

基于整体活性度模型的自适应分布式代理机制

孙永平, 马建国, 邢 玲

(西南科技大学信息工程学院, 四川 绵阳 621010)

摘 要: 将网络用户整体访问网络的行为视为个体用户兴趣的叠加, 以此建立用户整体活性度模型。基于该模型, 提出一种自适应分布式代理机制, 给出代理集群系统的结构与模型代理规则。实验结果表明, 与规则网络、聚集网络相比, 采用自适应分布式代理机制的网络对网络链接的处理能力较优。

关键词: 用户行为; 整体活性度; 分布式代理; 幂律分布

Adaptive Distributed Proxy Mechanism Based on Overall Activities Degree Model

SUN Yong-ping, MA Jian-guo, XING Ling

(School of Information Engineering, Southwest University of Science and Technology, Mianyang 621010, China)

【Abstract】 The overall behavior network users access the network is regarded as the superposition of the user interest, in order to establish the model on the degree of overall users' activities. Based on the model, an adaptive mechanism for distributed proxy is proposed, and the structure of the cluster system and model proxy rules are given. Experimental results show that the capacity of adaptive mechanism for distributed proxy network dealing with network links is better than regular network and gathering network.

【Key words】 user behavior; overall activities degree; distributed proxy; power-law distribution

DOI: 10.3969/j.issn.1000-3428.2012.02.030

1 概述

网络行为具有高度的复杂性, 但过去在社会、生态系统的研究中, 常把人的行为简化为简单的、随机的泊松系统。现今, 一些关于人类的行为统计发现, 人类行为存在长时间的静默与短时间的高频爆发现象。如文献[1-2]对电子邮件和平邮的统计, 这些统计特征说明人类的许多行为是不能简单地使用泊松分布来描述的。文献[3]基于单个个体提出广泛的人类动力学兴趣模型。文献[4]是从网络整体路由角度分析了网络相变的成因与分布式代理机制, 缺乏个人活性因素对整个网络表现的整体性能探讨。文献[5]讨论了一种基于用户行为和会话的 Web 应用测试方法。本文结合用户整体活性因素对人类行为的影响, 提出一种基于用户整体活性度模型的自适应分布式代理机制。

2 用户整体活性度模型

2.1 模型规则

个人兴趣可能促使人的行为出现准周期震荡变化。文献[6]指出人行为的周期性可能会使人类行为时间间隔出现胖尾分布。网络性能可以看成是成千上万的网络用户个体在网络上综合作用的结果。本文把网络用户整体访问网络的行为看作是由个体用户兴趣叠加导致的, 以此定义用户整体活性度。

定义 1 用户个体兴趣度(Degree of Individual User's Interest, DIUI): 访问网络的每一个个体在访问中所表现出的喜好程度, 记为 I 。

定义 2 用户整体活性度(Degree of Overall Users' Activities, DOUA): 访问网络的用户, 其个体访问兴趣度之和叠加定义为整体活性度, 记为 U 。

$$U = I_1 + I_2 + \dots + I_n \quad (1)$$

其中, n 为个体数量。

假设:

(1)把网络行为表现量化为离散时刻的集合。对某种网络用户整体行为, 用户整体的活性度 U 在某一时刻的表现都会影响下一个时刻该行为的概率。考虑最简单的情况, 即每一次概率改变的比例是相同的。

(2)整体表现的离散时刻存在 2 个阈值, 当间隔过小或过大, 达到一个值时会自动改变特定行为的概率变化趋势逆转为反方向, 会使做该事的概率以固定比例减小, 反之则使概率以相反比例增大。

据此, 模型规则^[3]为:

(1)时间为离散的, 在 t 时刻表现为某种行为的概率为 $s(t)$ 。

(2)在每个统计时刻, $s(t)$ 要进行一次更新, 更新规则为: $s(t+1) = A(t)s(t)$, 其中, $A(t)$ 只有 2 个值: 如果这一个时刻点与上一时刻点时间间隔小于或等于 T_1 , 则 $A(t) = A_0$; 如果该时间间隔大于 T_2 , 则 $A(t) = A_0^{-1}$; 对于其他情况, $A(t) = A(t-1)$; 其中, T_1 、 T_2 为固定的正整数; A_0 为固定的参数, 并且分别满足 $T_1 \ll T_2$, $0 < A_0 < 1$ 。

基金项目: 国家“863”计划基金资助项目(2007AA01Z151); 国家自然科学基金资助重点项目(60932005)

作者简介: 孙永平(1984—), 男, 硕士研究生, 主研方向: 网络数据分析, 智能建模; 马建国, 教授、博士生导师; 邢 玲, 副教授、博士

收稿日期: 2011-04-25 **E-mail:** sunyongping503@163.com

2.2 数值模拟

根据以上的规则, 网络整体行为一次表现与上一次表现的时间间隔小于或等于 T_1 的情况大多出现在当 $s(t)$ 接近或者等于 T_1^{-1} 时。因此, T_1 的取值决定了 $s(t)$ 所能达到的最高值; 同样的 T_2 的取值决定了 $s(t)$ 所能够达到的最小值在 T_2^{-1} 附近。为简单起见, 本文固定 $T_1=1$ 。

在模型的数值模拟中, 取 $s(t)$ 的初始固定值为 1.0。通过数值模拟, 对于 $T_2=1000$, 发现曲线逼近幂律分布, 如图 1、图 2 所示, 其中, t 是不同时间间隔的表示; $p(t)$ 是相应不同时间间隔下的行为概率。

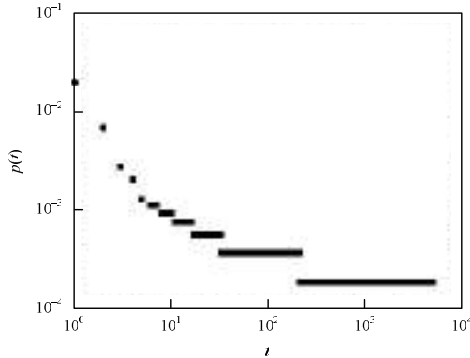


图 1 行为概率与时间间隔的分布

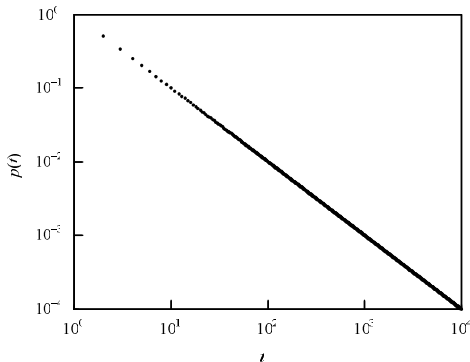


图 2 幂指数为-1的幂律分布

2.3 模型解析

在这个模型中, $s(t)$ 只能在一定范围内变化, 而且变化具有准周期性, 仅考虑一个周期内 $s(t)$ 减小的阶段, 假设 s_m 为该周期初始的(也是最高的) $s(t)$ 值, 显然其取值在 T_1^{-1} 附近, 但在该减小阶段的第 i 次事件后, $s(t)$ 值为 $s_m A_0^i$, $i=1,2,\dots$, 则在第 i 次和第 $i+1$ 次 2 次之间的时间间隔为 t 的概率为:

$$p(t) = (1 - s_m A_0^i)^{t-1} s_m A_0^i \quad (2)$$

在这半个周期内的相邻 2 个时刻之间时间间隔 t 的平均概率为:

$$P(t) = I^{-1} \sum_{i=1}^{I-1} (1 - s_m A_0^i)^{t-1} s_m A_0^i \quad (3)$$

其中, I 表示在这半个周期所包含的事件次数; 当 $T_2 \gg T_1$ 时, 可以得到 s_m 的平均值为:

$$\langle s_m \rangle = \sum_{k=0}^{I-1} s_0 A_0^k \omega(k) = \sum_{k=0}^{I-1} (s_0 A_0^k)^2 \prod_{i=0}^{I-k-1} (1 - s_0 A_0^{I-i}) \quad (4)$$

在此函数下, 赋以 $T_2=1000$ 、 $s_0=1$ 、 $A_0=0.5$ 即可得到与式(3)类似的结果, 两者基本一致。

3 自适应分布式代理机制

简单的应用层超链接系统如图 3 所示。以 A 节点为例,

链接向它的节点假设为 $1, 2, \dots, i$, 有 i 个节点, 而它链接的节点假设有 j 个, 分别是 $1, 2, \dots, j$, 超链接节点输入及输出间的这种关系造就了非线性的特性。随着链接数的加大, 会造成资源网络拥塞而不能正常工作。

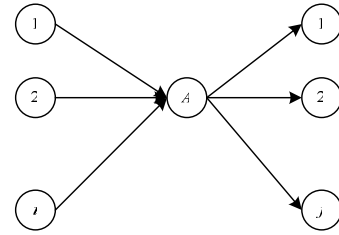


图 3 简单的应用层超链接系统

实际用户的网页访问需求统计如图 4 所示。本文所分析的数据源于西南科技大学校园网总出口。包含整个校园网用户群, 规模近似一个小社区, 数据具有一定的样本代表性。

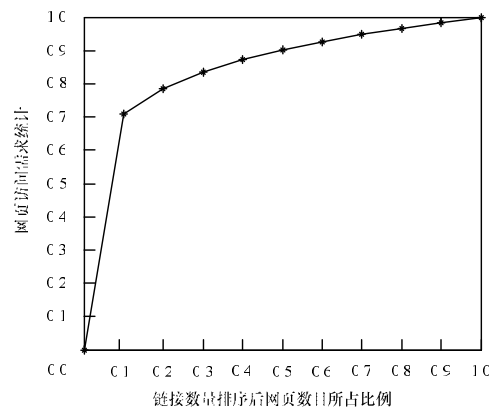


图 4 实际用户的网页访问需求统计

通过对 10 天的日志数据包含 39 272 730 条记录的统计分析发现, 超文本链接中 20% 的资源节点吸引了近 80% 用户的网络行为, 具有显著的聚集特性。

3.1 自适应代理集群系统的结构

自适应代理集群系统的结构如图 5 所示。

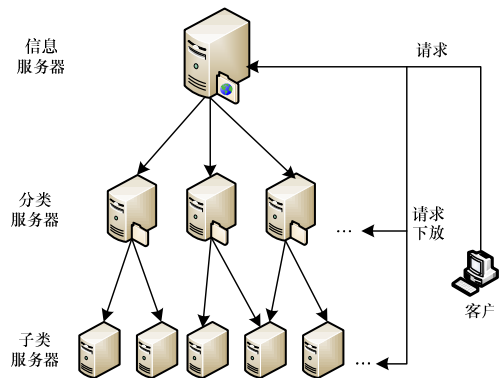


图 5 自适应代理集群系统的结构

代理结构是分层次的结构, 当一个客户进行服务请求时, 首先给出应答的是第 1 代信息服务器, 如果第 1 代信息服务器响应的客户群正处于客户访问某一类信息的上升期而且接近活性最大值时, 第 1 代信息服务器将响应权利下放, 由第 2 类已经将信息进行归类后的分类服务器来对客户进行应答, 一方面可以缓解第 1 代信息服务器的访问压力, 另一方面使用分类代理服务器的一个好处是可以多个子服务器满足远距离用户的访问效率, 以此类推, 当第 2 代分类服务器出现类似情况时, 将客户的请求继续下放, 下放到第 3 层

子类服务器，以此类推。

3.2 模型代理规则

为简单起见，定义任何2个资源节点都是等性能的，而且节点与节点之间的链接也是等概率随机，如果定义网络规模为M，则网络就有M²个节点，而且每一个节点与其上下左右节点距离相等，且每个节点都可以链入、链出和存储。规则网络的相邻节点示意图如图6所示，其中，不同的阴影表示相邻的节点。

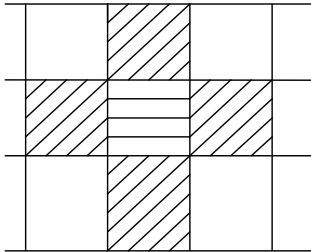


图6 规则网络中的相邻节点示意图

定义3 网络规模：表示网络中每行或每列的节点数，记做M。

上述规则网络模型中节点度值均为4，近似看作路由拓扑网络。在这种简单模型基础上刻画超文本链接网络的链接关系，并且带有驱动网络的用户和资源节点，假设密集的资源 and 用户节点落在核心路由器网络中，并且资源节点与核心路由器节点间由主干链路相连，在第1种规则网络上添加了这种用户访问超文本链接资源节点和网络拓扑关系的网络可以近似为简化的聚集网络^[4]，用户网络度表现为幂律分布^[7]。

分布代理网络的节点示意图如图7所示，假设最初只有A节点吸引了区域内所有用户的访问链接。

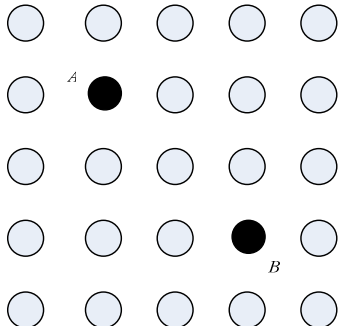


图7 分布代理网络的节点示意图

采用代理机制，现将B节点设置为A的代理。那么根据数据转发就近原则，图7中靠近原来以A为目的并且靠近B的链接就有部分路由到了B节点，增加了代理后，不仅使原来少数节点得到缓解，而且使原来以A为目的的节点的链接在空间上产生了分流，降低了A及其附近节点的负载。并且代理节点以及代理量的选取是根据用户访问的活性度U来度量的。假设处于活性低点时，U的值为0，处于活性高点时，U的值为1，也就是U始终在0~1之间波动，同时也代表了服务器受访问容量与活性容量之间的百分比。代理启动的先决条件是当用户的整体活性度U超过0.5，而A、B同时也代表A、B2个节点参与的链接数量，具体的代理函数为：

$$B = \frac{U}{2} \times A \tag{5}$$

当用户的活性U为0时，不代理，当U达到0.5时，启动代理，U代理原始A链接数据量的1/4，当U到达最大值1时，代理达到最大值，此时代理刚好为1/2，这时A、B平分分担链接数。

4 实验结果与分析

本文比较规则网络、聚集网络以及本文的自适应分布式代理网络这3种不同网络的性能。假设网络的网络规模M=10，每个核心路由器周围分布着10个用户节点，同时U取0.5。3种网络对超链接的处理能力比较如表1所示，X代表的是归一化后超链接节点的链接数目，表中数据是在此链接数目下3种网络的无法及时响应链接的概率。

表1 3种网络对超链接的处理能力比较

网络	无法及时响应链接的概率										
	X=0.0	X=0.1	X=0.2	X=0.3	X=0.4	X=0.5	X=0.6	X=0.7	X=0.8	X=0.9	X=1.0
规则网络	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.7	0.9	1.0
聚集网络	0.0	0.0	0.0	0.1	0.3	0.4	0.5	0.7	0.8	0.9	0.9
自适应 分布式网络	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.3	0.5	0.8

可以看出，在链接数目小于0.3时都差不多，但是当链接数目大于0.5时，自适应分布式网络滞留无法响应的链接数目明显小于规则网络和聚集网络，网络性能得到了很大的提高。

5 结束语

本文提出一种基于整体活性度模型的自适应分布式代理机制。实验结果证明，该机制分散了节点的链接压力，在一定程度上减少了节点能力与实际负荷之间的不匹配尺度，改善了网络的性能。下一步将从区域化、小集团化方向进行系统研究，进而得出更精确、细致的定量结果。

参考文献

- [1] Barabasi A L. The Origin of Bursts and Heavy Tails in Human Dynamics[J]. Nature, 2005, 435: 207-211.
- [2] Oliveira J G, Barabasi A L. Darwin and Einstein Correspondence Patterns[J]. Nature, 2005, 437: 1251-1252.
- [3] 韩筱璞, 周涛, 汪秉宏. 基于自适应调节的人类动力学模型[J]. 复杂系统与复杂性科学, 2007, 4(4): 1-5.
- [4] 李一鹏, 任勇, 袁坚, 等. 网络相变的非线性成因与分布式代理对策[J]. 清华大学学报: 自然科学版, 2009, 39(8): 1131-1134.
- [5] 武晋南, 高建华. 基于用户行为和会话的Web应用测试方法[J]. 计算机工程, 2010, 36(8): 83-85.
- [6] Hidalgo C A. Conditions for the Emergence of Scaling in the Inter-event Time of Uncorrelated and Seasonal Systems[J]. Physica A: Statistical Mechanics and its Applications, 2006, 369(2): 877-883.
- [7] 王磊, 周淑华, 袁坚, 等. 虚拟网络行为对互联网整体特性的影响[J]. 物理学报, 2007, 56(1): 36-42.

编辑 顾姣健