

视频点播系统用户行为模型的构建与应用

凌 强¹ 张逸成¹ 严金丰¹ 徐理想¹ 赵 峰² 郝诗海²

¹(中国科学技术大学 自动化系, 合肥 230027)

²(安徽三联交通应用技术股份有限公司, 合肥 230081)

E-mail: qling@ustc.edu.cn

摘要: 通过研究视频点播系统中用户的行为, 构建出用户行为的数学模型, 可以为视频媒体数据缓存算法改进提供依据. 本文对常用的建立用户访问模型的数学模型进行了研究, 对给出的一批媒体访问数据进行了建模分析, 指出广延指数模型能够比 Zipf 模型更接近地描述实际影片点播的频率; 并且用广延指数模型对影片播放长度进行了分析和拟合, 验证了其有效性; 进一步, 基于所得的用户点播频率模型和点播长度模型, 给出了一种缓存算法命中率上界的计算方法, 从而为评价视频媒体数据缓存算法的性能提供了重要指导.

关键词: 视频点播; 用户行为; Zipf; 广延指数

中图分类号: TP391

文献标识码: A

文章编号: 1000-1220(2013)03-0548-05

Construction and Application of Users' Behavior Model in the Video on Demand System

LING Qiang¹ ZHANG Yi-cheng¹ YAN Jin-feng¹ XU Li-xiang¹ ZHAO Feng² HAO Shi-hai²

¹(Department of Automation, University of Science and Technology of China, Hefei 230027, China)

²(Anhui Sanlian Applied Traffic Technology Co. LTD, Hefei 230081, China)

Abstract: By studying the behavior of Video-On-Demand users, a mathematical behavior model can be built, which may guide the improvement of the video data caching algorithms. This paper investigates the common mathematical models of the users' behavior, analyzes a batch of video data and finds that the stretched exponential model can better match the real demanding frequencies of videos than the Zipf model. It also implement the stretched exponential model to model the playback length of videos and gets very good fitting, which confirms the correctness of the obtained playback length model. Moreover, it proposes a method to predict the upper bound of the hitting rate of video data caching algorithms based on the achieved video demanding frequency model and the playback length model. That upper bound plays an important role in quantitatively evaluating video data caching algorithms.

Key words: VOD; users' behavior; Zipf; stretched exponential distribution

1 引言

Internet 上网络服务的重心, 已由原先的传送基于文本和图像的 Web 内容转向传送含有丰富信息的多媒体内容^[1]. 日渐增多的网络多媒体应用形式(如: 网络教学、视频点播、网络电视等)带来了呈爆炸式增长的网络业务量. 这种高质量的音频和视频的传输, 以及对用户访问体验的要求, 对网络传输带宽和服务器性能是一种新的更严苛的挑战.

Amigot M 在他为 IBLNEWS 撰写的报告中提到, 半年间在线观看视频的互联网用户数增加了 18%, 而互联网上的视频访问流量每三到四个月就会增加一倍^[2]. 另一份由 Alexa Internet 出具的报告显示, 类似于 youtube 等大型的点播网站的访问量增速更快^[3]. 在中国, 由中国互联网络信息中心 2011 年 1 月出具的《中国互联网络发展状况统计报告》中显示, 在各类应用中, 网络视频行为以 62.1% 的普及率居于表单前列^[4], 网络多媒体服务已成为中国最为广泛的网络应

用之一. Cisco 的一篇关于互联网增长的报告《走近泽它字节时代》预测, 到 2012 年, 所有形式的视频流量总和将占到网络流量的 90% 左右, 仅互联网视频将在 2012 年占有所有互联网流量的 50%^[5].

由于网络流量和带宽的限制, 媒体数据在经由代理服务器时会被代理服务器缓存; 当用户对于该媒体数据的访问到达时, 如果其数据已存储于代理服务器的缓存中, 就可以不从远端的服务器获取数据而直接从代理服务器中读取数据, 这样就可以降低网络上的流量, 起到降低网络运营成本的作用^[6]. 基于这种现状的考虑, 研究媒体数据的用户访问行为有助于改善服务器端缓存算法的效率, 提高服务器缓存的性能, 降低网络流量.

对于网络中的用户媒体数据进行行为分析是网络服务的重要工具, 它对提高网站服务质量、改善网络运行效率、为用户提供个性化服务都起到了重要作用. 随着网络管理的重要性日益突显, 服务意识在校园网管理中的日益增强, 在高校校

收稿日期: 2011-09-22 基金项目: 中央高校基本科研业务费专项资金新世纪优秀人才项目(NCET-10-0917)资助; 安徽省科技攻关计划项目(09010306042)资助. 作者简介: 凌 强, 男, 1975 年生, 博士, 副教授, 博士生导师, 研究方向为无线传感器网络、网络化控制; 张逸成, 男, 1990 年生, 硕士研究生, 研究方向为网络传播与控制; 严金丰, 男, 1987 年生, 博士研究生, 研究方向为无线传感器网络; 徐理想, 男, 1990 年生, 博士研究生, 研究方向为网络传播与控制; 赵 峰, 男, 1964 年生, 高级工程师, 研究方向为交通视频处理; 郝诗海, 男, 1967 年生, 工程师, 研究方向为交通视频处理.

园网络管理中引入用户行为分析,将成为提高网络管理质量的新思路.此外,针对校园网内用户行为模型的分析,对于掌握和预测校园网络运行状态也有着重要意义^[7].

用户行为分析指在获得网站访问量基本数据的情况下,对有关数据进行统计、分析,从中发现用户访问网站的规律,并将这些规律与网络营销策略等相结合,从而发现目前网络营销活动中可能存在的问题,并为进一步修正或重新制定网络营销策略提供依据^[8].网络环境下,用户行为分析所侧重的多数在于访问 web 的行为,由于网页的缓存通常较小,数据量有限,对网络带宽的要求不高,故而其用户行为与媒体系统的用户行为有一定的差异^[9].

流媒体系统因不同的使用环境、分发手段、受众面和用户群而千差万别,例如按不同的使用环境可划分为校园环境、企业环境和商用环境几种.不同的应用环境和用户会使得在某个特定情况下的用户行为有所不同,加之现有工作对用户行为的理解不够全面,这就导致了现有研究工作中多种用户行为模型的出现^[10].

目前针对流媒体系统用户访问行为的研究,按研究方法可分为主动测量分析法与被动测量分析法两大类;主动测量分析法关注的是互联网流媒体传送机制、质量以及资源利用效率等,这方面的研究工作有助于对流媒体现有技术的理解,并能对流化服务的设计提供指导.被动测量分析法从用户角度出发,观察用户的访问模式,对其进行分析建模.这方面的研究对流媒体缓存系统和内容分发网络的设计,以及进行性能评估的用户行为发生器的设计均有帮助.我们采用的是被动测量分析法,即从用户的访问数据出发,通过建立数学模型来分析用户访问的行为.

本文研究和分析的工作主要有以下几点:

对现有用户行为分析的数学模型进行调研,分析建模所使用的两种数学模型,即 Zipf 模型和广延指数模型;使用两种模型分析实际用户数据,分别对影片点播频率和访问长度进行建模,找出合理的参数进行拟合,发现采用广延指数模型建立的模型与原数据模型具有较好的一致性.

完成对用户访问数据的建模后,给出了用户访问行为模型的一种实际应用,即将该模型应用于流媒体缓存服务器的缓存算法设计中,分析其流化命中率的上限,得出一个近似的上界值,为缓存算法的改进提供了方向和标线.

2 媒体数据用户行为的数学模型分析

2.1 概述

对现有的流媒体视频点播系统的用户数据进行分析后得出,用户的访问行为表现出重尾现象. Zipf 模型对于常见的重尾现象可以较好地描述,但流媒体的用户行为具有其特有的时变性和随机性,使得 Zipf 模型不能很好地刻画用户行为;而广延指数模型对用户行为可较好地刻画.下面对两种模型进行介绍.

2.2 Zipf 模型

Zipf 分布在自然语言语料库中的表述如下:在自然语言的语料库里,一个单词出现的频率与它在频率表里的排名成反比.这个定律被作为任何与 power law probability distributions 有关的事物的参考^[11].

具体到数学上的表述为:设数据集 S 中有 N 个元素,将这些元素按出现次数降序排列,如果第 i 个元素 E_i 的出现次数 f_i 满足下面的公式:

$$f_i = \frac{|S|}{i^\alpha c} \quad (1)$$

则称这个数据集中的 N 个元素服从参数为 α 的 Zipf 分布,其中 $|S|$ 代表所有元素出现的总次数, c 为一个归一化常数.

2.3 广延指数模型

广延指数模型最早用于对玻璃纤维结构弛豫现象的经验描述中,又称为 Kohlrausch-Williams-Watts 模型.此后, Laherrere^[12] 在其论文中指出,广延指数模型具有两个可调参数,可用于具有重尾现象的自然和社会经济现象的描述中,补充了 Zipf 模型的不足之处. Wilson^[13] 也指出,广延指数模型在研究 Zipf 模型无法准确刻画的某些自然现象中具有其独特的作用. Guo^[14] 首先将广延指数模型用于对流媒体系统的研究中,指出了传统使用的 Zipf 模型较为适用于基于文本内容的 Web 网络系统,而较难适用于流量较大的流媒体系统,其引入了广延指数模型来刻画流媒体系统中媒体对象的访问次数分布,取得了较好的效果.

在广延指数模型下,互补累积分布函数(Complementary Cumulative Distribution Function, CCDF)为:

$$P_c(x) = \exp\left[-\left(\frac{x}{x_0}\right)^c\right] \quad (2)$$

式(2)中有 x_0 和 c 两个常量参数,其中 x_0 是尺度参数, c 是形状参数(也称广延参数).

在一个 n 部影片的系统,互补累积分布函数可以用下面的方法计算:

将 n 部影片的访问概率由高到低排序,依次记为 p_1, p_2, \dots, p_n , 则互补累积概率为:

$$\begin{aligned} P_1 &= p_1 + p_2 + p_3 + \dots + p_n \\ P_2 &= p_2 + p_3 + \dots + p_n \\ &\vdots \\ P_i &= p_i + p_{i+1} + \dots + p_n \end{aligned} \quad (3)$$

2.4 实际数据建模

在实验中,我们从实际的视频点播系统日志中提取了部分数据,这批数据的基本特征如下:影片总数为 56000 部,访问的规律基本满足二八模型,每一次用户访问都是从影片的头开始.

2.4.1 影片点播次数数学模型的构建

首先我们依据 Zipf 模型和广延指数模型,对用户点播影片的访问次数进行数学建模.具体做法为:采用现有的用户访问数据,分别统计出每一部影片的访问次数,计算出每一部影片的访问概率,再按概率高低依次排序,得出如前所述计算互补累积分布函数所需要的一串概率序列 $p_1, p_2, \dots, p_{56000}$, 由此计算出影片的互补累积分布函数(如下页图 1 中实线所示).

已知广延指数模型的互补累积分布函数为:

$$P_c(x) = \exp\left[-\left(\frac{x}{x_0}\right)^c\right] \quad (4)$$

公式(4)两边取两次对数得到:

$$\ln[-\ln P_c(x)] = c \ln x - c \ln x_0 \quad (5)$$

设

$$y = \ln[-\ln P_c(x)] - b - c \ln x_0 \quad (6)$$

则(6)可表示为:

$$y = c \ln x - b \quad (7)$$

在给定的数据中, $x = 1, 2, \dots, 56000$. 依据公式(3)可得到不同 x 下的 $P_c(x)$, 从而得到相应的 y 的值. 基于最小二乘法, 能够求出公式(7)中 c, b 两个参数的最佳拟合值, 就确定了一个广延指数模型. 而 Zipf 模型则可根据实际数据 ($p_1, p_2, \dots, p_{56000}$) 基于最小二乘法拟合出公式(1)中最佳的 α 和 c 的值. 所得结果如图1所示, 其中点线为 Zipf 模型, 虚线为广延指数模型, 比较二者与原始数据的契合程度.

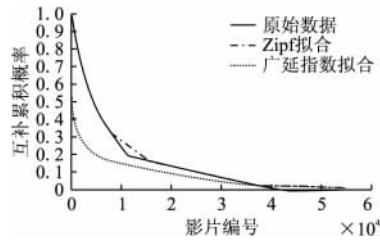


图1 对比使用 Zipf 模型和广延指数模型拟合的互补累积分布函数图线

Fig. 1 Comparison of the complementary cumulative distribution function (CCDF) fitting curves by the Zipf model and the stretched exponential model

始数据和广延指数模型的拟合程度较好, 而在广延指数模型拟合的尾部与实际数据的差别较大. 这是由于在这批实际数据中, 访问的影片大体只有四万部左右, 到 $x = 40000$ 整个图线就会截止; 而理论得出的广延指数模型的互补累积分布函数是在无穷远处才会趋于零, 这就导致了有一个较大的误差带的出现.

鉴于这一现象, 我们提出使用一种截尾的广延指数模型. 其思想如下, 将基于公式(7)所得的广延指数模型在 1 到 56000 上的概率值求和 (该值必然小于 1), 作为一个基准值; 然后将所有概率值都除以这个基准值, 得到一个新的广延指数模型; 再强制截尾, 即当 $x > 56000$ 时, $P_c(x) = 0$.

由图2和图3可以看出, 截尾的广延指数模型较原先的广延指数模型在数据的尾部有更好的拟合精度, 更加符合实际的数据分布情况.

2.4.2 用户点播行为与影片访问长度的关系

对影片的访问长度建模之前需要先判断用户对影片的点

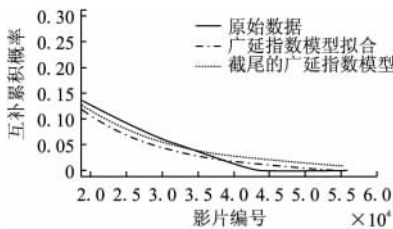


图3 使用截尾的广延指数模型拟合互补累积分布函数的图线(尾部放大图)

Fig. 3 Zoom in the CCDF fitting curve by the truncated stretched exponential model

如图5所示, 将图4中的两条图线进行归一化处理之后, 可以发现两条图线基本重合, 这说明用户对影片的访问长度基本上是一致的, 与这部影片的点播次数无关. 这就说明了用户对影片的访问长度与用户对影片的点播行为二者是相互独

立的. 我们可以分开讨论两个问题而不需要考虑两者之间的相互影响.

由图1可以看出, 广延指数模型相比 Zipf 模型有更好的拟合精度, 我们根据广延指数模型建模来模拟用户对影片的访问行为, 是较为合理的.

但是, 从图1中我们还可以看到, 在图线的前一部分, 原

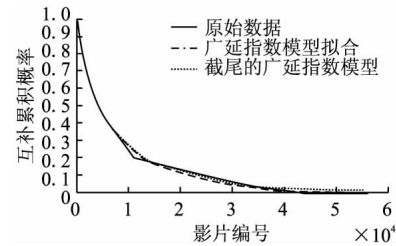


图2 使用截尾的广延指数模型拟合互补累积分布函数

Fig. 2 CCDF fitting curve by the truncated stretched exponential model

播次数和用户的访问长度是否是相互独立的. 如果这二者不独立, 我们就需要区别对待不同点播次数下的影片, 为它们分别建立访问长度的数学模型.

具体的做法如下, 将影片按点播次数多少分为点播次数最多的前 20% 影片和点播次数较少的后 80% 影片两大类; 再将每部影片等长分为 20 小片, 分别统计每一小片被访问的次数. 此处的被访问是指只要访问到了某一小片的内容, 不论访问长度的长短, 均算访问; 另外在每一类中, 计算小片访问次数时忽略影片的区别, 比如访问到该类中任何影片的第 1 个小片时, 对应于第 1 个小片的次数变量都简单加 1. 这样得到的就是两大类影片中各个小片的总的访问次数, 将总的访问次数除以各类中的影片数, 即分别为 56000×0.2 和 56000×0.8 , 得到两大类影片各小片的平均访问次数, 绘制于同一坐标系下. 再将图4中的两条曲线分别归一化处理, 即将两条曲线上的所有点的值都除以该条曲线上第一个点的值, 得到图5.

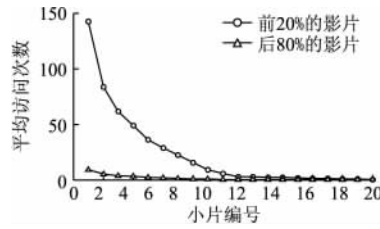


图4 用户对于影片各小片的平均访问次数统计图

Fig. 4 Average visting times of slices of all videos

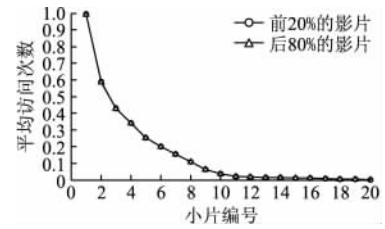


图5 用户对于影片各小片的平均访问次数统计图(归一化后)

Fig. 5 Normalized average visting times of slices of all videos

立的, 我们可以分开讨论两个问题而不需要考虑两者之间的相互影响.

2.4.3 影片访问长度数学模型的构建

由前一节得出的结论, 用户对影片的访问长度与点播次

数无关,我们就可以为其访问长度建立一个普适的数学模型.具体做法如下:按影片长度的百分数分片,每个小片的长度为影片总长度的 0.05%,共有 2000 个小片,其编号为 $x = 1, \dots, 2000$. 统计出每个小片的访问次数,计算出其访问概率,按高低依次排序,计算其互补累积概率 ($P_c(x)$).

再由广延指数模型的互补累积分布函数为:

$$P_c(x) = \exp\left[-\left(\frac{x}{x_0}\right)^c\right] \quad (8)$$

公式 (8) 两边取两次对数得到:

$$\ln[-\ln P_c(x)] = c \ln x = x \ln x_0 \quad (9)$$

设 $y = \ln[-\ln P_c(x)] = b = c \ln x_0$ (10)

则式 (10) 可表示为:

$$y = c \ln x - b \quad (11)$$

类似地,基于最小二乘法,拟合出访问长度模型下公式 (11) 的最佳 c, b 两个参数,就可以确定一个广延指数模型.采用与前面类似的截尾的做法,我们将模型在 $x > 2000$ 时进行截尾,即 $P_c(x) = 0$. 绘制其互补累积概率图线(图 6 实线)和采用截尾的广延指数模型来拟合的图线(图 6 虚线).如图 6 所示.

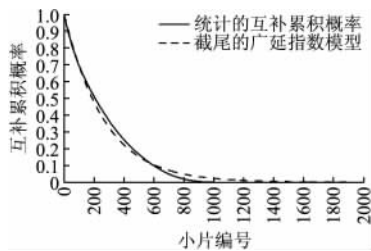


图 6 使用截尾的广延指数模型对影片内部小片访问概率进行拟合的曲线

Fig. 6 CCDF fitting curve of the playback length by the stretched exponential model

由图 6 可见,截尾的广延指数模型较好的拟合了影片的访问长度,有较好的精度.

3 媒体数据用户模型的应用—缓存算法中命中率上界的计算

在流媒体视频点播系统中研究用户的访问行为,并得出其数学模型,很重要的一个应用即为基于该模型、从理论上计算一定大小的缓存空间所能提供的用户访问命中概率的上限值;该上界在设计和改进缓存算法时可以作为标杆,越向这个值逼近说明设计的算法性能越优良.

下面先就视频点播系统缓存的概念做一个概述,然后给出一个缓存命中率的定义,最后基于所得的广延指数模型、从理论上计算缓存命中率所能达到的可能的上界值.

3.1 视频点播系统缓存的概念

视频点播系统 (Video On Demand, VOD) 是利用视频服务器 (服务端系统) 存储的视频节目和其他多媒体信息,通过网络系统和客户端系统,在用户需要时向用户传送其点播的高质量视频的系统.其构架如图 7 所示.

在视频点播系统中,近端服务器缓存媒体数据库中的部分数据;当用户点播行为到达时,优先由近端服务器提供用户

所访问的数据;只有当近端服务器中未缓存用户所请求的那部分数据时,才由远端的媒体数据库中拉取数据.这种缓存策略可以大大地降低网络流量、降低系统的运营成本.

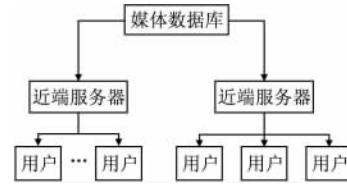


图 7 典型视频点播系统构架示意图

Fig. 7 Schematic diagram of a typical video-on-demand system

流媒体视频点播系统中的缓存算法就是在近端服务器中决定哪些影片应该进入缓存,哪些缓存过的影片应该被踢出去,目的是尽可能使得缓存获得最大的利用率.

3.2 缓存命中率定义

视频点播系统缓存算法的效率由缓存命中率来定量地衡量,一种典型的缓存命中率定义如下:

$$\text{缓存命中率} = 1 - \frac{B_{in}}{B_{out}} \quad (12)$$

式 (12) 中,

B_{in} —用户访问到达时,由于访问的数据不在缓存中从而需要从远端服务器取的数据量;

B_{out} —用户请求的总的访问数据量.

据以上定义可看出,命中率越高,由缓存所提供的用户访问量也就越高,缓存的利用效率越高,系统的性能也就越良好.

3.3 利用广延指数模型计算上界

根据 2.4 节的分析,可知用户的影片点播特性和影片访问长度特性分别遵从于两个广延指数模型.基于这两个广延指数模型,我们可以来求取缓存命中率的理论上限值.

其原理如下:每一部影片都有两种概率值,即其被点击的概率和内部各部分被访问的概率(注:此处每部影片被等长地分成若干个小片,比如 2000 个小片),也就是每一部影片的不同小片的访问概率值就可以由式 (13) 表示:

$$P_{mn} = p_m q_n \quad (13)$$

式 (13) 的 p_m 表示第 m 部影片的第 n 个小片被访问的概率, p_m 表示第 m 部影片被访问的概率, q_n 表示第 n 个小片被访问的概率. p_m 和 q_n 的值由前面 2.4 节的数据拟合出的两个广延指数模型给出.据此,可以计算出每一部影片每一个小片被访问的概率,如表 1 所示.

表 1 给影片的小片的点播概率

Table 1 Demanding probabilities of slices of all videos

$p_1 q_1$	$p_1 q_2$...	$p_1 q_{1999}$	$p_1 q_{2000}$
$p_2 q_1$	$p_2 q_2$...	$p_2 q_{1999}$	$p_2 q_{2000}$
$p_3 q_1$	$p_3 q_2$...	$p_3 q_{1999}$	$p_3 q_{2000}$
...
$p_{56000} q_1$	$p_{56000} q_2$...	$p_{56000} q_{1999}$	$p_{56000} q_{2000}$

表 1 中,不同行代表不同的影片,不同列代表不同的小片,共有 $N = 56000 \times 2000$ 个概率值.将表 1 中的各概率值排序,找出最大的前 20% 的小片,然后把这些小片的访问概率

相加 就得到缓存比例为 20% 时缓存命中率的一个理论上限. 当然对于其它缓存比例下的命中率的理论上限可以类似求取.

为了对所得命中率上限值的保守性有一个定量了解, 我们进行了仿真验证. 基于前面所得的两个广延指数模型(影片点播模型和影片访问长度模型), 产生了 600 万次的用户访问数据. 首先, 计算不同缓存比例下的命中率的理论上限; 其次, 在不同缓存比例下应用我们提出的一种缓存算法, 统计缓存命中率. 图 8 给出了不同缓存比例下, 缓存命中率的上限值

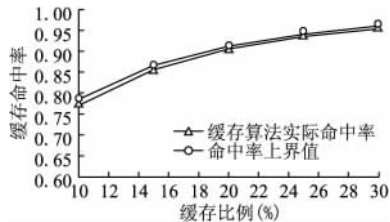


图 8 不同缓存比例下缓存算法实际命中率和命中率上界值的比较

Fig. 8 Comparison of the achieved hitting rates of a caching algorithm and the predicted upper hitting rate bounds under different caching ratios

与所用缓存算法的实际命中率值, 可见这两个值比较接近 (约为 1% 左右) 这说明本文给出的命中率上界值是比较有效的.

4 结论

本文对研究媒体数据用户行为常见的两种数学模型进行了讨论, 并使用实际流媒体视频点播系统的用户访问数据进行了分析, 验证了广延指数模型能够更好地拟合出影片的点播概率和访问长度分布. 并进一步, 将所得到的影片点播模型和访问长度模型用来估计缓存算法命中率的上限值, 为改进算法提供指导.

References:

- [1] Raghavan S V, Tripathi S K. Networked multimedia systems: concepts, architecture & design [M]. Upper Saddle River, N. J.: Prentice Hall, 1998.
- [2] Amigot M. Internet video traffic is doubling every 3 to 4 months [EB/OL]. <http://iblnews.com/story.php?id=17429> 2006.
- [3] Internet [EB/OL]. http://www.alexa.com/data/details/traffic_details/youtube.com 2007.
- [4] China Internet Network Information Center. China Internet development statistics report [EB/OL]. <http://www.cnnic.cn/dtygg/dtgg/201101/P020110119328960192287.pdf> 2011.
- [5] Approaching the Zettabyte Era [EB/OL]. http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481374_ns827_Networking_Solutions_White_Paper.html 2008.
- [6] Dong Fu-qiang. Study on network users behaviors analysis and its application [D]. Xi'an: Xi'an University of Electronic Science and Technology, 2005.
- [7] Liang Wei. Campus network user behavior analysis system research and implementation [D]. Beijing: Beijing Jiaotong University, 2009.
- [8] Liu Peng. Behavior of network users research based on time-varying & services [D]. Beijing: Beijing University of Posts and Telecommunications, 2010.
- [9] Ma Wei-liang. User behavior analysis system applied in broadband networks [J]. Science & Technology Information, 2008, (23): 403-404.
- [10] Zhou You. Study on access pattern of VOD system: measurement, analysis, and modeling [D]. Hefei: University of Science and Technology of China, 2009.
- [11] Hu Hai-bo, Wang Lin. A brief history of power law distributions [J]. Physics, 2005, 34(12): 889-896.
- [12] Laherrere J D. Stretched exponential distributions in nature and economy: "fat tails" with characteristic scales [J]. The European Physical Journal B-Condensed Matter and Complex System, 1998, 2(4): 525-539.
- [13] Wilson P S, Toumi R. A fundamental probability distribution for heavy rainfall [J]. Geophysical Research Letters, 2005, 32(14): 1-4.
- [14] Guo Lei, Tan En-hua, Chen Song-qing, et al. The stretched exponential distribution of Internet media access patterns [C]. PODC'08: Proceedings of the Twenty-seventh ACM Symposium on Principles of Distributed Computing, Toronto, Canada, New York, NY, USA: ACM, 2008.

附中文参考文献:

- [4] 中国互联网络信息中心. 中国互联网络发展状况统计报告 [EB/OL]. <http://www.cnnic.cn/dtygg/dtgg/201101/P020110119328960192287.pdf> 2011.
- [6] 董富强. 网络用户行为分析研究及其应用 [D]. 西安: 西安电子科技大学, 2005.
- [7] 梁伟. 校园网用户行为分析系统研究与实现 [D]. 北京: 北京交通大学, 2009.
- [8] 刘鹏. 网络用户行为分析的若干问题研究 [D]. 北京: 北京邮电大学, 2010.
- [9] 马维亮. 用户行为分析系统在宽带网络中的应用 [J]. 科技信息, 2008, (23): 403-404.
- [10] 周铀. 视频点播系统访问行为研究: 测量、分析与建模 [D]. 合肥: 中国科技大学, 2009.
- [11] 胡海波, 王林. 幂律分布研究简史 [J]. 物理, 2005, 34(12): 889-896.