

井下险情模糊位置辨析方法

孙长春^{1,4}, 黄晓凡², 苏令印³, 徐金陵¹, 汤效平², 王兹尧², 林泽东³

(1. 北京华电力拓能源科技有限公司, 北京 102400; 2. 华电电力科学研究院有限公司, 浙江 杭州 310012; 3. 山东科技大学 计算机科学与技术学院, 山东 青岛 266590; 4. 中国矿业大学 矿业工程学院, 江苏 徐州 221116)

摘要: 煤矿安全生产一直备受关注, 如何确保矿工人身安全、确定隐藏险情模糊位置, 一直是煤矿安全生产领域的研究重点。针对如何在井下复杂的环境中确定险情位置, 提出一种井下险情模糊位置辨析方法。将收集到的矿工语音数据在移动端转化为文本数据, 并对该数据进行模糊数据补全, 通过优先级的策略补全数据集。最后, 结合矿工的状态, 通过质心算法确定险情位置。

关键词: 煤矿安全; 隐藏险情; 位置辨析; 文本数据; 数据处理

DOI: 10. 11907/rjdk. 222049

中图分类号: TP391; TD76

文献标识码: A

开放科学(资源服务)标识码(OSID):

文章编号: 1672-7800(2023)008-0099-05



Discrimination Method of Fuzzy Position of Underground Danger

SUN Changchun^{1,4}, HUANG Xiaofan², SU Lingyin³, XU Jinling¹, TANG Xiaoping², WANG Ziyao², LIN Zedong³

(1. Beijing Huadian Leadtop Energy Technology Co., Ltd., Beijing 102400, China; 2. Huadian Electric Power Research Institute Co., Ltd., Hangzhou 310012, China; 3. College of Computer Science and Technology, Shandong University of Science and Technology, Qingdao 266590, China; 4. School of Mines, China University of Mining and Technology, Xuzhou 221116, China)

Abstract: Coal mine safety production has always been concerned. How to ensure miners' personal safety and determine the fuzzy position of hidden danger has always been the research focus in the field of coal mine safety production. Aiming at how to determine the dangerous position in the complicated underground environment, this paper puts forward a method of distinguishing the fuzzy position of underground dangerous situation. The collected miners' voice data is converted into text data at the mobile terminal, and fuzzy data is completed for this data, and the data set is completed by priority strategy; Finally, combined with the status of miners, the dangerous position is determined by centroid algorithm.

Key Words: coal mine safety; hidden danger; position discrimination; text data; data processing

0 引言

随着我国经济体制的不断改革, 现代化进程不断加快, 煤矿资源的安全生产一直备受人们关注^[1]。目前, 煤矿是我国主要能源生产企业, 对推动社会建设和发展起着巨大作用。但是, 煤炭储存条件较为复杂, 很多灾害与之并存, 导致开采难度加大、危险程度增高, 加上其他不确定性危险因数, 使得煤矿事故率极高。特别是在井下险情发现方面, 如何更精确地确定险情位置一直是研究的重点。

在国内已有研究中, 针对险情位置定位的研究大多采用传感器器件实现险情定位, 例如森林火灾的定位采用温度传感器^[3]、危险气体定位依据各类气体传感器^[4]等, 但这些方法都是依靠环境本身发生的变化而确定险情位置, 忽略了人在其中的作用。

近年来, 随着人工智能(Artificial Intelligence, AI)与大数据的发展, 作为AI的重要研究方向之一, 自然语言处理(Natural Language Processing, NLP)成为研究热点^[5-6]。利用矿工语音数据转化的文本数据进行自然语言处理, 语言终端转化的文本数据主要是无结构化的数据^[7], 因此要对

收稿日期: 2022-09-07

基金项目: 华电集团重点科技项目(CHDKJ21-01-22); 山东省泰山学者工程专项基金项目(ts20190936); 山东省高等学校青创科技支持计划项目(2019KJN024)

作者简介: 孙长春(1977-), 男, 北京华电力拓能源科技有限公司高级工程师, 中国矿业大学矿业工程学院硕士研究生, 研究方向为煤矿智能化信息系统集成与智能化开采; 徐金陵(1972-), 男, 北京华电力拓能源科技有限公司高级工程师, 研究方向为煤矿过程自动化与综合监控。

初始的文本数据进行中文分词、去除停用词、噪声处理、关键词提取、格式化等预处理操作,提取有用的文本信息。

在已有井下险情研究中,无论是传统的矿井作业,还是现代高度智能化的矿井工程,险情数据主要来源于井下负责人,普通矿工将发现的险情报告给负责人,负责人向井上人员传递险情信息^[8-9]。但是这种险情发现方式不仅浪费时间,而且由于是负责人单向传递信息,导致整个险情发现过程中信息不完整,缺少矿工的信息反馈。再者,井下工作人员文化知识有限导致的险情错误上报、险情描述不清等问题,也容易带来险情无法定位的问题。

由于是在复杂的环境中,井下人员对险情往往描述不清,因而提供给计算机的数据存在大量稀疏数据。还有险情的不同类别、有关险情的数据数量、险情发布者身份的重要性等因素对险情位置辨析结果的影响,并没有得到很好的处理。

为解决以上问题,本文提出了一种井下险情模糊位置辨析方法。当井下发生险情时,矿工报告险情发生情况,移动终端将收集的语音数据转化为文本数据;之后,对文本数据进行预处理,将预处理后的数据提供给计算机,同时根据不同优先级对从井下收集到的数据进行补充,结合质心算法确定大致险情位置。

1 险情模糊位置辨析方法

险情模糊位置辨析方法具体流程如图1所示。由此可知,矿工通过移动端内置的传感器和语音转换工具将语音转换为初始文本数据,初始文本数据发送到服务器端,

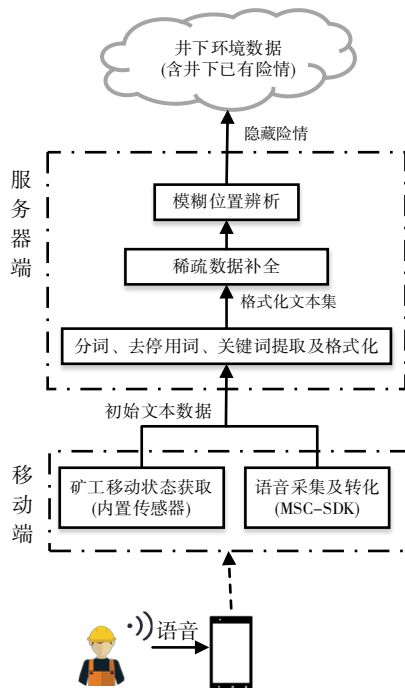


Fig. 1 Fuzzy position discrimination process of danger
图1 险情模糊位置辨析流程

对数据进行预处理,并且对数据集进行属性补充,继而结合矿工的状态,通过质心算法,确定险情的模糊位置。

1.1 数据预处理

当前,语音处理技术已非常成熟,许多第三方平台也提供了成熟的语音开发工具^[10]。本文涉及的语音开发工具为科大讯飞移动语音终端(MSC SDK)^[11],MSC SDK语音识别成功率在98%以上,确保了实验的准确性。

当井下发生险情时,矿工将发现的险情情况通过语音形式快速上报,语音数据通过MSC SDK转化为短文本数据,再对短文本数据作预处理操作,生成能被计算机直接处理的数据^[12]。

设 T_0 为完整的数据属性集:

$$T_0 = \{U, U_i, U_d, U_p, D_t, D_d, D_l, M_n\} \quad (1)$$

其中: U 表示矿工编号, U_i 表示发布者身份, U_p 表示矿工位置, U_d 表示矿工脸部朝向; D_t 表示险情种类; D_d 表示险情与矿工的相对方位角; M_n 表示险情的信息数量; D_l 表示矿工与险情发生位置之间的距离。

设 T' 为每条数据中必含的属性集合: $T' = \{U, U_i, U_p, U_d, M_n\}$ 。

每个员工的数据集为: $T_0(U) \subseteq T(Z_u)$, $T_0(U) = \{T_0(1), T_0(2), \dots, T_0(M_n)\}$ 。

则经过预处理后得到的数据集 $T(Z_t)$ 为:

$$T(Z_t) = \{T(Z_{t1}), T(Z_{t2}), \dots, T(Z_{tm})\}, T_{ic} \subseteq T_0 \text{ 且 } T' \subseteq T_{ic} \quad (2)$$

其中, T_{ic} 表示经过预处理后的稀疏数据集, $T(Z_u)$ 表示经过预处理后 t 时刻 Z_t 区域的数据集。

1.2 基于优先级的模糊数据补充方法

以往数据补充方法大多采用自编码器和对抗网络^[13]的相关算法,但这种方法在模型训练中会出梯度消失的情况,缺乏稳定性。同时,针对短期内的大量数据,对处理器要求过高且无法及时处理,为此本文提出基于优先级的模糊数据补充方法,先对数据进行优先级处理,再进行补充,极大缩短了数据处理时间,提高了效率。

由于原始数据的不完整性,矿工上传的数据经处理后不可避免地与实际数据存在一定误差^[14],因而需要补充数据集 $T(Z_t)$ 中的稀疏数据。同时,当井下发生险情时,矿工短期内会发送大量数据,造成井下数据井喷式增长。因此,在处理数据时,要对数据进行优化,以提高数据处理效率。

1.2.1 基于不同影响因子的数据排序

在数据处理中,由于 T_0 中包含着不同的属性值,不同属性值的重要性也不同,则根据属性值的不同,对数据处理的顺序也不同。

设置数据处理优先级的影响因子:

$$\langle \varepsilon \rightarrow U_i, \gamma \rightarrow D_t, \delta \rightarrow M_n \rangle \quad (3)$$

对不同的影响因子进行排序,并计算不同影响因子序列下的数据优先级权重,得到具有优先级权重的数据集

$T_1(Z_i)$ 。具体算法流程如算法 1 所示。

算法 1 $T(Z_i)$ 数据集基于优先级的排序算法

输入: t 时刻井下预处理后的数据集 $T(Z_i)$
 输出: 井下各区依据优先权重进行排序的待处理中间数据集 $T_1(Z_i)$

- 1 FOREACH($T(Z_{i_u}) \in T(Z_i), i \in N$)
- 2 依据险情类别权重大小, 对 Z_i 区域内的数据集 $T(Z_{i_u})$ 进行降序排列。获得不同类别处理顺序的数据集:
- 3 $T'(Z_{i_u}) = \{T'(Z_{i_u}(D_k)), i, k \in N\}$; 其中, 单位时间内, 按照一个整体处理关于同一险情的一条或多条数据;
- 4 FOREACH($T'(Z_{i_u}(D_k)) \in T'(Z_{i_u}), k \in N$)
- 5 对不同险情的数据集 $T'(Z_{i_u}(D_k))$, 依据发布者身份 U 的权重降序排序。得到相同险情下不同发布者身份的数据集:
- 6 $T''(Z_{i_u}(D_k \cdot U)) = \{T''(Z_{i_u}(D_k \cdot U_s)); i, k, s \in N\}$
- 7 FOREACH($T''(Z_{i_u}(D_k \cdot U_s)) \in T''(Z_{i_u}(D_k \cdot U)), s \in N$)
- 8 依据员工发布数量 M_n , 对于相同类型的险情、相同发布者身份发布的数据, 由大到小进行排序。其中, 如果发布数量 M_n 相同, 则按上传时间的先后顺序排序, 得到数据处理序列 $T_1(Z_{i_u})$ 。
- 9 RETURN $T_1(Z_i)$;

对数据进行 3 次算法循环操作, 按照不同的影响因素, 获得最终具有处理次序的数据集。

1.2.2 稀疏数据补全

当井下险情发生时, 矿工情绪紧张, 不能完全具体地描述整个险情, 导致大量稀疏数据存在于数据集 $T_1(Z_i)$ 内。因此, 需要对数据集中的稀疏数据采取属性值补全或删除操作, 转化为可直接用于险情位置辨析的数据集 $T_2(Z_i)$ 。具体算法流程如算法 2 所示。

经过基于优先级的模糊数据补全, 获得了完整的数据集, 为险情模糊位置辨析提供了基础。

1.3 险情模糊位置辨析

险情位置的确定还依靠矿工的坐标位置、脸部朝向等发布语音数据时的状态, 以及矿工与险情的距离、相对方位角等数据。

依据 T_j 中的 U_p, U_d, D_d, D_i 进行险情位置计算, 其中, $D_d \in [-\frac{\pi}{3}, 0]$ 。计算得到险情位置可能点的集合, 再通过质心算法求出险情位置, 具体算法如算法 3 所示。

2 仿真实验验证与结果分析

2.1 实验数据

在模拟仿真实验中, 为验证险情位置辨析方法的作用, 使用 SuperMap iDesktop 8C^[15] 工具, 绘制简易井下地图。如图 2 所示, 在地图中标记了 50 个定位点, 从而确定矿工位置。

根据井下发生险情时矿工的状态, 模拟了 1 000 条语

算法 2 补全有序数据集 $T_1(Z_i)$

输入: 待补全的有序数据集 $T_1(Z_i)$
 输出: 可直接用于险情位置辨析的数据集 $T_2(Z_i)$

- 1 FOREACH($T(Z_{i_u}) \in T(Z_i), i \in N$)
- 2 FOREACH($T(U^k) \in T_1(Z_{i_u})$)
- 3 依据对事件的描述流程, 补全员工 U^k 中的数据集 $T(U^k)$
- 4 设当前 Z_i 区域的可用于位置辨析的数据集为: $T_{temp}(Z_i)$
- 5 FOREACH($T_j \in T(U^k), j \in N$)
- 6 IF(T_j is complete)
- 7 将险情数据 T_j 保存到数据集 $T_{temp}(Z_i)$ 的末尾;
- 8 IF(T_j is incomplete)
- 9 抽取自身同险情数据关键属性值进行补全。如有新属性值, 则更新 T_j ; 将更新的 T_j 保存到 $T_{temp}(Z_i)$;
- 10 IF(T_j is incomplete)
- 11 从当前 Z_i 区域寻找其他人员同类险情数据 T_j ,
- 12 如果有, 则补全 T_j 保存到 $T_{temp}(Z_i)$ 中; 如果没有, 则删除 T_j 。
- 13 将数据集 $T_{temp}(Z_i)$ 添加到 $T_2(Z_i)$ 中;
- 14 RETURN $T_2(Z_i)$;

算法 3 险情位置辨析

输入: 可直接用于险情位置辨析的数据集 $T_2(Z_i)$
 输出: 险情位置集数据集 $F_r(Z_i)$

- 1 FOREACH($T_r(Z_i) \in T_2(Z_i), i \in N$)
- 2 FOREACH($T_j \in T_r(Z_i), j \in N$)
- 3 IF($D_d \neq 0$)
- 4 取矿工当前位置为 $A(x_a, y_a)$, 险情位置为 $D_p(x_d, y_d)$,
- 5 且 $A = U_p$, 设当前矿工的朝向为 \overline{AC} 。其中 $C(x_c, y_c)$ 为 U_d 上的任意一点。则矿工和险情的位置向量为
- 6 $\overline{AD} = (x_d - x_a, y_d - y_a)$,
- 7 由 $D_i^2 = (x_d - x_a)^2 + (y_d - y_a)^2$
- 8 $D_d = \arccos((\overline{AC} \cdot \overline{AD}) / (|\overline{AC}| * |\overline{AD}|))$
- 9 得出险情可能位置点集: $P_{temp} = \{p_1, p_2, \dots, p_n\}$;
- 10 通过质心算法求解点集 P_{temp} 的质心 D_p , 获得隐藏险情位置数据 $T_j = \{U, D_p, D_i\}$, 将 T_j 保存到结果集 $F_r(Z_i)$ 中
- 11 IF($D_d = 0$)
- 12 确定当前位置点为险情发生点, 将数据 T_j 保存到 $F_r(Z_i)$
- 13 RETURN $F_r(Z_i)$

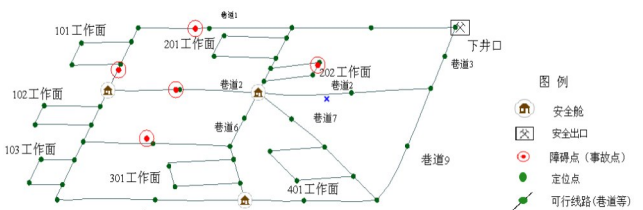


Fig. 2 A simple underground map

图 2 井下简易地图

音数据, 并将语音数据通过 MSC SDK 转换成短文本数据, 用于险情模糊位置辨析实验。模拟数据如表 1 所示, 数据

处理过程中,发布者身份优先级权重设计如表2所示,不同险情种类优先级权重设定如表3所示。

Table 1 Simulation data

表1 模拟数据

种类	示例	备注	数量(条)
单点险情数据	我前方30米处有石头滑落,有人受伤!	石头滑落	20
范围险情数据	我这里有大面积渗水	渗水	5
不完整数据	在左前方差不多60米处	-	950

Table 2 Priority weight of publisher identity

表2 发布者身份优先级权重

发布者身份	权重
区长、副区长	0.50
班组长、技术员、放炮员、安全员	0.30
普通矿工	0.20

Table 3 Priority weight of dangerous types

表3 险情种类优先级权重

险情种类	权重
失火灾害	0.14
地下水泄露	0.16
瓦斯泄露	0.14
煤尘爆炸	0.14
瓦斯突出	0.14
人身伤害	0.15
石头滑落	0.10
机电事故	0.03

2.2 实验评价指标

设 T_m 为数据补全时间,不同影响因子计算次序不同, T_m 也不同, T_m 越小,说明补全效率越高。同时,准确率 (Accuracy, A1) 越高,也能说明补全效果越好。

单位时间内,某区域内的准确率为:

$$A1 = \frac{M_r}{M_f} \times \frac{M_s}{M_0} \quad (4)$$

其中, M_0 为数据总量, M_s 为稀疏数据量, M_r 为最终补全结果量, M_f 为补全结果参考数据量。

以成功率 (Success, S1)、准确率 (Accuracy, A2) 作为险情模糊位置辨析结果的评价指标。成功率 S1 为:

$$S1 = \frac{M'_p + M'_{sp}}{M_p + M_{sp}} \quad (5)$$

其中, M_p 为参考点的险情数量, M'_p 为正确点的险情数量, M_{sp} 为范围险情数量, M'_{sp} 为正确范围的险情数量。

加权平均准确率 A2 为:

$$A2 = \frac{M'_p}{2 * M_p} + \frac{M'_{sp}}{2 * M_{sp}} \quad (6)$$

2.3 实验与结果分析

2.3.1 稀疏数据补全实验与结果分析

本文进行了多次仿真实验,统计了不同影响因子计算顺序下的数据补全操作耗时 T_m ,如图3所示。

可以看出,在同一实验中,不同影响因子计算顺序在数据补全中的耗时不同,当计算顺序为 $\gamma \rightarrow \epsilon \rightarrow \delta$ 时,数据

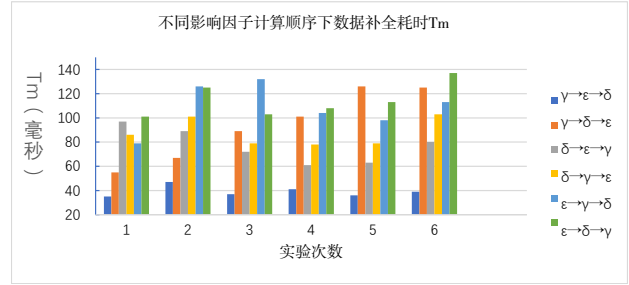


Fig. 3 Data completion time under different order of influence factor calculation

图3 不同影响因子计算顺序下的数据补全耗时

补全时间最低、补全效率最好。随着实验次数的增加,不同影响因子计算顺序在数据补全中的耗时也发生了变化。

稀疏数据补全操作准确率统计结果如图4所示。控制 M_0 不变,设定变量 M_s 和 M_f 的不同数据。

由图4可以看出,经过数次实验,稀疏数据补全操作准确率 A1 呈现平稳状态并均大于 63%。结合图3可以看出,不同数据补全耗时并不影响稀疏数据补全操作的准确率。由此可以证明,基于优先级的补全方法在稀疏数据补全方面是有效的。

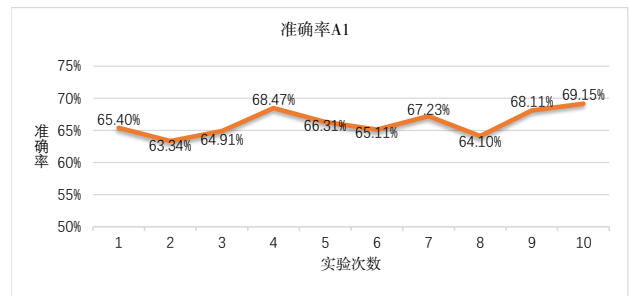


Fig. 4 Accuracy statistics of sparse data completion operation

图4 稀疏数据补全操作准确率统计

2.3.2 险情模糊位置辨析结果分析

经过多次实验,改变预设点险情数据量和预设范围险情数据量,统计险情模糊位置辨析 S1 和 A2,结果如图5所示。

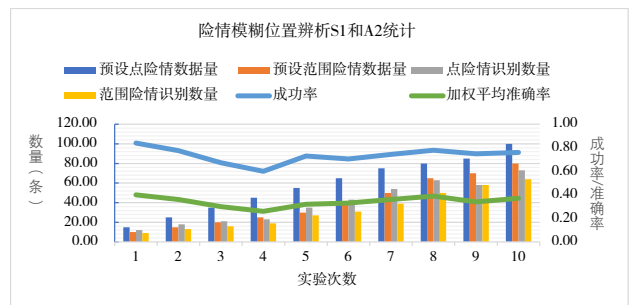


Fig. 5 S1 and A2 statistics of the fuzzy position of the dangerous situation analysis

图5 险情模糊位置辨析 S1 和 A2 统计

由图5可知,本文方法对于危险源的辨析成功率 S1 大于 60%,平均准确率大于 26%。并且,当预设点险情数据量与预设范围险情数据量较少时,实验仍有较高的成功

率。上述结果证明,本文方法具有较好的稳定性。

综上可知,对于范围险情识别成功率大于点险情识别成功率,原因是范围险情覆盖一定的面积,容易识别到险情位置。为此,本文针对险情识别进行可视化实验,如图 6 所示。

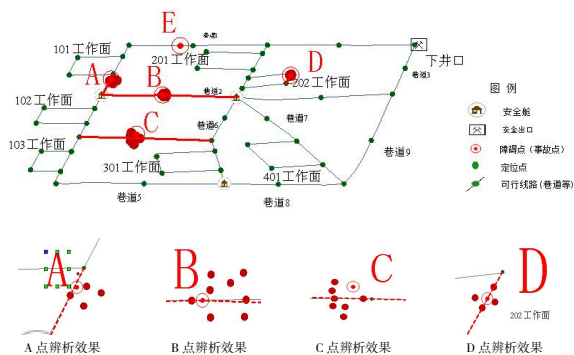


Fig. 6 Result of danger discrimination

图6 险情辨析效果

在图 6 中,空心圈代表真实险情位置,实心圈代表预测的险情位置。可以看出,虽然实验结果与真实险情位置存在误差,但都在真实险情的周围,因而能较好地确定险情位置。

实验结果表明,本文提出的险情模糊位置辨析方法能够有效地辨析出隐藏险情位置。

3 结语

针对井下发生险情时,如何在复杂的井下环境中确定险情位置这一问题,本文提出了一种井下险情模糊位置辨析方法。通过优先级策略补全矿工数据集,并结合质心算法确定险情位置,实验结果证明了本文所提出方法的有效性。但研究仍存在一些不足,如在语音识别和转换过程中,没有考虑 MSC SDK 移动语音终端识别准确性对实验结果的影响,直接对 MSC SDK 移动语音终端转换的文本数据进行处理;同时,在基于优先级补全数据的策略中,没有考虑到更多的影响因子,没有提出更加科学的优先级策略。在未来工作中,对于更加复杂嘈杂的环境,语音终端是否可以准确提取到有用信息,更多的影响因子对实验结果是否有影响,对于这些问题需作进一步研究,设计更加科学的计算方法。

参考文献:

[1] WANG N, LYU X H, MIAO Z B. Design of coal safety risk early warning management system based on cloud computing [J]. Coal Technology, 2022,41(9):204-207.
王诺,吕晓华,苗志彬.基于云计算的煤炭安全风险预警管理系统设计[J].煤炭技术,2022,41(9):204-207.

[2] WANG J N, LIANG Y Q. A summary of Chinese word segmentation research[J]. Software Guide, 2021,20(4):247-252.

王佳楠,梁永全.中文分词研究综述[J].软件导刊,2021,20(4):247-252.

[3] MO Y J. System research on forest firefighters danger detection and location[D]. Beijing:Beijing Forestry University, 2020.
莫燕君.林火消防员险情检测与定位系统研究[D].北京:北京林业大学,2020.

[4] LI Z H. Study on on-line monitoring system of hazardous gas leakage location[D]. Tianjin:Tianjin University of Technology, 2019.
李子豪.危险气体泄露定位在线监测系统的研究[D].天津:天津理工大学,2019.

[5] JI H L, HUANG J D. Research on the development and application of artificial intelligence industry in China [J]. Software Guide, 2019, 18(3):34-38.
纪汉霖,黄嘉冬.我国人工智能产业发展及应用研究[J].软件导刊,2019,18(3):34-38.

[6] ZHANG D. Research on the identification and prevention and control of hazard sources of water disaster in coal mines [J]. Energy and Energy Conservation, 2017(5):8-9.
张端.煤矿水灾危险源辨识及防控研究[J].能源与节能,2017(5):8-9.

[7] DENG D J, YAO L. Improved TF-IDF-based feature word extraction algorithm for Weibo short text [J]. Software Guide, 2016, 15(6):48-49.
邓丹君,姚莉.基于改进TF-IDF的微博短文本特征词提取算法[J].软件导刊,2016,15(6):48-49.

[8] ZHAO A X, SHI X G, LIU K. Technical path analysis of data-driven intelligent mine construction [J]. Software Guide, 2021, 20(10):13-18.
赵安新,史新国,刘柯.数据驱动的智能化工矿建设技术路径分析[J].软件导刊,2021,20(10):13-18.

[9] SU L Y. Research on fuzzy recognition of danger position and method to voice escape navigation under the underground coordination mechanism [D]. Qindao:Shandong University of Science and Technology, 2018.
苏令印.井下协同疏导机制下的险情模糊位置辨析与语音逃生导航方法的研究[D].青岛:山东科技大学,2018.

[10] XIE H, LIANG J G, CHENG T. Research on localization algorithm in wireless sensor networks [J]. Software Guide, 2015, 14(9):80-82.
解皓,梁建国,程涛.无线传感器网络定位算法研究[J].软件导刊,2015,14(9):80-82.

[11] iFLYTEK. Xunfei open platform [EB/OL]. <http://www.xfyun.cn>.科大讯飞.讯飞开放平台 [EB/OL]. <http://www.xfyun.cn>.

[12] YANG X D, WU Y H, ZHU H B, et al. A review of the application of natural language processing in financial text information analysis [J]. Sci-tech Finance Monthly, 2022(8):69-72.
杨晓丹,吴延晖,朱浩滨,等.自然语言处理在金融文本信息分析的应用综述[J].科技与金融,2022(8):69-72.

[13] WU F X, CHENG J. Configurable text-based image editing by autoencoder-based generative adversarial networks [J]. Journal of Software, 2022,33(9):3139-3151.
吴福祥,程俊.基于自编码器生成对抗网络的可配置文本图像编辑[J].软件学报,2022,33(9):3139-3151.

[14] LIU L Q, LI Y K, SUO H J. Research and development of cloud service security platform and application of speech recognition [J]. Software Guide, 2014, 13(1):7-8.
刘琳茜,李永康,索红军.云服务安全平台研究开发与语音识别应用[J].软件导刊,2014,13(1):7-8.

[15] CUI F. Design of a map display system based on SuperMap [J]. China Science and Technology Information, 2022(8):59-60,63.
崔凤.一种基于SuperMap的地图展示系统设计[J].中国科技信息,2022(8):59-60,63.

(责任编辑:孙娟)