

基于数据挖掘的非结构化工单信息关键要素自动标识

罗润祥 陈春雪 李 雄

(贵州电网有限责任公司, 贵州 贵阳 550002)

摘 要 不同渠道的非结构化工单信息缺乏统一格式约束, 格式存在异构性, 增加了工单信息关键要素的标识难度, 为此本文提出基于数据挖掘的非结构化工单信息关键要素自动标识方法。对非结构化工单信息进行泛化处理, 通过建立正向最短编辑距离路径并进行聚类, 以消除格式异构性。利用预训练语言模型(BERT)提取泛化工单信息特征, 并将其与工单元数据进行结构化映射, 得到混合特征向量。将该混合特征向量输入数据挖掘算法中的最小二乘支持向量机模型中, 优化模型参数, 实现非结构化工单信息关键要素的自动标识。实验以国家电网 95598 服务热线工单为例, 结果表明, 该方法泛化后领域适应差距远低于阈值, 格式异构性较低。特征提取的斯皮尔曼等级相关系数接近 1, 结构化映射长尾实体覆盖率良好, 对同地点报修工单的服务地点、故障核心要素和特殊影响等关键要素标识精准, 能够有效标识电力运维非结构化信息的关键要素。

关键词 数据挖掘; 非结构化工单; 关键要素; 自动标识; 结构化映射

DOI: 10.27024/j.wlygc.2025.09.03.02

AUTOMATIC IDENTIFICATION OF KEY ELEMENTS IN UNSTRUCTURED CHEMICAL ORDER INFORMATION BASED ON DATA MINING

LUO Runxiang CHEN Chunxue LI Xiong

(Guizhou Power Grid Co., Ltd., Guiyang, Guizhou 550002)

Abstract The unstructured work order information from different channels lacks unified format constraints, and the format is heterogeneous, which increases the difficulty of identifying key elements of work order information. Therefore, this paper proposes an automatic identification method for key elements of unstructured work order information based on data mining. Generalization processing is applied to unstructured single information by establishing a forward shortest editing distance path and clustering to eliminate format heterogeneity. Using a pre trained language model (BERT) to extract generalized single information features, and structurally mapping them with work unit data to obtain a mixed feature vector. Input the mixed feature vector into the least squares support vector machine model in the data mining al-

收稿日期: 2025-09-03; 修回日期: 2025-09-03; 出版日期: 2026-03-30

基金项目: 基于大数据和 AI 技术的 95598 服务风险管控总体技术研究与应用(GZKJXM20232616)。

通信作者: 罗润祥, lrxiaoyou@163.com。

引文格式: 罗润祥, 陈春雪, 李雄. 基于数据挖掘的非结构化工单信息关键要素自动标识[J]. 物理与工程, 2026, 36(2): 179-185.

Cite this article: LUO R X, CHEN C X, LI X. Automatic identification of key elements in unstructured chemical order information based on data mining[J]. Physics and Engineering, 2026, 36(2): 179-185. (in Chinese)

gorithm, optimize the model parameters, and achieve automatic identification of key elements of unstructured single information. The experiment takes the work order of State Grid 95598 service hotline as an example, and the results show that the domain adaptation gap after generalization of this method is much lower than the threshold, and the format heterogeneity is low. The Spearman rank correlation coefficient of feature extraction is close to 1, and the structured mapping long tail entity coverage is good. It accurately identifies key elements such as service location, fault core elements, and special effects of repair orders in the same location, and can effectively identify key elements of unstructured information in power operation and maintenance.

Key words data mining; unstructured chemical order; key elements; automatic identification; structured mapping

在数字化转型加速背景推动下,工单已成为现代企业与客户或者企业内部各个部门之间信息交互的载体,在客服支持、故障申报、运维任务下发等业务场景中广泛应用^[1]。通常工单内容以文本形式呈现,其中包括问题描述、业务诉求以及时间地点等关键信息,但信息记录格式、位置等不同,使工单信息具备非结构化特征。对非结构化工单信息关键要素进行自动标识,是业务流程自动化提升服务效应效率的核心要素^[2,3]。但非结构化工单信息句式多样且表述随意,同时混杂了大量专业术语,因此在人工标识工单信息关键要素的过程中,容易出现漏标识或错误标识现象。

对此,艾青林等人^[4]提出特征融合的非结构化文本语义分割方法。该方法将工单信息输入轻量级卷积模块内,提取工单信息语义特征。将其输入注意力机制模块内,生成小目标增强特征。通过全连接层+Softmax 预测每个位置的标签,得到非结构化工单信息关键要素自动标识结果。但该方法未消除非结构化数据的格式异构性,针对不同类型文本工单的适配性受限。唐瑞雪等人^[5]提出工单文本实体关系联合抽取模型。该方法定义工单关键要素,对非结构化文本进行清洗处理。通过 CNN 提取局部语义片段,构建实体—关系—实体候选三元组。使用关系分类器输出有效的三元组,实现非结构化工单信息关键要素自动标识。但该方法缺乏连续数值离散处理化手段,容易混淆标注结果。李玉聪等人^[6]提出基于 WAI-ARIA 的工单文本属性标识方法。该方法将工单文本按段/分句拆解为逻辑节点,通过依存句法分析识别主谓宾结构。构建主谓宾结构之间

的文本节点间引用关系,依据该引用关系,得到非结构化工单信息关键要素自动标识结果。但该方法仅凭文本节点间引用关系难以处理大规模离散化数据的标注。王景慧等人^[7]提出关系导向型实体关系抽取方法。该方法通过标点符号将工单文本分割为短句集合。使用 BiLSTM-CRF 模型识别实体边界后,输出非结构化工单信息关键要素自动标识结果。但对于新业务线工单,当缺乏标注的依存关系数据时,预训练模型的关系分类准确率不够理想。

数据挖掘是从海量、复杂的数据中自动识别隐藏模式和关联关系知识的过程,其能够打破传统数据处理的标称描述方式^[8],获得数据之间潜在的结构和规律,实现从数据到洞察的转化^[9]。本文基于数据挖掘算法,提出非结构化工单信息关键要素自动标识方法,以提升工单处理效率与准确性,并驱动业务流程自动化。

1 非结构化工单信息关键要素自动标识

1.1 非结构化工单信息泛化处理

在实际业务中,非结构化工单可能通过客服对话、用户邮件、现场记录等多种渠道产生,不同渠道的信息输入方式缺乏统一格式约束;同时,用户或业务人员在描述问题(如故障现象、服务诉求等)时,倾向于使用自然语言自由表述,句式结构、专业术语使用习惯存在显著差异^[10],难以被预定义的数据字段直接捕获。因此在对非结构化工单信息关键要素进行自动标识之前,需要对非结构化工单信息进行泛化处理,以消除非结构化数据的格式异构性,将非结构化工单信息转换为后续

数据挖掘模型可以理解的向量。

按照非结构化工单信息距离,依据待泛化非结构化工单信息模式,建立正向最短编辑距离路径,依此获得非结构化工单信息最佳泛化模式。令 $G(i, j)$ 表示非结构化工单信息编辑距离,其表达式如下

$$G(i, j) = \min\{g(i, j) + G(i-1, j-1), 1 + G(i-1, j), 1 + G(i, j-1)\} \quad (1)$$

其中, i, j 表示 2 个非结构化工单信息; $g(\cdot)$ 表示编辑距离数值。

在式(1)内,当非结构化工单信息 i 和 j 类型相同时, $h(i, j)$ 数值设置为 1, 反之, $g(i, j)$ 数值设置为 0。

在式(1)基础上,对非结构化工单信息进行泛化处理的步骤如下:

Step 1: 使用 1 表示非结构化工单信息 i 和 j 的初始序号,将其取值范围设置为 $[1, n]$ 。

Step 2: 令非结构化工单信息 i 和 j 在取值范围 $[1, n]$ 内按照顺序进行遍历,然后利用式(1)获得非结构化工单信息编辑距离。

Step 3: 依据非结构化工单信息编辑距离,使用聚类方法对非结构化工单信息进行聚类处理,能够获得 n 个类别的非结构化工单信息 z_n 。

经过上述过程,实现非结构化工单信息泛化,将泛化后的非结构化工单信息作为后续工单信息特征提取的基础。

1.2 基于 BERT 的泛化工单信息特征提取与结构化映射

将 1.1 节泛化后的非结构化工单信息 z_n 输入预训练语言模型(BERT)中,提取非结构化供电信息特征。将该特征与工单元数据进行结构化映射,得到非结构化工单信息混合特征向量。

该模型将泛化后的非结构化工单信息转换为 token 序列的隐藏表示,其表达式如下

$$Z = [z_1, z_2, \dots, z_n] \in \mathbb{R}^{n \times d} \quad (2)$$

其中, Z 表示所有 token 的隐藏层表示矩阵; $\mathbb{R}^{n \times d}$ 表示实数空间,其中 n, d 表示 token 数量和空间维度。

在所有 token 的隐藏层表示矩阵 Z 内,提取非结构化工单信息文本级表示,即非结构化工单信息文本向量特征^[11],其由 x_{text} 表示,提取方法分别为 CLS 池化、平均池化和注意力池化,表达

式分别如下

$$\dot{x}_{text} = z_{CLS} \in \mathbb{R}^d \quad (3)$$

$$\ddot{x}_{text} = \frac{\sum_{i=1}^n z_i}{n} \in \mathbb{R}^d \quad (4)$$

$$\ddot{x}_{text} = \sum_{i=1}^n \left[\frac{z_i \exp(w^T z_i)}{\sum_{j=1}^n \exp(w^T z_j)} \right] \quad (5)$$

其中, $\dot{x}_{text}$ 、 $\ddot{x}_{text}$ 、 $\ddot{x}_{text}$ 分别表示 CLS 池化、平均池化和注意力池化后得到的非结构化工单信息文本向量,其共同组成最终非结构化工单信息文本向量,由 $x_{text} = (\dot{x}_{text}, \ddot{x}_{text}, \ddot{x}_{text})$ 表示; w^T 表示可学习的注意力权重向量; z_{CLS} 表示对内 Z 元素进行池化操作。

经过上述方程得到非结构化工单信息文本向量特征后,对非结构化工单信息单元数据进行编码。非结构化工单信息单元数据通常包含数值型、类别型^[12],对于数值型的工单信息单元数据(提交时间、处理时长等) x ,需要将连续数值进行离散处理,使用分桶编码形式对其进行独热编码,表达公式如下

$$x \rightarrow OneHot(Bucket(x)) \quad (6)$$

其中, $OneHot(\cdot)$ 表示独热编码函数; $Bucket(\cdot)$ 表示分桶操作。

对于类别型非结构化工单信息单元数据,则需要使用映射方式将其嵌入低维连续空间内^[13,14],则类别型非结构化工单信息单元数据编码表达公式如下

$$x_c = Embedding(c) \in \mathbb{R}^{d_{embed}} \quad (7)$$

其中, x_c 为第 c 类工单信息单元数据; d_{embed} 表示嵌入维度; $Embedding(\cdot)$ 表示低维映射函数。

对式(6)、(7)结果进行融合处理,得到非结构化工单信息单元数据特征,由 $x_{meta} = x \oplus x_c$ 表示,其中 $\oplus$ 为融合操作符。将最终非结构化工单信息文本向量与 x_{meta} 进行结构化映射处理,获得混合特征

$$x_{concat} = [x_{text}; x_{meta}] \quad (8)$$

其中, x_{concat} 表示非结构化工单信息混合特征向量。

1.3 基于数据挖掘的工单信息关键要素自动标识实现

最小二乘支持向量机是数据挖掘算法中的核心算法,能够处理大规模数据,且对数据的解释性

较强,适用于工单信息挖掘^[15]。因此使用最小二乘支持向量机实现非结构化工单信息关键要素自动标识。

将非结构化工单信息混合特征向量构建为向量集合 $X = \{x_{concat,1}, x_{concat,2}, \dots, x_{concat,N}\}$, 其中 $i \in N$ 为非结构化工单信息混合特征向量索引。令表示 $y = [y_1, y_2, \dots, y_N]^T$ 工单信息关键要素标签, 其中 $y_i \in \{-1, +1\}$, 其中 $+1$ 表示含有关键要素, -1 表示不含关键要素。最小二乘支持向量机通过核函数将非结构化工单信息关键要素自动标识的线性问题扩展到非线性空间, 得到其自动标识表达式为

$$f(X) = \omega^T \phi(X) + \eta \quad (9)$$

其中, $\phi(\cdot)$ 表示非线性映射函数; η 为偏置项; ω^T 表示高维空间权重向量。

为了提升最小二乘支持向量机的收敛速度, 最小化求解自动标识过程中的非线性参数

$$\min_{\omega, b, \xi} \frac{\|\omega\|^2 + \Phi \sum_{i=1}^N \xi_i^2}{2} \quad (10)$$

其中, ξ 表示松弛变量; Φ 表示正则化参数。

设置式(10)约束条件如下

$$y_i (\omega^T \phi(x_{concat,i}) + \eta) = 1 - \xi_i \quad (11)$$

利用拉格朗日函数解决非结构化工单信息关键要素自动标识过程中的松弛变量对偶问题

$$\mathcal{L} = \frac{\|\omega\|^2 + \Phi \sum_{i=1}^N \xi_i^2}{2} - \sum_{i=1}^N \lambda_i [y_i (\omega^T \phi(x_i) + \eta) - 1 + \xi_i] \quad (12)$$

其中, $\mathcal{L}$ 表示拉格朗日函数; λ_i 表示第 i 个拉格朗日乘子。

对参数偏导并令其数值为 0, 求解带约束优化问题的必要条件

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial \omega} = 0 \Rightarrow \omega = \sum_{i=1}^N \lambda_i y_i \phi(x_{concat,i}) \\ \frac{\partial \mathcal{L}}{\partial \eta} = 0 \Rightarrow \sum_{i=1}^N \lambda_i y_i = 0 \\ \frac{\partial \mathcal{L}}{\partial \xi_i} = 0 \Rightarrow \lambda_i = \Phi \xi_i \end{cases} \quad (13)$$

将式(13)带入式(12)内, 替代 ω 和 ξ_i , 得到关于拉格朗日乘子和偏置项的线性方程组, 表达式如下

$$\begin{bmatrix} 0 & y^T \\ y & K + \frac{1}{\Phi} \end{bmatrix} \begin{bmatrix} \eta \\ \lambda \end{bmatrix} = \begin{bmatrix} 0 \\ 1_N \end{bmatrix} \quad (14)$$

其中, K 表示核函数矩阵; $1_N = [1, 1, \dots, 1]^T$ 。

利用求解出的最优拉格朗日乘子、偏置项、权重对自动标识式(9)进行训练, 利用训练好的自动标识表达式, 对新的非结构化工单信息混合特征向量 $\tilde{x}_{concat,i}$ 进行训练, 输出其非结构化工单信息关键要素自动标识结果

$$\hat{y} = \text{sign} \left(\sum_{i=1}^N \lambda_i y_i k(x_{concat,i}, x_{concat,j}) + \eta \right) \quad (15)$$

其中, $\hat{y}$ 表示新 $\tilde{x}_{concat,i}$ 特征非结构化工单信息混合特征向量对应的自动标识结果; $\text{sign}(\cdot)$ 表示二分类标签函数; $k(x_{concat,i}, x_{concat,j}) \in K$ 。

式(15)计算结果即为非结构化工单信息关键要素自动标识结果。

2 实验分析

以国家电网有限公司的全国统一电力服务热线 95598 作为实验对象。该热线平台是一站式电力服务通道, 其主要功能包括: 提供 24 小时电力服务咨询、电费查询、停电报修、故障处理、投诉建议等。95598 服务热线每天会产生大量的电力服务咨询、停电报修、故障处理、投诉等多种类型非结构工单。应用本文方法对 95598 服务热线非结构工单信息内的关键要素重复来电、多次催办、重复诉求等类型工单进行标识, 降低 95598 服务热线投诉风险。从平台数据库采集 2021—2023 年的 50 万条文本工单(含语音转文本记录)作为实验数据。工单平均长度为 85 字/条(最短 12 字, 最长 512 字), 噪声率约 15%(含方言、错别字、非标准表述)。加载本文方法程序的深度学习计算节点采用 Dell PowerEdge R750xa (双路 Intel Xeon Gold 6338, 512GB DDR4)。存储系统为华为 OceanStor 6810(50TB SSD+200TB HDD); 操作系统为 Ubuntu 20.04 LTS(Linux 5.15 内核)。采用 LTP 4.2.7 作为文本数据处理工具。PyTorch 2.1+Transformers 4.35 搭建数据挖掘算法的机器学习框架。

对非结构化工单信息进行泛化处理是实现其关键要素自动标识的基础。以领域适应差距作为

衡量指标,以 95598 服务热线内的紧急服务类、意见建议类、投诉举报类、故障报修类、咨询查询类非结构化工单作为实验对象。使用本文对以上类型非结构化工单信息关键要素进行泛化处理,计算其领域适应差距数值。领域适应差距数值阈值为 3%。测试结果如图 1 所示。

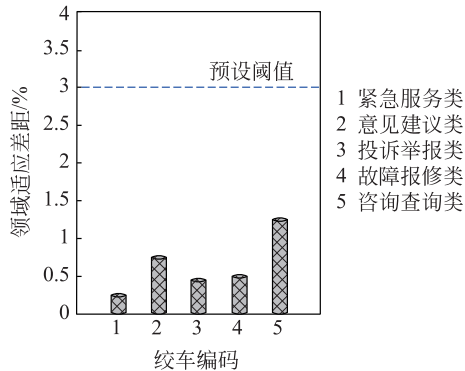


图 1 非结构化工单信息泛化处理领域适应差距

分析图 1 可知,五类 95598 服务热线非结构化工单(紧急服务、意见建议、投诉举报、故障报修、咨询查询类)泛化处理后的领域适应差距数值均远低于阈值。这表明本文方法能有效泛化非结构化工单信息。低差距数值意味着该方法对不同类型工单适配性佳,其泛化能力强且可靠性高。

验证本文方法提取非结构化工单信息特征能力,以斯皮尔曼等级相关系数作为衡量指标。以提取到的 500 个非结构化工单信息特征作为实验对象,统计其斯皮尔曼等级相关系数,以散点形式呈现实验结果,如图 2 所示。

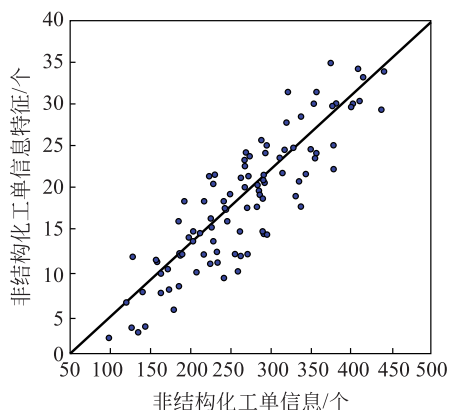


图 2 特征提取的斯皮尔曼等级相关系数

分析图 2 可知,使用本文方法对非结构化工单信息特征提取时,其信息与特征之间呈强相关

关系,斯皮尔曼相关系数接近 1 的曲线,说明本文方法提取的非结构化工单信息特征能体现非结构化工单信息的内在联系。

长尾实体是指在非结构化工单信息中出现频率低、语义多样性高,但对业务具有重要价值的实体。以长尾实体覆盖率(Long-Tail Entity Coverage)作为衡量本文方法对工单信息特征结构化映射能力指标,以 1000 个非结构化工单作为实验对象,分析其特征结构化映射后得到混合特征向量的长尾实体覆盖率,实验结果如图 3 所示。

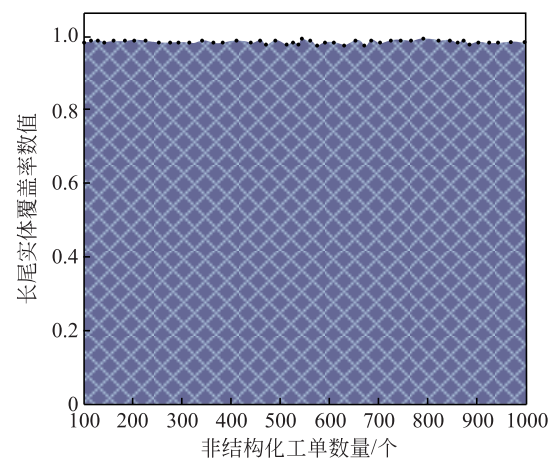


图 3 特征结构化映射混合特征长尾实体覆盖率

分析图 3 可知,随着非结构化工单数量的增加,长尾实体覆盖率数值始终维持在较高水平,数值均接近 1.0,该数值表明:本文方法在处理不同数量的非结构化工单时,对长尾实体的覆盖能力较为稳定且出色,能够有效识别和提取非结构化工单信息中的长尾实体,说明此方法在对工单信息特征结构化映射方面具备良好且稳定的能力。

以 3 个 95598 服务热线同地点非结构化报修工单作为实验对象,使用本文方法对工单内关键要素进行自动标识,该 3 个工单详情如表 1 所示,其关键要素自动标识结果如表 2 所示。

结合表 1 与表 2 分析,本文方法能够准确自动标识非结构化工单信息的关键要素。首先,服务地点解析精准:根据表 1 的自然语言描述(如“15 栋·2·单元”“步行街·8·号商铺”“地下车库负·2·层”),表 2 清晰提取地点并区分功能场景(居民、商业、公共设施),实现空间维度的结构化标识。其次,故障核心要素提取透彻:针对表 1 中零散的故障描述(如“电缆分支箱故障”“专用线路断路”“保护装置跳闸”),表 2 不仅定位故障设备(电

表 1 95598 服务热线同地点非结构化报修工单内容

项目		内容	
工单编号	Z001	Z002	B001
受理时间	2024-06-11 08:33	2024-06-11 08:26	2024-06-11 08:37
客户姓名	王女士	陈先生	刘女士
联系电话	135 **** 1231	189 **** 6721	150 **** 9011
工单类型	停电报修	故障处理	停电报修
服务地点	望京花园 15 栋 2 单元	望京花园步行街 8 号商铺	望京花园地下车库负 2 层
服务内容描述	单元楼突然停电,自查户内空开无跳闸,物业反馈单元总配电箱无异常,怀疑楼前电缆分支箱故障	商铺及周边 3 家门店停电,电表显示正常但无电压,隔壁住宅楼供电正常,推测商铺专用供电线路断路	车库照明及通风系统停电,车辆进出困难(部分电动车充电中断),保安亭监控显示供电干线保护装置跳闸
处理状态	已受理	待派工	待派工
备注	家中有医用制氧机,急需恢复供电	正值促销活动期,需尽快恢复商业用电	已启动应急照明,维修人员排查电缆接头过热问题

表 2 非结构化工单信息关键要素自动标识结果

工单编号	服务地点	故障核心要素	特殊影响
Z001	望京花园 15 栋 2 单元	电缆分支箱故障(公共线路)	医用制氧机供电中断(医疗紧急)
Z002	望京花园步行街 8 号商铺	商铺专用供电线路断路(商业专用)	促销活动期商业用电中断(经济紧急)
B001	望京花园地下车库负 2 层	供电干线保护装置跳闸(公共设施)	车辆进出/电动车充电中断(安全隐患)

缆分支箱、供电干线保护装置),还明确线路属性(公共、商业专用),将模糊语义转化为技术化故障类型,支撑运维决策。最后,特殊影响识别敏锐:从表 1 备注的“医用制氧机”“促销活动”“车辆进出”等非结构化信息中,表 2 提炼出“医疗紧急”“经济紧急”“安全隐患”等优先级标签,为工单调度提供核心依据。上述结果表明:本文方法对 95598 服务热线非结构化工单关键要素进行自动识别,能够实现故障根源、场景属性、影响等级的自动化解构,为 95598 服务的快速响应(如医疗优先派工、商业用电加急处理、公共安全联动)提供数据支撑,也呈现了本文方法在电力运维场景中对复杂非结构化信息的强大标识能力。

3 结语

本文针对非结构化工单信息关键要素标识难

题,提出基于数据挖掘的自动标识方法。通过泛化处理消除数据格式异构性,利用预训练语言模型提取特征并与工单元数据结构化映射,结合最小二乘支持向量机实现关键要素自动标识。实验结果可知,本文方法具备非结构化工单信息关键要素自动标识能力,能够为企业提升工单处理效率、驱动业务流程自动化提供有效的解决方案。

参 考 文 献

[1] 周显春,喻佳. 基于图神经网络的人工自然语言语义挖掘仿真[J]. 计算机仿真,2024,41(1):344-348.
ZHOU X C, YU J. Simulation of artificial natural language semantic mining based on graph neural network[J]. Computer Simulation, 2024, 41(1): 344-348. (in Chinese)
[2] 安然,储继华,洪先锋. 面向非结构化数据的情报分析方法体系框架研究[J]. 情报理论与实践,2024,47(2):143-150.
AN R, CHU J H, HONG X F. Research on the system framework of intelligence analysis methods for unstructured data [J]. Information Studies: Theory & Application,

- 2024, 47(2): 143-150. (in Chinese)
- [3] 王章宇, 陈阳, 周彬, 等. 一种面向非结构化道路的点云语义分割方法[J]. 北京航空航天大学学报, 2025, 51(2): 457-465.
- WANG Z Y, CHEN Y, ZHOU B, et al. A point cloud semantic segmentation method for unstructured roads[J]. Journal of Beijing University of Aeronautics and Astronautics, 2025, 51(2): 457-465. (in Chinese)
- [4] 艾青林, 张俊瑞, 吴飞青. 基于小目标类别注意力机制与特征融合的 AF-ICNet 非结构化场景语义分割方法[J]. 光子学报, 2023, 52(1): 189-202.
- AI Q L, ZHANG J R, WU F Q. AF-ICNet unstructured scene semantic segmentation method based on small target category attention mechanism and feature fusion[J]. Acta Photonica Sinica, 2023, 52(1): 189-202. (in Chinese)
- [5] 唐瑞雪, 秦永彬, 陈艳平. 融合限定关系和交互信息的实体关系联合抽取模型[J]. 中文信息学报, 2024, 38(10): 106-116.
- TANG R X, QIN Y B, CHEN Y P. A joint entity relation extraction model fusing limited relations and interactive information[J]. Journal of Chinese Information Processing, 2024, 38(10): 106-116. (in Chinese)
- [6] 李玉聪, 汪士钦, 张梦玺, 等. 基于 WAI-ARIA 的网页导航栏地标属性的标识方法[J]. 吉林大学学报(理学版), 2024, 62(3): 697-703.
- LI Y C, WANG S Q, ZHANG M X, et al. A method for identifying landmark attributes of web navigation bars based on WAI-ARIA[J]. Journal of Jilin University (Science Edition), 2024, 62(3): 697-703. (in Chinese)
- [7] 王景慧, 卢玲, 段志丽, 等. 融合依存信息的关系导向型实体关系抽取方法[J]. 计算机应用研究, 2023, 40(5): 1410-1415, 1440.
- WANG J H, LU L, DUAN Z L, et al. Relation-oriented entity relation extraction method fusing dependency information[J]. Application Research of Computers, 2023, 40(5): 1410-1415, 1440. (in Chinese)
- [8] 姜艳杰, 东春浩, 刘辉. 一种基于词法特征和数据挖掘的无意义变量名检测方法[J]. 计算机科学, 2024, 51(6): 23-33.
- JIANG Y J, DONG C H, LIU H. A method for detecting meaningless variable names based on lexical features and data mining[J]. Computer Science, 2024, 51(6): 23-33. (in Chinese)
- [9] 张咏华, 邬开俊. 基于随机森林的非常规突发事件结构化情景体系模型[J]. 公路交通科技, 2024, 41(2): 139-147, 190.
- ZHANG Y H, WU K J. A structured scenario system model for unconventional emergencies based on random forest[J]. Journal of Highway and Transportation Research and Development, 2024, 41(2): 139-147, 190. (in Chinese)
- [10] 孙熠衡, 刘茂福. 基于知识提示微调的标书信息抽取方法[J]. 计算机应用, 2025, 45(4): 1169-1176.
- SUN Y H, LIU M F. A bid information extraction method based on knowledge prompt tuning[J]. Journal of Computer Applications, 2025, 45(4): 1169-1176. (in Chinese)
- [11] 温清华, 朱洪银, 侯磊, 等. 多策略中文开放关系抽取方法[J]. 中文信息学报, 2023, 37(1): 88-96.
- WEN Q H, ZHU H Y, HOU L, et al. A multi-strategy method for open Chinese relation extraction[J]. Journal of Chinese Information Processing, 2023, 37(1): 88-96. (in Chinese)
- [12] 邢季, 刘瑾, 张建伟. 基于双步抽取的低资源中文工业领域术语抽取方法[J]. 武汉大学学报(理学版), 2024, 70(3): 329-340.
- XING J, LIU J, ZHANG J W. A low-resource Chinese industrial term extraction method based on two-step extraction[J]. Journal of Wuhan University (Natural Science Edition), 2024, 70(3): 329-340. (in Chinese)
- [13] 庞娜, 袁钺, 薛秋红. 基于迁移学习的化学键能数据自动抽取[J]. 现代情报, 2023, 43(1): 19-28.
- PANG N, YUAN Y, XUE Q H. Automatic extraction of chemical bond energy data based on transfer learning[J]. Journal of Modern Information, 2023, 43(1): 19-28. (in Chinese)
- [14] 董家慧子, 谢忠, 邱芹军, 等. 融合容错机制的基于 Attention-Mask RCNN 地质表格信息抽取方法[J]. 地质科学, 2023, 58(3): 1147-1163.
- DONG J H Z, XIE Z, QIU Q J, et al. A geological table information extraction method based on Attention-Mask RCNN with fault-tolerant mechanism[J]. Geological Science, 2023, 58(3): 1147-1163. (in Chinese)
- [15] 周炫余, 刘林, 卢笑, 等. 多模态信息增强表示的中文关键词抽取方法[J]. 清华大学学报(自然科学版), 2024, 64(10): 1785-1796.
- ZHOU X Y, LIU L, LU X, et al. A Chinese keyword extraction method with multimodal information enhanced representation[J]. Journal of Tsinghua University (Science and Technology), 2024, 64(10): 1785-1796. (in Chinese)