

移动社交网的生命周期评估模型研究

赵国锋 李 兵 徐 川 唐 红

(重庆邮电大学未来网络研究中心 重庆 400065)

摘 要 类似于自然界中的其它系统,移动社交网(MSN)也会存在一个从生到死的“生命周期”。文中着重研究 MSN 的生命周期及其评估模型,并探究处于不同生命周期阶段的 MSN 用户行为特征。提出了一个基于波士顿矩阵的移动社交网生命周期模型。该模型通过分析用户点击流数据,判断某个移动社交网站所处的生命周期阶段:幼年期、成长期、成熟期或衰老期。模型从移动互联网的用户访问行为中提取出两个参数指标,分别表征竞争地位与竞争实力,避免了以往需要通过长期市场调研的方法来判断分析对象所处生命周期阶段的困难,可以快速有效地判断一个移动社交网所处的生命周期阶段。根据实地采集的某移动互联网运营商的数据集,分析和验证了文中模型的有效性。此外,从6个方面分析了移动社交网用户在不同生命周期阶段的行为特征。

关键词 移动社交网;用户行为;生命周期理论;波士顿矩阵;数据挖掘
中图法分类号 TP393 **DOI号** 10.3724/SP.J.1016.2013.00727

Research on Lifecycle of Mobile Social Network

ZHAO Guo-Feng LI Bing XU Chuan TANG Hong

(Network and Computation Research Center, Chongqing University of Posts and Telecommunications, Chongqing 400065)

Abstract Similar to other systems in nature, Mobile Social Network (MSN) also have a lifecycle from birth to death. This research is on MSN lifecycle model. Also, user behavior characteristics of MSN at different life stage are explored. This paper proposes a BCG matrix based lifecycle model of mobile social network. By analyzing the click-stream set, this model can identify the right life stage for a MSN: infancy, adolescence, adulthood or aging. The method, which can identify the right phase for a mobile social network site simply and effectively, reduces difficulty and complexity of acquiring market share and sales growth of a product by the traditional BCG matrix. According to the real-data set of a mobile Internet operators, we analyze and verified the effectiveness of the proposed model. In addition, we analyze user behavior in different phase of the lifecycle from 6 aspects.

Keywords mobile social network; user behavior; lifecycle theory; Boston matrix; data mining

1 引 言

移动社交网(Mobile Social Network, MSN)是指用户使用无线移动终端访问社交网站,以实现与

好友互动与信息传输的新型社交网络。移动社交网是 社会网络服务与移动互联网不断发展并结合的产物,它将社会计算与移动计算无缝地结合起来^[1]。

随着移动互联网技术的迅速发展,国外 Bebo、MySpace 和 Facebook 等早期社交网络已经实现了

收稿日期:2011-12-08;最终修改稿收到日期:2013-03-06。本课题得到国家“九七三”重点基础研究发展规划项目基金(2012CB315806)、重庆自然科学基金重点基金(CSTC. 2009BA2089, CSTC. 2012JJB40008)资助。赵国锋,男,1972年生,博士,教授,主要研究领域为未来互联网、移动互联网和网络管理。E-mail: zhaogf@cqupt.edu.cn。李 兵,男,1986年生,硕士研究生,主要研究方向为移动互联网与社交网络。徐 川,男,1981年生,博士,主要研究方向为移动互联网。唐 红,女,1957年生,教授,主要研究领域为互联网。

向 MSN 的过渡. 在 3G 网络和智能手机的共同作用下, 国内的移动社交网进入一个快速发展期, 大量的 MSN 网站涌现, 但仅有小部分得以存活并能够继续发展. 在自然界中, 植物、动物乃至人类, 以及各种机构系统, 如企业、商品、软件等等, 都有自己的生命周期. MSN 作为人类设计与开发的系统, MSN 也会存在着一个从生到死的“生命周期”, 且生命周期的演变是由其用户行为决定的.

针对 MSN 生命周期的研究具有重要的意义: (1) 理解 MSN 的生命周期规律, 有助于更好地设计和运维 MSN, 避免发展的盲目性. (2) 从管理角度来说, 有助于 MSN 服务提供者识别其系统所处阶段, 采取相应的措施来管理并调整 MSN 系统的生命周期, 以获取更多的利益. (3) 社交网络的核心是用户行为, 将生命周期与用户行为的研究相结合, 有助于深入研究用户在不同生命周期阶段所表现出来的行为特征.

但是, 当面对市场上大量的、不同的 MSN, 如何找到一种简便可行且有效的方法, 能够对它们所处的生命周期阶段进行评估, 以便为市场、商业、管理、投资分析等服务, 是本文的目标. 因此将着重研究 MSN 的生命周期评估模型, 并探究不同生命周期阶段 MSN 用户的行为特征.

波士顿矩阵^①(BCG matrix)是一种经典的、研究产品、系统、企业等生命周期理论的模型, 通过划分矩阵空间来区分对象的不同阶段. 虽然该模型没有深厚的理论支持, 但是被广泛应用到市场、工业、经济管理等诸多领域的决策分析. 本文也将基于 BCG 矩阵模型来研究 MSN 的生命周期.

在传统的 BCG 矩阵模型中, 利用市场占有率与市场增长率两个参数, 分别代表研究对象的竞争地位与竞争实力并据此来划分矩阵空间. 在过去的应用中, 两个参数的取得主要是通过市场调研方法, 但是对于 MSN 来说, 行业发展时间较短, 且用户具有极强的动态性, 很难利用传统的市场调研手段来快速准确地获得一个 MSN 的市场占有率与市场增长率. 为解决该问题, 本文思路是首先研究分析 MSN 的用户访问行为特性; 然后从中提取出两个参数指标, 作为竞争地位与竞争实力的代表; 最后据此建立 MSN 的生命周期评估模型, 判定各个 MSN 网站所处的成长阶段.

本文的主要工作有:

(1) 在传统 BCG 矩阵模型的基础上, 提出了适用于 MSN 的生命周期模型, 用来评估判断某个

MSN 所处的生命周期阶段.

(2) 基于从某移动互联网运营商的 WAP 网关上采集的实际点击流数据, 利用文中模型对当前几个主流的 MSN 进行分析, 检验模型的有效性.

(3) 以上述的几个 MSN 为对象, 从 6 个方面研究分析其用户行为特征, 揭示了不同成长阶段的 MSN 用户行为特征.

(4) 刻画了 MSN 生命周期转移图, 并为 MSN 管理者提出了一些建设性的意见与策略, 以便合理有效地利用 MSN 生命周期规律来增加运行效益.

本文第 2 节描述 MSN 生命周期的各个阶段及其转移图; 第 3 节给出一个基于 BCG 矩阵的 MSN 生命周期模型; 在第 4 节, 基于实地采集的运营商数据集, 检验模型的有效性并对结果进行分析; 在第 5 节, 通过对数据集的分析, 揭示不同成长阶段的 MSN 用户行为特征; 第 6 节分析相关工作; 第 7 节总结全文.

2 MSN 的生命周期

2.1 BCG 波士顿矩阵与生命周期

波士顿矩阵最初由 Bruce Henderson 在 1970 年提出, 并于 1983 年由波士顿咨询集团发展而来, 用于企业的决策支持分析. 该矩阵中 x 轴作为企业产品的相对市场占有率, y 轴作为其市场销售增长率, 将企业产品划分为 4 个象限: 问题类、明星类、奶牛类和劣狗类. 近几十年来, BCG 矩阵作为分析工具被广泛应用到多个研究领域, 刻画产品、应用、企业等的生命周期. 同理, 我们应用它来刻画 MSN 的生命周期模型.

2.2 MSN 的生命周期转移

MSN 作为一种由人类创造出来的事物, 必将遵循其自然发展规律, 如图 1 所示, 也将经历幼年、成长期、衰老期直至消失的基本生命周期. 但 MSN 毕竟不是生物体而是一个人为的系统, 它并不具有必然会走向衰老并死亡的特性; 如果 MSN 的管理者能采取恰当的措施, 就可能延缓它们死亡的时间, 甚至使其不断恢复活力并取得再次成功, 这也正是研究 MSN 生命周期模型的目的.

MSN 的用户可分为 3 类: 随意(Easy-going)用户、指那些活跃度(流量、点击量)相对很低的用户; 普通(Ordinary)用户指活跃度一般的用户; 深度

^① http://en.wikipedia.org/wiki/Bcg_matrix

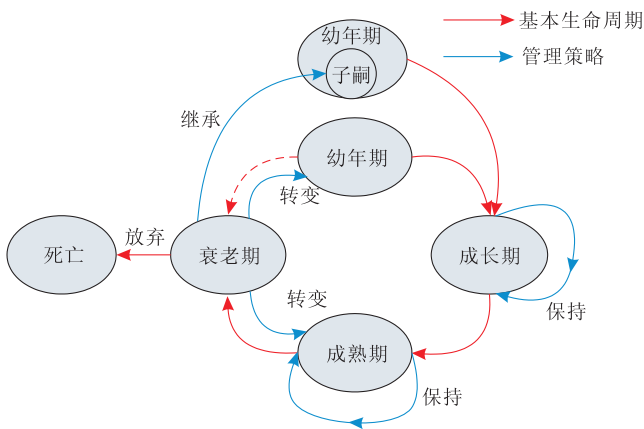


图 1 MSN 生命周期转移图

(Heavy)用户,指那些活跃度很高的用户(点击量较大,流量也明显地高于其它用户)。

MSN 生命周期包括 4 个阶段:

(1) 幼年期. 对应于 BCG 矩阵中的问题类, 开始于 MSN 网站的发布, 结束于 MSN 进入成长期. 多数情况下, 此阶段的 MSN 拥有较快的用户增长速率, 但用户群还很小, 利润较低. 该阶段的用户主要由 Easy-going 和 Heavy 用户组成, 而 Ordinary 用户较少. 处于此阶段的 MSN, 具有发展的未知性: 如果有较高的用户增长速率, 并获得较大的用户群, 将会进入成长期; 否则会直接进入衰老期, 如图 1 虚线所示. 此阶段的 MSN 网站需要进行大量的投资与网站内容及结构的改进, 如何保持或取得更快的用户增长率是管理者在该阶段要考虑的主题.

(2) 成长期. 对应于 BCG 矩阵的明星类, 拥有较大规模的用户群、较高的市场份额. MSN 的收益开始从投资向盈利转化, 处于生命周期的最佳状态. 在用户组成上, 3 类用户所占有的比例相对较均衡. 此时管理者应加大对 MSN 投资并保持用户增长率, 因为即将获取更大的利益.

(3) 成熟期. 对应于 BCG 矩阵的奶牛类, 其用户市场份额比较大, 收益率也最好. 3 类用户所占有的

的比例非常均衡, 但用户市场增长低迷, 管理者应尽可能维持现状. 如何延长 MSN 成熟期的时间以产生长期的高利润收益, 成为这阶段的主要目标.

(4) 衰老期. 对应于 BCG 矩阵的劣狗类, 此阶段的用户市场占有率与增长率都处于较低状态, 收益直线下滑, 甚至亏损. 在用户组成结构上, Heavy 用户占有率很高, Ordinary 与 Easy-going 用户很少. 管理者需要对运营状况进行认真检查, 可以考虑将此 MSN 融入一个全新的 MSN, 作为其“子嗣”以利用已有的市场份额, 或者直接放弃.

3 MSN 生命周期建模

3.1 MSN 用户行为特性分析

MSN 不同于传统非移动的社交网络, 而 MSN 网站也不同于非社交类的普通移动互联网网站.

与传统互联网相比, 移动互联网的用户行为在各个方面都有很大的差异^[2]. 而相比于非移动的社交网络, 使用移动终端的 MSN 首先在网络架构方面有明显区别, 如图 2 所示. 它具有高度分布、自治、拓扑动态变化等移动无线网络的特性, 用户的虚拟交友空间也由“以计算机为中心”向以“用户为中心”转变. 其次, MSN 用户访问的视觉界面、网络接入方式、网络协议等有别于基于 Web 的社交网络. 此外, 移动社交网络具有: (1) 感知的实时连续性. 用户情境数据可实时、连续地得以反馈. (2) 数据的真实时空性. 即相对于传统社交网络主体的在线虚拟行为而言, 移动社交网络主体更趋于真实世界人类行为参与的社会交互. (3) 服务的即时即地性. 移动社交网络能提供包括基于位置推荐服务的即时即地服务. (4) 社群在特定情境下演化, 移动社交网能动态演绎群体网络进化模型等特征. 现有研究报告^[3]表明: 用户通过手机等移动终端访问社交网的频率较非移动的社交网用户更高, 更具粘性.

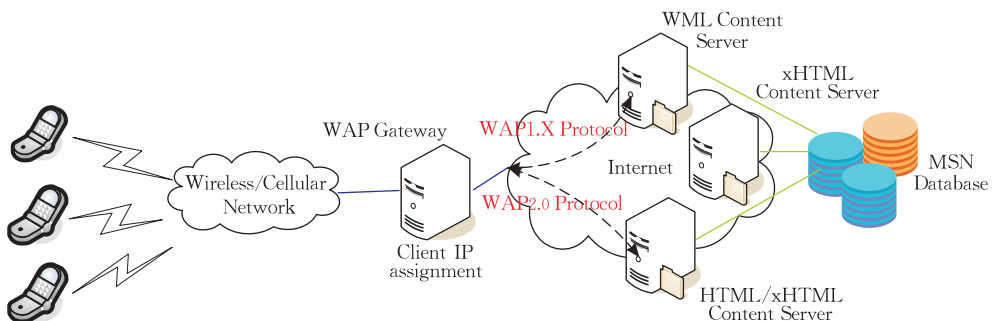


图 2 移动终端访问 MSN 内容服务器框架图

再者,对非社交类的普通移动互联网(如移动门户、校企官网等),其网站内容对所有用户基本上都是公共且一致的,用户行为也大多是信息浏览。但是 MSN 用户在登陆之后将进入以个性化内容为主的虚拟空间,一方面是阅读,另一方面是回复、发布新信息、参与社交类游戏等。因此访问次数、会话次数会更多更频繁,且各不同 MSN 的用户行为也有较大的差异性。据文献[3],MSN 用户的平均访问频率、在线时长比非社交类移动网站更高,而且会话中具有更多互动,点击量也明显高于一般手机互联网站。

因此,考虑到 MSN 用户访问行为的上述特性,文中以会话数和会话中的请求数为基础建立 MSN 用户行为模型,其中会话数表征该 MSN 用户的访问频率,而会话中的互动及点击量则可以由会话中的请求数来表征;并进而对 MSN 生命周期建模。

3.2 MSN 用户行为模型

MSN 用户行为模型是用来刻画 MSN 网站服务器与其用户之间的信息交互,如图 3 所示,MSN 用户行为模型可划分为 3 个层次:会话层、事务层、数据包层^[4]。

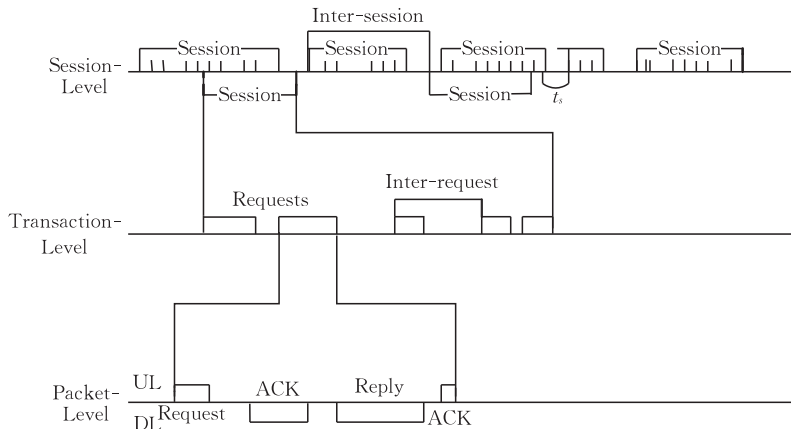


图 3 MSN 用户行为模型

(1) 会话层 (Session Level)

一个会话 (Session) 是一个用户打开移动终端浏览器后发起一系列的 MSN 访问请求,并阅读或下载文档,最后退出浏览器的过程。一个会话包括若干个以时间为顺序的访问请求,其中会话时长是指在同一个 Session 内,第一个与最后一个请求之间的时间间隔。

由于来自不同用户的会话在时间尺度上会有交叉,从服务器的角度来看,Inter-session 代表时间上连续的两个会话请求 MSN 服务器的时间间隔,尽管这两个 Session 可能会属于不同的用户。Benevenuto 等人^[5]指出:Inter-session 的平均时间值越小,表示社交网的服务器就越活跃,而当一个 MSN 服务器越活跃时,则说明该 MSN 网站会拥有相对更大的用户群。

(2) 事务层 (Transaction Level)

一个事务 (Transaction) 指包含于一个 Session 内的若干个时间上连续的访问请求。类似的,Inter-request 是指归属于同一会话中两个连续的 MSN 访问请求到达时刻的时间间隔。Inter-request 平均值越短,意味着用户在其访问的 Session 中越活跃。换言之,Inter-request 平均值越小,意味着用户对该 MSN 网站更加感兴趣^[5]。

(3) 数据包层 (Packet Level)

当用户发起一次访问请求时,移动终端的浏览器会向 WAP 网关发送一个 GET 请求(上行流量),随后该请求被 WAP 网关转送到 MSN 服务器。服务器接收到请求后,再通过 WAP 网关返回移动终端一个 Reply(下行流量),随后传送内容(下行流量)。

3.3 MSN 生命周期评估模型

利用 BCG 矩阵研究生命周期时,核心是确定 BCG 矩阵两个坐标轴所对应的指标。在研究产品的生命周期时,通常以市场占有率与增长率分别作为产品目前的竞争地位与竞争实力的代表指标,即分别作为 BCG 矩阵的 x 和 y 坐标轴。然而 MSN 用户具有极强的动态性,其用户市场占有率与增长率难以及时准确获得,因此将选取其它适合的指标来描述 MSN 的竞争地位与竞争实力。

基本思路:在建立 MSN 生命周期模型时,从 MSN 用户行为模型中提出相应的指标来描述 BCG 矩阵的两个坐标轴。在 3.1 节中提到:Inter-session 与 Inter-request 的平均值分别与 MSN 用户数和用户兴趣度密切相关。因此,我们将利用这两个指标来代表 MSN 的竞争地位与竞争实力。

假设对于任意的一个 MSN 网站,在观察周期

T 内, 总共接受到 N 个 Session, 其中 t_i 为第 i 个 Session 的到达时间, 则 $t_{i+1} - t_i$ 即为第 i 个与 $i+1$ 个 Session 之间的 Inter-session. 因此, 在周期 T 内的平均 Inter-session 值可表示为

$$\bar{t}_s = \frac{\sum_{i=1}^{N-1} (t_{i+1} - t_i)}{N-1} \quad (1)$$

同理, 假设在任意一个会话 s_i 中, 包括 M_i 个访问请求, 其中 t_j 代表会话 s_i 中第 j 个请求的到达时刻, 则 $t_{j+1}^i - t_j^i$ 即为第 j 与 $j+1$ 请求之间的 Inter-request, 对于整个周期 T 内的 N 个 Session, 其平均 Inter-request 值可以表示为

$$\bar{t}_q = \frac{\sum_{i=1}^N \sum_{j=1}^{M_i-1} (t_{j+1}^i - t_j^i)}{\sum_{i=1}^N (M_i - 1)} \quad (2)$$

于是, 可得到 $\alpha = 3600/\bar{t}_s$, $\beta = 3600/\bar{t}_q$, 其中 α 表示在观察周期 T 内, 到达 MSN 服务器的每小时会话数平均值; β 表示会话中每小时请求数目的平均值. 考虑到平均 Inter-session 与服务器活跃度的相关性, 即 α 值越大, 服务器则越活跃, 表示 MSN 所拥有的用户数也就更多. 而一个 MSN 的竞争地位与其拥有的用户数量正相关, 所以 α 能够很好地映射 MSN 的竞争地位. 同样的, 平均 Inter-request 与用户的活跃度之间存在相关性, 即 β 值越大, 表示用户访问的行为更加活跃, 也说明 MSN 服务器中的内容对用户具有更强的吸引力, 即 MSN 具有更强的竞争实力, 所以 β 可以用来反映 MSN 的竞争实力.

为描述 MSN 生命周期模型, 使用类似的 BCG 矩阵图; 其中 x 轴代表 MSN 的相对竞争地位, 用 γ 表示; y 轴代表 MSN 的相对竞争实力, 用 λ 表示.

假设有 N 个 MSN 网站, 对于第 i 个 MSN, 其参数 γ_i 与 λ_i 通过下式计算:

$$\gamma_i = \alpha_i / \max(\alpha_1, \alpha_2, \dots, \alpha_N) \quad (3)$$

$$\lambda_i = \beta_i \times N / \sum_{i=1}^N \beta_i \quad (4)$$

对于第 i 个 MSN, 式(3)中的 $\max(\alpha_1, \alpha_2, \dots, \alpha_N)$ 代表观测的 α_i ($i=1, 2, \dots, N$) 中最大值. 因此, γ_i 表示第 i 个 MSN 相对于占有最大用户市场的那个 MSN 的相对竞争地位(相对比率); 在式(4)中, λ_i 表示第 i 个 MSN 相对于所有 MSN 平均竞争实力的相对竞争力.

3.4 模型门限值的选取

传统的 BCG 矩阵主要面向市场分析应用, 由于

一个产品的年市场增长率达到 10%, 就会被认为非常高, 因此其矩阵对应的门限值 λ 设置为 0.1, 而根据经验 γ 一般取为 0.2. 如图 3 所示, 在本模型中, 由于 λ 表示一个 MSN 相对于所有 MSN 平均竞争实力的相对竞争力, 而平均竞争力对应 $\lambda=1$, 因此设定临界值为 1; 而 γ 根据传统应用经验依然取为 0.2. 然后就可以利用该模型对某个 MSN, 根据对应的指标来确定其在 BCG 矩阵中的位置, 明确其当前所处的生命周期阶段.

4 模型性能评估

本节基于从某运营商处采集到的移动互联网用户点击流数据, 分析 Qzone、kaixin001、51 社区和 renren 4 个 MSN 的生命周期.

4.1 数据集描述

原始数据采集于重庆市某一主流运营商的 WAP 网关, 记录时段为 2010.4.5~2010.4.11 共 1 周的 log 日志记录, 整个数据集共 130 GB, 包括了 80690 个用户的 17316616 条记录. 每条记录包括了请求时间、电话号码、客户端 IP、目的 IP、客户端端口、目的端口、上、下行流量、请求内容格式、DNS、URL、用户代理等信息.

通过对 MSN 网站 URL 的分析, 从中识别并提取了目前中国几个主流的 MSN 用户通过移动终端产生的点击流数据, 包括 Qzone、kaixin001、51 社区和 renren. 其详细的统计数据如表 1 所示.

表 1 各 MSN 点击流数据统计

MSN	用户数	会话数	HTTP 请求数	上行流量/ MB	下行流量/ GB
Qzone	9142	294383	2089745	374.77	9.6953
kaixin001	868	29851	214699	1.08	1.3052
renren	997	6243	135253	1.46	0.3841
51	169	2755	13974	0.21	0.0734
Total	11176	333232	2453671	377.52	11.4580

需要强调的是, 我们的数据集仅仅是重庆地区使用某一运营商的数据, 不是访问相应 MSN 网站的所有用户记录. 然而, 全重庆共有 3000 多万的人口, 经过该运营商接入移动互联网的用户量很大, 这些数据集虽然是局部的, 但是也能够揭示出 MSN 的现状. 分析报告^[6]指出: Qzone 依托即时通信工具 QQ 的庞大用户量, 用户市场份额最大, 而 renren 与 kaixin001 份额分别为 29% 与 26.83%, 两者之比约为 1.08. 从表 1 数据可以看出, renren 用户数为 997, kaixin001 用户数为 868, 两者的用户数之比约

为 1.1, 该结果说明我们的测量数据可以较准确地反映出当时 MSN 市场的实际状况。

我们的数据采集于 2010 年 4 月, 通过文中方法可以迅速得到当时各个 MSN 的生命周期发展状况, 并且结果与市场分析报告的结果极其相近。一方面本文方法可以比市场分析方法更快速地获得结果; 另一方面, 通过持续地进行网络监测与数据分析, 本文方法可以及时地了解各个 MSN 的变化状况, 且时间粒度可以更小, 其潜在的商业应用价值更好。这充分说明采用用户行为数据来分析 MSN 的生命周期是一个非常迅捷且有效的途径, 具有良好的市场应用价值。

4.2 模型应用

在数据处理时, 我们以电话号码来区分用户; 对于同一用户的不同 Session, 则用记录中的客户端 IP 地址来区分, 这是因为每次用户重新接入移动互联网时, WAP 网关都会给移动终端动态分配一个 IP 地址, 而同一用户的连续两个 Session 出现相同 IP 地址的概率很小。

基于每个 MSN 的用户数据集, (1) 设定观察时间周期 T 为 1 周, 计算得到两个参数: 平均 Inter-session 与平均 Inter-request. (2) 由式(1)~(4), 计算出每个 MSN 的 4 个参数值(α 、 β 、 γ 和 λ), 如表 2 所示。

表 2 MSN 的生命周期模型参数表

MSN	平均 Inter-session/s	平均 Inter-request/s	α	β	γ	λ
Qzone	2.61	18.33	1378.8	202.2	1.00	0.56
kaixin001	20.26	5.37	177.6	670.2	0.13	1.85
renren	80.68	12.52	44.4	288.6	0.03	0.80
51	219.08	12.58	16.2	286.2	0.01	0.79

刻画 BCG 矩阵并按临界值把矩阵空间分成 4 个象限, 然后按照表 2 的参数值, 确定每个 MSN 在 BCG 矩阵中的位置, 如图 4 所示。可以看出各

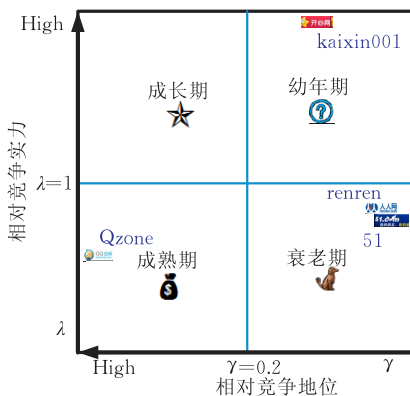


图 4 MSN 的生命周期识别图

MSN 中, kaixin001 处于幼年期, renren 与 51 社区正处于衰老期, 而 Qzone 处于成熟期。

4.3 结果分析

首先, 根据数据集, 刻画出访问各 MSN 的用户数, 如图 5 所示。可以发现: 81.8% 的用户访问 Qzone, 显然它占据了绝大部分移动社交网用户市场。此外, 它还占据了约 85% 移动社交网用户流量 (包括上行与下行流量), 如表 1 所示。由此说明, Qzone 在竞争地位上遥遥领先于其它的 MSN, 这也和当前中国的 MSN 现状相符^[3,6]。

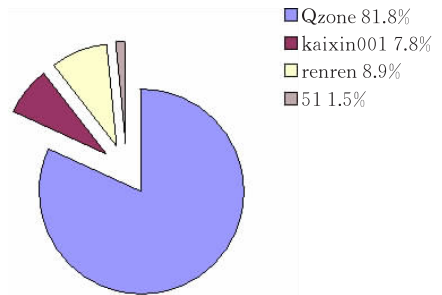


图 5 各 MSN 用户的比率

其次, 从图 6 发现一个有趣的现象, Qzone 的用户平均登录数 (会话数) 与请求数目却都少于 kaixin001。该现象的一个解释是: 在 Qzone 服务器上的内容并不如 kaixin001 服务器的内容对用户更有吸引力, 即虽然 Qzone 的用户群远远大于 kaixin001, 但是它的竞争强度却不如后者。在模型分析结果中, Qzone 处于成熟期, kaixin001 处于幼年期, 该现象与模型分析的结果相符。

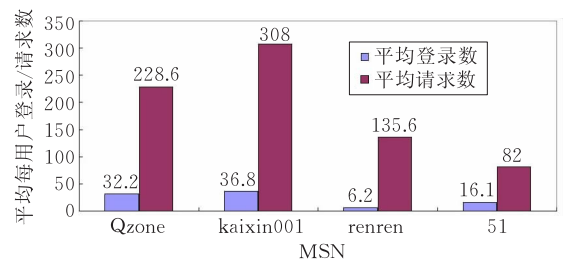


图 6 MSN 用户平均登录/请求数

最后, 根据历史调查^[6]可知, Qzone、51 社区、renren 网都是 2005 年进入移动互联网市场, 而 kaixin001 是 2008 年进入移动互联网市场。但是 renren 与 51 社区在用户市场占有率上远低于 Qzone; 虽然有更长的发展时间, 而用户占有率却与 kaixin001 相当或者更少。kaixin001 有很强的市场竞争力, 但其用户群还远远低于 Qzone。本文的 MSN 生命周期模型分析结果中, renren 与 51 社区处于衰老期, kaixin001 处于幼年期, 这与我们对

3 个网站的直接分析结果相吻合. 显然, 根据第 2 节的生命周期转移图, 如果 renren 与 51 社区意欲重新获得市场发展生命力, 需要注入新的内容、营销方法等革新手段.

4.4 模型讨论

该模型针对 MSN 建立, 基于 Inter-session 与 Inter-request 两个指标能较好地反映出服务器的活跃度与用户会话活跃度, 进而映射出用户占有率与网站吸引力的基础上, 模型要求分析对比的网站具有相同或相近的业务, 拥有共同的用户和市场. 此外, 还要求网站的吸引力通过自身, 而不依赖其外链接网站来体现, 所以模型还适用于其它一些类型移动网站, 但并不能普遍地适用于移动互联网任意类型的网站, 举例如下:

(1) 针对主要内容由外链接组成的手机导航类网站(如 wap. uc123.com、hao123.com、1616.net), 用户在导航网页的 Inter-request 越小, 反映用户对该导航网站外链接网站内容的兴趣度反而越低, 才可能会反复返回到该导航网站. 因为这些内容网站的吸引力是通过用户从该网站跳入外链接网站后的 Inter-request 来体现. 各导航网站的参数如表 3 所示, 3 个导航网站的用户数、Inter-session 值变化明显, 但 Inter-request 值相差太小, 应用文中模型难以进行区分.

表 3 各移动互联网网站信息统计

网站	用户数	平均 Inter-session/s	平均 Inter-request/s	流量/MB
wap. uc123	843	8.43	166.78	213.7
hao123	546	13.55	172.93	183.5
1616.net	239	19.83	159.55	103.4
qidian	72	41.77	26.74	51.3
3g.xs.cn	418	9.43	67.42	439.2

(2) 对于面向文本阅读为主的手机小说类网站, 不同网站对小说单页显示的字数可能不一样(例如, 同一部小说的某一章节在起点小说网 qidian.com 中分 5 页来响应和显示, 而在小阅网 3g.xs.cn 中只需一页即可全部响应并显示), 统计信息如表 3 所示, 小阅网的用户量与总流量都远大于起点网, 但其 Inter-request 却比起点大得很多, 这并不能说明起点网的吸引力就强于小阅网, 相反小阅网每用户的周平均流量(1.05 MB)反而高于起点网(0.71 MB), 因此应用文中模型并不能客观地进行结果评估.

对于该模型对其它类型网站是否适用, 限于篇幅不再一一讨论.

5 不同阶段的 MSN 用户行为特征

通过前面的 MSN 生命周期模型, 在识别出 Qzone、kaixin001、renren 及 51 社区分别所处的阶段后, 本节将研究处于不同生命周期阶段的 MSN 用户行为特征, 从用户在线时长、点击数、访问流量、会话长度、Inter-session 与 Inter-request 共 6 个方面进行分析和讨论.

5.1 MSN 用户访问特征

(1) 用户在线时长、点击数及流量特征

用户在线时长与点击数分别指用户在时间周期 T 内在某个 MSN 停留的时间总和与点击总次数; 用户流量是指用户点击某个 MSN 所产生的主网页流量(包括上行流量与下行流量).

根据我们的数据集, 分别作出了上述 4 个 MSN 用户在一天的平均在线时长、点击数和访问总流量的累积分布函数(CDF 函数), 如图 7 所示.

从图 7(a)可知, Qzone 的平均在线时长最大, renren 的平均在线时长最小, Qzone、kaixin001、51 和 renren 的在线时长超过 1000 s 的用户比例依次为 36.7%、25.2%、13.5% 及 10.8%, 据此可推测, MSN 处于成熟期的 Heavy 用户在线时间最长, 依次递减为幼年期、衰老期的用户. 同理对于在线时长小于 100 s 的情况, renren 的用户比例最高, kaixin001 和 51 的差异不大, Qzone 的比例最小, 据此说明衰老期的 Easy-going 用户比例大于成熟期的, 同时幼年期的此类用户比例也较大.

图 7(b)给出了各个 MSN 的用户点击次数的 CDF 曲线. 统计点击次数在 3 个区间中的用户比例, 即 $[1, 10]$ 、 $[11, 100]$ 、 $[101, 101 \text{ 以上}]$, 如表 4 所示. 假设这 3 个区间的用户分别对应 Easy-going, Ordinary, Heavy 这 3 类用户. Qzone 在 3 个区间的用户比例比较均衡, 说明处于成熟期的 MSN 的 3 类用户所占比例相当; kaixin001 在区间 $[11, 100]$ 的用户比例(21%)远小于其它两个区间(42%与 37%), 即说明处于幼年期 MSN 的 Ordinary 用户较少, 而 Easy-going 用户与 Heavy 用户较多且这两者比例相当. 51 与 renren 在区间 $[101, 101 \text{ 以上}]$ 比例较小(均为 15%), 说明处于衰老期的 Heavy 用户比例较小. 由于成长期的 MSN 处于幼年期与成熟期之间, 我们推测: 成长期的 MSN 的 3 类用户分布不如成熟期的均衡, 但 Ordinary 用户比例比幼年期的要大, Easy-going 与 Heavy 用户比较也比较均衡.

表 4 MSN 各生命阶段的用户结构

区间	用户比例/%			
	Qzone	kaixin001	51	renren
[1,10]	30	42	46	73
[11,100]	34	21	39	12
[101,101+]	36	37	15	15

此外,图 7(c) 给出了各个 MSN 网站用户的访问流量(不包括内嵌对象的流量) CDF 曲线. 可以看出:处于衰老期的 renren 与 51 的 CDF 曲线值一直

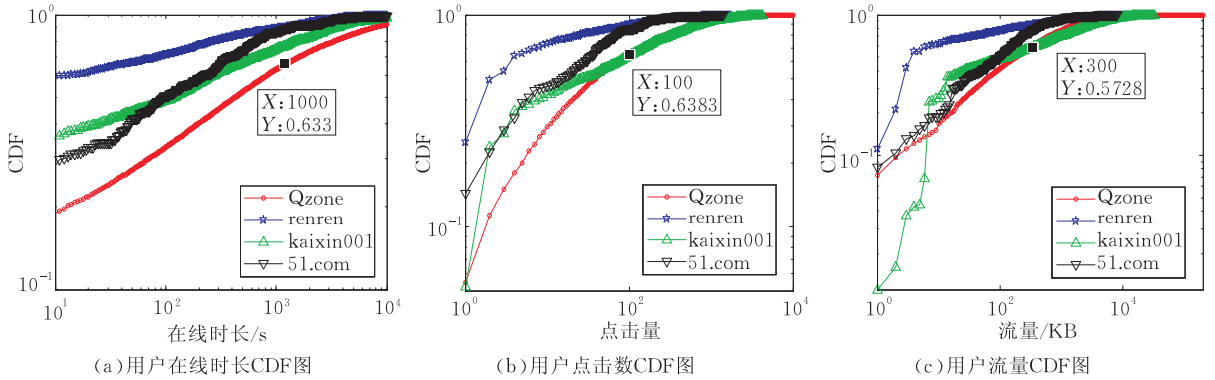


图 7 各 MSN 用户访问特性

(2) MSN 用户分布特征

以上研究发现了 MSN 用户结构存在着较大的差异,这里利用 Gini 系数^[4]刻画 MSN 用户行为的不均匀性,计算出 Qzone、kaixin001、51、renren 这 4 个 MSN 用户流量的 Gini 系数分别为 0.715、0.734、0.798 及 0.878;用户点击数的 Gini 系数分别为 0.744、0.786、0.821 及 0.873. 两组 Gini 系数值表明:不管 MSN 处于何种生命阶段,用户流量与点击量都具有较严重的不均衡性,但处于衰老期的 MSN 流量与点击流分布最不平衡,其次分别为幼年期、成熟期.

(3) 用户在线时长与会话次数、流量的相关性

通过对各个 MSN 的用户研究,我们计算出 Qzone、kaixin001、51、renren 这 4 个 MSN 的用户在线时长与访问次数的皮尔逊相关系数^[7]分别为 0.114、0.049、0.276、0.325,这说明 MSN 用户在线时长不因其用户访问频率而发生变化. 然而,上述 4 个 MSN 用户在线时长与其产生的流量的皮尔逊相关系数却相差甚大,分别为 0.315、0.783、0.264、0.216,这说明处于幼年期的 MSN 用户在线时长与其产生的流量的相关性较大,即有着更大比例的 Heavy 用户,其它 MSN 生命阶段的用户在线时长与其产生的流量的相关性不大.

5.2 MSN 用户会话特征

(1) MSN 用户会话长度的分布特征

高于成熟期 Qzone,说明成熟期 MSN 用户所产生的流量要普遍高于衰老期的用户流量. 处于幼年期的 kaixin001 在流量值大于 300 KB 及流量小于 7KB 的用户所占比例也高于其它的 MSN,这也说明,幼年期的 MSN 用户中,Heavy 与 Easy-going 用户比例大,而 Ordinary 用户比例较小. 这些结论与第 2 节中的生命周期分析结果相符合,也验证了本文模型的有效性.

为分析 MSN 会话的活跃度,首先定义:一个会话中的点击次数为该会话的会话长度. 然后定义会话长度序列 $L = \{L_1, L_2, \dots, L_n\}$, L_i 表示该 MSN 中第 i 个会话的会话长度.

作出 4 个 MSN 的各个会话长度的边际分布频率(两轴都取对数),如图 8(a)~(d)所示. 利用线性回归分析^①对图中的点进行拟合,发现 4 个 MSN 均能很好地服从 Zipf 分布^[8] ($\beta x^{-\alpha}$),其中各 MSN 的相应 α 、 β 参数值如图标示, R^2 为拟合优度 (R^2 介于 0~1 之间,值越大说明拟合越成功,一般 $R^2 > 0.5$ 即可大致认为拟合成功), α 值越大说明会话长度的差异越大. 可以看出,各生命阶段的 MSN 用户的会话长度相差很大,即大部分的会话都涉及较少的点击次数,但小部分的会话却占有较大的点击次数. 然而处于不同生命阶段的 MSN 会话长度差异性还是有所不同,其中处于成熟期的 MSN 会话长度差异最大(Qzone 的 α 值最大),其次分别是幼年期、衰老期.

(2) MSN Inter-session 与 Inter-request 分布特征

为了研究处于不同生命阶段的 MSN 服务器的活跃性与用户会话活跃性的规律,可用两个指标 Inter-session 与 Inter-request 分别对应,其互补累积分布 (CCDF 分布) 如图 9 与图 10 所示.

① http://en.wikipedia.org/wiki/Linear_regression

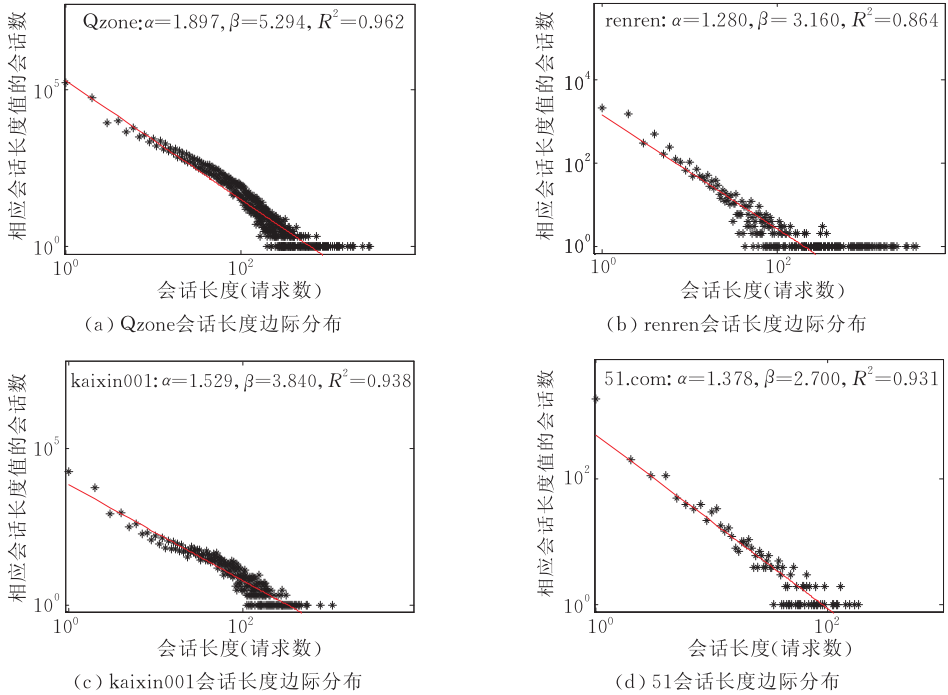


图 8 各 MSN 的会话长度分布

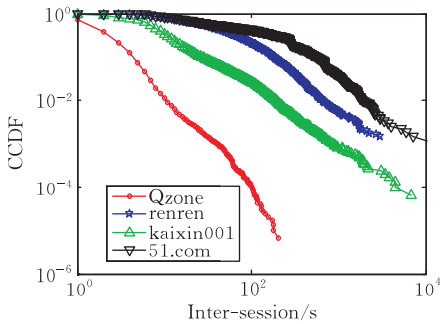


图 9 Inter-session 分布

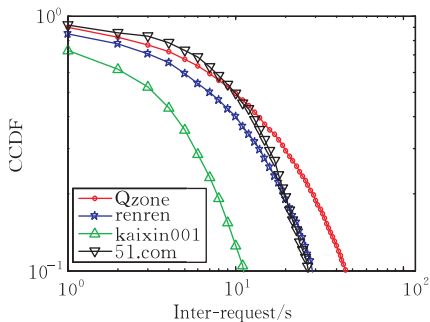


图 10 Inter-request 分布

利用 K-S^[9] 测试,我们发现 4 个 MSN 的所有 CCDF 曲线都能完美地服从对数正态^[10] (Lognormal) 分布. 对数正态分布的概率密度函数如式(5):

$$f(x) = \frac{1}{\sigma x \sqrt{2\pi}} e^{-\frac{(\log(x)-\mu)^2}{2\sigma^2}} \quad (5)$$

其中, μ 代表各 MSN 的 Inter-session 与 Inter-request 对数的平均值, σ 为曲线服从对数正态分布

的方差,我们列出服从对数正态分布的 4 个 MSN 的相关参数 μ 和 σ ,如表 5 所示.

表 5 对数正态分布参数表

MSN	Inter-session 参数		Inter-request 参数	
	μ	σ	μ	σ
Qzone	0.61	0.71	2.27	1.27
renren	3.61	0.75	1.98	1.06
kaixin001	1.97	1.12	1.27	0.85
51	3.80	1.68	3.91	1.54

对 Inter-session 与 Inter-request 两个参数的分析来看,尽管都服从对数正态分布,但处于不同成长阶段的参数值有所不同. 对 Inter-session 来说, μ 值小,即到达的会话间隔小,说明其服务器的活跃性高;同理对 Inter-request, μ 值小,即用户在会话中的点击请求间隔小,说明用户会话的活跃性高.

从表 5 可知, Qzone 的 Inter-session 参数 μ 最小,其次是 kaixin001、renren、51,说明处于成熟期的 MSN 服务器最为活跃(生命周期模型中映射为竞争地位最高);其次分别是幼年期、衰老期;同时 kaixin001 的 Inter-request 参数 μ 值最小,说明处于幼年期的 MSN 用户会话的活跃度最高(生命周期模型中映射为竞争实力最强).

(3) 模型指标 Inter-session 与 Inter-request 分析

从图 9 和图 10 可以看出,处于不同周期的 MSN 的 Inter-session 和 Inter-request 曲线具有明显的特性. 所有 MSN 都服从对数正态分布,但是处

于不同生命周期阶段的 MSN 具有显著不同的参数 μ 和 σ 以及曲线特性,但是处于相同生命周期的 MSN,如 renren 和 51,却具有近似相同的参数特性.该结果也说明模型所用的指标是可以用来区分不同 MSN 的.

6 相关工作

近年来,随着移动社交网服务的迅速发展,大量有关移动社交网的商业模式、网络安全、社交联系、网络结构和服务优化等方面的研究在国内外得到了广泛开展. Fourli^[11] 为 MSN 的发展提出了一个商业模式;谈嵘等人^[12] 提出了一种结合策略与算法的位置隐私保护机制,用于保护移动社交网应用中的用户位置隐私;Nguyen 等人^[13] 提出了一个基于模块的自适应方法用于识别并跟踪移动社交网的社区结构;此外, Niyato 等人^[14] 也提出了一个可控制的联合游戏模型用于研究服务提供商的决策支持;王玉祥等人^[15] 提出了一个基于上下文感知的移动社交网络服务选择机制,目的是提高服务选择的准确性和可靠性.但是,关于移动社交网的生命周期的研究还未有报道.

BCG 矩阵已经被广泛应用于多个学科领域的生命周期规律研究;Costa 等人^[16] 使用 BCG 矩阵用于刻画他们的信息系统生命周期;Zhu 等人^[17] 使用 BCG 矩阵,为工业发展提供了相关的策略与建议,其它的研究还有很多.

7 总结

本文基于经典的 BCG 矩阵模型和 MSN 的用户访问行为特性,建立了 MSN 的生命周期评估模型,能够方便有效地为市场、商业、管理、投资分析等应用服务.模型可以避免传统的市场调研方法的实时性较差以及人为主观因素的影响,通过网络测量手段对 MSN 用户访问数据进行采集和分析,使得分析结果具备良好的真实性和客观性.利用在运营商实地采集的数据集,分析和验证了文中模型的有效性并进行了相关适用性讨论.此外,基于该数据集,本文从 6 个方面分析和讨论了处于不同生命周期阶段的 MSN 用户行为特征,结果发现不同阶段的 MSN 用户行为表现出很强的差异性.

由于该模型能反映各 MSN 的即时竞争地位与竞争力,采用持续的数据采集方式,就可以了解各

MSN 的竞争地位与竞争力的动态变化(例如可以每个月给出一个分析结果).MSN 管理者可以以此来评估过去一段时间内采取的增强竞争力的策略与方法的有效性.另外,市场分析者也能实现对 MSN 市场发展的动态分析与预测.

针对 MSN 生命周期的研究才起步,进一步的工作包括:(1) 本文的生命周期模型中,MSN 竞争地位与竞争实力分别由两个用户访问行为指标来刻画,考虑到用户在线时长、流量与会话的相关性,下一步可以综合这些指标到模型中;(2) 目前的测试数据集只有一周,如有更多的数据对模型进行评价,结果将会更有说服力,并且也能体现各 MSN 生命周期状况的动态变化.但是就现有的数据集,其依然比较准确地反映了市场实际状况.由于这些数据是在相同条件下各个 MSN 进行测量所得,因此模型分析的结果具备良好的横向比较性,在应用于 MSN 市场分析时会有帮助.

参 考 文 献

- [1] Luo Jun-Zhou, Wu Wen-Jia, Yang Ming. Mobile Internet: Terminal devices, network and services. Chinese Journal of Computers, 2011, 34(11): 2029-2051(in Chinese)
(罗军舟, 吴文甲, 杨明. 移动互联网: 终端、网络与服务. 计算机学报, 2011, 34(11): 2029-2051)
- [2] Zhao Guo-Feng, Shan Qing, Xiao Sha-Sha, Xu Chuan. Modeling Web browsing on Mobile Internet. IEEE Communications Letters, 2011, 15(10): 1081-1083
- [3] Research report on social network application of netizen in China. China Internet Network Information Center (CNNIC), 2011(in Chinese)
(2010 年中国网民社交网站应用研究报告. 中国互联网络信息中心(CNNIC), 2011)
- [4] Hu H B, Wang L. The gini coefficient's application to general complex networks. Advances in Complex Systems, 2005, 8(1): 159-167
- [5] Benevenuto F, Tiago R, Meeyoung C, Almeida A. Characterizing user behavior in online social networks//Proceedings of the SIGCOMM. Barcelona, Spain, 2009: 49-61
- [6] Jin Yu. Mobile communities: One-stop service platform. China International Capital Corporation Limited, 2011 (in Chinese)
(金宇. 移动社区: 高粘度/一站式服务平台. 中国国际金融有限公司证券报告, 2011)
- [7] Benesty Jacob, Chen Jing-Dong, Huang Yi-Teng. On the importance of the pearson correlation coefficient in noise reduction. IEEE Transactions on Audio, Speech and Language Processing, 2008, 16(4): 757-765
- [8] Breslau L, Cao Pei, Fan Li. Web caching and Zipf-like distributions: Evidence and implications//Proceedings of the INFOCOM. NY, USA, 1999: 126-134

- [9] Varga T, Haverkamp T, Sanders B. Analysis and modeling of WAP traffic in GPRS networks//Proceedings of the ITC Specialist Seminar. Antwerp, Belgium, 2004: 122-129
- [10] Liu Z, Almhana J, McGorman R. Approximating lognormal sum distributions with power lognormal distributions//Proceedings of the IEEE Vehicular Technology Society. Calgary, Canada, 2008: 2611-2617
- [11] Fourli Irene. Business model for mobile social network [M. S. dissertation]. Athens, the Greek; Athens Information Technology, 2010
- [12] Tan Rong, Gu Jun-Zhong, Yang Jing et al. Designs of privacy protection in location-aware mobile social networking applications. Journal of Software, 2010, 21(Supplement): 298-309(in Chinese)
(谈嵘, 顾君忠, 杨静等. 移动社交网中的隐私设计. 软件学报, 2010, 21(增刊): 298-309)
- [13] Nguyen N P, Dinh T N, Ying Xuan, Thai M T. Adaptive algorithms for detecting community structure in dynamic social networks//Proceedings of the INFOCOM. Shanghai, China, 2011; 2282-2290
- [14] Niyato D, Han Zhu, Saad W, Hjorungnes A. A controlled coalitional game for wireless connection sharing and bandwidth allocation in mobile social networks//Proceedings of the GLOBECOM. Miami, USA, 2010: 1-5
- [15] Wang Yu-Xiang, Qiao Xiu-Quan, Li Xiao-Feng, Meng Luo-Ming. Research on context-awareness mobile SNS service selection mechanism. Chinese Journal of Computers, 2010, 33(11): 2126-2135(in Chinese)
(王玉祥, 乔秀全, 李晓峰, 孟洛明. 上下文感知的移动社交网服务选择机制研究. 计算机学报, 2010, 33(11): 2126-2135)
- [16] Costa C J, Aparicio M. Managing the information system life cycle in construction and manufacturing//Proceedings of the World Academy of Science, Engineering and Technology Volume. Bhubaneswa, India, 2005; 143-147
- [17] Zhu Yonghua, Wang Yanyan, Ding Houchun. Research for evaluation of regional strategy industry and selection of develop strategy based on rough theory and BCG matrix//Proceedings of the 2009 International Conference on Electronic Commerce and Business Intelligence (ECBI). Beijing, China, 2009; 345-352



ZHAO Guo-Feng, born in 1972, Ph. D., professor. His research interests include future Internet, mobile Internet and network management.

LI Bing, born in 1986, master candidate. His research interests include mobile Internet, social network.

XU Chuan, born in 1981, Ph. D.. His current research interests include mobile Internet.

TANG Hong, born in 1957, Ph. D., professor. Her research interests include Internet.

Background

With rapid rising of cellular phone users, mobile Internet has become an important factor in modern society, which is more closed to users' daily life. PC-based Social network site(SNS), which is an import part of Internet website, turns to mobile. In parallel, numerous native mobile social network sites have popped up, such as Foursquare, Gowalla et al.

However, new mobile social networks emerge but most of them vanish with a transient life process from birth to death. So, we argue that mobile social networks may also have a lifecycle, liking human and other creature. Research on lifecycle of mobile social network is becoming popular. BCG matrix, which has been applied to portray the traits of products, enterprise, information system, motivates us to

give this research.

We stand in the managers' point of view in this paper to research the lifecycle process of mobile social network, in order to give some useful suggestion and decision-support on behalf of them.

In this paper, we give a sample and effective lifecycle model for mobile social networks by extracting two important factors from user behavior pattern and verify this lifecycle model by real data from a WAP gateway of a Mobile Telecom Carriers in Chongqing Province. Having given some decision-support for the managers of the MSN sites, we conduct detailed analysis on the data and reveal key features of the mobile social network workloads. At last, we summarize characteristics for each phase of the lifecycle.