软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2014,25(Suppl.(2)):44−52
E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

内嵌人格分析的社交关系强度层次模型及算法*

李艳兵[1,2], 叶 剑[1,2], 朱珍民[1,2]

[1](中国科学院 计算技术研究所,北京 100190)

[2](移动计算与新型终端北京市重点实验室,北京 100190)

通讯作者: 叶剑, E-mail: jye@ict.ac.cn

摘 要: 社交网络中用户关系强度计算对于个性化社交服务呈现具有重要意义.同时,心理学研究表明人格特征是影响用户关系强度的关键因素之一.基于社会心理学中人与人之间的关系产生原理,提出一种内嵌人格分析的社交关系强度层次模型及计算方法.通过社交网络行为建模,建立用户大五人格特征预测模型,实现用户人格倾向性演算.同时结合偏好相似性和交互熟悉性计算,实现嵌入人格特征的用户关系强度的求解算法.最后,本文通过构建人人网社交关系仿真实验平台,验证了该方法的合理性和有效性.

关键词: 社交网络;关系强度;人格特征

中文引用格式: 李艳兵,叶剑,朱珍民.内嵌人格分析的社交关系强度层次模型及算法.软件学报,2014,25(Suppl.(2)):44−52. http://www.jos.org.cn/1000-9825/14022.htm

英文引用格式: Li YB, Ye J, Zhu ZM. Design of TPM chip based on signal integrity analysis. Ruan Jian Xue Bao/Journal of Software, 2014,25(Suppl.(2)):44−52 (in Chinese). http://www.jos.org.cn/1000-9825/14022.htm

Design of TPM Chip Based on Signal Integrity Analysis

LI Yan-Bing[1,2], YE Jian[1,2], ZHU Zhen-Min[1,2]

[1](PLA Information Engineering University, Zhengzhou 450001, China)

[2](State Key Laboratory of Mathematical Engineering and Advanced Computing, Zhengzhou 450001, China)

Corresponding author: YE Jian, E-mail: jye@ict.ac.cn

Abstract: It is important for personalized social services to calculate the relationship strength between users in a social network. Meanwhile, the psychological studies has shown that the personality traits is one of the key factors affecting the user's relationship strength. Based on the relationship generation principle in the social psychology, this paper proposes a personality embedded social relationship strength hierarchical model and algorithm. With the analysis of a user's behavior in social network, this paper predicts the Big Five personality traits of the user to calculate the propensity of personality. The propensity of personality is combined with the similarity of preference and the familiarity of interaction to formalize the personality embedded user relationship strength calculation. At the end of this paper, the proposed algorithm is demonstrated to be reasonable and effective in a simulation experiment of RENREN social network.

Key words: social network; relationship strength; personality trait

个性化社交服务的目标是不断发现和满足用户对社交服务的个性化需求及其变化,并准确提供、推荐、呈现其真正感兴趣的社交服务及其信息内容.而现实生活中,熟人推荐、好友推荐[1]则是促成用户消费的行为的一个非常重要的手段.个性化社交服务若是能够结合社交网络服务中真实人际关系,便可以将现实生活中的好友推荐扩展到电子商务中.因此,社交网络中的真实社会关系及其关系强度的衡量和计算显得尤为重要.

随着社交网络的快速增长和流行,针对社交网络的研究已成为学术界和产业界的热点问题.Singla 等人[2]

* 基金项目: 国家自然科学基金(61076109); 国家高技术研究发展计划(863)(2009AA011902)
收稿时间: 2013-06-15; 定稿时间: 2013-08-21

的研究表明用户之间的个性化特征影响着他们的交互行为.Gilbert 等人[3]利用 Facebook 上用户的个人信息、交互信息以及关系网络的全局信息建立了用户关系强度预测模型.Rongjing 等人[4]基于用户的个人信息和交互信息,提出无监督的隐变量模型,估计用户间关系强度.Backstrom 等人[5]则发现同质性是影响 Myspace 和 LinkdIn 上网络成员预测的重要因素.所谓同质性[6]一方面表现在社会影响过程中,用户趋于接受交互对方的行为爱好,这种影响使得交互双方越来越相似;另一方面,用户趋于选择已经与其相似的人作为潜在好友,这种现象便是社会学中的选择.然而这里的相似性更多地考虑用户间个人信息的相似:年龄相近、籍贯相同、兴趣相似等,忽略了用户自身的人格特征的影响.

从心理学的角度出发,人格特征与现实社会中的行为有显著联系.人格研究的大五人格模型[7]认为,人格特征由 5 个维度组成:开放性、尽责性、外向性、和善性以及神经质.外向性得分高的人更善于社交[8].和善性得分高的人更值得被人信任,更积极维护与他人的关系[9].最近 Golbeck 等人[10]分析了 167 位 Facebook 用户的人格特征,并通过用户的个人信息和发表的帖子,成功预测了用户的五项人格特征,但计算过程中更多地侧重于自然语言处理,必然会涉及到用户的隐私保护问题.Querci[11]等人则分析了 Facebook 上用户交互人数与人格特征之间的关系.他们发现 Facebook 上受欢迎的用户在现实生活中同样是活跃的,这表明线上交互和线下交互没有显著的不同.此外,人格特征至少可以通过 3 种方式影响关系的形成[12],首先,每个用户由于其人格特征的差异,使得用户的交友数量存在严格的不同,这主要取决于用户自身的交友意愿,即外向性指标;其次,用户人格特征也影响着用户被选作好友的难易程度,即和善性指标;最后,用户还趋向于选择与自己相似的用户作为潜在好友.Lu[14]和乔秀全[15]则从信任产生的机制出发,从熟悉度产生的信任度、相似性产生的信任度等多角度计算用户之间的信任度.社交网络以真实的人际关系为依托,延伸了现实生活中的关系网络.因此,本文首先对用户社交网络行为建模,建立大五人格特征拟合模型,获取用户的人格特征;然后,借鉴社会学和心理学中人与人之间的关系形成原理,提出了一种内嵌人格分析的社交关系强度层次模型及其计算方法.

1 社交关系强度层次模型

本文借鉴社会学中人与人的信任产生原理[14],将社交网络中用户之间的关系强度计算划分为 3 个层次:原始数据层(original data layer,简称 ODL)、抽象数据层(abstract data layer,简称 ADL)和关系数据层(relational data layer,简称 RDL)(如图 1 所示),并从人格倾向性、偏好相似性以及交互熟悉性 3 个维度描述社交网络中用户间的社交关系强度,更能体现出现实生活中关系强度形成的过程,也更贴近真实世界中用户之间的关系.

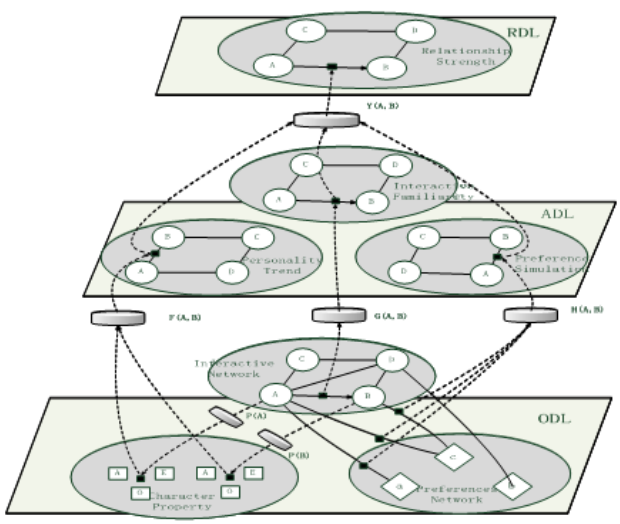


图 1 用户关系强度计算框架

1.1 原始数据层ODL

原始数据层 ODL 由用户 3 类原始信息组成:用户交互网络(IN)、用户的人格特征属性(PC)和用户偏好网络(PN)这 3 类用户原始信息组成,即 $ODL=\langle IN,PC,PN\rangle$.

定义 1(交互网络(IN)). IN 表示为有向带权图 $G_{in}=(U,E_{in},W_{in})$,其中 $U=\{1,2,\ldots,N\}$为用户节点集合,E_{in} 为用户交互集合 $E_{in}=\{e_{ij}=\langle u_i,u_j\rangle|u_i,u_j\in U\}$,$W_{in}$ 为网络中用户间的交互频率集合 $W_{in}=\left\{c_{e_{ij}}\mid e_{ij}\in E_{in}\right\}$. $c_{e_{ij}}$ 表示用户 i 与用户 j 的交互频率.

定义 2(人格属性(PC)). PC 表示为集合 $S_{pc}=\left\{\vec{p}_i\mid i=1,2,...,n\right\}$,$n$ 为用户个数, $\vec{p}_i=(o,e,a)$ 为第 i 个用户的人格特征属性,分别表示开放性、外向性以及和善性的人格得分.

定义 3(行为人格映射 P). 设 S_{his} 为用户的社交记录集合 $S_{his}=\left\{\vec{h}_i\mid u_i\in U\right\}$. $\vec{h}_i$ 为用户 i 的社交记录向量,社交记录包括好友数、来访数、状态数等.P 为定义于 ODL 中用户社交记录集合 S_{his} 到 ODL 中人格属性集合 S_{pc} 之间的映射,即 $P:S_{his}\to S_{pc},\forall u\in U,\exists h_u\in S_{his},\exists\vec{p}_u\in S_{pc}$,使 $\vec{p}_u=P(h_u)$.基于此,能够获得任意用户的人格属性向量.

定义 4(偏好网络(PN)). PN 表示为二分图 $G_{pn}=(U,R,E_{pn},W_{pn})$,其中 U,R 分别为用户节点集合和兴趣节点集合,$E_{pn}=\{e_{ij}=\langle u_i,r_j\rangle|u_i\in U,r_j\in R\}$为用户对兴趣节点的偏好集合,为边对应的权集,表示用户 i 对兴趣 j 的偏好程度.

1.2 抽象数据层ADL

抽象数据层 ADL 由交互熟悉性(IF)、人格倾向性(PT)和偏好相似性(PS)组成,即 $ADL=\langle IF,PT,PS\rangle$,分别对应于原始数据层 ODL 中信息,是它们的抽象.

定义 5(交互熟悉性). IF 表示为有向带权图 $G_{if}=(U,E_{if},W_{if})$,其中,U 为用户集合,$E_{if}=\{e_{ij}=\langle u_i,u_j\rangle|u_i,u_j\in U\}$表示网络中用户的交互熟悉度集合,$W_{if}$ 为边对应的权集 $W_{if}=\{f_{e_{ij}}\mid e_{ij}\in E_{if}\}$, $f_{e_{ij}}$ 表示用户 i 对用户 j 由于交互频繁而产生的熟悉程度.

定义 6(交互熟悉映射 *G*). G 为定义于 ODL 中交互集合 W_{in} 到 ADL 中交互熟悉性集合 W_{if}之间的映射,即 $G:W_{in},\to W_{if},\forall u_1\in U,\forall u_2\in N_{u_1}$ (N_{u_1} 为用户 u_1 的邻居用户), $\exists c_{u_1u_2}\in W_{in},\exists f_{u_1u_2}\in W_{if}$,使得 $f_{u_1,u_2}=G(c_{u_1u_2})$.在此基础上,可以计算出网络中任意两个用户基于交互而产生的熟悉度.

定义 7(人格倾向性 PT). PT 表示无向带权图 $G_{pt}=(U,E_{pt},W_{pt})$,其中,U 为用户集合,$E_{pt}=\{e_{ij}=(u_i,u_j)|u_i,u_j\in U\}$表示网络中用户间人格倾向性集合,$W_{pt}=\{t_{e_{ij}}\mid e_{ij}\in E_{pt}\}$ 为对应权集,其中, $t_{e_{ij}}$ 表示用户 i 与用户 j 由于人格相似而可能产生的人格倾向性.

定义 8(人格倾向映射 *F*). F 为定义于 ODL 中人格属性集合 S_{pc} 到 ADL 中人格倾向性集合 W_{pt} 之间的映射,即 $F:S_{pc}\times S_{pc}\to W_{pt},,\forall u_1\in U,\forall u_2\in N_{u_1},\exists\vec{p}_{u_1},\vec{p}_{u_2}\in S_{pc},\exists t_{u_1,u_2}\in W_{pt}$ 使得 $t_{u_1,u_2}=F(\vec{p}_{u_1},\vec{p}_{u_2})$.在此基础上,可以计算出网络中任意两个用户基于人格相似而产生的人格倾向性.

定义 9(偏好相似性 PS). PS 表示为无向带权图 $G_{ps}=(U,E_{ps},W_{ps})$,其中,U 为用户集合,$E_{ps}=\{e_{ij}=(u_i,u_j)|u_i,u_j\in U\}$表示网络中用户由于偏好相似而产生的相似似性集合,$W_{ps}=\{s_{e_{ij}}\mid e_{ij}\in E_{ps}\}$ 为对应的权集,其中, $s_{e_{ij}}$ 表示用户 i 与用户 j 偏好之间的相似程度.

定义 10(偏好相似映射 H). H 为定义于 ODL 中偏好集合 W_{pn} 到 ADL 中偏好相似性集合 W_{ps} 之间的映射,即 $H:W_{pn}\times W_{pn}\to W_{ps},\forall u_1\in U,\forall u_2\in N_{u_1},\exists n_{u_1^*},n_{u_2^*}\in W_{pn},\exists s_{u_1,u_2}\in W_{ps}$,使得 $s_{u_1,u_2}=H(n_{u_1^*},n_{u_2^*})$, $n_{u_1^*}$ 为用户 u_1 对所有兴趣*的偏好集合.在此基础上,计算出网络中任意两个用户的偏好相似程度.

1.3 关系数据层RDL

关系数据层 RDL 综合抽象数据层 ADL 中的 3 类信息,得出用户间的关系强度 STR.

定义 11(关系强度 STR). STR 表示为有向带权图 $G_{str}=(U,E_{str},W_{str})$,其中 U 为网络中所有用户集合,$E_{str}=\{e_{ij}=\langle u_i,u_j\rangle|u_i,u_j\in U\}$表示网络中用户之间的关系强度集合,$W_{str}$ 为对应的权集 $W=\{str|e\in W\}$,其中 $str_{e_{ij}}$ 表示用

户 i 与用户 j 的关系强度.

定义 12(关系强度映射 Y). Y 为定义于 ADL 中交互熟悉性集合 W_{if}、人格倾向性集合 W_{pt} 和偏好相似性集合 W_{ps} 到 RDL 中关系强度集合 W_{str} 之间的映射,即 $Y:W_{if}\times W_{pt}\times W_{ps}\to W_{str},\forall u_1\in U,\forall u_2\in N_{u_1}$,使得 $f_{u_1u_2}\in W_{if}$, $t_{u_1u_2}\in W_{pt},S_{u_1u_2}\in W_{ps},\exists str_{u_1u_2}\in w_{str}$ 并且 $str_{u_1u_2}=Y(t_{u_1u_2},t_{u_1u_2},s_{u_1u_2})$.在此基础上,可以计算出网络中任意两个用户之间的关系强度.

综上所述,通过上述 3 个层次,综合用户的人格倾向性、偏好相似性以及交互熟悉性 3 部分,最终计算出用户之间的关系强度,更能体现出真实社会中关系强度形成的过程,也更贴近真实生活中用户的行为方式.

2 社交关系强度计算方法

社交关系强度层次模型基于用户的社交行为,分别从人格倾向性、偏好相似性以及交互熟悉性 3 个维度描述用户之间的社交关系强度,其重点在于层次模型中 5 个映射的实现:行为人格映射 P、人格倾向映射 F、交互熟悉映射 G、偏好相似映射 H 以及关系强度映射 Y,本部分将详细介绍用户社交关系强度计算模型 5 个映射的具体实现.

2.1 行为人格模型P

由于用户在社交网络的使用过程中会无形地渗透其人格特征,因此本文针对社交网络中用户行为建模,提取用户的大五人格特征.建模的方法许许多多,可以采用线性回归、非线性回归、偏最小二乘等,然而为了充分利用数据集,本文采用线性回归建模:

$$\vec{p}_i(t)=\alpha+\sum_{k\in\vec{h}_i}\beta_k\vec{h}_i(k)+\varepsilon_t \tag{1}$$

其中,$t\in\{\text{OCEAN}\}$, $\vec{p}_i\in S_{pc}$ 是用户 i 的人格属性向量, $\vec{p}_i(t)$ 为第 t 个人格特征分量, $\vec{h}\in S$ 是用户 i 的社交行为向量, $\vec{h}_i(k)$ 为用户 i 的第 k 个行为分量,α为常数项,ε_t 为误差率.以人人网为例,对人人网用户的社交行为数据使用逐步回归法对用户的开放性、和善性以及外向性人格特征[12]建立回归拟合模型,模型指标见表 1~表 3.

从模型结果来看,在 95%的置信度的前提下,开放性的变化与朋友数、日志分享数、留言数和发布照片数有线性回归关系(显著程度 p<0.05).此外,朋友数,日志分享数,发布照片数的回归系数β均大于 0,说明随着用户的朋友数,日志分享数,发布照片数的增加,其开放性就越高,符合基本常识:开放的人拥有较多的朋友,愿意分享自己的信息.此外,模型中朋友数的标准系数 r 的绝对值最大,说明朋友数对开放性的影响最大.总体来看,开放性预测模型的复相关系数 R=0.870,决定系数 R^2=0.757($0\le R^2\le 1$)说明模型对数据的拟合程度较好,表明开放性变异的 75.7%可由朋友数,日志分享数,留言数和发布照片数的变化来解释.外向性和和善性的模型结果解释与开放性类似.因此,我们完全可以基于用户的社交行为预测其人格特征,完成 ODL 层行为人格映射 P 的提取.

表 1 开放性拟合模型

Variables	Regression coefficient β	Standard coefficient r	t-test	Significance level p
Number of friends	0.125	0.806	11.104	0.000
Number of logs	0.089	0.121	2.490	0.014
Number of messages	–0.006	–0.144	–2.457	0.015
Number of albums	0.594	0.118	2.162	0.032

表 2 外向性拟合模型

Variables	Regression coefficient β	Standard coefficient r	t-test	Significance level p
Number of friends	0.104	0.654	11.941	0.000
Number of logs	0.125	0.166	3.747	0.000
Number of albums	0.876	0.170	3.399	0.001

表 3 和善性拟合模型

Variables	Regression coefficient β	Standard coefficient r	t-test	Significance level p
Number of friends	0.145	0.777	10.571	0.000
Number of logs	1.161	0.193	3.471	0.001
Number of messages	−0.008	−0.170	−2.877	0.005
Number of albums	0.087	0.099	2.003	0.047

2.2 用户关系强度计算

2.2.1 人格倾向映射 F

对根用户 u_1 和目标 u_2 分别使用行为人格模型 P 预测得出其人格特征 $\vec{p}_{u_1}=(o_{u_1},e_{u_1},a_{u_1})$ 和 $\vec{p}_{u_2}=(o_{u_2},e_{u_2},a_{u_2})$,然后基于人格特征的影响,针对开放性、外向性以及和善性计算用户的人格交友倾向度:

$$t_{u_1u_2}=F(\vec{p}_{u_1},\vec{p}_{u_2})=\sum_{i=o.e.a}\gamma_i\cdot\left(1-\frac{|u_{1i}-u_{2i}|}{100}\right) \tag{2}$$

其中,u_{1i},u_{2i} 分别是用户 u_1 与 u_2 在开放性、外向性和和善性方面的每一个得分值.另外,为了表示每项人格特征对总体人格相似度的影响权重的不同,γ_i 采用如下公式来计算:

$$\gamma_i=\frac{u_{1i}+u_{2i}}{\sum_{i=o.e.a}(u_{1i}+u_{2i})} \tag{3}$$

2.2.2 偏好相似映射 H

设社交网络中用户 u_1 关注的偏好集合为 $R_{u_1}=\{r_{11},r_{12},...,r_{1l_1}\}$,$l_1$ 表示用户 u_1 关注的偏好个数;用户 u_2 关注的偏好集合为 $R_{u_2}=\{r_{21},r_{22},...,r_{2l_2}\}$,$l_2$ 表示用户 u_2 关注的偏好个数;用户 u_1 和用户 u_2 关注的偏好交集为 $R_c=\{r_1,r_2,...,r_{l_c}\}$,$l_c$ 为共同关注的偏好个数.在这 l_c 个共同偏好中,用户 u_1 转发的信息(状态、日志、视频等)次数为 $\{n_{11},n_{12},...,n_{1l_c}\}$,用户 u_2 的转发的信息次数为 $\{n_{21},n_{22},...,n_{2l_c}\}$.由于用户浏览偏好信息时,都会对不同的信息产生不同的偏好,就会转发其较为喜好的信息,包括状态、日志、照片、视频、链接、音乐等等,于是,本文将用户转发信息的次数作为用户的偏好度,于是得到

$$T(u_1,u_2)=\frac{l_c^2}{l_{u_1}\times l_{u_2}}\cdot\sum_{i=1}^{l_c}\min(n_{1i},n_{2i}) \tag{4}$$

其中 $\min(n_{1i},n_{2i})$表示用户 u_1 和用户 u_2 对第 i 个共同偏好转发信息次数的较小值.

设与 u_1 拥有共同偏好的用户集合为 $V_r(u_1)=\{v_1,v_2,...,v_m\}$,$V_r(u_1)\subset U$,$m$ 为与用户 u_1 拥有共同偏好的用户个数.对 $T(u_1,u_2)$进行归一化,

$$s_{u_1u_2}=H(u_1,u_2)=\frac{T(u_1,u_2)}{\sum T(u_1,u_x)}(u_x\in V_r(u_1)) \tag{5}$$

该式计算的是 $T(u_1,u_2)$在与用户 u_1 拥有共同偏好的所有用户中转发次数所占的比例.

2.2.3 交互熟悉映射 G

设社交网络中与用户 u_1 交互、聊天等的用户集合为 $V_1(u_1)=\{v_1,v_2,\ldots,v_m\}$,$V_I(u_1)\subset U$,$m$ 为 u_1 与交互过的用户个数.本文采用双方交互次数中的最小值作为两者的交互度,因为人与人之间的交互是双向的,若是交互过程中仅仅只有一方发送消息,另一方却没有回复任何信息,则不可以称这两个用户是交互熟悉的.只有两个方向都有消息流动的用户对,才能称其为交互熟悉的.于是得到与用户 u_1 相关的用户交互次数为

$$IC_{u_1}=\{c_{u_1v_1},c_{u_1v_2},...\},\ IC_{u_1}\subset W_{in},$$

$c_{u_1v_i}$ 表示用户 u_1 与用户 v_i 的较小交互次数.采用归一化方法,针对每个用户 $u_2\in V_I(u_1)$得到用户 u_1,u_2 的交互熟悉度:

$$f_{u_1u_2}=G(u_1,u_2)=\frac{IC(u_1,u_2)}{\sum IC(u_1,u_x)}(u_x\in V_I(u_1)) \tag{6}$$

该式计算的是用户 u_1 和用户 u_2 的交互的次数在与 u_1 有交互过的所有用户中所占的比例.

2.2.4 关系强度映射 Y

综合上述人格倾向性 PT、交互熟悉性 IF 和偏好相似性 PS 三类信息计算用户的社交关系强度,得到根用户 u_1 对于目标用户 u_2 的关系强度计算公式:

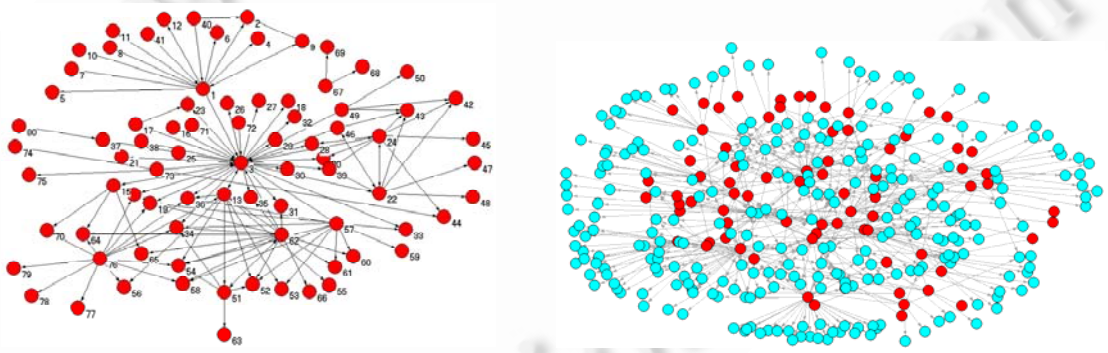
$$str(u_1,u_2)=\alpha\cdot F(u_1,u_2)+\beta\cdot G(u_1,u_2)+\gamma\cdot H(u_1,u_2) \tag{7}$$

式中,$str(u_1,u_2)$为根用户 u_1 与目标用户 u_2 的关系强度,$F(u_1,u_2)$为 u_1,u_2 间的的人格倾向性,$G(u_1,u_2)$为 u_1,u_2 间基于交互的熟悉度,$H(u_1,u_2)$为用户间偏好的相似性.其中,α,β,γ为调整系数,当用户 u_1 与用户 u_2 为熟人时,他们之间拥有更高的交互度,因此α,γ作用更显著;而当用户 u_1 与用户 u_2 之间都是兴趣相投的人时,γ的作用便不是那么重要了.因此,这些参数可以根据实际情况动态调整,鉴于目前还没有用户人格特征在关系强度计算中的研究,本文视这 3 种属性同等重要,进而分配相同的权重$\alpha=\beta=\gamma=1$.

3 实验及分析

3.1 实验数据集

为了获取真实的用户情况和大量实验样本需求,课题组在社交网站人人网上建立站内应用 Personality,并组织课题组人员基于自己的社交圈,邀请好友对自己及其他人的大五人格特征进行自我评价以及好友评价,获得 300 个用户的个人信息、大五人格得分以及来访者数量、朋友留言数、朋友数等,并通过第 2.1 节建立行为人格模型,使用好友数、留言数、日志数、相册数来预测用户的人格特征.此外,为了计算人人网中用户的关系强度,还获取了用户间的交互情况(包括留言、评论)以及用户关注转发话题(人人主页)信息.为了直观的显示这个社会化网络,图 2(a)为随机提取的 80 名用户交互数据图,节点为用户,边的权为用户间的交互次数.图 2(b)为 80 名用户的转发人人主页数据图,深色节点为用户,浅色节点为人人主页,边的权为用户转发对应话题的次数.



(a) 用户-用户交互关系图 (b) 用户-话题关注关系图

图 2 人人网采集数据关系图

3.2 社交关系强度的有效性验证原理

为了验证本文提出的社交网络中用户社交关系强度模型及方法的有效性.采用绝对的关系强度值对用户来说意义并不是很大,相对的关系强度则更有意义,因为用户只要能够区别开不同的用户关系强度顺序即可,因此,本文将考虑 3 个特征因素的关系强度计算方法和只考虑兴趣偏好度的排序方法以及考虑兴趣偏好和交互熟悉的排序方法进行了比较.

为了评估社交网络中用户社交关系强度计算方法的准确性.本文采用评价指标 nDCG(normalized discounted cumulative gain)[16]来衡量信任度排序的准确率.nDCG 是搜索系统中广泛使用的一种排序评价手段,它不仅考虑了搜索结果自身的重要性,还考虑了搜索结果所在的相对位置,强相关的结果在列表中出现的位置越靠前(rank 越高)说明其越有用,相反一个强相关的结果排名靠后,则应该受到相应的惩罚.nDCG 的定义为

$$nDCG_p = \frac{DCG_p}{IDCG_p} \tag{8}$$

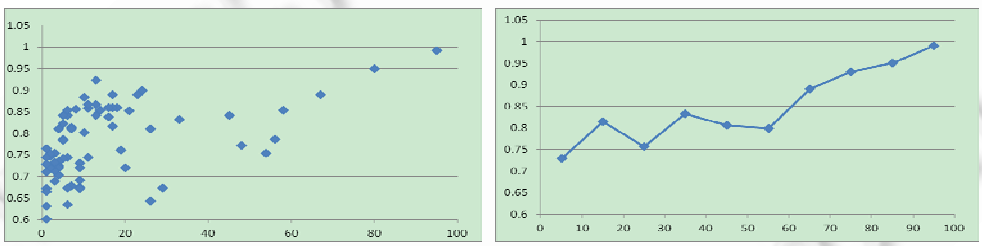
而 DCGp 的计算公式为

$$DCG_p = G[1] + \sum_{i=1}^{p} \frac{G[i]}{\log_2(i+1)} \tag{9}$$

其中 $G[i]$表示第 i 个目标节点用户与根用户 A 的关系强度评分,p 目标用户数.$IDDG_p$ 表示完美排序情况下的 DCG_p 结果,完美的关系强度计算结果形成的排序结果会使 DCG_p 和 $IDCG_p$ 值相同,从而使 $nDCG_p$ 值为 1.实际情况下 $nDCG$ 值不会达到 1,其取值在 0~1 之间.$nDCG$ 值越接近 1,则说明关系强度计算方法的计算结果越准确.

3.3 实验结果及分析

本文社交关系强度计算主要基于假设用户人格会影响社交网络中用户之间的关系强度,而不同关系强度的用户间交互频度也不同.因此,本文从人人网数据集中随机抽取 80 对用户的交互记录,并计算其人格倾向性,得出人格与社交行为的关系如图 3(a)所示,图中横轴表示用户对在社交网络中的交互次数,纵轴表示用户对的人格倾向性.图中有大量数据集中在左侧,这主要是由于这些用户对之间在社交网络中缺乏交互所致,属于噪音点.为了降低噪音点影响,本文对交互次数进行聚类,聚类步长为 10 次交互,聚类后的关系如图 3(b)所示.根据关系图的走向可以发现,交互越频繁,其人格倾向性越高,即从数据层面说明人格越相似,其成为好友的可能性越大,验证了人格影响社交关系强度的合理性.



(a) 聚类前对比图 (b) 聚类后对比图

图 3 用户人格倾向性与社交交互关系对比

为了验证社交关系强度计算的有效性,本文随机选取了 20 个用户作为根节点,并针对每个根节点分别随机选取 30 个目标用户节点做 3 组实验分析;然后号召 100 个志愿者参与实验评价,针对每一个志愿者,我们提供如下信息:(1) 基于根用户的社交信息,使用行为人格模型计算的大五人格得分;(2) 根用户关注及转发的人人主页信息;(3) 根用户与目标节点的交互记录(人人留言、状态、日志、评论等).

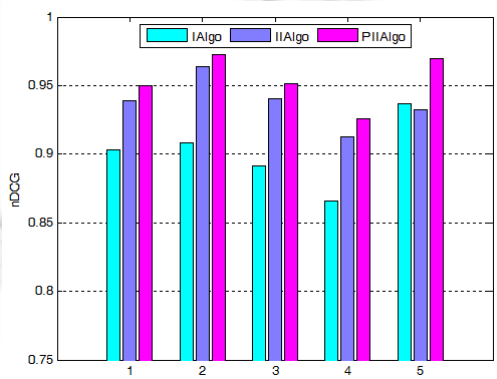


图 4 3 种方法比较

每个志愿者针对每组实验依据给定的信息,对根用户及其目标用户的关系强度进行打分(0~10),0 分代表一

点不匹配;10 分代表非常匹配.本文进行了 3 组实验,第 1 组只考虑信息 2,即只考虑用户偏好相似性,称之为 IAlgo 方法;第 2 组考虑信息 2 和 3,即综合考虑用户的偏好相似性及交互频繁性,称为 IIAlgo 方法;最后一组则考虑全部信息称之为 PIIAlgo 方法,于是得到了 3 组 nDCG,结果图 4 所示.从图中的实验结果可以看出,总体看来,IIAlgo 计算的匹配度优于 IAlgo,PIIAlgo 计算的相似度优于 IIAlgo,亦即计算用户匹配度时,考虑的因素越多,匹配度越准确.然而,第 5 组数据中 IIAlgo 算法的准确性低于 IAlgo 算法,可能是由于该用户与目标用户之间的消息交互量比较少,使得依据用户之间的交互度和兴趣相似度综合计算的关系强度低于其他值.

3 种方法的 nDCG 盒图如图 5 所示,盒图能够方便、直观地显示数据的特征:最大值(max)、上四分位数(Q3)、中位数(median)、下四分位数(Q1)、最小值(min).通过盒图可以直观的识别数据集中的异常值点.图中 3 个盒图从左到右分别表示 IAlgo、IIAlgo 和 PIIAlgo 方法得出的 nDCG 向量.从盒图也可以得出 IIAlgo 计算的匹配度优于 IAlgo,PIIAlgo 计算的相似度优于 IIAlgo.

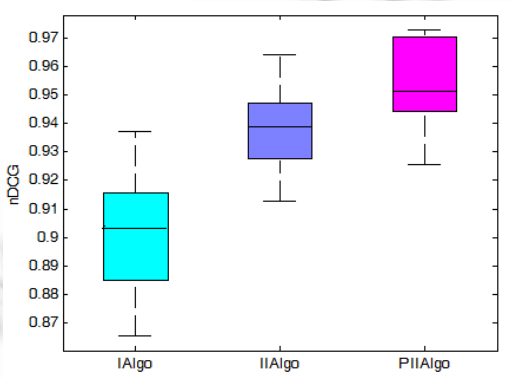


图 5 3 种方法的 nDCG 盒图

4 结束语

本文提出了一种内嵌人格分析的社交关系强度层次模型及其计算方法,将社会学和心理学中人际关系形成的原则引入到用户社交关系强度的计算中,将用户之间的社交关系强度分为人格倾向性、偏好相似性和交互熟悉性 3 个部分,充分提高了用户关系强度计算的合理性和准确性.除此之外,为了方便获取社交媒体中用户的人格特征,本文对社交网络中用户社交历史信息建模预测其人格特征,避免了传统人格特征提取的弊端,优化了关系强度计算.此外,使用用户人格特征信息还可以为用户提供更加个性化的服务,例如好友推荐、商品推荐、人机交互界面的定制等.

Reference

[1] Liu F, Lee HJ. Use of social network information to enhance collaborative filtering performance. Expert Systems with Applications, 2010,37(7):4772−4778.

[2] Singla P, Richardson M. Yes, there is a correlation from social networks to personal behavior on the Web. In: Proc. of the 17th Int'l Conf. on World Wide Web. 2008.

[3] Gilbert E, Karahalios K. Predicting tie strength with social media. In: Proc. of the 27th Int'l Conf. on Human Factors in Computing Systems. 2009.

[4] Xiang R, Neville J, Rogati M. Modeling relationship strength in online social networks. In: Proc. of the 19th Int'l Conf. on World Wide Web. 2010.

[5] Backstrom L, Huttenlocher D, Kleinberg J, Lan XY. Group formation in large social networks: Membership, growth, and evolution. In: Proc. of the 12th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. 2006.

[6] Lauw H, Shafer C, Agrawal R, Ntoulas A. Homophily in the digital World: A LiveJournal case study. Internet Computing, 2010,14(2):15−23.

[7] Jr. Costa PT, Mccrae RR. NEO Personality Inventory-Revised. SAGE Publications, 2005.

[8] Anderson C, John OP, Keltner D, Kring M. Who attains social status? Effects of personality and physical attractiveness in social groups. Personality and Social Psychology, 2001.

[9] Jensen-Campbell LA, Graziano WG. Agreeableness as a moderator of interpersonal conflict. Personality and Social Psychology, 2004.

[10] Golbeck J, Robles C, Turner K. Predicting personality with social media. In: Proc. of the 2011 Annual Conf. Extended Abstracts on Human Factors In Computing Systems. Vancouver, 2011.

[11] Quercia D, Lambiotte R, Stillwell D, Kosinski M, Crowcroft J. The personality of popular facebook users. In: Proc. of the ACM 2012 Conf. on Computer Supported Cooperative Work. 2012.

[12] Selfhout M, Burk W, Branje S, Denissen J, van Aken M, Meeus W. Emerging late adolescent friendship networks and big five personality traits: A social network approach. Journal of Personality, 2010,78(2):509–538.

[13] Asendorpf BJ, Wilpers S. Personality effects on social relationships. Journal of Personality and Social Psychology, 1998,74(6): 1531–1544.

[14] Lu YB, Zhao L, Wang B. From virtual community members to C2C e-commerce buyers: Trust in virtual communities and its effect on consumers' purchase intention. Electronic Commerce Research and Applications, 2010, 346–360.

[15] 乔秀全,杨春,李晓峰,陈俊亮.社会网络服务中一种基于用户上下文的信任度计算方法.计算机学报,2011,34(12).

[16] Manning D, Raghavan P, Schütze H. Introduction to Information Retrieval. Cambridge University Press, 2008.

李艳兵(1987－),男,山西太原人,硕士,主要研究领域为普适计算,社会计算.
E-mail: liyanbing1987@gmail.com

朱珍民(1962－),男,博士,正研级高级工程师,主要研究领域为普适计算,嵌入式系统,服务计算.
E-mail: zmzhu@ict.ac.cn

叶剑(1974－),男,博士,高级工程师,主要研究领域为普适计算,情境感知.
E-mail: jye@ict.ac.cn