基于主题树的BBS论坛用户互动行为分析

胡雪娇 李 慧 马国栋

(首都师范大学教育技术系 北京 100048)

摘 要 用户之间的互动对研究BBS论坛有着重要意义。为准确描述BBS树状论坛用户之间的互动过程,引入了主题树的概念,并根据自主定义的主题广度系数W和主题综合深度系数D等统计指标,对主题中用户的互动情况进行了详细描述。研究结果表明:主题总帖数的大小并不能全面描述主题中用户的互动情况,用户互动频繁的主题帖不一定具有较高的主题总帖数。因此,根据主题广度系数W和主题综合深度系数D对用户的互动行为进行划分,得到5种分类结果。

关键词 BBS论坛,互动行为,主题树

中图法分类号 TP399 **文献标识码** A

Analysis of User's Interacting Behavior in BBS Forum Based on Subject Tree

HU Xue-jiao LI Hui MA Guo-dong

(Department of Educational Technology, Capital Normal University, Beijing 100048, China)

Abstract The interactions between users have great significance for studying BBS forum. The subject tree is introduced to accurately describe the process of interaction between users of BBS tree forum, at the same time the theme breadth coefficient W and the theme depth coefficient D are introduced in the paper to detailed describe the user interaction in the subject. The results show that the total number of posts does not fully describe the user interaction, and the post which has frequent user interaction does not necessarily have a higher total topic posts. Therefore, the user interaction is divided into five results according to the theme breadth coefficient W and the theme depth coefficient D.

Keywords BBS forum, Interactive behavior, Subject tree

1 引言

随着互联网和Web2.0的飞速发展,许多供人们交流和沟通的虚拟空间也随之产生,例如博客、BBS论坛等。BBS论坛是一种依托互联网网站的虚拟互动空间,是虚拟社区的一种表现形式[1-3]。在BBS论坛中,用户之间通过发帖和回帖的形式建立联系,形成虚拟社区中一种新的互动形式。由于虚拟社区与现实社区存在不同特性,国内外不少研究者都对BBS用户行为进行了相关研究。文献[4]通过研究帖出、帖入和主帖3种帖子的比例分布,将BBS的互动模式分为单中心互动模式、多中心互动模式、跨网互动模式、两两互动模式和宣告-阅读互动模式。文献[5]对BBS论坛用户的交互行为进行实证分析,通过考察复杂网络的人度、出度、度相关系数和簇系数等统计特性,发现隐式的虚拟社区网络是有向、不对称、异配的无标度网络。文献[6]通过BBS成员交互方式构建成员回复网络,并利用K-means聚类算法把BBS成员划分为5大类用户,且通过卡方检验证实5类用户之间存在特定的交互模式。

目前有关BBS用户互动的研究都仅限于宏观方面,这些研究得出的现象和规律并不能对BBS用户互动进行全面描述,所以本文从微观角度出发,以新浪树状论坛婚姻家庭版块的数据为研究对象,以用户之间回复关系构建的回复网络为基础,引入主题深度、主题广度等参数来描述用户的互动行为。通过分析主题广度系数W和主题综合深度系数D[7]综合评价主题的互动情况,并根据互动特征将用户的互动行为分为5类。

2 BBS论坛的结构描述

2.1 BBS论坛中帖子的树形结构

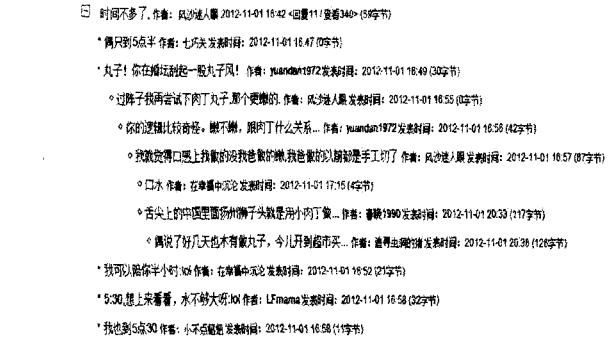

图1 婚姻家庭版块帖子列表的树形模式

本文受国家社会科学基金(10CTQ012),北京市教育委员会科技计划面上项目(KM201210028021),北京市属高等学校人才强教计划项目(PHR201108137)资助。

胡雪娇(1989—),女,硕士生,主要研究方向为复杂网络,E-mail:xiaofeng2440@163.com;李 慧(1977—),女,博士,副教授,主要研究方向为网络与计算智能;马国栋(1987—),男,硕士生,主要研究方向为复杂网络。

本文数据来源于新浪论坛婚姻家庭版块，其树形结构如图 1 所示，其中树根为主题帖，回复帖在被回复帖的下一行缩进显示。树状论坛可以清晰地表示主题帖和回帖之间的隶属关系，使用户快速地掌握该主题帖的讨论状态。

2.2 主题树及其参数定义

为了更简便地表达树状论坛的帖子结构，本文引入主题树的概念，将主题的互动情况用树状结构表示出来。在主题树中，每篇帖子都由一个节点表示，主题帖为根节点，其它子节点代表回帖，有向弧线代表回复方向，用于表示用户的回复指向。图 2 即为图 1 的主题树结构，横向表示层级，纵向表示该层级的回复，其中节点 B 指向节点 A 的有向弧线表示 B 对 A 的回复。

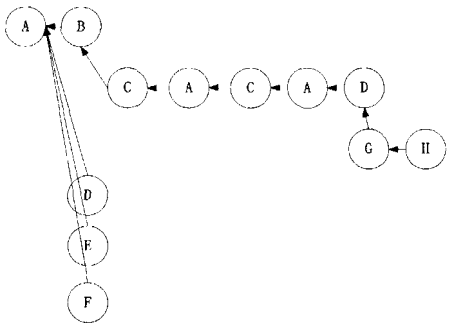

图 2 婚姻家庭版块帖子的主题树

根据主题树图，定义描述主题树图的 4 个参数：

- 主题参与度：主题树中的用户 id 数(不重复计数)。
- 主题总帖数：主题树中所有节点个数之和。
- 主题广度：主题树中与根节点直接相连的节点个数。
- 主题深度：主题树中从根节点到终端节点路径上的节点个数(包括重复节点)，节点数越多主题深度越深。每个节点属于一层，根节点是第 0 层。

在论坛用户的互动过程中经常出现主题总帖数相同、主题广度和深度不同的帖子，现有研究中往往把它们进行同质化处理[8-10]，但主题广度和主题深度表达的实际意义并不同。主题的互动过程是 BBS 论坛中用户传播信息的过程，对主题帖的回复与对回帖的回复是两种截然不同的互动行为，前者代表对主题的直接关注，后者则代表主题的深入讨论。因此，本文提出综合评价主题树中高层互动的 4 个参考指标：新用户数 N、老用户数 O、主题广度系数 W、主题综合深度系数 D。

(1)新老用户

对用户 id 进行重新编码，按照注册时间的先后顺序对用户进行排序，在爬取时间段内新注册的用户称为新用户，否则称为老用户。其中，新用户数用 N 表示，老用户数用 O 表示。

(2)高层回帖数量 R

将主题帖中 2 层及以上回帖定义为高层回帖，主题帖的高层回帖数量 R 越大，表明用户对主题的讨论越深入。

(3)主题层数 L

即主题所获得的最大回帖层级数，主题层数 L 越大，说明用户在主题高层讨论得越激烈。

(4)主题广度系数 W

主题广度系数定义为直接回复主题帖的用户数，即处于主题树中第 1 层的用户 id 数(重复发布的用户只统计一次)。在互动过程中，广度表示主题的实际传播效果，以主题帖为根的树形图的分支越多，表明该主题受到的关注越多。

(5)主题综合深度系数 D

对于主题树中高于第 1 层的主题帖，其发展情况多种多样，每个节点下都有可能增加新的节点，即使是同样的主题树结构也可能代表不同的互动行为。在主题树中与根节点直接相连的用户表示对主题的回复，这些用户是对主题感兴趣的直接受众。位于主题树中第 2 层及以上的节点表示对该主题回帖的回复，这些用户有可能是位于主题树中第 1 层的用户，他们针对别人的回复发表自己的看法；也可能是之前没有参与讨论的用户直接对某个回复的看法，还可能是主题的发布者对其他用户的回应等。

本文定义主题综合深度系数为一维向量 $D(O,N,R,L)$，用以综合评估主题中用户的互动情况。

3 基于主题树的 BBS 论坛用户互动行为分析

3.1 主题特征向量

根据树状论坛的结构特性，首先对所有论坛用户发布的主题数据进行预处理，删除重复数据和无效数据，并将整理好的数据存储到新的数据表中，对新数据表进行统计得到主体特征向量，描述如下：

(1)低层回帖用户数(d)：即第一层回帖用户的 id 数，用于描述主题的广度。

(2)高层回帖用户数(g)：二层及以上回帖用户的 id 数，如果主题有高层回帖用户则如实描述，否则全部为 0。

(3)高层回帖层级(l)：即用户在高层发表回帖时对应的层数。

(4)主题序号(i)：即按照帖子序号重新分布的主题 id。

3.2 主题综合深度系数

将上述特征向量输入一个四元数组 $hy=\{d,g,l,i\}$，把相同主题 id 的特征向量划分到 $i*4$ 的矩阵(i 代表主题总帖数)，然后将每个主题的 d、g、l 分别输入 A、B、C 3 个一维数组，计算主题综合深度系数 D。步骤如下：

(1)计算数组 A 中用户 id 数(不重复计算)，并将其记为主题广度系数 W。

(2)判断数组 C 是否全为 0，如果为 0，表示该主题没有二层及以上回复，则 $R=0$、$L=1$、$O=0$、$N=0$，结束计算。

(3)计算数组 B 中的新用户 id 数(不重复计算)，并将其记为 N 值。

(4)计算数组 B 中的老用户 id 数(不重复计算)，并将其记为 O 值。

(5)计算数组 B 中不为 0 的用户 id 数，将其记为高层回帖数量 R。

(6)找出数组 C 中的最大值，并将其记为 L。

(7)得出结果主题综合深度系数 $D(O,N,R,L)$。

3.3 BBS 论坛用户互动行为分类

(1)无互动行为：即用户所发表的主题帖没有得到任何回复，此时主题广度系数 W 与主题综合深度系数 D 取值均为 0。

(2)单中心互动行为：即主题的发表者处于中心位置，回帖者环绕在发帖者四周，与发帖者进行单向或双向的互动。此时，主题广度系数 $W>0$，主题综合深度系数 D 中 $L=1$、O、N、R 取值为 0，或 $L\geqslant 1$、O、N、R 取值不为 0。

(3)多中心互动行为：即主题中存在两个以上的用户处于

中心位置，其他用户的回复也围绕这些中心展开。这类帖子的大部分用户在高层进行互动，所以此类主题帖的特征是主题广度系数 W 值较小，主题综合深度系数 D 中 O、N、R、L 值较大。

(4)两两互动行为：即多用户参与的主题互动过程中用户之间两两互动较多。此时，主题广度系数 $W>0$，主题综合深度系数 D 中 R 与 L 的值要大于 O 与 N 之和(高层用户 id 数)。

(5)单向互动行为：即在回复关系中呈现单向回复状态，说明参与主题讨论的用户在发表自己的观点后便脱离讨论，不再进行关注。此时，主题综合深度系数 D 中 $O+N=R$。

4 实证分析

本文的数据来源于新浪论坛婚姻家庭版块和婆媳关系版块，分别简称为版块 1 和版块 2，两个版块的帖子信息如表 1 所列。

表 1 版块 1 和版块 2 的帖子信息

版块	主题帖数	用户数	时间
版块 1	4894	4011	2012 年 5 月－2012 年 8 月
版块 2	461	2220	2012 年 5 月－2012 年 8 月

利用算法可以计算出每个帖子的主题综合深度系数 $D(O,N,R,L)$，两个版块中各种情况的主题数量分布如表 2、表 3 所列。

表 2 基于 D 和 W 的主题数量分布

		版块 1			版块 2	
主题帖数			4894		461	
D=0			516		80	
	版块 1	版块 2		版块 1	版块 2
D≠0(L=1)	1006	239	D≠0(L>1)	3372	142
W=1	66	66	O>0,N=0	2307	34
1<W<10	626	158	O=0,N>0	172	56
W≥10	13	15	O>0,N>0	893	52

表 3 基于互动模式的主题数量分布

版块	无互动行为	单中心互动行为	多中心互动行为	两两互动行为	单向互动行为
版块 1	516	1006	760	1698	914
版块 2	80	239	37	23	82

如表 2 所列，版块 1 中大部分的主题回帖层级是大于 1 的，说明版块 1 中的用户在高层回帖中互动频繁，而版块 2 中 $L=1$ 的帖子数大于 $L>1$ 的帖子数，说明在版块 2 中用户对主题的直接关注较多，对主题的深入讨论较少。在版块 1 中，高层级的回帖中老用户的 id 数要远远大于新用户，老用户与新用户之间的互动也较少，表明该版块中老用户比新用户活跃，老用户之间的互动也较为频繁。而在版块 2 中，高层级的回帖中新用户的 id 数大于老用户，新老用户之间的互动也较多，这表明版块 2 中新用户的活跃程度较高，在注册该版块之后能够积极地参与主题讨论。

如表 3 所列，在两个版块中，单中心互动行为和两两互动行为在主题互动中占大多数，说明这两个版块中的大部分主题不能引起较多用户的关注。同样，多中心互动行为在两个版块所占比例较少，表明能够引起用户激烈讨论、频繁交流的主题在社区中并不多，且这类主题帖大部分是精华帖或是热帖。

详细分析用户之间的互动特性，并将用户之间的互动情况分为 5 类。

(1)无互动行为模式

示例选择版块 1 第 2 号主题帖和版块 2 第 3 号主题帖，D 值均为(0,0,0,0)，W 值也均为 0，如图 3、图 4 所示。

图 3 版块 1 无互动行为模式图例

图 4 版块 2 无互动行为模式图例

图 3 与图 4 中均只有一个节点，表明用户发表的主题帖没有得到任何关注，不能引起其他用户的讨论兴趣。

(2)单中心互动行为模式

对主题综合深度系数 D 的求解结果进行分析，如果 $L=1$，说明当前主题不存在 2 层及以上的回帖，表明该主题只具有广度上的讨论意义而没有值得探讨的深度。若主题广度系数 W 的值很大，说明这是个能够引起其他用户关注、但内容不值得深入讨论的主题。

示例选择版块 1 第 683 号主题帖(D 值为(0,0,0,1)，W 值为 16)和版块 2 第 193 号主题帖(D 值为(0,1,9,2)，W 值为 10)。从图 5 中可以看出所有用户都指向 id 号为 3605 的发帖者，表明该主题引起了一部分用户对该话题的直接关注，而在版块 2 的示例中，尽管 $L>1$，但在高层回复中是主题发布者对用户的回帖的回应，同样形成以主题发帖者为中心的单中心互动话题。

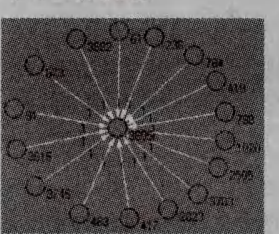

图 5 版块 1 单中心互动行为模式图例

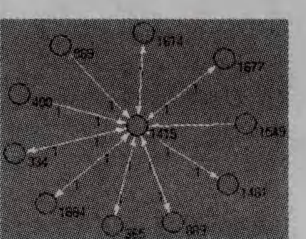

图 6 版块 2 单中心互动行为模式图例

(3)多中心互动行为模式

示例选择版块 1 第 3105 号主题帖(D 值为(10,0,71,11)，W 值为 5)和版块 2 第 283 号主题帖(D 值为(6,1,44,9)，W 值为 13)，如图 7 和图 8 所示。

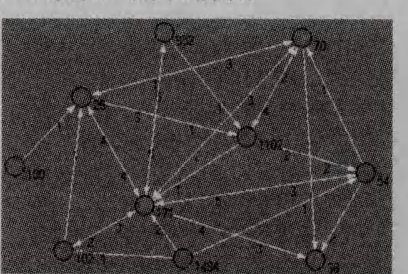

图 7 版块 1 多中心互动行为模式图例

从 D 值中可以看出在帖子高层回复中，用户数少，但是高层回帖数多，说明用户在高层互动频繁，帖子有可能出现了新的互动中心，脱离了原主题。如图 7 所示，节点 171 是发帖

人，但是与节点 70、54 的用户相连的节点也比较多，形成了以节点 161、70、54 为中心的主题互动，由于节点 70 和 54 的用户不是主题的发起者而是回复者，表明以节点 70 和 54 为中心的互动可能偏离了原主题中心，产生了新话题的互动。从图 8 中也可以看出，除了以 id 号为 1265 的发帖者为中心的互动以外，该主题还形成了以 id 号为(929，1265)、(1265，1326)等用户为中心的两两互动中心，同时还包括以(27，421，1265)、(121，772，1265)和(129，1265，1328)为中心的用户之间相互互动的情况。

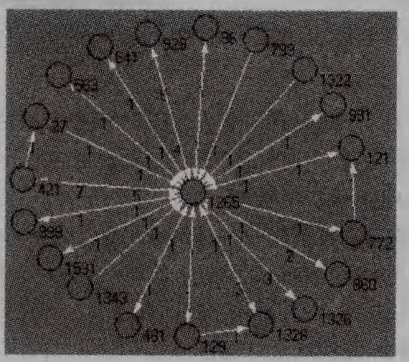

图 8　版块 2 多中心互动行为模式图例

(4)两两互动行为模式

示例选择版块 1 第 1617 号主题帖(D 值为(2，1，23，12)，W 值为 4)和版块 2 第 67 号主题帖(D 值为(0，1，2，2)，W 值为 2)，如图 9、图 10 所示。

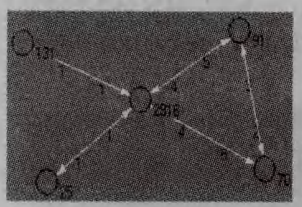

图 9　版块 1 两两互动行为模式图例

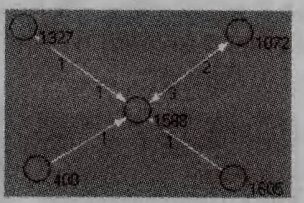

图 10　版块 2 两两互动行为模式图例

从图 9 和图 10 中可以看出，两个主题帖的高层回复中分别只有 id 号为 91、70、2916 和 1072、1327、1588 这 3 个用户参与，但是 D 值中高层回帖数 R 和主题帖层数 L 的值却相对较大，说明在高层回复中只有一个用户反复发帖，或是发帖者与回复者之间频繁互动。并且在互动过程中没有其他用户的加入，虽然这类帖子的主题帖数较大，在树状论坛中占取的版面也大，但是这类帖子并没有引起其他用户的关注。

(5)单向互动行为模式

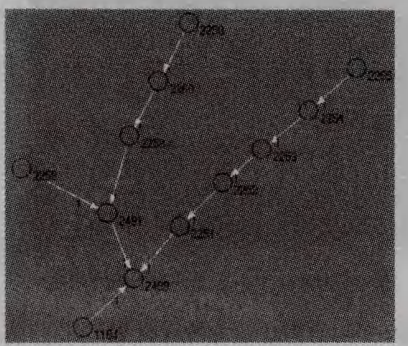

图 11　版块 1 单向互动行为模式图例

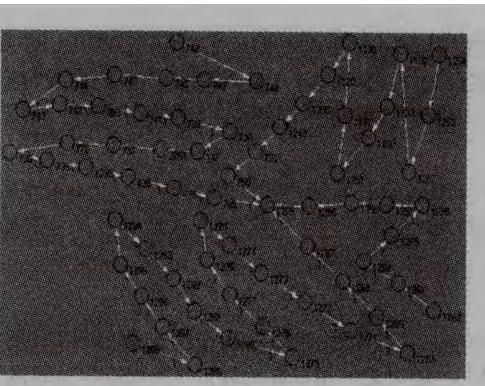

图 12　版块 2 单向互动行为模式图例

示例选择版块 1 第 2610 号主题帖(D 值为(0，8，8，5)，W 值为 3)和版块 2 第 290 号主题帖(D 值为(20，43，63，23)，W 值为 4)，如图 11、图 12 所示。

如图 11、图 12 所示，图中每个节点的回复关系都是单向连接，从 D 值中可以看出 $O+N=R$，即帖子中新用户的个数加上老用户的个数等于高层回帖数。说明参与这种帖子讨论的用户在该主题中只发过一个帖子，之后对这个帖子再也没有进行关注，且用户与用户之间也没有交流，这种主题具有时效性，只能够引起一时的关注，且主题不具争议性，不能引起用户之间的讨论。

结束语　本文通过树形论坛的特殊结构构建用户之间的回复关系，并引入主题树、主题广度系数 W 和主题综合深度系数 D，综合评价用户在主题上的互动情况。通过分析主题综合深度系数 D，得知主题总帖数很突出的主题，用户之间不一定有很高的互动，而在用户互动很频繁的主题中很可能出现偏离中心的讨论。因此，在分析 BBS 用户互动过程中，应着重把握主题是否具有较高的广度和深度、用户互动是否偏离主题中心，这对 BBS 网络信息管理起到积极的指导意义。

参 考 文 献

[1] Rheingold H, Community V. Homesteading on the Electronic Frontier [M]. Boston: Addison-Wesley, 1993: 2-15

[2] Scott A, Kuehn. Communication Innovation on A BBS: A Content Analysis [J/OL]. http://www.helsinki.fi/science/optek/1993/n2/kuehn.txt, 2013-02-07

[3] Ding F, Liu Y. Modeling opinion interactions in a BBS community[J]. European Physical Journal B - Condensed Matter, 2010 (78): 245-252

[4] 白淑英，何明升. BBS 互动的结构与过程[J]. 社会学研究，2003 (5): 8-18

[5] 司夏萌，刘云. 虚拟社区中人际交互行为的统计分析研究[J]. 物理学报，2011，60(7): 078903

[6] 荣波，夏正友. 基于聚类的 BBS 成员交互网络特性研究[J]. 重庆科技学院学报：自然科学版，2009，11(6): 165-169

[7] 井昱. Web 社区的互动分析[D]. 北京：中国人民大学，2007

[8] Kou Z, Zhang C. Reply networks of a bulletin board system [J]. Physical Review E, 2003, 67(3): 123-130

[9] 荣波，夏正友，朱永真，等. BBS 在线网络及其成员交互特性研究[J]. 复杂系统与复杂性科学，2009，6(4): 57-65

[10] 黄少华. 虚拟社区中的互动网络——以“广穆社区”为例[J]. 淮阴师范学院学报：自然科学版，2012，32(3): 400-409