

智能分类算法在游戏故障告警中的应用^①

王月瑶^{1,2}, 胡琴敏¹, 刘伟², 陈乃华², 程洁³

¹(华东师范大学 计算机科学与软件工程学院, 上海 200062)

²(腾讯公司 腾讯公司 IEG, 上海 200233)

³(上海智臻智能网络科技股份有限公司, 上海 201803)

通讯作者: 胡琴敏, E-mail: qmhu@cs.ecnu.edu.cn

摘要: 在游戏运维领域, 为了充分挖掘游戏用户在反馈故障时的意图, 运用自然语言处理的手段对记录故障事件的文本进行分类, 可以减少告警的误判发布更具价值的告警. 本文提出一种面向突发告警事件的智能分类算法, 首先分析用户反馈文本数据的各项特征, 然后对数据利用半自动构建的语义映射工具进行概率计算, 从而形成有效的告警类别. 将此方法应用到企业级业务故障定位中, 能更有针对性地发布告警, 收敛无关告警, 支撑运维人员更快速准确地定位到故障原因. 实验结果表明, 基于语义映射的智能分类算法能够有效地提高告警的针对性, 比简单的文本分类算法具有更好的结果.

关键词: 游戏运维; 语义映射; 智能分类算法; 自然语言处理

引用格式: 王月瑶, 胡琴敏, 刘伟, 陈乃华, 程洁. 智能分类算法在游戏故障告警中的应用. 计算机系统应用, 2018, 27(7): 133-138. <http://www.c-s-a.org.cn/1003-3254/6417.html>

Application of Intelligent Classification Algorithm in Game Alarm

WANG Yue-Yao^{1,2}, HU Qin-Min¹, LIU Wei², CHEN Nai-Hua², CHENG Jie³

¹(School of Computer Science and Software Engineering, East China Normal University, Shanghai 200062, China)

²(Tencent IEG, Tencent, Shanghai 200233, China)

³(Xiao Robot, Shanghai 201803, China)

Abstract: Natural language processing techniques have been confirmed to well classify the alarm messages based on the user intents. In this study, we propose an intelligent classification algorithm for burst alarm events. First, we analyze the features of user feedback text data. We then present a semi-automatic semantic mapping tool to calculate the probability, so as to form a valid alarm category. By application this method into the enterprise fault location, it takes advantages of (1) releasing alarms upon different user intents; (2) reducing the alarm miscalculation; and (3) helping game operation personnel accurately and quickly locating the reasons of the fault. The experimental results show that the intelligent classification algorithm based on semantic mapping can effectively improve the performance of the alarm distribution, and achieves better results than the simply algorithm for text categorization.

Key words: game operation maintenance; semantic mapping; intelligent classification; natural language processing

随着游戏产业的蓬勃发展, 各类游戏层出不穷, 游戏用户的选择日渐增多, 稳定的运营与良好的服务渐渐成了游戏能否胜出的重要因素. 而随着人工智能风

潮的掀起, 数据挖掘技术在游戏行业中的应用获得了广泛的关注^[1]. 越来越多的企业已经意识到了这一点, 开始重视起智能化运维服务^[2], 从数据中获取有用的信

① 基金项目: 上海市科学技术委员会资助 (16511102702)

Foundation item: Funded by Science and Technology Commission of Shanghai (16511102702)

收稿时间: 2017-11-02; 修改时间: 2017-11-29; 采用时间: 2017-12-01; csa 在线出版时间: 2018-06-27

息,辅助用户与维护人员更好地解决问题。

在企业级游戏运维中,客服人员每天会对游戏用户的反馈信息进行记录.由于游戏发生故障时会产生大量的用户反馈,因此工作人员可以从大量的用户反馈中提炼有效信息,人工确认后形成游戏故障单,交由对应的运维人员处理.每个故障单记录了一个突发的游戏故障告警事件,在传统的告警处理方法中,一旦发现故障告警事件就会将所有告警信息发送给运维人员.然而,由于游戏故障原因的多样性,告警数据库信息量巨大^[3],故障发生的同一时间会产生各类告警.这样的处理方式实际上忽略了用户的反馈,损失了宝贵的信息资源,同时存在很多告警误判误报的问题.不相关的告警信息会干扰运维人员做原因诊断,因而告警信息需要进行分析。

在对告警信息的分析中,过去的研究者已经取得一定的成果.但他们的工作主要聚焦于利用关联规则收敛电网的数据^[4],或根据网络异常流量曲线进行网络告警的分类^[5].当然,针对用户的反馈信息,也有研究者根据体验价值、交易成本和服务质量三个维度建立用户服务模型^[6],或从中找出用户忠诚度与价格关系^[7].然而,这些方法都不关注文本信息的分类,不适用于分析故障的文本信息,而从用户反馈中提炼出的文本信息是十分重要的。

因此,我们提出一种面向运维突发故障告警事件的智能分类算法.首先,我们分析故障事件文本的各项特征;然后,对文本信息利用半自动构建的语义映射工具进行概率计算;最终将记录突发故障事件的文本自动分为三类:网络级突发故障、程序主机类突发故障、平台支付类突发故障.根据判断的类别向运维人员发布该类别相关的告警,排除大类上的误判,减少无关告警信息的轰炸。

我们的贡献在于将情感分类的算法创新性地利用在游戏运维的文本分类上,并提出了一种语义映射工具供后续企业级的故障文本分类使用。

我们在企业近三年的故障单上做了实验,实验表明:本文提出的智能分类算法,对比简单的基准方法决策树算法和朴素贝叶斯算法,准确率分别有13%和9%的提升.将此智能分类方法应用到企业级业务故障诊断中,能实现更有针对性地发布告警,减少告警的误判,支撑运维人员更快速准确地定位故障原因。

1 面向故障告警事件的智能分类模型

本文提出了一种面向故障告警事件的智能分类模型,去分析描述突发告警事件的文本信息,从中判断出故障事件的大类,收敛无关告警,减少告警的误判,从而使发布告警更具有针对性.整个模型的结构如图1所示,分为以下三个模块,分别是数据预处理模块,构建映射工具模块和类情感分类模块.第一层数据预处理模块主要分为数据清洗和jieba分词.第二层构建语义映射工具模块指的是从已标注类别的故障告警事件文本中,抽取三个类别对应的关键词,建立每个类别下的运维词典,经过人工确认后半自动构建语义映射工具.第三层类情感分类模块主要是利用所构建的语义映射工具进行类别分析计算,选择概率最高的类别作为最终的告警类别,完成分类过程。

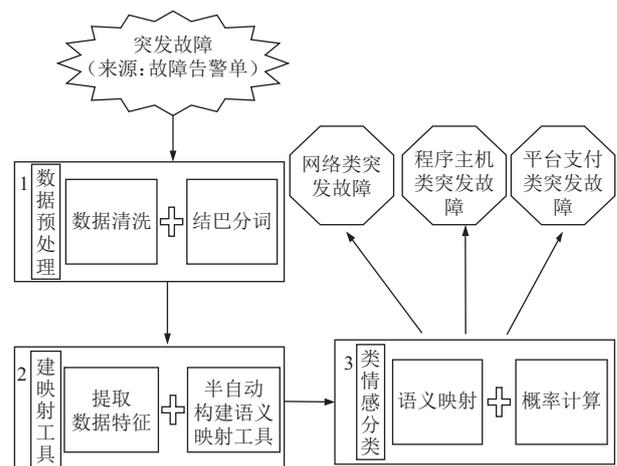


图1 告警智能分类模型

描述突发故障事件的文本经过这个智能分类模型,便可自动分类到以下三类故障中:网络级突发故障、程序主机类突发故障、平台支付类突发故障.分类后能收敛无关告警,以实现告警信息的针对性发布。

1.1 故障告警事件文本的预处理

描述游戏故障告警事件的文本是从用户反馈中提取的信息,其中包括一些游戏用户的个人信息和相关问题描述.这些信息包含很多与故障告警无关的内容,因此需要对数据进行清洗,并进行jieba分词^[8,9],从而最终形成描述该故障事件的词典集合。

假设 $sentence(j)$ 为清洗后的故障事件 j 的所有描述语句, $word(i)$ 为 $sentence(j)$ 中的第 i 个描述词,数据预处理的方法如下:

$$\begin{aligned} \text{sentence}(j) &= \text{clean}(\text{data}) \\ \text{word}(i) &= \text{jieba}(\text{sentence}(j)) \end{aligned} \quad (1)$$

公式(1)中的 clean 指的是数据清洗过程, jieba 指的是数据分词过程, 通过公式(1), 将所有的故障事件文本数据进行转换, 获得一个分词后的词典集合, 进而提供给下一步构建语义映射工具使用。

1.2 构建语义映射工具

本文构建的语义映射工具是基于已标注类别的故障文本数据, 采用了 Term Frequency-Inverse Document Frequency (TF-IDF) 算法^[10]提取文本中相应类别的词典, 形成智能分类算法的映射工具. TF-IDF 由 Salton^[10]在 90 年代提出, 并且在文本检索中不断发挥着重要作用, TF-IDF 的主要思想是指一个词在特定文档中出现的频率越高 (TF), 并且在文档中出现的范围越广 (IDF), 这个词就越关键权重越高. 其经典计算公式如公式(2)所示:

$$w_{ij} = tf_{ij} \times idf_j = tf_{ij} \times \log(N/n_j) \quad (2)$$

在公式(2)中, t_j 是特征项, tf_{ij} 指的是 t_j 在文档 d_i 中出现的次数; idf_j 表示逆文档频率, N 表示总文档数, n_j 表示 t_j 的文档数目. 在本文中利用该算法思想来提取每个类别下每条故障描述中的关键词。

半自动构建语义映射工具的整个过程如图 2 所示, 主要分为两步, 首先针对已存在并已让专家标注类别的故障事件文本, 利用 TF-IDF 算法提取每个故障事件描述文本中前三名的词, 构建相应类别的词典表述, 确保词典的专业性. 第二步是建立每个类别下的特有词典, 因为在故障表述中, 不同类别下的故障描述内容可能相同, 不同的词典可能包含相同词, 所以需要对比相同词语在不同词典下的 TF-IDF 值, 将这个词划分到其 TF-IDF 值较高的类别词典中. 同时采用人工确认的方法, 确保词典的独特性与正确性。

例如网络类的故障描述会涉及到“客户端”这个词, 而程序主机类的故障也会频繁地提及“客户端”这个词. 因此“客户端”这个词在两个类别下的 TF-IDF 值都很高, 被同时选入网络类的词典和程序主机类的词典, 这时通过比较得到“客户端”在网络类的 TF-IDF 总值, 和在程序主机类的 TF-IDF 总值, 由于“客户端”这个词在程序主机类的 TF-IDF 总值较高, 这个词被归为程序主机类词典, 依次类推, 再通过人工确认最终确定每个类别下的语义映射工具。

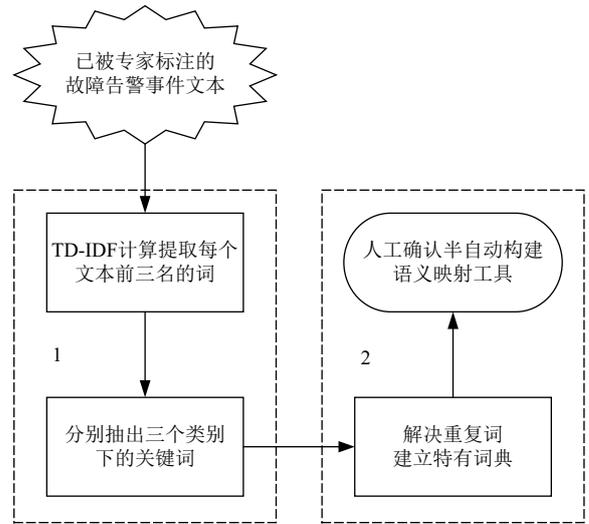


图 2 半自动构建语义映射工具

通过以上过程, 我们半自动构建了针对网络类、程序主机类和平台充值类这三个类别的语义映射工具, 提供给下一过程的分类使用. 该语义映射工具已可以直接使用, 但后续故障描述如果有更新, 可以通过如图所示的方法将语义映射工具进行扩展。

1.3 类情感分类

情感分类中一种常用算法^[11]是利用构建好的情感词典进行情感分类, 分为正向负向和中性的情感. 借鉴这样的思路, 本文在最后一步利用已经构建好的语义映射工具进行类情感分类, 分类过程如公式(3):

$$P_k = \frac{\text{count}(\text{word}_i \in \text{Class}_k)}{\text{length}(\text{sentence}_j)} \quad (k = 1, 2, 3)$$

$$\text{Label} = \begin{cases} 1 & P_1 > (P_2 \& P_3) \\ 2 & P_2 > (P_1 \& P_3) \text{ or 存在相同概率} \\ 3 & P_3 > (P_1 \& P_2) \end{cases} \quad (3)$$

应用以上公式进行类情感分类, P_k 为每个类别下的概率, k 对应各个类别的顺序 (网络类顺序为 1, 程序主机类顺序为 2, 平台充值类顺序为 3, Class_k 为半自动构建语义映射工具中词典的类别, 此处标记网络类词典为 Class_1 , 程序主机类词典为 Class_2 , 平台充值类词典为 Class_3 , Label 指最终分类形成的类别。

首先, 将每个故障事件中文本 sentence_j 分词后的词典 word_i 通过语义映射工具映射, 计算该故障事件文本属于每个类别的概率. 每个类别下的概率 P_k 是通过映射到每个类别 Class_k 的个数, 除以该故障事件 sentence_j 的长度。

然后,比较每个类别下的概率 P_k 大小,如果是网络类 P_1 的概率最高,则判断为网络类, *Label* 为 1. 如果程序主机类概率 P_2 最高,则为程序主机类, *Label* 为 2. 如果平台充值类的概率 P_3 最高,则为平台充值类, *Label* 为 3. 如果存在相同的概率,这样的情况占极少数,允许这一小部分的容错,考虑到程序主机类的故障最多,因而相同概率被标记程序主机类, *Label* 为 2, 由此完成分类的这个过程.

2 智能分类算法的应用

在企业级的应用中,面向游戏故障告警事件的智能分类算法可以直接应用在游戏运维平台上,排除大类的故障,发布与类别相关的告警,收敛无关告警.

在应用场景中,接入该智能分类算法作为一层滤网,运维平台在建立新的故障单文本后,可以直接利用已构建的语义映射工具,自动判断平台上新形成的故障单文本类别,根据类别自动去寻找类相关的告警,收敛无关告警,将相关告警推送给运维人员,整体过程自动化方便快捷,不需要人工干预,实现案例见图 3.

[影响业务]: XX业务
 [影响情况]: XX大区掉线掉包网络波动有问题是否是服务器问题
 [相关告警]:
 ✓发送: XX时间XX地区机房网络波动
 X不发送: XX时间服务器进程被关闭(这个原因通过智能分类后,属于网络类,不发送程序类告警)

图 3 智能分类算法的应用案例

在后续的使用中,由于语义映射工具是可更改扩展的,因而可以根据实际需求,三个月或半年更新一次语义映射工具,保证及时性和有效性.

3 实验设置与分析

我们在企业近三年的历史故障事件单上做实验,历史故障单由专家标记类别,实验数据集为去除专家标记为自定义类别无法区分的单据,共有 2456 条故障事件文本. 在该数据集上,我们应用了本文提出的智能分类算法,同时与决策树分类算法^[12]与朴素贝叶斯分类算法^[13]做对比. 实验结果采用了准确率、召回率、*F1* 值这三个评价指标.

3.1 实验前期准备

整个数据集为企业近三年的历史故障事件单,是

企业的私有数据积累,历史故障事件单由客服反馈收集,数据集均以文本形式呈现,包含故障单号、故障来源、故障标题、故障描述、以及发生时间这几个字段. 在本文实验中,我们抽取故障描述与内容作为用户反馈的文本信息进行分类.

实验前期准备为对故障文本数据的预处理,以及专家需要对故障单人工标记. 数据预处理过程为对数据清洗后,利用 Python 的 *jieba* 包^[8]进行分词,并在分词中加了自定义的词典,修改了一些错误的分词. 实验的评价指标为准确率 (*Precision Rate*)^[14], 召回率 (*Recall Rate*)^[15], *F1* 值^[15], 相关的公式如 (4) 所示:

$$\begin{aligned} Precision &= \frac{tp}{tp+fp} \\ Recall &= \frac{tp}{tp+fn} \\ F_1 &= 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \end{aligned} \quad (4)$$

公式中, *tp* 是将正类预测为正类数, *fn* 是将正类预测为负类数, *fp* 是将负类预测为正类数.

3.2 实验结果对比

本文在各项评价指标下,测试了智能分类算法,以及对比了简单的机器学习算法: 决策树算法与朴素贝叶斯算法.

在文本分类中,通常采用的做法是对文本进行向量化后,利用监督学习的算法来分类. 监督学习算法中常用的是决策树算法与朴素贝叶斯算法,文献^[16]介绍了利用决策树进行文本分类的多种方法,文献^[17]阐述了朴素贝叶斯在文本分类中的应用与改进. 这两种方法在文本分类中均被证明了有效性. 因此本文的对比实验首先采用 *TFIDF* 文档特征权重表示^[18]对文本进行向量化,然后采用决策树算法与朴素贝叶斯算法. 由于两者均是监督学习,需要将数据集进行划分,本文采用四六分. 随机抽取相同 60% 的数据作为训练集,抽取相同 40% 的数据作为测试集.

由于智能分类算法为无监督算法,因此需要在抽取的相同测试集上验证各项指标,所有的实验在同一数据集上利用相同评价指标进行. 智能分类与决策树文本分类、朴素贝叶斯文本分类的实验结果如表 1 所示,我们将在下一小节对实验结果做出进一步的分析.

表1 智能分类与决策树、朴素贝叶斯分类算法的比较

	类别	准确率	召回率	F1 值	实验数目
智能分类	网络类	0.76	0.53	0.62	170
	主机程序类	0.72	0.70	0.71	536
	平台充值类	0.56	0.69	0.62	277
	avg/total	0.68	0.69	0.67	983
决策树分类	网络类	0.44	0.29	0.35	170
	主机程序类	0.60	0.77	0.68	536
	平台充值类	0.52	0.35	0.42	277
	avg/total	0.55	0.57	0.55	983
朴素贝叶斯分类	网络类	0.71	0.10	0.18	170
	主机程序类	0.57	0.92	0.71	536
	平台充值类	0.55	0.19	0.28	277
	avg/total	0.59	0.57	0.49	983

3.3 实验结果分析

如表1所示,对比的三个算法在网络类、主机程序类、平台充值类、平均/总体 (avg/total) 均有对应评价指标的结果,表格每一列分别对应类别、准确率、召回率、F1 值以及所支持的实验数目 (Support)。Support 是实验支持的数目,相同的测试集共计 983 条故障文本:包括网络类文本 170 条,主机程序类文本为 536 条,平台充值类文本 277 条。各项评价指标的结果均在表1中可见。

在较为常用的评价指标 avg/total 准确率上,本文提出的智能分类算法比决策树文本分类要高出 13%, 而比朴素贝叶斯高出 9%。

整体的 avg/total 的实验结果对比如图4所示,可以看出目前智能分类算法在故障事件文本中的应用是简单而有效的,智能分类在准确率、召回率、以及 F1 值上,均比决策树分类与朴素贝叶斯分类要高。

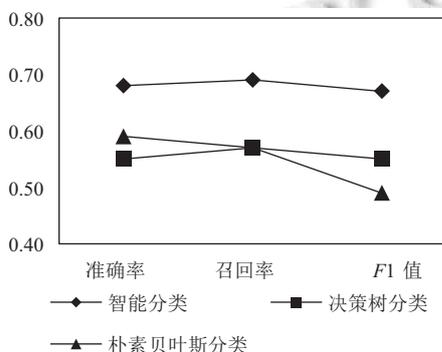


图4 实验结果对比图

同时,本文提出的智能分类算法基于半自动构建的语义映射工具,是无监督的学习算法。我们可以将智

能分类算法应用在全部的数据集上,测试总体的实验效果。此时数据集支持的数目为数据集全部数目 2456 条文本。为了进一步验证智能分类算法的有效性,在这里我们也加入了无监督学习算法的实验对比,经常被采用的无监督学习算法主要是 K-means 分类^[19]算法和 DBSCAN 分类算法^[20],由于 K-means 分类算法需要人工指定类别数 N,若直接将 N 设置为 3 加入了太多人工干扰的成分,因此在这里无监督的对比实验选用 DBSCAN 分类算法, DBSCAN 的参数 eps 选用比较常用的 0.5。

在整个实验集上,采用相同的评价指标,智能分类得出的总体结果如表2所示。

表2 智能分类在整体数据集的结果

	类别	准确率	召回率	F1 值	实验数目
智能分类	网络类	0.64	0.53	0.58	452
	主机程序类	0.71	0.67	0.59	1312
	平台充值类	0.56	0.68	0.62	692
	avg/total	0.66	0.65	0.65	2456
DBSCAN 分类	网络类	0.35	0.10	0.15	452
	主机程序类	0.53	0.70	0.60	1312
	平台充值类	0.34	0.30	0.32	692
	avg/total	0.44	0.47	0.44	2456

从上表可以看出,智能分类算法在整个数据集上的 avg/total 准确率为 66%,比无监督的分类算法 DBSCAN 算法准确率高 12%,进一步证明了智能分类算法的有效性,并且,分类错误的主要在平台充值类,由于故障描述来自用户反馈存在语义模糊的问题,这类问题只有运维专家能够发现,因而目前的准确率已基本可以在实际场景中应用,去处理新生成的故障单,减少一部分告警误判。

4 结论与展望

针对故障告警文本,本文提出一种智能分类算法,创新性地将从情感分类的思路应用在了处理故障文本上,能够挖掘用户反馈的意图,减少对告警的误判,发布更具有针对性的告警,辅助运维人员做故障诊断。该算法在企业级的应用中能实现较好的效果,并具有扩展性。实验证明本文提出的智能分类算法,与简单的监督学习算法以及无监督学习算法相比,更适用于故障事件文本中。

但告警智能分类算法不涉及文本语义上的识别,

另外由于游戏故障文本记录并不广泛,因而采用实验的数据集还不够充分.在后续工作中,可以关注文本语义上的问题,使文本分类结果更准确,更靠近人类的思维.同时寻找挖掘更多相关数据集,进一步分析验证智能分类算法的效果.

参考文献

- 1 张彦俊. 游戏运营中的数据挖掘[硕士学位论文]. 上海: 复旦大学, 2011.
- 2 曹逸峰, 陈晓伟. 基于知识分层提取模型的服务台知识库建设. 计算机系统应用, 2015, 24(2): 261–265.
- 3 李金凤, 王怀彬. 基于关联规则的网络故障告警相关性分析. 计算机工程, 2012, 38(5): 44–46.
- 4 石永革, 梅玉洁, 石峰. 通信网网管告警过滤机制的研究与应用. 计算机工程与设计, 2008, 29(9): 2169–2171.
- 5 吴冰, 云晓春, 陈海永. 基于统计的网络流量模型及异常流量发现. 全国网络与信息安全技术研讨会'2005论文集(上册). 北京, 中国. 2005. 1007–1012.
- 6 Yang HE, Wu CC, Wang KC. An empirical analysis of online game service satisfaction and loyalty. *Expert Systems with Applications*, 2009, 36(2): 1816–1825. [doi: 10.1016/j.eswa.2007.12.005]
- 7 Martín-Consuegra D, Molina A, Esteban Á. An integrated model of price, satisfaction and loyalty: An empirical analysis in the service sector. *Journal of Product & Brand Management*, 2007, 16(7): 459–468.
- 8 Sun JY. Jieba 中文分词. <https://github.com/fxsjy/jieba>, 2012.
- 9 陶伟. 警务应用中基于双向最大匹配法的中文分词算法实现. 电子技术与软件工程, 2016, (4): 153–155.
- 10 Salton G, Fox EA, Wu H. Extended Boolean information retrieval. *Communications of the ACM*, 1983, 26(11): 1022–1036. [doi: 10.1145/182.358466]
- 11 王勇, 吕学强, 姬连春, 等. 基于极性词典的中文微博情感分类. *计算机应用与软件*, 2014, 31(1): 34–37, 126.
- 12 栾丽华, 吉根林. 决策树分类技术研究. *计算机工程*, 2004, 30(9): 94–96, 105.
- 13 李静梅, 孙丽华, 张巧荣, 等. 一种文本处理中的朴素贝叶斯分类器. *哈尔滨工程大学学报*, 2003, 24(1): 71–74.
- 14 Olson DL, Delen D. *Advanced data mining techniques*. Berlin Heidelberg: Springer, 2008.
- 15 Powers DMW. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation. *Journal of Machine Learning Technologies*, 2011, 2(1): 37–63.
- 16 田苗苗. 基于决策树的文本分类研究. *吉林师范大学学报(自然科学版)*, 2008, 29(1): 54–56.
- 17 邸鹏, 段利国. 一种新型朴素贝叶斯文本分类算法. *数据采集与处理*, 2014, 29(1): 71–75.
- 18 张玉芳, 彭时名, 吕佳. 基于文本分类 TFIDF 方法的改进与应用. *计算机工程*, 2006, 32(19): 76–78. [doi: 10.3969/j.issn.1000-3428.2006.19.028]
- 19 Xu JH, Liu H. Web user clustering analysis based on KMeans algorithm. *Proceedings of the 2010 International Conference on Information, Networking and Automation (ICINA)*. Kunming, China. 2010, 2. V2-6–V2-9.
- 20 Ester M, Kriegel HP, Sander J, *et al.* A density-based algorithm for discovering clusters in large spatial databases with noise. *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96)*. Portland, OR, USA. 1996.