

8-14-2020

Real-time Visualization Method for Log Data Stream

Li Hui

1. College of Management, Beijing Union University, Beijing 100101, China;;

Hongqian Chen

2. Beijing Key Laboratory of Big Data Technology for Food Safety, School of Computer and Information Engineering, Beijing Technology and Business University, Beijing 100048, China;

Qianyu Yang

2. Beijing Key Laboratory of Big Data Technology for Food Safety, School of Computer and Information Engineering, Beijing Technology and Business University, Beijing 100048, China;

Chen Yi

2. Beijing Key Laboratory of Big Data Technology for Food Safety, School of Computer and Information Engineering, Beijing Technology and Business University, Beijing 100048, China;

Follow this and additional works at: <https://dc-china-simulation.researchcommons.org/journal>



Part of the [Artificial Intelligence and Robotics Commons](#), [Computer Engineering Commons](#), [Numerical Analysis and Scientific Computing Commons](#), [Operations Research](#), [Systems Engineering and Industrial Engineering Commons](#), and the [Systems Science Commons](#)

This Paper is brought to you for free and open access by Journal of System Simulation. It has been accepted for inclusion in Journal of System Simulation by an authorized editor of Journal of System Simulation.

Real-time Visualization Method for Log Data Stream

Abstract

Abstract: To discover the network cheating behavior in the online examination room in real time, *a log data flow visualization method was proposed*. The monitoring log data flow was used as the data source of the method. The candidates, the examination machine and other information were extracted according to the rules of the log. The correct candidate/machine corresponding relationship was established according to the first related log record information. *The main information elements were mapped to the visual elements for each log data*. All students' status and behavior were presented in the same visual result. *A clear reminder will be achieved once there are cheating and other abnormal behavior*. The experimental results show that the method is able to show the log data stream data in real time. The data processing method can adapt to real-time interaction.

Keywords

data visualization, data stream visualization, log visualization, time series data, online examination

Recommended Citation

Li Hui, Chen Hongqian, Yang Qianyu, Chen Yi. Real-time Visualization Method for Log Data Stream[J]. Journal of System Simulation, 2016, 28(9): 2133-2138.

一种针对日志数据流的实时可视化方法

李慧¹, 陈红倩², 杨倩玉², 陈谊²

(1. 北京联合大学管理学院, 北京 100101; 2. 北京工商大学计算机与信息工程学院食品安全大数据技术北京市重点实验室, 北京 100048)

摘要: 针对网络考场中的常规监考手段无法及时发现网络作弊的问题, 提出了一种针对网络考场监控日志数据流的可视化方法。该方法以网络考场的监控日志数据流作为处理对象, 对数据流进行实时可视化, 以及时发现考试过程中的异常行为。首先将实时采集到的日志数据, 根据日志规则分解为考生、考试机、题目等主要信息; 根据考生的首次日志记录信息, 建立考生/考试机对应信息表; 然后针对每一条接收到的日志, 将其主要信息元素与可视化元素进行映射, 将整个考场中所有学生的考试状态呈现在同一可视化结果中。一旦有考生存在作弊等异常行为, 可视化结果中能够实现明显的提醒, 并能对考生进行行为分析。实验结果表明, 文中方法能够对日志数据流数据进行实时展示, 文中的数据处理方法能够满足实时交互需求。

关键词: 数据可视化; 流数据可视化; 日志可视化; 时序数据; 在线考场

中图分类号: TP391.9 文献标识码: A 文章编号: 1004-731X (2016) 09-2133-06

Real-time Visualization Method for Log Data Stream

Li Hui¹, Chen Hongqian², Yang Qianyu², Chen Yi²

(1. College of Management, Beijing Union University, Beijing 100101, China; 2. Beijing Key Laboratory of Big Data Technology for Food Safety, School of Computer and Information Engineering, Beijing Technology and Business University, Beijing 100048, China)

Abstract: To discover the network cheating behavior in the online examination room in real time, a log data flow visualization method was proposed. The monitoring log data flow was used as the data source of the method. The candidates, the examination machine and other information were extracted according to the rules of the log. The correct candidate/machine corresponding relationship was established according to the first related log record information. The main information elements were mapped to the visual elements for each log data. All students' status and behavior were presented in the same visual result. A clear reminder will be achieved once there are cheating and other abnormal behavior. The experimental results show that the method is able to show the log data stream data in real time. The data processing method can adapt to real-time interaction.

Keywords: data visualization; data stream visualization; log visualization; time series data; online examination

引言

近年来随着互联网应用的发现,在教育领域



收稿日期: 2016-05-30 修回日期: 2016-07-11;
基金项目: 北京市自然科学基金(9164028, 4154066);
作者简介: 李慧(1983-), 女, 山东, 博士, 讲师, 研究方向为数据挖掘与知识发现; 陈红倩(通讯作者 1982-), 男, 山东, 博士, 副教授, 研究方向为虚拟现实与可视分析。

中, 基于网络的在线考试与评估系统越来越应用广泛, 给教学评估带来巨大的变革。在线考试的优点包括避开人为因素带来的负面影响, 减少了物力人力的开销, 效率较高^[1]。与传统的方式相比较, 通过计算机技术和网络技术支持的在线考场, 可以动态地管理各种考试信息, 按照考试要求, 从题库中抽取题目生产各种试卷, 考试在计算机上进行答题。

<http://www.china-simulation.com>

• 2133 •

尤其是在线考场中的考试评分系统^[2],更是对在线考试系统的一大促进。通过计算机快速阅卷辅助实现高效的考试流程,成为现代考试方式的有力补充。

在在线考场的考试形式中,通过网络技术实现的作弊方式是最为难以防范的。该作弊形式通过交换考生账号信息进行代答或通过网络技术实现。该方式能够通过网络技术手段隐蔽完成,由于作弊考生没有身体动作方面的行为异常,因此通过常规的人工监考基本无法发现异常。即便通过考场中的摄像头进行视频监控,也难以实现准确高效的防范作弊。

对于这种新的作弊形式,必须针对在线考场研制一种作弊行为实时检测方法,能够在作弊行为发生时,尽快采取措施,并能够进行作弊行为举证,以降低作弊行为的发生。随着在线考场应用的增多,对在线考场的作弊防范与作弊举证方法的探讨,成为一个急需解决的问题。

2013 年孟小峰教授^[3]明确指出:“可视化技术是数据分析与信息获取的重要手段。”信息可视化能有效抓住人们的注意力。

针对在线考场中的考生网络操作行为监测问题,提出一种对在线考场日志进行实时分析,并实现针对考场状态进行实时监控与报警的方法。该方法从考场的考试服务器中实时截取考试机操作日志数据,并对日志数据进行即时可视化处理,使考务人员能够从可视化图形中一目了然了解当前考场状态,以及考生中是否存在潜在作弊行为。该系统的使用结果表明,文中所提出方法能够对日志数据流数据进行实时展示,能够实现有效的考场状态分析与监控,数据处理方法能够满足实时交互需求。

1 相关工作

国内外许多大学和研究机构致力于流场可视化技术理论及应用的研究,如美国国家宇航局埃姆斯研究中心、奥地利虚拟现实与可视化研究中心、德国斯图加特大学可视化与交互系统研究小组、清华大学、浙江大学 CAD&CG 国家重点实验室等。

在可视化方法的实现原理方面,Neumann^[4]给出了一个可视化信息分析技术的框架,根据可视化方法的实现原理,可视化技术主要可分为基于几何的可视化技术、面向像素的可视化技术和基于图标的可视化技术。Keim^[5]同时考虑数据类型、表现功能和交互功能三个方面,增加了标准 2D/3D 显示和重叠显示,并将文本/Web 页、算法/软件作为可视化的数据类型。

在传统考试形式中,为了防止作弊舞弊等违纪行为,一般通过监考教师对考生身体动作行为进行监控,以防止考生通过传递答案或卷面抄袭的形式进行作弊。

而在在线考试系统中,主要通过两方面降低作弊概率。一是根据题库测试试题数量和考生数量,提供随机测试试题搭配和随机试卷搭配两种方式进行检测。二是在设计中随机出题、选择题随机打乱顺序外,还需要实现防止程序间的切换,以及禁止考生关闭浏览器、禁止刷新当前网页、禁止下载、测试试题复制、鼠标右键屏蔽等等。然而现在各种浏览器中具有“解除网页对右键菜单的限制”、“解除网页对复制粘贴等操作的限制”等各种功能,而且是多标签页显示网页,对全屏也是一个限制。如胡世清^[6]利用 Silverlight 技术,通过鼠标与键盘事件处理,实现防作弊行为的在线考试系统。

在线考场中的日志数据是随考试过程实时产生的,因此是一种数据流的形式,对日志数据需要在考试过程中实施可视化处理,才能实现对作弊行为的及时发现和纠正。

对于数据流可视化来说,所主要面临的难点在于两方面。一方面是采集到的原始过程数据往往并不能直接用于最终展示过程。另一方面,数据是以数据流的形式不断产生的,而非一次性提前获取到的,因此必须以流式过程数据为基础进行与真实过程或仿真过程同步的实时可视化。

针对数据流可视化,朱扬勇^[7]总结了序列数据的分类和特点,给出了几种序列数据相似性度量和随机序列之间距离分布的统计信息。陈红倩^[8]提出

了一种基于静态数据载体的数据可视化方法, 实现过程状态数据的实时可视化。刘晓平^[9]提出针对数据流信息的可视化方法, 用于动态展示非线性系统的变化规律。夏晓忠^[10]基于 NetFlow 流技术通过提取园区网边界数据流的地址、端口、协议和流量等特征属性在三维空间中建立流的几何可视化模型, 简化了网络流量的显示。Charles^[11]提出一种渐进式可视分析方法, 能够对整个数据集进行分步探索式可视化。

2 数据流实时可视化方法

本文针对基于日志数据流的考场状态监控与分析问题, 设计了一种实时可视化方法, 该方法主要包含如下几个步骤:

- 1) 根据所需可视化的日志数据特征, 提取日志数据中所包含的主要信息内容。
- 2) 根据考生的首次日志记录信息, 建立考生/考试机对应信息表。

在本文所进行的在线考场日志分析过程中, 存在两个假设: 第一个假设是一般情况下, 在一场考试中, 一名考生在一台考试机上进行试卷作答, 偶尔会发生更换考试机的情况; 第二个假设是在考试过程中, 考生的第一次登录一般是考生本人的正常登录行为。

因此, 基于上述两个假设, 根据考生的首次日志记录信息, 建立考生/考试机对应信息表。

- 3) 针对日志数据流中每一条数据, 将日志中的信息与可视化元素进行对应, 形成可视化结果。

2.1 在线考场的日志数据实时采集

本文所使用的在线考场日志数据流数据, 其获取流程如图 1 所示。首先由考试服务器记录考试系统与考试机的考试操作, 形成考场日志; 然后由可视化监控计算机从服务器上实时采集该日志数据。

2.2 日志数据内容提取

在在线考场日志数据中, 所包含的主要内容有: 考试服务器编号、操作发生时间、考生唯一身

份标识、考试机 IP 地址、考试操作类型、所操作试题编号等。

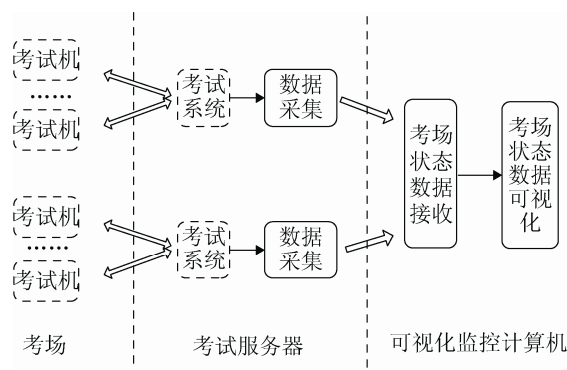


图 1 日志数据实时采集流程

提取到的信息, 需要进行数据预处理, 以便于可视化过程, 所需进行的预处理包括:

- (1) 将操作发生时间转换为相对时间。转换方法为: 将考试开始时间设置为 0, 考试过程中的时间转换方式如公式(1)所示:

$$RT = (H - H_0) * 60 + (M - M_0) \quad (1)$$

式中, RT 为相对时间; H 和 M 是当前考试操作发生时间的小时数与分钟数; H_0 和 M_0 是考试起始时间的小时数与分钟数。

- (2) 考生身份信息、考试机 IP 地址, 转换为编码。

- (3) 根据考生相关的首条日志, 将考生的编码与其对应的考试机进行绑定。

2.3 日志信息与可视化元素对应

本文提出一种状态折线的形式, 使用折线形式上的各可视化元素表达日志信息, 对应关系如表 1 所示。

表 1 日志信息与可视化元素映射关系

日志信息	可视化元素
操作开始时间	起始点 X 坐标
操作结束时间	结束点 X 坐标
考生/考试机信息	Y 值坐标
操作持续时长	对应线段长度
试题信息	线段颜色

2.4 可视化元素参数计算

操作开始的时间和结束时间, 对应起始点和结

束点的横坐标, 转换方法如公式(2)所示:

$$X = RT * \frac{W_{WIN} - 2 * X_M}{T_M} + X_M \quad (2)$$

式中: X 为当前交互数据在可视化图形中的映射点的横坐标值; W_{WIN} 为可视化窗口的整体宽度像素数, 一般设定 $W_{WIN} \geq 600$ 像素; T_M 为所设定的考试时间的总分钟数; X_M 为横坐标方向上可视化图形与可视化窗口边缘间的预留空白宽度, X_M 的值在 10~50 像素之间。

将考生信息通过公式(3)转换为当前交互数据在可视化图形中的映射点的纵坐标值:

$$Y = N_C * \text{int} \left(\frac{H_{WIN} - 2 * Y_M}{NUM_C} \right) + Y_M \quad (3)$$

式中: Y 为当前交互数据在可视化图形中的映射点的纵坐标值; N_C 为考生信息日志数据中首次出现的次序号, 每个考生在整个考试过程中, 其所对应的次序号将保持不变, 因此整个考试过程中 Y 值也会保持不变; $\text{int}()$ 为取整函数; H_{WIN} 为可视化窗口的整体高度像素数, $H_{WIN} \geq 400$ 像素; Y_M 为纵坐标方向上可视化图形与可视化窗口边缘间的预留空白高度, Y_M 的值在 10~30 像素之间; NUM_C 为在线考场中考生总数。

2.5 可视化元素着色

根据考场日志数据中的考试操作类型以及试题编号进行排列组合, 组合数如公式(4)所示:

$$C_P^1 + C_Q^1 \times C_T^1 \quad (4)$$

式中, P 表示考试题目无关的考试操作类型的总数, 用 Q 表示考试题目相关的考试操作类型总数, 用 T 表示试题的总数。

考试操作类型以及试题编号的每一种组合对应色彩对照表里的一条数据, 同时为在色彩对照表里为交互数据中的考试操作类型以及试题编号的每一种组合设置一个 RGB 颜色值。RGB 颜色值, 根据 24 颜色轮进行设置, 该颜色值设置方法能够有效区分相邻可视化元素。

2.6 日志数据可视化结果绘制

根据设定的绘制可视化图形的刷新频率绘制可视化图形。在每次绘制过程中, 针对映射点存储库中的每一个映射点队列中的所有映射点, 使用折线绘制模式进行绘制。每一个映射点队列中的映射点连接为一条折线, 两个映射点之间的连线颜色设置为从前点颜色渐变至后点颜色。

当同一映射点队列中的映射点所绘制出的折线绘制结果为水平, 则判断该折线所代表的考试机不存在潜在作弊行为; 否则, 则判断该折线所代表的考试机存在潜在作弊行为, 对该折线的宽度增加至原宽度的 2~3 倍。

3 实验结果与分析

根据本文提出的网络考场监控日志数据可视化方法, 已经实现了一个可视化原型系统, 并用于在线考场日志实时分析, 检验可视化方法的有效性。所采用的日志数据对应的考试环境包括 2 台考试服务器以及 400 台考试机, 其中每台考试服务器与 200 台考试机连接。

图 2 为基于本文所提出的方法, 实现的考场日志可视化展示效果图。从图 2 中能够直观准确的展示每位学生考试过程中的状态变化。

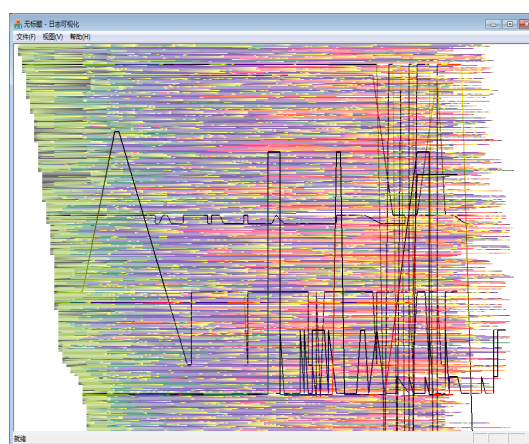


图 2 考场所有考生日志可视化效果

在该图中, 无潜在作弊行为的考生的交互数据映射点, 连接后获得的线为一条水平线, 水平

线宽度为 1 px, 水平线中的每一个不同颜色的线段对应一个考试操作的交互数据; 线段的不同颜色对应不同操作类型和试题编号。有潜在作弊行为的考生的交互数据映射点, 连接后获得的线为一条折线, 折线宽度为 2 px, 折线中的不同高度的线段对应的是使用不同考生信息的考试操作的交互数据。

图 3 为学生异常状态图, 考务人员根据可视化图形中的非水平加粗折线快速找到考场中存在潜在作弊行为的考生。

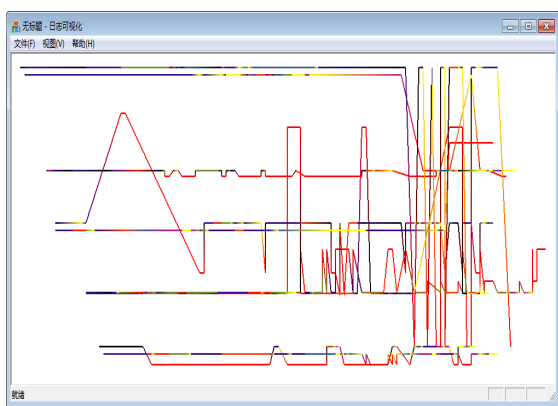


图 3 存在异常状态的考生日志可视化效果

根据图 3 中可以发现, 多名考生存在潜在作弊行为。我们选取考生编号为 207 的考生所在考试机, 进一步分析其考试操作行为。

图 4 是考生编号为 207 的考生所在考试机的操作分析图, 从图中可以分析学生的具体行为。

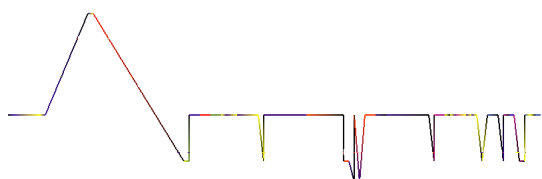


图 4 某个异常行为学生分析图

该考试机的部分具体行为描述如下:

(1) 在相对时间为 11 时, 以考生编号 207 的考生(考生本人)身份信息“登录”考试系统;

(2) 在相对时间为 26 时, 以考生编号 73 的考生身份信息“登录”考试系统, 对 73#考生的第 1

题和第 2 题进行“查看答案”操作;

(3) 在相对时间为 44 时, 以考生编号 268 的考生身份信息“登录”考试系统, 并在相对时间为 44 至 45 时间段内对 268#考生的第 1、2、3 题进行“查看答案”操作;

(4) 在相对时间为 46 时, 以考生编号 207 的考生身份信息“登录”考试系统; 并在相对时间为 46 至 51 时间段内, 依次对第 1、2、3 题进行“提交答案”操作;

(5) 在相对时间为 51 至 58 时间段内, 以考生编号 207 的考生身份信息, 对第 4、5、6、7 题进行“查看答案”操作;

(6) 在相对时间为 59 时, 以考生编号 268 的考生身份信息“登录”考试系统, 并在相对时间为 59 时对 268#考生的第 4、5 题进行“查看答案”操作;

(7) 在相对时间为 59 时, 以考生编号 207 的考生身份信息“登录”考试系统; 并在相对时间为 46 时对第 4 题进行“提交答案”操作;

由上述考试操作可以分析得出, 该生存在明显的考试抄袭行为。通过图 3~4 能够看出本文提出的方法, 能够快速便捷的对考场日志数据进行可视化, 并能有效的展示学生异常行为等, 能够为考务人员监考提供直观、易于理解的可视化效果。

4 结论

本文为解决网络考场中网络作弊手段隐蔽不容易发现的问题, 基于考场服务器上的日志, 实现了一种实时可视化的方法。该方法针对实时采集到的日志文件, 分析分析内容提取日志信息, 根据日志记录建立考生/考试机的对应关系; 基于每一条日志数据, 将其信息内容映射到可视化结果中。该方法能快速及时的发现网络考场中的异常操作行为, 对考务人员实现即时报警。

实验结果表明, 文中方法能够在考试过程中, 实时的展示日志数据流数据, 绘制效率能满足实时绘制和交互需求。

(下转第 2145 页)