

## 高校网络沉迷与防沉迷系统的研究与实现

林海卓<sup>1,2</sup>, 王继龙<sup>1,2</sup>, 张颐哲<sup>1,2</sup>, 朱晶<sup>1,2</sup>

(1. 清华大学 网络科学与网络空间研究院, 北京 100084; 2. 清华大学 计算机科学与技术系, 北京 100084)

**摘要:** 基于教育网流量分析, 提出游戏服务商 IP 地址发现机制, 包括 Boilerpipe 正文抽取算法以及 Stanford Chinese NLP 中文分词算法, 通过流量分析技术发现与分析大学生网络沉迷现象的一般方法和计算框架。统计了学生在网络游戏、视频观看、社交网站访问这 3 个常见的网络沉迷行为, 包括总时长、连续沉迷时间、频率、沉迷时间段选择等关键指标。提出网络沉迷指数的概念并通过层次分析法对沉迷现象进行量化研究。进一步基于网络沉迷指数模型设计和实现了网络防沉迷系统。

**关键词:** 网络沉迷; 网络游戏; 社交网络; 防沉迷系统; 层次结构分析法

中图分类号: TP393

文献标识码: A

文章编号: 1000-436X(2014)Z1-0170-08

## Research and implementation on college students internet addiction disorder and anti-addict system

LIN Hai-zhuo<sup>1,2</sup>, WANG Ji-long<sup>1,2</sup>, ZHANG Yi-zhe<sup>1,2</sup>, ZHU Jing<sup>1,2</sup>

(1. Academy of Network Science and Cyber Space, Tsinghua University, Beijing 100084, China;

2. Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China)

**Abstract:** Based on the traffic analysis on Cernet2, a new discovery mechanism of games service provider IP address is proposed, including Boilerpipe text extraction algorithm and Stanford Chinese NLP, realize this new general method and computing architecture of college student internet addiction. Three kinds of Internet addiction behaviors-online games, video viewing and social networking sites visit, including the total hours spend, length of continuous addiction, frequency, selected time of addiction are analyzed. Concept of Internet addiction disorder index is proposed with quantization on the internet addiction disorder made through AHP. Finally, Internet addiction prevention system is designed and implemented.

**Key words:** internet addiction disorder; online games; social network; anti-addict system; AHP

### 1 引言

网络沉迷 (IAD, internet addiction disorder) 最早由 Federioco 与 Lucio 提出, 其含义包括对网络的操作出现时间失控、欲罢不能、难以自拔的状况; 多沉湎于网上自由说谈或网上互动游戏, 并因此忽视了现实生活的存在, 或对现实生活不再满足; 初时只是精神上的依赖, 尔后可发展成为躯体的依赖<sup>[1,2]</sup>。随着互联网飞速发展, 很多当代大学生沉浸于此难以自拔, 并极大在影响了正常的学业, 因此造成的家庭、校园甚至是社会的种种不安问题。

人大代表在十八大中多次针对学生网络沉迷现象的提案已引起社会的广泛关注。

然而由于当前互联网的高度普及以及购物、资讯、交友、多元化教育资源、教育渠道的开展, 大学生每天的学习生活已经与互联网密不可分, 这也使得大学生网络沉迷的现象, 特别是游戏沉迷的现象越发难以被监管。当前高校所采用的各类监管章程均不可避免地具有滞后性和执行力低等问题。其根本原因是缺少一种实时的监管和反馈技术, 能够通过挖掘大学生上网时的流量行为特征来第一时间判断大学生的网络沉迷状态。因此, 本文采用流

收稿日期: 2014-10-18

基金项目: 国家重点基础研究发展计划(“973”计划)基金资助项目

Foundation Item: The National Basic Research Program of China (973 Program)

量识别与分析的手段建立网络沉迷定义与发现的一般判别方法和流程架构，同时基于网络游戏沉迷、视频沉迷和社交网站沉迷 3 个维度，通过层次结构分析法建立网络沉迷指数的理论模型，并开发与实现了高校防沉迷系统。

## 2 游戏服务商 IP 地址发现机制

当前学生对于网络游戏的选择集中度非常高，主要集中在魔兽世界、英雄联盟、地下城与勇士等游戏中。传统的 3 大流量识别与分析方法在网络沉迷的研究中效果欠佳，其主要原因包括以下几点。

1) 按端口进行流量识别。该方法理论上需要各个应用在 IANA 上注册其使用的端口，因此可以通过流量中的端口号判断流量类型，这种方法虽然具有简单、快捷、准确的特点，但目前的应用很多采用随机端口，以及很多应用并没有在 IANA 注册，加之会出现多种应用使用同一端口地问题，这使得传统按端口进行流量识别的方法难以有效的执行。对于客户端类网络游戏和对战平台而言，其客户端的端口、服务器端的端口和 IP 一般较为固定。但由于教育网与外网的连接速度问题，很多学生在进行网络游戏时，会开启各种代理，有部分代理会改变本地端口，甚至使用随机端口，服务器端的端口也不十分固定。因此，为保证准确性，不以端口作为识别标准。2) 根据特征字进行匹配<sup>[3]</sup>。该方法适用于连接模式固定，特征字出现频率较高的网络应用，例如 FTP。部分关键词会重复出现，通过分析流量包即可判断。但此种方法需要分析数据分组内容，时间开销较大，面对大规模流量时处理有困难，此外还面临隐私权等问题，在高校网络环境中应用特征词的匹配方法不具备可行性。3) 特征识别<sup>[4]</sup>。通过用户行为特征、流特征，使用机器学习等技术，通过人工标注的训练集学习到判断标准，用于自动识别流量。这种方法是目前主流的算法，可以应用于各种流量的识别，但应用技术较为复杂，准确率难以达到特别高的要求。

本文采用的网络沉迷流量分析的技术依托于游戏服务商的服务器 IP 地址及代理商的服务器 IP 地址。出于安全考虑，并没有任何一家运营商对外公布其服务器列表和使用端口列表，因此本文的难点在于设计游戏服务商的服务器 IP 地址及代理商的服务器 IP 地址的发现机制与计算框架。由于玩网络游戏具有持续性，即用户在进行游戏后，会有较

大的概率在短时间内再次进行，如果他访问了其他游戏服务器，可以找到该游戏其他服务器的 IP。进而通过统计每名学生与游戏服务商服务器 IP 地址集的连接情况统计出每名学生每周的游戏总时长、平均连续时间、频率以及时间概率分布。

本文采用分光技术从 CERNET2 清华对外出口的线路中抽取数据，抽样比为 10:1，并从中按照 IP 信息筛选出校园网本科生的网络流量信息。使用 NetFlow 每 15 min 制成一份报告，报告由一条条的流量记录组成，每条流量记录包含源地址及端口、目的地址及端口、协议类型、分组大小等信息。完整的网络沉迷流量分析框架如图 1 所示。

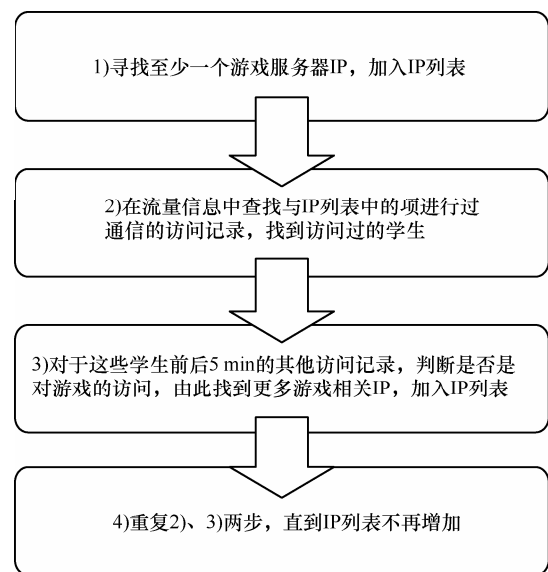


图 1 网络沉迷流量识别流程

其中，第 3 步中需要判断某个 IP 是否与网络游戏相关，这里采用的方式如下。首先，将此 IP 导入搜索引擎。通过 Boilerpipe 中文抽取算法<sup>[5]</sup>对该 IP 地址在 Google 搜索引擎搜索排名前 10 位的正文进行提取，获得提炼后的主题短文。其次，采用 Stanford Chinese NLP 分词工具<sup>[6]</sup>对提炼后的短文进行分词处理，最后用 TF-IDF 算法<sup>[7]</sup>将分词后的关键词组与网络游戏名称词组进行名称匹配，若包含网络游戏名称，则判定此 IP 地址与游戏相关。但有时由于用户使用代理，或在搜索结果中不直接出现游戏名称，而是代以玩家对此游戏的俗称等，关键词可能并非直接为游戏名称。因此，这里采用两级搜索引擎分词关联。将搜索 IP 结果中的关键词再次放入搜索引擎进行搜索，做二次关联，如果第二次搜索结果中再次提取的关键词中包括游戏名称，则也判定

此 IP 地址为游戏相关 IP。

下面以网络游戏魔兽世界为例,进行一次上述的流量识别工作。

**第 1 步** 首先对比在开启游戏客户端前后本机的连接状态,找到魔兽世界的连接如图 2 所示。

TCP	192.168.1.105:2135	203.66.119.22:3724	TIME_WAIT
TCP	192.168.1.105:2136	203.66.119.23:3724	ESTABLISHED
TCP	192.168.1.105:2142	203.66.119.44:3724	ESTABLISHED

图 2 本机连接状态

将上述服务器 IP: 203.66.119.22、203.66.119.23、203.66.119.44, 加入到 IP 列表, 作为流量识别的起始点。

**第 2 步** 在所有网络流量中, 查找与 203.66.119.22、203.66.119.23、203.66.119.44 这 3 个 IP 地址相关的流量记录, 结果找到若干 IP 与上述地址进行过通信, 其中包括 59.66.130.161 (为保护个人隐私, 本文中出现的具体学生 IP 均为示例)。

**第 3 步** 对 59.66.130.161 同次的流量信息进行搜索, 将其访问的外网 IP 放入搜索引擎进行搜索。对于搜索的前 10 个结果都做正文提取以及分词的处理后, 统计各个词语在搜索结果中出现的频率, 结合语料库中各个词语出现的频率, 得到每个词语的 TF-IDF 值。例如, 在流量信息中包含 121.254.200.130 这个 IP 地址, 将其放入搜索引擎后, 共有 60 多个搜索结果。对前 10 个结果进行分析, 最终得到的关键词包含以下几个: 艾泽拉斯、代理、台服等, 将若干个关键词再次放入搜索引擎进行搜索, 对结果再次进行正文提取、分词的处理, 结果中包含了游戏名称“魔兽世界”, 由此推断这个 IP 与游戏直接相关, 此网络用户是在进行游戏。在这里会对游戏名称在 TF-IDF 算法计算词频给予一定的“优惠”, 将其词频乘以 3 倍后再与其他词语进行比较, 这是由于与纯粹判断文章主题不同, 本文的目的是判断一篇文章主题是否为某个特定词语, 因此对这个词的出现更加敏感。在这个例子中, 由于 121.254.200.130 并非是魔兽世界游戏服务器, 因此, 在直接搜索 IP 时, 并没有多少结果中直接包含游戏名称, 但该 IP 为玩此游戏用到的代理服务器, 通过间接的二次关联搜索, 就得到了 IP 与游戏相关的信息。由于每次搜索并加以提取正文、分词、关键词认定过程的时间开销较大, 需要避免重复搜索, 一是对于已搜索过的用户网络记录以及 IP 进行记录, 当需要进行搜索时先判断是否已经搜索过, 二是对

于进行过的二次关联搜索词汇, 记录是否与游戏相关, 这样避免多次对同一词汇进行重复搜索。

**第 4 步** 由于 IP 列表增加, 重复第 2 步和第 3 步, 如果 IP 列表又有增加, 则继续重复, 直到列表不再变化。网络游戏运营商和代理运营商使用多个服务器有 2 种原因。第一是为满足不同地域玩家的网速需要, 例如分为电信服务器和网通服务器, 或者更精细地按照地域划分, 在这种情况下, 由于清华大学在全国网络中相当于一个点, 因此在选择服务器时会非常集中, 很少出现访问多个服务器的问题。第二是由于网络游戏内容的庞大或游戏人数的过多, 将游戏的不同部分运行分摊到不同的服务器, 在这种情况下, 即便是同一地点的玩家进行游戏时, 也可能连接到不同的服务器, 这会造成 IP 列表的多次增长, 带来更大的时间开销。不过在实际中, 重复进行第 2 步和第 3 步的次数基本不会超过 5 次。

### 3 识别结果与改进

本文实验环节的一个难点是如何对识别结果进行准确性的评估。幸运的是, 魔兽世界的访问端口固定为 3724, 并且在进行游戏时, 会同时与至少两台服务器进行通信, 即使使用代理, 代理种类也只有 3 种, 且端口固定, 这就提供了判断上述识别方法准确性的机会。需要注意的是, 并非所有网络游戏都有此良好的固定规则, 传统通过端口识别应用的方式仅仅适用于其中一部分网络游戏流量的鉴定, 这里使用此方法对魔兽世界的流量识别结果进行检验。

降低识别准确率的原因有 2 种: 一为遗漏, 即本应为游戏流量, 但在识别准则下被错过, 其中包括游戏相关 IP 未能被加入 IP 列表, 以及在语义识别时认定为并非游戏相关; 二为错判, 即识别为网络游戏流量的记录本并非游戏流量。

基于上述校验原理进行识别准确率的评估, 结果如图 3 所示。

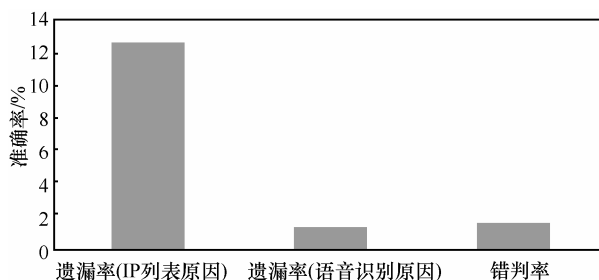


图 3 识别准确率

从图 3 中可以看到，判断的准确性并不让人十分满意。影响判断准确率的主要原因是没有搜索到游戏运营商和代理的全部服务器列表，从而造成了超过 10% 的遗漏率。这里采用 2 个方案来进一步降低遗漏率和错判率。

1) 游戏运营商的服务器放置较为集中，很多情况下 IP 地址十分接近。例如在上面的例子中，203.66.119.22、203.66.119.23、203.66.119.44 同属同一 C 类地址网段。因此通过对方法进行改进，在找到新的服务器后，将其所属 C 类网段加入 IP 列表，这样就大幅减少了遗漏服务器的可能。

2) 同宿舍的同学往往容易互相影响，而网络游戏多数为多人对战式的游戏。在高校大学生的宿舍同学中，同时玩同一款游戏的可能性较高，因此，可以当找到玩特定类别游戏的同学时，对其本人及宿舍同学的当期流量记录进行判断，这样可以比较有针对性地增加发现游戏相关 IP 的概率。虽然从表面上增加了每次搜索的数量，但同时减少了 IP 列表增加的轮数，因此总时间开销并未显著增多。

进行上述 2 种改进后，对数据进行重新分析判断，并与准确方法重新对比的结果如图 4 所示(改进 2 中包含改进 1)。

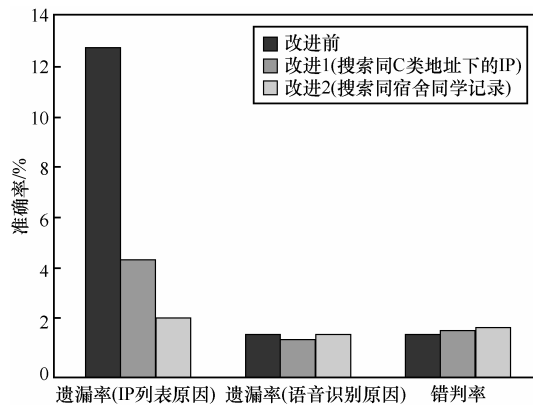


图 4 改进后的识别准确率

从图 4 中可以看出，在基本不影响错判率和由于语义识别原因造成遗漏地情况下，2 种改进都对发现游戏相关 IP 产生了显著地功效。最终魔兽世界的流量识别准确率可以接近 95%，为后续网络沉迷指数研究以及防沉迷系统的设计与实现提供了准确的信息。

#### 4 网络视频与社交网站沉迷

在常见网络沉迷现象中，除网络游戏沉迷外，

其他主要的形式包括频繁浏览社交网站、长时间观看视频等<sup>[10]</sup>。由于用户行为的识别依据为浏览器访问记录，因此，需要依靠域名来识别网络行为。

伴随视频压缩技术的进步和网络带宽的提高，更多的人有条件选择网络视频进行观看。近年来多数的电视台都陆续推出了自己的网络直播平台，很多热门的电视剧也会在网络同步播出，这些都进一步扩大了网络视频的观众人数。由于在内容选择、点播回看等方面具有无与伦比的优势，网络视频甚至有取代电视成为主流视频媒介的趋势。特别是在大学校园，宿舍基本没有电视可供观看，网络视频、特别是美剧、韩剧视频成为了学生的首选。但过于方便的点播方式也创造了沉迷的客观条件。目前国内的主要视频网站包括优酷、土豆、爱奇艺等和几大门户网站的视频站点。与此同时，刷微博、逛论坛等社交网络活动也是造成大学生网络沉迷的主要诱因。目前在国内最主要的社交网站有 QQ 空间、人人网、开心网、新浪微博、腾讯微博、豆瓣、百度贴吧等，在本文统计中也以以上社交网站为标准。

视频与社交网站沉迷数据的来源为搜狗浏览器的访问数据。由于教育网和公网间的访问速度问题，具有高速模式的搜狗浏览器成为最受学生欢迎的浏览器，因此搜狗浏览器的记录能够较为准确地反映学生的网络行为。数据分组括用户每次执行访问的网址、时间、用户 IP 地址等。数据格式如图 5 所示。

<http://user.qqzone.qq.com/304402128/infocenter>  
1330591749 59.66.130.161 21:15:06 00AFAF39C0903B17BB8248F764CA575F

图 5 数据格式

#### 5 基于 AHP 的沉迷指数量化模型

层次分析法 (AHP, analytic hierarchy process) 是一种用于定量解决复杂决策问题的结构性数学方法。它基于数学和心理学理论，于 20 世纪 70 年代由 Thomas L. Saaty<sup>[8,9]</sup>提出，并很快成为一种重要的决策理论。这里应用 AHP 对网络沉迷指数进行量化研究，沉迷指数的量化模型结构如图 6 所示。

在量化过程中，单位均为每周，如每周总时长、每周平均连续时长等，旨在将网络沉迷程度量化，得到每个学生的网络沉迷指数。本文的研究对象包括 3 个内容，即网络游戏、视频和社交网站，沉迷指数也包含这 3 个方面。考虑到网络游戏的危

害性更高，特别是对周围同学的影响更大，因此设定权重为 2，其他 2 种沉迷现象为 1。设  $\kappa$  为网络沉迷指数，从准则层到目标层的计算方式为

$$\kappa = (\kappa_{\text{游戏}} \times 2 + \kappa_{\text{视频}} + \kappa_{\text{社交网站}}) / 4$$

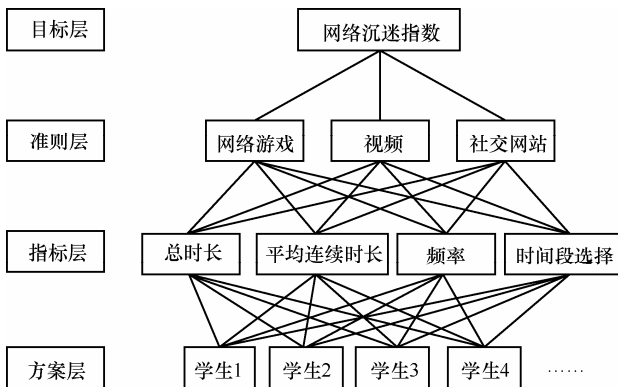


图 6 沉迷指数量化模型结构

对于 3 个分项的沉迷指数，由 4 个沉迷属性决定，包括总时长、平均连续时长、频率和时间段选择。在这部分中属性的相互权重难以准确量化，采用一致矩阵法。学生在这 4 个属性上的评分规则，即方案层到指标层的定量化，将由群体的统计规律而定。因此，在评价学生的网络沉迷程度时，需先统计学生在网络游戏、视频、社交网站 3 个沉迷现象中的总时长、平均连续时长、频率和时间段选择 4 个属性，将其按照设定好的评分规则进行打分，继而根据判定矩阵得到在 3 种网络沉迷现象中的评分，最后加权求和得到个人的网络沉迷指数。

这里采用一致矩阵法对网络沉迷程度进行量化。首先分别设定网络游戏沉迷矩阵  $Z_{\text{游戏}}$ 、视频沉迷矩阵  $Z_{\text{视频}}$ 、社交网站沉迷矩阵  $Z_{\text{社交}}$  如下

$$Z_{\text{游戏}} = \begin{matrix} & \begin{matrix} A1 & A2 & A3 & A4 \end{matrix} \\ \begin{matrix} A1 \\ A2 \\ A3 \\ A4 \end{matrix} & \begin{bmatrix} 1 & 5 & 7 & 2 \\ 1/5 & 1 & 4 & 1/3 \\ 1/7 & 1/4 & 1 & 1/5 \\ 1/2 & 3 & 5 & 1 \end{bmatrix} \end{matrix}$$

$$Z_{\text{视频}} = \begin{matrix} & \begin{matrix} A1 & A2 & A3 & A4 \end{matrix} \\ \begin{matrix} A1 \\ A2 \\ A3 \\ A4 \end{matrix} & \begin{bmatrix} 1 & 3 & 8 & 1/2 \\ 1/3 & 1 & 6 & 1/4 \\ 1/8 & 1/6 & 1 & 1/9 \\ 2 & 4 & 9 & 1 \end{bmatrix} \end{matrix}$$

$$Z_{\text{社交}} = \begin{matrix} & \begin{matrix} A1 & A2 & A3 & A4 \end{matrix} \\ \begin{matrix} A1 \\ A2 \\ A3 \\ A4 \end{matrix} & \begin{bmatrix} 1 & 2 & 1/4 & 1/2 \\ 1/2 & 1 & 1/7 & 1/3 \\ 4 & 7 & 1 & 3 \\ 2 & 3 & 1/3 & 1 \end{bmatrix} \end{matrix}$$

其中， $A1 \sim A4$  分别为总时长、平均连续时长、频率和时间概率分布。不同沉迷类型中，沉迷属性的侧重点有所不同，其数值的设定参考前面对群体网络沉迷特点的统计。对于沉迷较为严重的方面，权重设定相对较高。

对上面设定的判断矩阵进行一致性检验，网络游戏沉迷判断矩阵最大特征根为  $\lambda = 4.1291$ ， $CI = 0.043$ ， $CR = 0.0478 < 0.1$ ，满足一致性要求，特征向量归一化后为  $\omega_{\text{游戏}} = [0.5193, 0.1326, 0.0532, 0.2950]$ ，分别为总时长、平均连续时长、频率和时间段选择 4 项属性对于网络游戏沉迷程度定量化的权重。视频沉迷和社交网站沉迷的判断矩阵也通过一致性检验， $\omega_{\text{视频}} = [0.3169, 0.1464, 0.0381, 0.4985]$ ， $\omega_{\text{社交}} = [0.1330, 0.0737, 0.5693, 0.2240]$ 。

为得到学生的网络沉迷指数，需要设定指标层的评分规则，即如何通过学生的网络行为对其在总时长、平均连续时长、频率和时间段选择 4 个指标上进行数值评价。这里采用人工设定指标层的评分规则，规则的设定将参考各种沉迷属性的统计结果，使用相对指标。各属性的沉迷分值范围为 0~100，数值越高表示沉迷程度越深，实际评分由个人在此沉迷属性中的情况在群体中的排名位置决定，例如某学生的网络游戏总时长在所有学生中位于前 7%，则其评分为 93 分；如果位于 48%，则得分为 52，以此类推。本文实验以 2011 年 9 月到 2012 年 3 月的校园网对外流量为研究样本。

下面以某学生和整个群体的沉迷行为统计结果为例，计算该学生的网络沉迷指数，统计结果如表 1 所示，表中的百分比表示该沉迷属性位于群体中的位置，括号内为相应的评分。

类型	总时长	平均连续时长	频率	时间分布
网络游戏沉迷	6%(94)	36%(64)	14%(86)	78%(22)
网络视频沉迷	84%(16)	55%(45)	91%(9)	80%(20)
社交网站沉迷	21%(79)	43%(57)	15%(85)	64%(36)

最后计算得到网络沉迷指数  $\kappa = (68.3658 \times 2 + 21.9713 + 71.1624) / 4 = 57.4663$

这样就得到了该学生的网络沉迷指数，通过计算每名学生的网络沉迷指数，可以得到每个人相对于群体的沉迷程度，并可以通过对比 3 个沉迷现象分别的沉迷指数，判断学生的沉迷特点和类别。

一个好的评价体系在于有良好的区分度，也就是应用此体系对群体进行评价后，所有个体的评分分布可以接近服从正态分布。使用 SPSS 对评价结果，包括综合沉迷指数和 3 个分项沉迷指数进行正态分布检验，检测方法为 *D* 检验，结果视频沉迷指数和社交网沉迷指数的 *P* 值大于 0.05，表明数据服从正态分布。而综合沉迷指数和网络游戏沉迷指数之所以不满足正态分布，是由于网络游戏的参与度只有 6 成左右，有大量学生在此项的沉迷指数为 0，参与学生的沉迷属性在群体中的排序位置普遍靠前，因此，沉迷指数较高，这样就出现了两头多，中间少的情况。受此影响，加之网络游戏沉迷指数在综合沉迷指数计算中所占权重较大，使得后者分布也没有通过正态分布检验。对此这里可以将网络游戏参与者和不参与者分离，对参与者的评分结果重新进行检验，结果表明它服从正态分布，综合沉迷指数也顺利通过检验。由此，本文的评价体系是合理有效的。

在下一章节，将基于沉迷指数进一步建立网络防沉迷系统，用于直观展示每位学生的沉迷状况。

## 6 高校大学生防沉迷系统

### 6.1 系统概述

网络成瘾与其他病态成瘾症相似，身陷其中的人们对于自己缺乏足够的认知。如果能让学生对自己的网络行为有更充分准确的了解，并能够与周围同学进行对比，将有助于提高自我认知，达到自我控制网络沉迷的效果。因此，亟需建立一套系统，用于实时采集网络沉迷数据，并随时反馈统计结果（统计均以周为单位）。用户点击相应统计时，将结果以直观的图表形式展示在系统页面，使用工具为 Google Chart。系统分为 3 种模式：学生模式、辅导员模式和研究模式。

1) 学生模式。当学生访问网络防沉迷系统后，通过 IP 识别进行登录，登录后系统从数据集调取该用户数据，内容包括总体及分项的网络沉迷指数，以及各个沉迷属性的评分，并对比群体的整体情况。为保护个人隐私，学生登录后将只能看到自己的记录，以及与整体情况的对比。

2) 辅导员模式。对于沉迷程度较为严重的和正

在逐渐陷入沉迷的学生，应将信息反馈到院系或相关人员处，以便被及时发现，避免影响整个大学生活。在学校里，对学生的情况较为了解和起监督引导职责的是辅导员，辅导员登录系统后，可以看到所管辖的每个学生的网络沉迷状况，通过排序和筛选，找到沉迷的学生，并结合其实际学习生活情况，进行相应的引导工作。

3) 研究模式。系统对于网络沉迷的研究工作也会大有裨益，这套系统充当了研究平台的作用。对原始的流量数据和浏览器数据做相应识别和提取工作后，将统计结果保存在服务器，为后面的研究做好数据的铺垫。通过此系统可以查看直观的图表，对沉迷行为作进一步的挖掘分析。此外，通过沉迷行为的时间序列，可以对个人或群体的沉迷趋势展开研究，这将使研究更加科学和全面。

### 6.2 系统流程

数据来源包括网络流量数据和浏览器访问记录。流量数据从校园网对外出口处抽样获取，经过流量识别后得到网络游戏沉迷信息，放入服务器数据集。浏览器访问记录通过信息提取后，以相同格式放入数据集。数据的收集和处理工作实时进行。

当用户访问网络防沉迷系统时，系统向服务器发出数据请求，根据一定的权限控制，将访问者自身数据及整体平均数据返还，并将结果展示给用户。展示内容有沉迷量化指数和沉迷属性统计结果，前者包括综合沉迷指数、分项沉迷指数及各沉迷属性评分，后者包括总时长、平均连续时长、频率及时间段选择。以上内容均可与选定群体进行对比，还可以按照时间序列查看变化趋势。具体流程如图 7 所示。

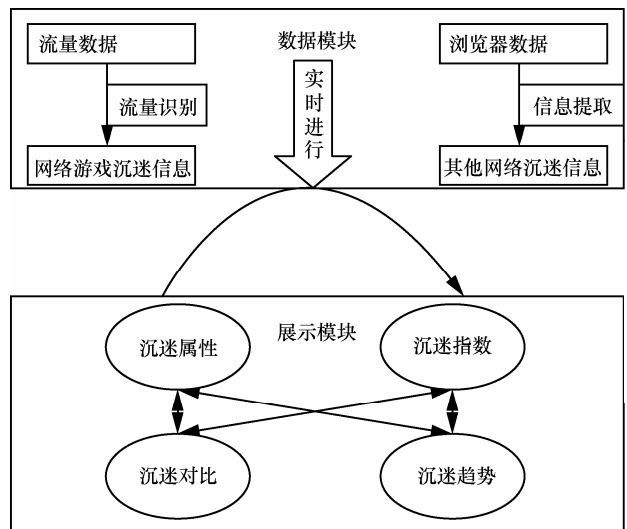


图 7 网络防沉迷系统流程

### 6.3 系统展示

部分展示内容如图 8 和图 9 所示。

内容	个人	整体	是否超过均值
网络沉迷指数	57	57	√
——网络游戏沉迷指数	68	73	×
——网络游戏总时长	94	72	√
——网络游戏平均连续	64	61	√
——网络游戏频率	86	69	√
——网络游戏时间段选择	22	81	×
——视频沉迷指数	22	47	×
——视频总时长	16	39	×
——视频平均连续时长	45	32	√
——视频频率	9	27	×
——视频时间段选择	20	58	×
——社交网站沉迷指数	71	36	√
——社交网站总时长	79	37	√
——社交网站平均连续时长	57	46	√
——社交网站频率	85	29	√
——社交网站时间段选择	36	51	×

图 8 网络防沉迷系统流程（沉迷指数）

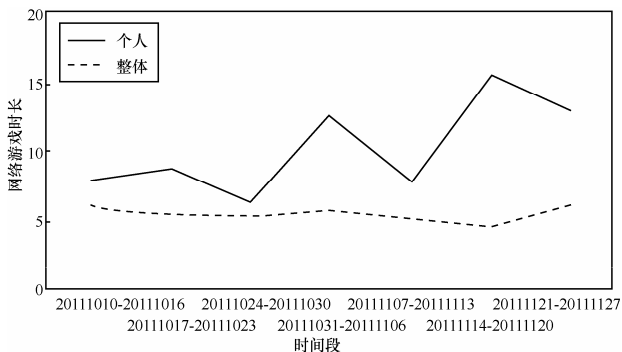


图 9 网络防沉迷系统流程（网络游戏时长）

从图 8 中可知，该学生总体的沉迷程度与整体相当，沉迷指数均为 57。与整体相比，共同的特点为偏好于网络游戏，但在游戏时间段选择上更加健康，较少出现熬夜游戏的情况。不同特点为该学生对视频不太感兴趣，除看视频平均连续时长外，其他指标都远小于平均值。该学生更沉迷于社交网站，除时间段选择指标外，其他均高于平均水平。通过上述量化的数值对比，可以一目了然个体相对于整体的沉迷特点。

图 9 为某学生的网络游戏时长趋势的统计结果，并与同时间段的整体情况进行对比。对比选项中，默认选择与全体学生对比，也可选择同性别、同年级、同院系进行更有针对性的对比。

通过图中的统计可以明显看到，该学生在网络游戏中花费的时间远远高于整体平均水平，需要得到充分的警惕。通过直观的图表对比情况，学生会对自己的网络沉迷情况以及变化趋势有更清晰的认识。

从图 10 中可以看到，相比整体情况，该学生在游戏的习惯上相对更为健康，其连续游戏小于 1 h、1~2 h 的比例超过整体均值，而长时间连续游戏的情况远小于整体均值。当在量化的沉迷指数中发现与整体情况差异较大的项时，便可点击查看统计详情，有益于了解自己在这一方面的行为特点，有针对性地进行改变。

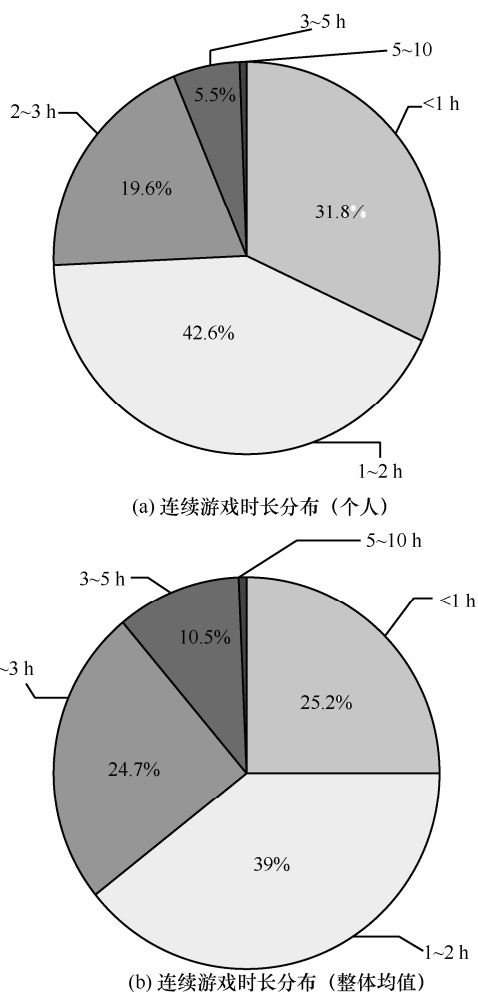


图 10 网络防沉迷系统（连续游戏时长分布）

图 11 为某学生的游戏时间段选择情况。从图 9 中可看出，该学生网络游戏总时长略低于平均水平，特别是在白天几乎不进行网络游戏。然而在晚上娱乐时间较多，特别是 0~3 点期间远超过整体情

况，这说明该学生不时玩到深夜，这对于正常的作息有着非常不利的影响。

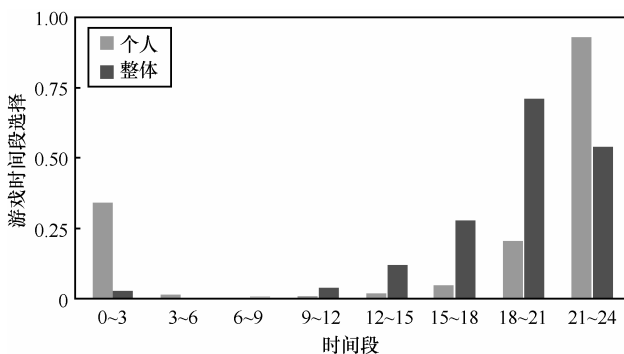


图 11 网络防沉迷系统流程（游戏时间段选择）

## 7 结束语

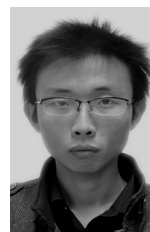
本文从网络管理与流量分析的角度研究了高校大学生网络沉迷的问题。量化得到了识别与判定大学生网络沉迷问题的一般分析方法与计算框架。并基于层次结构分析法，重点研究网络游戏、视频观看、社交网站 3 大类沉迷现象及其对应的包括总时长、连续沉迷时间、频率、沉迷时间段选择等指标的评价与建模标准，从而提出了大学生网络沉迷指数的概念及计算模型。最后基于大学生网络沉迷问题的流量统计与计算框架，设计并实现了高校大学生防沉迷系统，填补了当前社会学中棘手的网络沉迷问题在网络技术发现与分析领域的空白。

### 参考文献：

- [1] TONIONI F, D'ALESSANDRIS L, *et al.* Internet addiction: hours spent online, behaviors and psychological symptoms[J]. *General Hospital Psychiatry*, 2012,34(1):80-87.
- [2] PARK T W, PARK S H, CHUNG S K. Gender and Grade Differences in the Relationship between Psychopathology and Internet Usage in Children and Adolescents[J]. *J Korean Neuropsychiatr Assor*, 2011 50(5): 392-400.
- [3] TURMON M, JONES H P, MALANUSHENKO O V, *et al.* Statistical feature recognition for multidimensional solar imagery[J]. *Solar Physics*, 2010, 262(2): 277-298.
- [4] BABIC B R, NESIC N, MILJKOVIC Z. Automatic feature recognition using artificial neural networks to integrate design and manufacturing: Review of automatic feature recognition systems[J]. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 2011, 25(3): 289-304.
- [5] ROZMUS-GÓRNIKOWSKA M, Blicharski M, Kusiński J, *et al.* Influence of boiler pipe cladding techniques on their microstructure and properties[J]. *Archives of Metallurgy and Materials*, 2013, 58(4): 1093-1096.

- [6] LI M, ZONG C, NG H T. Automatic evaluation of Chinese translation output: word-level or character-level?[A]. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2. Association for Computational Linguistics[C]*. 2011.159-164.
- [7] SUGIYAMA K, HATANO K, YOSHIKAWA M, *et al.* Refinement of TF-IDF schemes for web pages using their hyperlinked neighboring pages[A]. *HYPertext '03 Proceedings of the fourteenth ACM conference on Hypertext and hypermedia[C]*. 2003.198-207.
- [8] SAATY T L. *What is the Analytic Hierarchy Process[M]*. Springer Berlin Heidelberg, 1988.
- [9] SAATY T L, VARGAS L G. *Criteria for Evaluating Group Decision-Making Methods[M]//Decision Making with the Analytic Network Process. Springer US*, 2013: 295-318.
- [10] SAATY T L, VARGAS L G. *Stabilizing Social Security for the Long-Term[M]*. *Decision Making with the Analytic Network Process. Springer US*, 2013.213-233.

### 作者简介：



林海卓（1988-），男，辽宁大连人，清华大学博士生，主要研究方向为网络管理与测量等。



王继龙（1973-），男，清华大学教授、博士生导师，主要研究方向为下一代互联网体系、网络管理与测量等。



张颐哲（1986-），男，辽宁大连人，清华大学硕士生，主要研究方向为网络管理与测量等。



朱晶（1989-），男，辽宁大连人，清华大学博士生，主要研究方向为域间路由等。