

# 基于专家系统的 Web 日志数据处理方法研究

李春生, 豆立宪, 张可佳, 刘 涛, 邹林浩

(东北石油大学 计算机与信息技术学院, 黑龙江 大庆 163318)

**摘 要:**企业员工访问网站的记录在企业服务器中形成了大规模日志数据,从这些海量数据中发掘出有意义的信息是管理企业员工重要的环节,并成为大数据时代关注的重点。完成数据挖掘的核心工作之一就是对原始数据的预处理,它影响到数据分析结果的准确度。但由于 web 日志是 Apache、Nginx 和 IIS 等 web 服务器运行时产生的系统日志,web 日志格式根据使用 web 服务器的不同而不同。所以,传统的日志数据清洗方式难以应对目前如此复杂的日志格式。由于专家系统其技术特点能应对复杂的日志格式,所以通过结合专家系统,推理出对应的 web 日志格式,从而自动进行日志数据清洗,可以使得当下复杂海量 web 日志数据预处理更加高效、智能。首先,介绍了专家系统;其次,介绍了 web 日志数据一般的数据预处理流程;最后,结合专家系统对海量复杂的 web 日志数据进行数据预处理,并通过实验验证了该方法的可行性。

**关键词:**专家系统;web 日志数据;数据预处理;复杂日志格式;规则库

**中图分类号:**TP182

**文献标识码:**A

**文章编号:**1673-629X(2022)07-0117-05

**doi:**10.3969/j.issn.1673-629X.2022.07.020

## Research on Web Log Data Processing Method Based on Expert System

LI Chun-sheng, DOU Li-xian, ZHANG Ke-jia, LIU Tao, ZOU Lin-hao

(School of Computer and Information Technology, Northeast Petroleum University, Daqing 163318, China)

**Abstract:** The records of enterprise employees visiting the website form large-scale log data in the enterprise server. It is an important link to manage enterprise employees to find meaningful information from these massive data, and it has become the focus of attention in the era of big data. One of the core tasks of data mining is to preprocess the original data, which affects the accuracy of data analysis results. However, web logs are generated by web servers such as Apache, Nginx and IIS. The format of web logs varies according to different web servers. However, the traditional way of cleaning log data is difficult to deal with such a complex log format. Because the technical characteristics of knowledge reasoning machine can deal with complex log format, the corresponding web log format can be inferred by combining with knowledge reasoning machine, so as to automatically clean the log data, which can make the current complex massive web log data preprocessing more efficient and intelligent. Firstly, the expert system is introduced. Secondly, the general data preprocessing process of web log data is introduced. Finally, the massive and complex web log data is preprocessed with knowledge inference engine, and the feasibility of this method is illustrated by experiments.

**Key words:** expert system; web log data; data preprocessing; complex log format; rule base

## 0 引 言

通过对企业员工的 web 日志的数据挖掘,能够更加科学有效地监督管理员工,数据挖掘的核心工作之一就是对原始数据的预处理,它影响到数据分析结果的准确度。但当前面向 web 日志的数据预处理方法没有考虑到复杂 web 日志数据格式的问题,浪费了过

多的人力与时间,缺乏智能性且不够高效。通过专家系统自动对复杂 web 日志数据格式的判别,将对应的 web 日志格式进行数据清洗,完成数据预处理。能够很大程度地节约时间与人力,相对于当下 web 日志数据的预处理方法更加智能高效。

由于专家系统其技术特点能应对复杂的日志格

收稿日期:2021-07-12

修回日期:2021-11-12

**基金项目:**国家自然科学基金项目(51774090);黑龙江省自然科学基金项目(F2015020);黑龙江省青年创新人才培养计划(UNPYSCT-2020144);黑龙江省教育厅科研计划项目(2017-YDL-12)

**作者简介:**李春生(1960-),男,博士,教授,博导,研究方向为数据挖掘与智能系统、软件集成技术、图像处理与模式识别、智能仪器与计算机控制系统;豆立宪(1996-),男,研究生,研究方向为智能计算与智能系统。

式,因此应用专家系统能很大程度地实现自动对复杂 web 日志数据格式的判别,弥补传统方法的不足。

## 1 专家系统概述

专家系统定义:经由推论引擎、知识库及接口为基础而组成的系统即是所熟知的专家系统。专家系统的职责就是对社会生活中的问题进行一定的判断解释及认知。只是在这个特别的领域中有着不同的对于认知定义的解释。虽然当下还没有明确的定义,但是只要当设计的系统在执行的信度及效度、对专业知识的需求、复杂性等能够与专家不相上下的时候,就可以把该系统称为专家系统<sup>[1]</sup>。

### 1.1 基于框架的专家系统

由于基于规则的专家系统衍生出的一种推广,即基于框架的专家系统,本身具有一种区别于其余专家系统的编程风格。当初提出用框架来表示数据结构的人是 Minsky。很多概念的槽、知识、名称包含于框架之中,当符合这个概念的某些实例出现时,就可以把和这个实例相关的特定值输入到当前框架中。编程语言中引入框架的概念后,就形成了面向对象的编程技术。所以目前可以得出一个结论:在某种程度上面向对象的编程技术等于基于框架的专家系统。但由于当下日志的总字段长短不一,属性不同,所以基于框架的专家系统不适用于日志类型的判别。

### 1.2 基于案例的专家系统

采用以前存在的案例来推理解答目前问题的系统,即基于案例推理的专家系统。求解的顺序:第一步,得到问题,并开始对以往的案例进行匹配,直到找到跟目前问题最匹配的案例;第二步,基于第一步找到的当前问题最合理的匹配案例,就将之前案例所用的解决方案运用到现在,如果并没有找到合适的匹配案例,将当前的作为新的案例并加入到案例库。综上所述,基于案例推理的专家系统能够不断地学习新的经验,并不断进步。

当然,如何从过往的案例中找出最合理的案例并与当前的问题进行匹配是当前基于案例推理的专家系统比较大的一个难点。在案例寻找过程中,需要将一些过分相似的案例进行删除,因为如图案案例库太大,将会影响案例匹配的效率。目前来说一些主流的匹配算法有:径向基函数网络、k-近邻法、最近邻法。但基于案例的分析相对于基于规则的推理有一定的不准确性<sup>[2]</sup>。

### 1.3 基于规则的专家系统

由于该类系统是在专家系统中最早被研发出来的,同样也是基本的结构被用来组成专家系统。由于使用的规则来源于产生式定义,所以,在一般的状况

下,在基于规则的专家系统中就会用“if……then……”的类型来解决问题。环境与行为的关系一般会被用一种心理学术语来进行描述,即产生式。而专家系统在最开始是从其中演变而来的。斯坦福大学设计的第一个专家系统,它的主要功能是判定物质的结构。而且在产生式的系统当中,知识被分割成了两个方面,静态的系统是一方面,用事实来表示,例如事件与事物及其相互之间的联系;推理的行为及其推理过程是另一方面,用产生式的规则来实现。因为该类系统的知识库主要用来存储规则,所以该类系统又被称为基于规则专家系统<sup>[3]</sup>。

当前最常使用的方法是基于规则的专家系统,其主要原因是拥有许多成功的案例和便捷轻快的开发工具。基于规则的专家系统不仅可以用各种各样的规则来实现专家系统,而且可以用来直接对人的心理过程进行模仿。基于规则的专家系统有两种基本的推理:第一种是将已经存在的知识作为专家系统推理开始点的正向推理;第二种是将假定的结果作为专家系统推理开始点的反向推理。且对于反向推理,如果找不到匹配的规则,改变假设,重新进行推理<sup>[4]</sup>。

知识与处理的相分离。基于规则的专家系统的结构为知识库和推理引擎提供了有效的分离机制。因此,能够使用同一个专家系统框架开发不同的应用,系统本身也容易扩展。在不干扰控制结构的同时通过添加一些规则,还能使系统更聪明。

大多数基于规则的专家系统都能表达和推理不完整、不确定的知识。因此可以采用基于规则的专家系统思路来构建专家系统。

## 2 web 日志数据预处理流程

在数据分析的过程中,如果数据没有进行过数据处理,原来的数据依然存在各种各样的异常,且存在不一致、不完整等问题,即会导致数据分析的结果不够准确。所以在数据分析之前,对原始数据进行数据预处理的步骤至关重要。

一条真实的 NCSA (ECLF) 日志记录为“27. 19. 74. 143 -- [30/May/2013:17:38:20 +0800] "GET /static/image/common/swfupload.swf? preventswfcache=1369906718144 HTTP/1.1" 200 13333”,这类日志主要由以下几个部分组成:

标识符 (ident): 目前几乎多数的浏览器已经取消了该功能,因为涉及到了用户邮箱等隐私信息。

授权用户 (authuser): 虽然目前多数的网站的该类日志项并不存在,但是该类字段对于那些需要身份验证或者访问时候个人密码保护的信息项网站来说,不能为空。

日期时间 (date): 即 [日期/月份/年份:小时:分钟:秒钟时区], 占用的字符位数也基本固定, 一般的格式形如 [22/Feb/2010:09:51:46 +0800]。

请求 (request): 在网站上通过使用哪些方式获取了哪些信息在日志数据格式中也扮演着重要的角色。

请求类型 (method): 目前来说, 主流的请求类型主要有以下三种: POST/HEAD/GET。

请求资源 (resource): 可以是 css、动画、图片等资源, 也可以是某网页的地址, 即是请求资源的相应的 UPL。

协议版本号 (protocol): 通常是 HTTP/1.0 或者 HTTP/1.1, 用于版本信息及显示的协议。

状态码 (status): 用于响应浏览器的请求。

传输字节数 (bytes): 在一次请求过程中的资源数量。

访问主机 (remotehost): 已经解析的域名或者用于显示主机的 ip。

各类 web 日志数据的处理流程基本一致。处理流程主要分为: 数据抽取、数据格式转换、数据重复值处理<sup>[5]</sup>。

## 2.1 web 日志数据抽取

web 日志分析过程中, 并不是每个字段都是日志分析中所需要的。比如, DNS 的 IP 在每行数据中都是相同的, 所以没有必要分析; 日志记录中也记录了两用户请求的主机地址, 实际分析中只需要一个即可; 还有一些无关的符号 (client、query、IN、+ 客户端端口号等)<sup>[6]</sup> 也都需要过滤。最终, 在数据表中保存了所需要的有用字段。

## 2.2 web 日志数据格式转换

web 日志数据格式的转换主要针对的是日期字段。在这里, 将字符型的日期格式 04. Jan. 2017 转换成了日期类型的数据格式 2017-01-04<sup>[7]</sup>。

## 2.3 web 日志数据重复值的处理

在海量的 web 日志数据行中, 有大部分数据行是重复的, 数据量很大, 会对数据分析的准确性带来一定的影响, 因此有必要对重复的数据进行去重处理<sup>[8]</sup>。

# 3 基于专家系统的 web 日志数据预处理

## 3.1 数据来源

本次实验的数据来源于某企业 Apache 服务器 2013 年的员工 web 日志数据。

## 3.2 专家系统设计

专家系统主要逻辑:

(1) 获取 web 日志数据, 将获取 web 日志数据某一行的前三列转换为已知事实。

(2) 输入目前已经存在的事实—>进行推理。

(3) 综合数据库 db 添加输入的已知事实。

(4) 规则库获取, 并在之后将存在对应关系的前提与结论分别各自存储于两个列表里<sup>[9]</sup>。

(5) 前提与已知知识库进行匹配:

如果有一个前提, 所有这些在已知的事实中出现, 那么最少也能够得出一个结论<sup>[10]</sup>。将结论添加到综合数据库中, 并对推理过程进行标记。在推理列表之中存在的数字是前提和推理出来结论的下标<sup>[11]</sup>。用于展示。

如果这样的前提并不存在, 就无法推理出想要的中间结论, 跳转(7)。

(6) 等循环完了, 因为至少存在一个中间结论, 所以直接输出推理过程和(中间或者最终)结论<sup>[12]</sup>。

(7) 提示什么也不能推出来, 询问是否进行补充, 如果选择“是”就回到主页面, 如果选择“否”就关闭程序。

界面设置:

框: 读取并显示 web 日志数据的框, 输入事实的框, 显示推理过程的框, 显示结论的框, 自动显示当前规则库的框, 用来添加规则库的框。

按钮: 打开 web 日志数据的按钮, 点击进行推理的按钮, 点击添加规则库并更新当前窗口的按钮。

对话提示框: 询问是否进行补充的框。

主界面如图 1 所示。

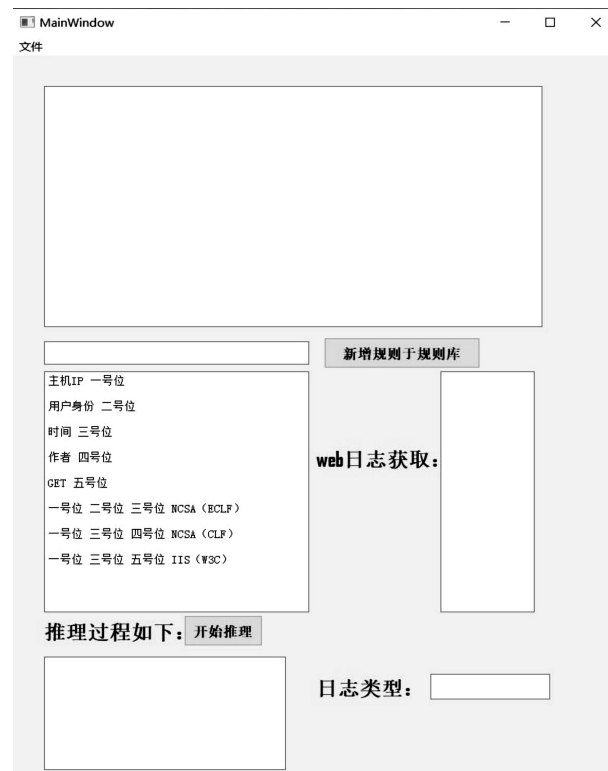


图1 主界面

点击文件, 打开需要处理的 web 日志数据, 如图 2 所示。





图 2 打开 web 数据

显示读取的 web 日志数据,将获取的 web 日志数据的第一行前三列,转换为已知事实,并显示于“获取 web 日志”,推理输出对应的日志格式,如图 3 所示。

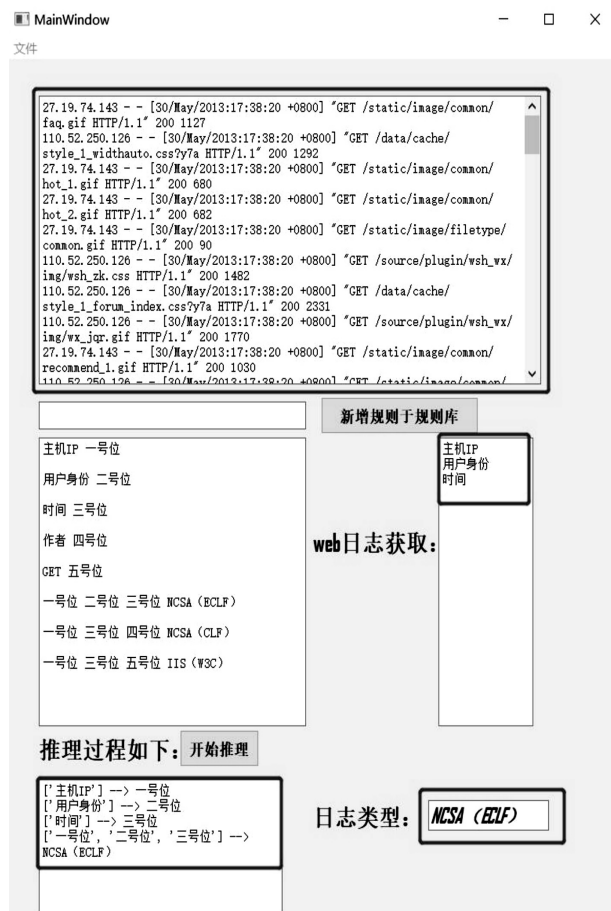


图 3 推理过程及结果显示

### 3.3 web 日志数据预处理

通过“2”得到输入的 web 数据格式是 NCSA

(ECLF)<sup>[13]</sup>,针对其特定的数据格式自动进行数据预处理,NCSA(ECLF)日志格式如下所示:

```
27.19.74.143 -- [30/May/2013:17:38:20 +0800] "GET /static/image/common/faq.gif HTTP/1.1" 200 1127
110.52.250.126 -- [30/May/2013:17:38:20 +0800] "GET /data/cache/style_1_widthauto.css?y7a HTTP/1.1" 200 1292
27.19.74.143 -- [30/May/2013:17:38:20 +0800] "GET /static/image/common/hot_1.gif HTTP/1.1" 200 680
27.19.74.143 -- [30/May/2013:17:38:20 +0800] "GET /static/image/common/hot_2.gif HTTP/1.1" 200 682
27.19.74.143 -- [30/May/2013:17:38:20 +0800] "GET /static/image/common/logo.png HTTP/1.1" 200 4542
110.52.250.126 -- [30/May/2013:17:38:20 +0800] "GET /source/plugin/wsh_wx/img/wsh_zk.css HTTP/1.1" 200 1482
110.52.250.126 -- [30/May/2013:17:38:20 +0800] "GET /data/cache/style_1_forum_index.css?y7a HTTP/1.1" 200 2331
27.19.74.143 -- [30/May/2013:17:38:20 +0800] "GET /source/plugin/wsh_wx/img/wx_jqr.gif HTTP/1.1" 200 1770
27.19.74.143 -- [30/May/2013:17:38:20 +0800] "GET /static/image/common/recommend_1.gif HTTP/1.1" 200 1030
110.52.250.126 -- [30/May/2013:17:38:20 +0800] "GET /static/image/common/logo.png HTTP/1.1" 200 4542
27.19.74.143 -- [30/May/2013:17:38:20 +0800] "GET /data/attachment/common/c8/common_2_verify_icon.png HTTP/1.1" 200 582
110.52.250.126 -- [30/May/2013:17:38:20 +0800] "GET /static/js/logging.js?y7a HTTP/1.1" 200 603
8.35.201.144 -- [30/May/2013:17:38:20 +0800] "GET /uc_server/avatar.php?uid=29331&size=middle HTTP/1.1" 301 -
27.19.74.143 -- [30/May/2013:17:38:20 +0800] "GET /data/cache/common_smilies_var.js?y7a HTTP/1.1" 200 3184
27.19.74.143 -- [30/May/2013:17:38:20 +0800] "GET /static/image/common/pn.png HTTP/1.1" 200 582
27.19.74.143 -- [30/May/2013:17:38:20 +0800] "GET /static/image/common/swfupload.swf?preventsrfwaching=1369906718144 HTTP/1.1" 200 13333
```

#### 3.3.1 数据抽取

企业员工的 web 日志中时区、方法、协议、状态码等字段对于分析没有作用。企业员工都在同一时区,所以时区这一字段对于企业分析员工行为没有意义,可以舍去;在企业分析员工 web 日志过程中,请求方法、协议、状态码等字段没有实际应用价值,亦可以舍去。因此将这些无意义字段都删除,只保留主机 IP、请求时间、资源、发送字节量等字段<sup>[14]</sup>。

```
0,27.19.74.143,30/May/2013:17:38:20,/static/image/common/faq.gif,1127
1,110.52.250.126,30/May/2013:17:38:20,/data/cache/style_1_widthauto.css?y7a,1292
2,27.19.74.143,30/May/2013:17:38:20,/static/image/common/hot_1.gif,680
3,27.19.74.143,30/May/2013:17:38:20,/static/image/common/hot_2.gif,682
4,27.19.74.143,30/May/2013:17:38:20,/static/image/common/logo.png,4542
5,110.52.250.126,30/May/2013:17:38:20,/source/plugin/wsh_wx/img/wsh_zk.css,1482
6,110.52.250.126,30/May/2013:17:38:20,/data/cache/style_1_forum_index.css?y7a,2331
7,110.52.250.126,30/May/2013:17:38:20,/source/plugin/wsh_wx/img/wx_jqr.gif,1770
8,27.19.74.143,30/May/2013:17:38:20,/static/image/common/recommend_1.gif,1030
9,110.52.250.126,30/May/2013:17:38:20,/static/image/common/logo.png,4542
10,27.19.74.143,30/May/2013:17:38:20,/data/attachment/common/c8/common_2_verify_icon.png,582
11,110.52.250.126,30/May/2013:17:38:20,/static/js/logging.js?y7a,603
12,8.35.201.144,30/May/2013:17:38:20,/uc_server/avatar.php?uid=29331&size=middle,-
13,27.19.74.143,30/May/2013:17:38:20,/data/cache/common_smilies_var.js?y7a,3184
14,27.19.74.143,30/May/2013:17:38:20,/static/image/common/pn.png,582
15,27.19.74.143,30/May/2013:17:38:20,/static/image/common/swfupload.swf?preventsrfwaching=1369906718144,13333
16,27.19.74.143,30/May/2013:17:38:20,/static/image/editor/editor.gif,13648
17,8.35.201.165,30/May/2013:17:38:21,/uc_server/data/avatar/000/05/94/42_avatar_middle.jpg,6153
18,8.35.201.164,30/May/2013:17:38:21,/uc_server/data/avatar/000/03/13/51_avatar_middle.jpg,5087
19,8.35.201.163,30/May/2013:17:38:21,/uc_server/data/avatar/000/04/87/84_avatar_middle.jpg,5117
20,8.35.201.165,30/May/2013:17:38:21,/uc_server/data/avatar/000/01/01/03_avatar_middle.jpg,5844
21,8.35.201.160,30/May/2013:17:38:21,/uc_server/data/avatar/000/04/12/85_avatar_middle.jpg,3174
22,8.35.201.164,30/May/2013:17:38:21,/uc_server/avatar.php?uid=53635&size=middle,-
23,8.35.201.163,30/May/2013:17:38:21,/static/image/common/anw_r.gif,65
24,8.35.201.166,30/May/2013:17:38:21,/static/image/common/pn.png,210
```

#### 3.3.2 数据格式转换

由于数据格式的转换主要针对的是日期字段,因此需要将字符型的日期格式转换成日期类型的数据格式。

```
0,27.19.74.143,2013-05-30 17:38:20,/static/image/common/faq.gif,1127
1,110.52.250.126,2013-05-30 17:38:20,/data/cache/style_1_widthauto.css?y7a,1292
2,27.19.74.143,2013-05-30 17:38:20,/static/image/common/hot_1.gif,680
3,27.19.74.143,2013-05-30 17:38:20,/static/image/common/hot_2.gif,682
4,27.19.74.143,2013-05-30 17:38:20,/static/image/common/logo.png,4542
5,110.52.250.126,2013-05-30 17:38:20,/source/plugin/wsh_wx/img/wsh_zk.css,1482
6,110.52.250.126,2013-05-30 17:38:20,/data/cache/style_1_forum_index.css?y7a,2331
7,110.52.250.126,2013-05-30 17:38:20,/source/plugin/wsh_wx/img/wx_jqr.gif,1770
8,27.19.74.143,2013-05-30 17:38:20,/static/image/common/recommend_1.gif,1030
9,110.52.250.126,2013-05-30 17:38:20,/static/image/common/logo.png,4542
10,27.19.74.143,2013-05-30 17:38:20,/data/attachment/common/c8/common_2_verify_icon.png,582
11,110.52.250.126,2013-05-30 17:38:20,/static/js/logging.js?y7a,603
12,8.35.201.144,2013-05-30 17:38:20,/uc_server/avatar.php?uid=29331&size=middle,-
13,27.19.74.143,2013-05-30 17:38:20,/data/cache/common_smilies_var.js?y7a,3184
14,27.19.74.143,2013-05-30 17:38:20,/static/image/common/pn.png,582
15,27.19.74.143,2013-05-30 17:38:20,/static/image/common/swfupload.swf?preventsrfwaching=1369906718144,13333
16,27.19.74.143,2013-05-30 17:38:20,/static/image/editor/editor.gif,13648
17,8.35.201.165,2013-05-30 17:38:21,/uc_server/data/avatar/000/05/94/42_avatar_middle.jpg,6153
18,8.35.201.164,2013-05-30 17:38:21,/uc_server/data/avatar/000/03/13/51_avatar_middle.jpg,5087
19,8.35.201.163,2013-05-30 17:38:21,/uc_server/data/avatar/000/04/87/84_avatar_middle.jpg,5117
20,8.35.201.165,2013-05-30 17:38:21,/uc_server/data/avatar/000/01/01/03_avatar_middle.jpg,5844
21,8.35.201.160,2013-05-30 17:38:21,/uc_server/data/avatar/000/04/12/85_avatar_middle.jpg,3174
22,8.35.201.164,2013-05-30 17:38:21,/uc_server/avatar.php?uid=53635&size=middle,-
23,8.35.201.163,2013-05-30 17:38:21,/static/image/common/anw_r.gif,65
24,8.35.201.166,2013-05-30 17:38:21,/static/image/common/pn.png,210
```

#### 3.3.3 重复字段删除

由于数据量庞大,因此以 10 秒钟为间隔定义员工日志。在这里,删除的对象针对的是同一 ip,访问时间

相差大于 10 秒的数据行。将员工日志数据重新定义完成后,删除其中的重复数据行。在一定程度上能提高之后数据分析的准确性<sup>[15]</sup>。

经过处理完的 ECLF web 日志数据如下所示:

```
0|27.19.74.143|2013-05-30 17:38:20,/static/image/common/faq.gif,1127
1,110.52.250.126,2013-05-30 17:38:20,/data/cache/style_1_widthauto.css?y7a,1292
2,8.35.201.165,2013-05-30 17:38:21,/uc_server/data/avatar/000/05/94/42_avatar_middle.jpg,6153
3,183.62.140.242,2013-05-30 17:38:22,/data/cache/style_1_widthauto.css?y7a,1292
4,211.97.15.179,30/May/2013:17:38:22,/static/image/common/background.png,1166
5,110.178.217.18,30/May/2013:17:38:21,/home.php?mod=space&uid=17496&do=profile,10159
6,157.56.93.187,30/May/2013:17:38:21,/thread-6921-1-1.html,32743
7,222.219.184.66,30/May/2013:17:38:24,/static/image/common/locked.gif,319
8,61.160.45.26,30/May/2013:17:38:24,/static/image/common/nv_a.png,2076
9,58.247.191.45,30/May/2013:17:38:23,/api.php?mod=js&bid=94,275
10,101.226.33.227,30/May/2013:17:38:25,/static/js/logging.js?y7a,603
11,182.48.112.82,30/May/2013:17:38:27,/thread-52573-1-1.html,144080
12|27.19.74.143|30/May/2013:17:38:31,/static/js/forum_moderate.js?y7a,3867
```

## 4 结束语

传统的日志数据清洗方式难以应对目前如此复杂的日志格式。由于专家系统其技术特点能应对复杂的日志格式,所以通过结合专家系统,推理出对应的 web 日志格式,从而自动进行日志数据清洗。该文通过对若干种日志数据格式建立规则库,并对其进行知识推理,基于推理得到对应日志类型,自动进行数据预处理。证明了基于专家系统的 web 日志数据预处理的可行性,使得当前的日志数据处理更加高效且智能,有效提高了企业的工作效率,节约了时间成本,推动了企业快速发展。

### 参考文献:

- [1] 毕学工,杭迎秋,李昕,等.专家系统综述[J].软件导刊,2008,7(12):7-9.
- [2] 张煜东,吴乐南,王水花.专家系统发展综述[J].计算机工程与应用,2010,46(19):43-47.
- [3] 王瑞通.基于框架知识结构的专家系统设计与实现[D].南京:南京邮电大学,2017.
- [4] 伍欣,刘自伟.提高基于规则专家系统效率的技术和方法[J].微计算机信息,2006,22(11):270-272.
- [5] 崔金国.基于蚁群算法的主题爬虫技术研究与实现[D].成都:成都理工大学,2010.
- [6] ZHU Lihua. Implementation of web log mining device under apriori algorithm improvement and confidence formula optimization[J]. International Journal of Information Technology and Web Engineering, 2020, 15(4):52-71.
- [7] 聂虎.基于 Elastic Stack 的 Web 日志分析系统设计与实现[J].中国新通信,2020,22(19):68-69.
- [8] HE Mingshu, JIN Lei, WANG Xiaojuan, et al. Web log classification framework with data augmentation based on GANs[J]. The Journal of China Universities of Posts and Telecommunications, 2020, 27(5):34-46.
- [9] 于琦. Web 日志挖掘中的数据预处理研究[J].河南科技,2018(19):18-20.
- [10] 胡晴,吕世超,石志强,等.基于专家系统的高级持续性威胁云端检测博弈[J].计算机研究与发展,2017,54(10):2344-2355.
- [11] 侯天祝.本体推理中组合推理机制的研究[D].成都:西南交通大学,2016.
- [12] 杨磊,赖志杰,陈皓,等.基于 Web 及知识推理的宠物狗疾病诊断专家系统[J].广东技术师范学院学报,2016,37(5):43-49.
- [13] BITAN Y, ABRAMOVICH I, JAFFE E, et al. Designing a first responders call center for a pandemic[J]. Ergonomics in Design: The Quarterly of Human Factors Applications, 2021, 29(1):25-28.
- [14] BUNIK M, GLAZNER J E, CHANDRAMOULI V, et al. Pediatric telephone call centers: how do they affect health care use and costs? [J]. Pediatrics, 2007, 119(2):e305-13.
- [15] PRAVEENA A D, SELVI V. The role of data preprocessing system on web log files for mining students access logs[J]. International Journal of Recent Technology and Engineering, 2020, 9(1):1045-1050.