model Based on Groups Recommendation in Social Network[J]. Computer Science, 2014, 41(5): 168-172(in Chinese)
张丰，王箭，赵燕飞，等. 社交网络中一种基于社区推荐的信任模型[J]. 计算机科学，2014，41(5)：168-172