

Calculation Modeling and Statistical Analysis of Network TV User Behavior

Songtao Wu¹, Yi Wang², Lan Wang³, Can Yang^{3*}

¹Guangzhou Digital Media Group Co., Ltd., Guangzhou Guangdong

²Tencent, Shenzhen Guangdong

³School of Computer Science & Engineering, South China University of Technology, Guangzhou Guangdong

Email: *cscyang@scut.edu.cn

Received: Jan. 7th, 2019; accepted: Jan. 22nd, 2019; published: Jan. 29th, 2019

Abstract

Network television (IPTV) services are becoming more and more popular in China, and in-depth exploration of audience behavior has become an emerging topic in IPTV. We define a state model to understand network TV users' behaviors, and conduct users' behaviors' computation and analysis from such 3 aspects as an individual user, groups and a single channel in a large-scale IPTV system. We also focus on statistical analysis sampling from the watching length, the number of the online and the distribution of 24 hours, the result of which will be provided some references for delivering web content and advertisement, resources allocation of television channels and the precise content recommendation.

Keywords

IPTV, Live Streaming, User Behavior, Statistics Analysis

网络电视用户行为计算建模与统计分析

吴颂涛¹, 王意², 王澜³, 杨灿^{3*}

¹广州珠江数码集团股份有限公司, 广东 广州

²深圳市腾讯计算机系统有限公司, 广东 深圳

³华南理工大学计算机科学与工程学院, 广东 广州

Email: *cscyang@scut.edu.cn

收稿日期: 2019年1月7日; 录用日期: 2019年1月22日; 发布日期: 2019年1月29日

*通讯作者。

摘要

网络电视(IPTV)业务在国内日趋普及,深度探索IPTV受众行为成为一个新兴的课题。我们定义一个网络电视用户行为计算状态模型,从单频道、个体用户和群体用户三个角度对一个大型IPTV系统的用户行为数据进行计算与分析,对观看时长、在线人数及其在24小时内的分布情况进行抽样统计分析,其结果为网络电视内容与广告投放、频道资源合理分配和精准推荐提供参考依据。

关键词

网络电视,直播,用户行为,统计分析

Copyright © 2019 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 介绍

网络电视(IPTV)是一种基于互联网通信协议(TCP/IP)的新媒体业务,其技术架构使实时精准获取用户行为数据成为可能,可为用户收视行为调查(User Behavior Investigation,简称UBI)提供完整的数据资料。相比于传统广播电视和数字电视(DVB),IPTV以其视听质量和交互能力方面的革命性提升迅速笼络一大批用户,成为引领视听媒体产业发展的新趋势。国外较早开展有关IPTV用户行为建模研究[1][2],我们发现,IPTV环境下的用户行为特征不仅涉及群体收视率,还可以从在线时长、频道切换特征、到达率、离开率、时频分布、子群特征及个性化特征等多个维度进行描述。作为“注意力经济”时代的重要量化指标,用户行为特征研究对IPTV的频道价值评估(CE)、有效广告投放(AD)、精准内容推荐(CR)、云媒体资源分配(CDN)以及故障检测(FD)均具有重要的应用价值。本文的目标是构建一个IPTV用户收视行为的基本计算架构和指标体系,并基于一个城市实际IPTV用户行为的数据子集进行切实的测量和计算,形成IPTV用户收视特征图谱。本文的数据集包括了133个频道、22.2万用户和7200万条记录,时长为 $31 * 24$ 小时。本文首先介绍用户收视行为分析的背景和意义,然后创建电视观众收视行为数学定义与计算模型,再进行多维度收视特征统计分析。最后是本文的工作总结与研究展望。

2. 定义与模型

2.1. 数据集与数据结构

本文研究的数据集中,我们定义每一条观看的数据记录为一个四元组 $r(u,c,s,d)$,这里 u 代表观众ID号, c 代表观众 u 本次观看电视的频道ID号, s 代表该次观看该频道 c 的起始时间, d 代表用户 u 本次观看频道 c 的时长。 $r(i)$ 表示一次第 i 次观看的记录,由此可衍生出与 $r(u(i),c(i),s(i),d(i))$ 相关的一系列观众用户行为特征参数,通过对它们的定义和计算可以获知网络电视用户的基本行为特征,详见表1。

2.2. 观看行为建模

为了进一步深入研究IPTV用户观看电视行为状态,我们创建一个IPTV用户收视行为状态转移模型。首先定义三种用户收视行为状态:

Table 1. IPTV audience user behavior characteristics basic parameter definition table
表 1. IPTV 观众用户行为特征基本参数定义表

序号	参数	参数意义	相关公式与备注
1	$r(i)$	$r(u(i),c(i),s(i),d(i))$ 表示第 i 条观看记录	$u(i) = n$ 代表第 i 条观看记录是第 n 个用户产生的; $c(i) = m$ 代表第 i 条观看记录中, 用户观看频道 m ; $s(i)$ 表示第 i 条记录的开始观看时间; $d(i)$ 表示第 i 条记录观看时长
2	e	观看频道结束时刻(即离开该频道时刻)	$e = d + s$ 或 $e(i) = d(i) + s(i)$
3	N	总用户数	例如: $N = 222,000$
4	M	总频道数	例如: $M = 133$
5	C	频道观看人次	若 $C = 5$, 代表所有频道观看总人次为 5
6	Δt	相邻两条观看记录之间的时间差	$\Delta t = s_n(k+1) - e_n(k) = s_n(k+1) - s_n(k) - d_n(k)$
7	W	时间窗口的宽度, 即一个统计时隙的宽度	
8	dt	微分时长	
9	U	在线观看人数	$U(t)$: 时刻 t 在线观看人数; $U(t,m)$: 当前时间 t 在线观看频道 m 的人数
10	u	在线率(系统收视率)	$u = U(t)/N$
11	$u(m)$	频道在线率(频道收视率)	$u(t,m) = U(t,m)/N$
12	v	在线人数变化率(系统收视率变化率)	$v(t) = [U(t+dt) - U(t)]/U(t)/dt$
13	$v(m)$	频道在线人数变化率 (频道收视率变化率)	$v(t,m) = \frac{U(t+dt,m) - U(t,m)}{U(t,m)dt}$
14	U_{in}	进入开启 TV 系统的人数	U_{out} : 关闭离开 TV 系统的人数
15	$U_{in}(m)$	进入频道 m 的人数	$U_{out}(m)$: 切换或离开频道 m 的人数
16	λ_{in}	单位时间内平均到达系统的人数	$\lambda_{in}(t) = U_{in}(t)/N/dt$
17	λ_{out}	单位时间内平均离开(逃逸)系统的人数	$\lambda_{out}(t) = U_{out}(t)/N/dt$
18	$\lambda_{in}(m)$	频道到达率	$\lambda_{in}(t,m) = U_{in}(t,m)/N(t,m)/dt$
19	$\lambda_{out}(m)$	频道离开(逃逸)率	$\lambda_{out}(t,m) = U_{out}(t,m)/N(t,m)/dt$

S: Surfing 浏览状态, 该状态下的用户表现为快速浏览多个电视频道, 一般采用顺序方法依次在相邻频道间进行切换, 发现其观看意愿, 其特征为每个被浏览的频道的停留时间极短。记作:

$\{S|0 < d < \varepsilon, C > 1, t < C\varepsilon\}$, 这里 ε 表示一个大于零的足够小的观看时长。

V: Viewing 欣赏状态, 该状态下的用户表现为停止切换, 停留在一个频道上观看, 不发生频道跳转, 且观看时间较长。记作: $\{V|d > \varepsilon, C = 1, \varepsilon < t < \ln f\}$ 。

I: Idle 状态, 该状态下用户表现为既不继续观看某一频道, 又不进行频道切换, 系统处于空闲状态。为了保障研究的一致性, 本文定义一个空频道 Null Channel, 若用户处于空频道(频道号为 0, ch0), 即用户处于空闲期, 没有观看任何频道内容。记作: $\{I|d = 0, C = 0, 0 < t \leq \ln f\}$ 。

综合考虑观众注意力、广告、节目等方面因素的动态变化, 观众状态维持时间长短和状态之间的转化存在多种可能。一个观众观看一个频道较长时间之后, 可能开始浏览其它频道, 也可能切换到另一个感兴趣频道观看。我们定义 4 种状态转移行为:

- ① 关闭 close, 包括两个子行为 $V \rightarrow I, S \rightarrow I$;
- ② 开启 start, 包括两个子行为 $I \rightarrow V, I \rightarrow S$;
- ③ 换台 change, 包括两个子行为 $V \rightarrow V, S \rightarrow V$;

④ 浏览 surfing, 包括两个子行为 $V \rightarrow S, S \rightarrow S$ 。

综上, 从用户角度出发, IPTV 行为状态转移模型包括 S 、 V 、 I 三种状态以及各个状态之间的转移行为, S 和 V 状态包括自我转移, I 状态不含自我转移。如图 1, 有向图 $g=(v,e)$ 表示一个 IPTV 典型用户的行为状态转移模型, 图 1 中节点分别代表用户的三种状态 S 、 V 、 I , 有向边连接不同状态, 表示观看行为的状态迁移动作。回路表示观众收视行为闭环状态链。

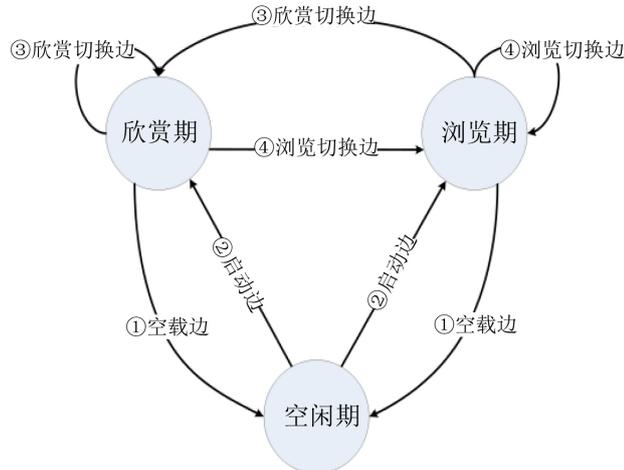


Figure 1. Network TV user viewing behavior modeling state diagram
图 1. 网络电视用户观看行为建模状态图

为了清晰展示用户观看行为时序, 图 2 给出一个用户典型的收视时序行为描述。用户从 I 状态开始, 也就是空频道 ch_0 , 表示未开启 IPTV 服务; t_0 时刻开启 IPTV 系统。在时间区间 $[t_0, t_3]$ 内, 用户进行了频道浏览切换, 即为状态 S , 用户从 ch_1 切换到 ch_4 , 最后观看 ch_4 至时间 t_4 , 即在时间区间 $[t_3, t_4]$ 为 V 状态。在时区 $[t_3, t_6]$, 该用户浏览过 ch_3 , 再切换至 ch_2 。从 t_6 到 t_7 , 用户又离线(关机)。从 t_7 到 t_8 , 看 ch_2 后结束观看。基于图 1 所示的状态转移模型, 该观看过程可描述为:

$$\{I(t < t_0), S[t_0, t_3], V[t_3, t_4], S[t_4, t_5], V[t_5, t_6], I[t_6, t_7], V[t_7, t_8], I(t > t_8)\}$$

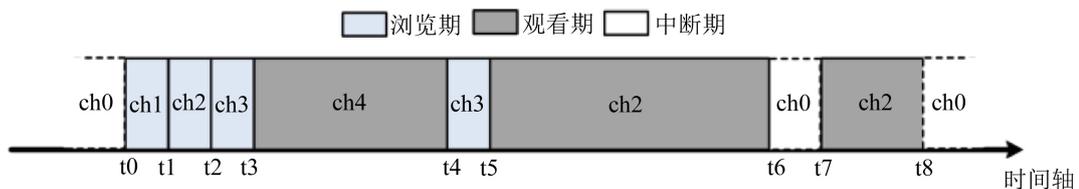


Figure 2. Schematic diagram of the user-based viewing process based on the timeline
图 2. 基于时间轴的用户观看过程示意图

基于实际 IPTV 用户数据集, 通过大量统计研究, 我们发现用户平均观看时长 D 有着与文献[3]中所描述的基本一致分布特征, 如图 3 所示。

图 3 为观看时长的累计概率密度分布图 CDF, 图 3(a), 图 3(b)和图 3(c)是我们统计计算用户数据的分布概率结果, 可见 56.06%的记录观看时长小于 1 m, 98.92%的记录观看时长小于 3 h。图 3(b)的 x 轴采用了图 3(a) x 轴的对数坐标, 以便清晰展示相对短的观看时长的分布变化特征; 我们也将放大 x 轴时间区间的统计结果展示在图 3(c)中, 以便展示更多的细节, 同时, 作为比对, 我们也展示了文献[3]的结果

于图 3(d), 这说明我们的研究跟历史和国外研究情况的相关一致性。在本文的研究中, 浏览状态 S 是指观看时长小于 60 s 的观看行为; 观赏状态 V 是指观看时长位于区间[60 s, 3 hour]的观看行为; 离开行为 Idle 状态是指关机或者观看时长大于等于三小时的观看行为。

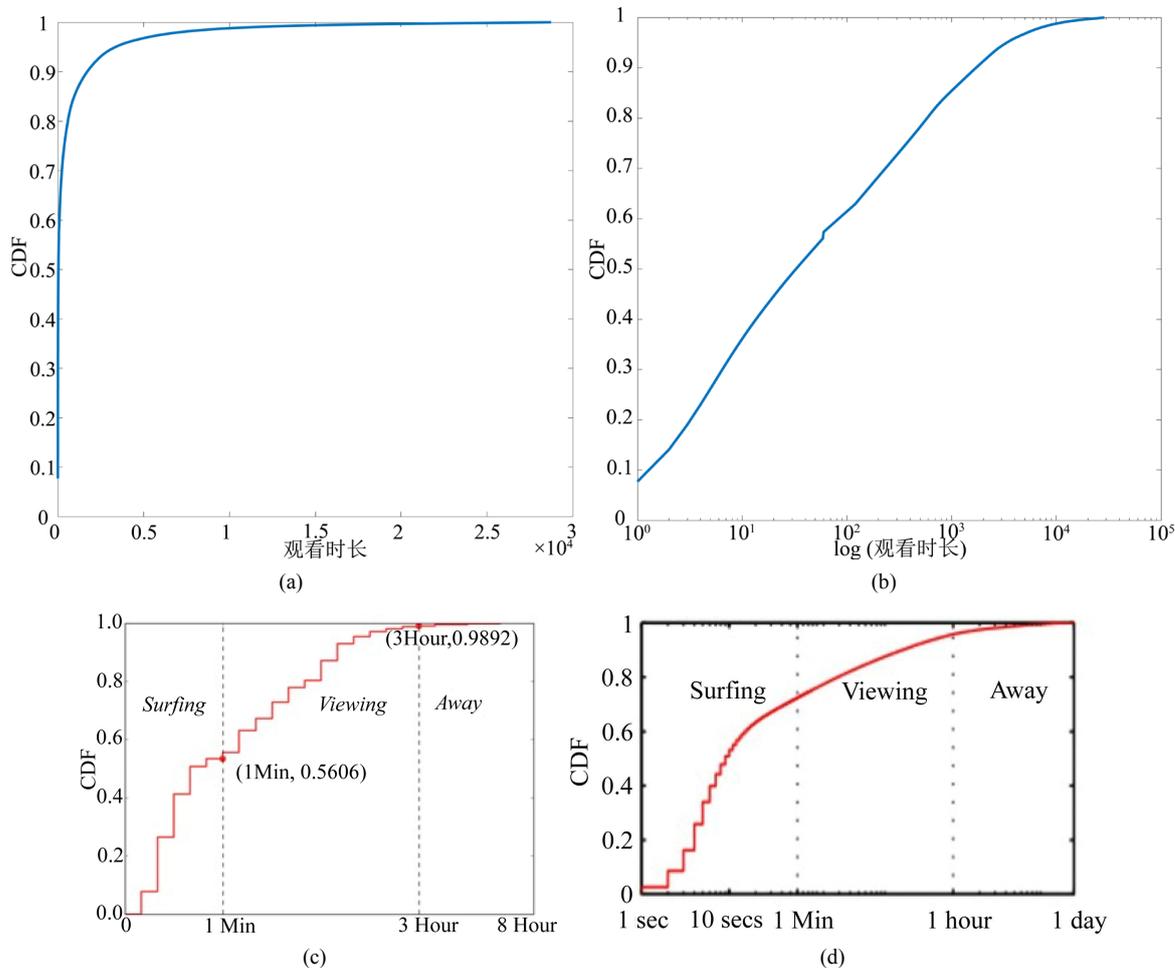


Figure 3. User watch time CDF chart. (a) x watch duration, y probability distribution; (b) pair (a) x-axis logarithm; (c) partition enlargement x-axis; (d) Meeyoung' Result

图 3. 用户观看时长 CDF 图。(a) x 观看时长, y 概率分布; (b) 对(a) x 轴取对数; (c) 分区放大 x 轴; (d) Meeyoung' Result

3. 网络电视数据统计分析

3.1. 群体行为的统计分析

图 4(a)是整个 IPTV 用户群体从 2014 年 8 月 25 日到 8 月 31 日一周的在线人数变化曲线。横坐标是时间, 纵坐标是在线人数。从图 4 可知, 整体在线人数具有较强的自相似性, 凌晨 5 点观众最少, 下午 1 点观看人数出现一个次高峰; 一天内的观看高峰出现在晚上 7 点到 10 点之间。文献[4]也揭示了境外 IPTV 用户在线人数分布的类似特征, 可知国内外用户观看直播电视的行为特征具有较强的相似性。图 4(b)、图 4(c)和图 4(d)展示了热频道、一般频道与冷频道 7 天的收视人数统计分布, 我们发现不同频道内部具有自相似性特征, 但是不同频道之间各向异性。图 4 所示不同频道每日在线人数呈明显的自相似性和周期性。

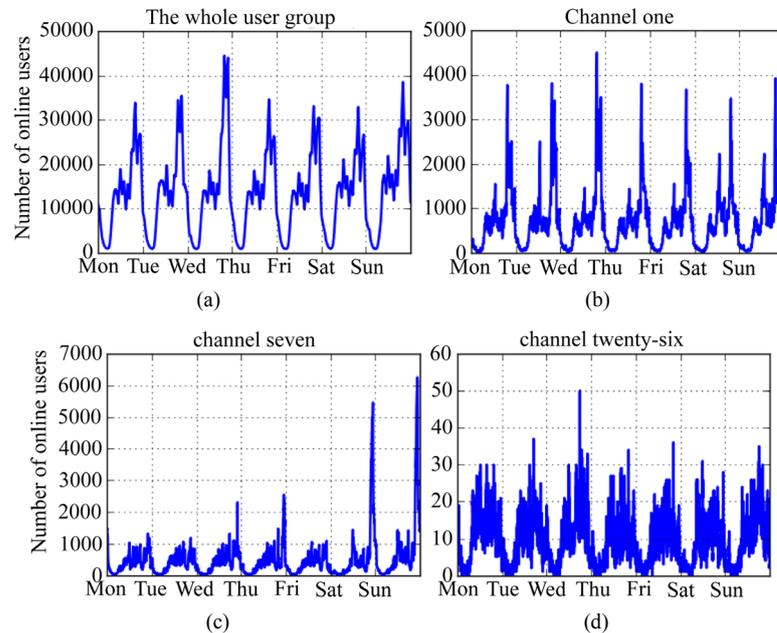


Figure 4. Online population statistics. (a) The number of online users of the entire IPTV user group; (b) Number of online users of ch1; (c) Number of online users of ch7; (d) Number of online users of ch26

图 4. 在线人数统计图。(a) 整个 IPTV 用户群在线人数；(b) ch1 在线人数；(c) ch7 在线人数；(d) ch26 在线人数

3.2. 典型频道统计分析

我们统计分析 ch1 在 2014 年 8 月 25 日一天的到达率(平均观看率)、离开率(平均逃逸率)、在线人数和在线人数变化率。用户到达率是指单位时间观看频道的用户次数,见表 1 序号 18;用户离开率是指单位时间离开频道的用户次数,见表 1 序号 19;用户在线人数变化率是指单位时间在线人数变化数,见表 1 序号 13。从图 5(a)和 5(b)中我们发现到达率和离开率具有强相似性,这与文献[4]的统计相似。图 5(a)峰值点相对较少,我们推测用户观看某一个频道没有太明显的时间相关性,即就是说人们在一个频道上的到达时间是相对随机的。图 5(b)峰值点较多,可以推测人们常常在相同的时间离开某一个频道,也就是说离开频道有强时间相关性。可能的原因是因为插播广告或节目播放完毕时,大量用户都放弃了该频道的观看。图 5(c)显示 24 小时内,早上 8 点在线人数达到极值点,下午 1 点又达到一个小峰值点,晚上 7 点和晚上 10 在线人数出现大峰值点。图 5(d)表示在线人数变化率,在线人数增加为+,在线人数减少为-,我们发现大正值较少,大负值较多,再次说明观众切换进入一个频道相对随机而切换离开一个频道呈现较强相关性。

3.3. 个体用户统计分析

图 6 呈现了两个用户在 2014 年 8 月一个月时间内观看次数时刻分布情况。 n 天内第 i 时刻总观看次数的计算公式为: $\text{count_all_watch}_{n,i} = \sum_{j=1}^n \text{watch}_{j,i}$,若用户在第 j 天 i 时刻观赏了电视节目,即这次的观看时长大于或等于一分钟,则 $\text{watch}_{j,i}$ 为 1,否则 $\text{watch}_{j,i}$ 为 0。

图 6(a)现实用户 1 在 22 点至次日凌晨 6 点无观看行为,可能为用户 1 睡觉时间;用户 1 早上 8 点、晚上 7 点和晚上 9 点出现峰值点,说明用户 1 在这些时间点附近经常看电视。该用户总体观看行为具备规律性,可推测该用户生活习惯规律,可能是生活较为规律的中老年电视观众用户。图 6(b)用户 2 在一天 24 小时内大多数时间段内都有电视观看行为发生,说明该用户可能熬夜观看电视节目,推测该用户作息时间和娱乐时间不规律,建议培养健康收视习惯,可能是生活节奏不太规律的年轻电视观众用户。

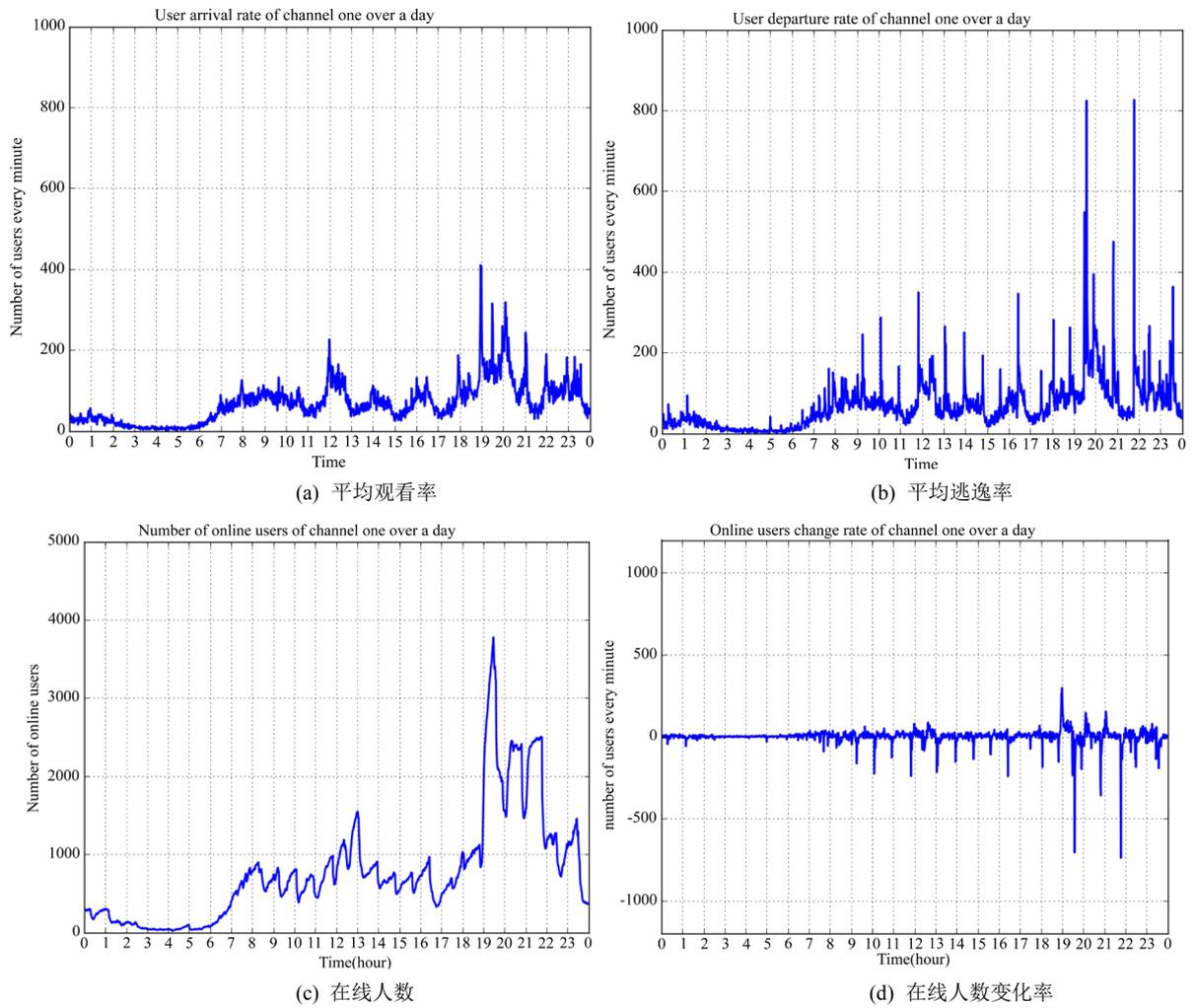


Figure 5. 24-hour online population statistics for ch1
图 5. 对 ch1 一天 24 小时在线人数统计图

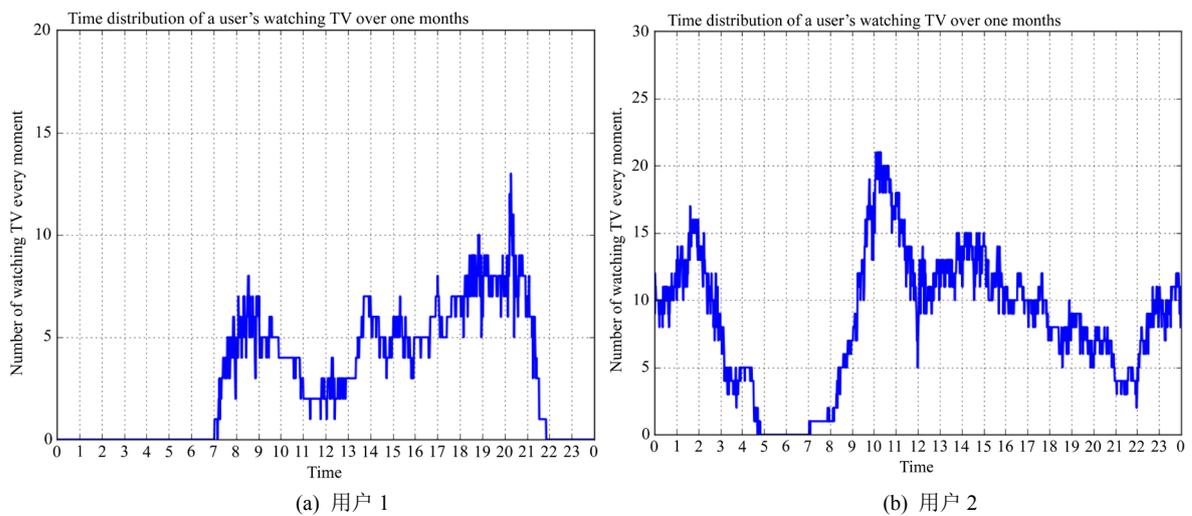


Figure 6. User 1 and User 2 time distribution of one month
图 6. 用户 1 和用户 2 一个月时间观看次数时刻分布

图 7 为两个用户在 2014 年 8 月一个月时间内观看某个频道的 24 小时分布。该图的计算方式为 n 天内第 i 时刻观看频道 k 的总次数，计算公式为： $\text{count_all_watch}_{n,i,k} = \sum_{j=1}^n \text{watch}_{j,i,k}$ ，若用户在第 j 天 i 时刻观看了频道 k ，则为 1，否则 $\text{watch}_{j,i,k}$ 为 0。图 7(a) 为用户 1 感兴趣频道(频道号为 131)观看次数分布图。如图所述，用户 1 观看频道 131 的时间为 18 点至 19 点，即该用户观看该频道具有较强时间相关性。图 7(b) 为用户 2 感兴趣频道(频道号为 3)观看次数分布图。如图所述，用户 2 观看频道 3 的时间主要为 14 点至 16 点，并且在一天内的其它时间，也看频道 3，可知该用户 2 观看频道 3 的时间相关性不显著。

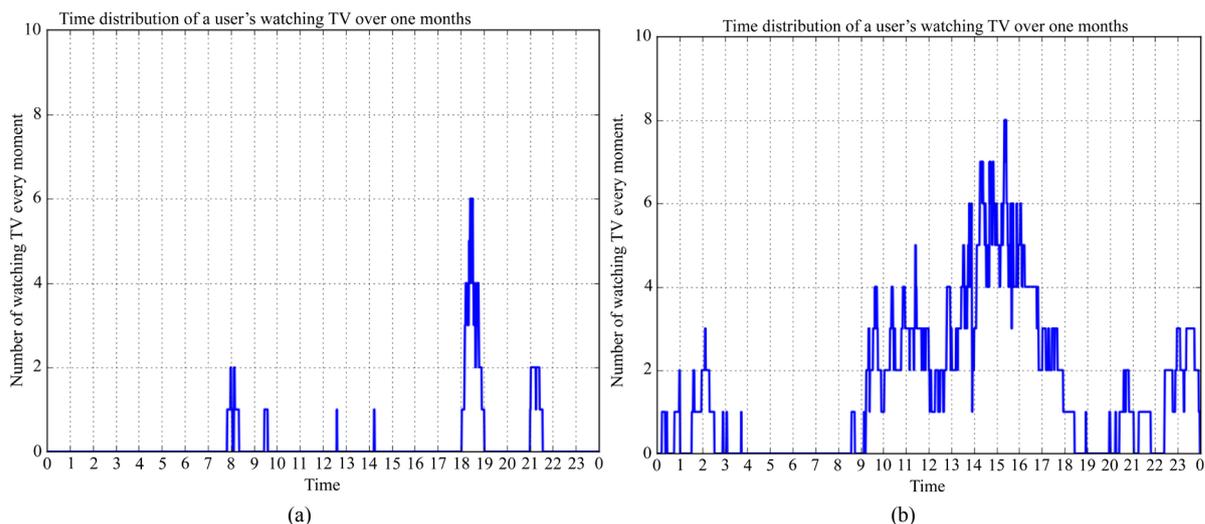


Figure 7. Time distribution of the number of views of user 1 time view and user 2 favorite channel. (a) User 1; (b) User 2
图 7. 用户 1 和用户 2 最喜好频道的观看次数时刻分布。(a) 用户 1; (b) 用户 2

通过对个体用户行为数据统计分析，发现个体用户观看行为具有各向异性特征，利用这些行为特点有助于开展预加载[5]、个性化推荐[6]、广告推荐等多方面的研究，以提高用户观看 IPTV 的体验和效率。

4. 结束语

本文从整个用户群体、单个频道和个体用户三个角度对网络电视用户数据进行统计分析，呈现网络电视用户观看电视的行为特点，为网络电视系统的设计、资源的合理分配以及个性化频道推荐等方面提供了依据。下一步的工作将会对个体用户进行更精细的数学建模，更加准确地呈现用户行为特征，通过深度学习，绘制用户画像，促进个性化用户精准频道推荐和广告推送。随着 IPTV 用户发展日趋饱和，基于用户行为统计特征分析，深度学习用户潜在兴趣，开启面向多尺度用户人群的增值业务研究对于 IPTV 产业发展意义重大。

基金项目

本研究获广州市科技计划产学研重大(国际合作)专项支持 No. 201704030124。

参考文献

- [1] Qiu, T., Ge, Z., Lee, S., et al. (2009) Modeling User Activities in a Large IPTV System. *Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement*, Chicago, Illinois, 4-6 November 2009, 430-441. <https://doi.org/10.1145/1644893.1644945>
- [2] Abrahamsson, H. and Nordmark, M. (2012) Program Popularity and Viewer Behavior in a Large TV-on-Demand System. *Proceedings of the 2012 Internet Measurement Conference*, Boston, Massachusetts, 14-16 November 2012,

199-210.

- [3] Wang, B., Wen, X.M., Sun, Y., *et al.* (2009) A New Approach Measuring Users' QoE in the IPTV. *IEEE 2009 Pacific-Asia Conference on Circuits, Communications and Systems*, Chengdu, 16-17 May 2009, 453-456. <https://doi.org/10.1109/PACCS.2009.63>
- [4] Cha, M., Rodriguez, P., Crowcroft, J., *et al.* (2008) Watching Television over an IP Network. *Proceedings of the 8th ACM SIGCOMM Conference on Internet Measurement*, Vouliagmeni, Greece, 20-22 October 2008, 71-83, 71-84. <https://doi.org/10.1145/1452520.1452529>
- [5] Yang, C. and Liu, Y. (2015) On Achieving Short Channel Switching Delay and Playback Lag in IP-Based TV Systems. *IEEE Transactions on Multimedia*, **17**, 1096-1106. <https://doi.org/10.1109/TMM.2015.2429552>
- [6] Yang, C., Ren, S.X., Liu, Y., Cao, H.W., Yuan, Q.H. and Han, G.Q. (2018) Personalized Channel Recommendation Deep Learning from a Switch Sequence. *IEEE Access*, **6**, 50824-50838. <https://doi.org/10.1109/ACCESS.2018.2869470>

知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2161-8801, 即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: csa@hanspub.org