



基元潜部特征元挖掘的智能方法研究

张丽芳, 李兴森

引用本文:

张丽芳, 李兴森. 基元潜部特征元挖掘的智能方法研究[J]. 智能系统学报, 2025, 20(2): 457-464.

ZHANG Lifang, LI Xingsen. Research on intelligent methods for latent features mining of basic element[J]. *CAAI Transactions on Intelligent Systems*, 2025, 20(2): 457-464.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202310039>

您可能感兴趣的其他文章

非结构化文档敏感数据识别与异常行为分析

Unstructured document sensitive data identification and abnormal behavior analysis

智能系统学报. 2021, 16(5): 932-939 <https://dx.doi.org/10.11992/tis.202104028>

“范式变革”引领与“信息转换”担纲：机制主义通用人工智能的理论精髓

Leading of paradigm shift and undertaking of information conversion: theoretical essence of mechanism-based general AI

智能系统学报. 2020, 15(3): 615-622 <https://dx.doi.org/10.11992/tis.202002019>

人工智能中的封闭性和强封闭性——现有成果的能力边界、应用条件和伦理风险

Criteria of closeness and strong closeness in artificial intelligence——limits, application conditions and ethical risks of existing technologies

智能系统学报. 2020, 15(1): 114-120 <https://dx.doi.org/10.11992/tis.202001001>

图像情境下的数字序列逻辑学习

Number sequence logic learning in image context

智能系统学报. 2019, 14(6): 1189-1198 <https://dx.doi.org/10.11992/tis.201905044>

重新找回人工智能的可解释性

Refining the interpretability of artificial intelligence

智能系统学报. 2019, 14(3): 393-412 <https://dx.doi.org/10.11992/tis.201810020>

因素空间理论——机制主义人工智能理论的数学基础

Factor space-mathematical basis of mechanism based artificial intelligence theory

智能系统学报. 2018, 13(1): 37-54 <https://dx.doi.org/10.11992/tis.201711034>

DOI: 10.11992/tis.202310039

网络出版地址: <https://link.cnki.net/urlid/23.1538.TP.20241223.1247.008>

基元潜部特征元挖掘的智能方法研究

张丽芳, 李兴森

(广东工业大学 可拓学与创新方法研究所, 广东 广州 510006)

摘要: 潜部特征元构建是可拓学基元理论的重要研究内容, 潜在信息挖掘对解决问题和激发创新思维至关重要。为了解决目前人工识别基元潜部特征元效率低、覆盖面窄和数量不足的问题, 研究基元潜部特征元显化理论与人工智能算法实现的结合点, 提出挖掘基元潜部特征元的流程化、系统性方法, 使用爬虫技术收集目标基元对象的相关信息, 清洗噪音数据并从句子中挖掘构成特征元的名称和描述, 用概率统计的定量方法筛选潜部特征元并通过 Python 代码实现智能挖掘功能, 最后通过案例对比分析进行论证。研究结果能有效提高基元潜部特征元的识别效率和智能化水平, 对从复杂多变的语料句中进行语义概括也有一定的参考作用, 为进一步提高特征名称及其量值智能提取的精确性积累训练集, 促进可拓展型人工智能理论的发展。

关键词: 可拓学; 潜部特征元; 特征元; 基元理论; 人工智能; 自然语言处理; 大语言模型; 可拓智能

中图分类号: TP18 **文献标志码:** A **文章编号:** 1673-4785(2025)02-0457-08

中文引用格式: 张丽芳, 李兴森. 基元潜部特征元挖掘的智能方法研究 [J]. 智能系统学报, 2025, 20(2): 457-464.

英文引用格式: ZHANG Lifang, LI Xingsen. Research on intelligent methods for latent features mining of basic element[J]. CAAI transactions on intelligent systems, 2025, 20(2): 457-464.

Research on intelligent methods for latent features mining of basic element

ZHANG Lifang, LI Xingsen

(Institute of Extenics and Innovation Methods, Guangdong University of Technology, Guangzhou 510006, China)

Abstract: Latent feature element construction is a key aspect of the basic-element theory of extenics, and mining latent information is crucial for problem solving and fostering innovative thinking. This study explores the integration of the basic-element latent feature element manifestation theory with artificial intelligence algorithms to address the current problems of low efficiency, narrow coverage and the insufficient number of manually identified basic-element latent feature elements. A process-oriented, systematic method for mining latent feature elements of basic elements is proposed. The method involves using crawler technology to collect relevant information regarding target basic-element objects, cleaning noisy data, and mining names and descriptions of constituent feature elements from sentences. A probability statistical approach is then used to filter latent feature elements, with the intelligent mining process implemented through Python code. Finally, a case study comparison is performed to demonstrate the effectiveness of this approach. Research results can notably improve the recognition efficiency and intelligence level of basic-element latent feature elements while also providing valuable insights for semantic generalization from complex and changeable dynamic corpus syntax. Additionally, it contributes to building a training set for enhancing the accuracy of intelligent extraction of feature names and their quantitative values, thus promoting the development of extensible artificial intelligence theory.

Keywords: extenics; latent feature element; feature element; basic-element theory; artificial intelligence; natural language processing; large language model; extension intelligence

收稿日期: 2023-10-27. 网络出版日期: 2024-12-23.

基金项目: 国家自然科学基金项目 (72071049); 广东省自然科学基金项目 (2024A1515011324).

通信作者: 李兴森. E-mail: lixingsen@126.com.

大数据时代人工智能技术的发展进一步促进了互联网、物联网的应用, 复杂多变交互的网络信息环境下, 需不断解决问题和持续创新。在此

背景下,与解决问题有关的事物潜在信息的挖掘和利用显得尤为重要,而互联网技术的发展在便利信息获取与交流的同时,也加剧了信息膨胀和结构复杂化,对信息处理和分析能力提出了更高要求。可拓学是以蔡文教授为首的中国学者创立的一门横断学科,阐明了显部与潜部等共轭部的概念,论述了共轭分析原理、可拓分析原理,是潜在信息挖掘的主要理论基础^[1-3]。可拓学将描述事物的潜在信息定义为潜部特征元,并形式化表达,为智能技术处理提供方法基础^[4-9]。潜部特征元在解决矛盾问题和研究创新规律方法中发挥关键作用,但其隐蔽性导致识别困难^[7]。爬虫技术、数据预处理技术、数据挖掘技术、自然语言处理技术和 ChatGPT(chat generative pre-trained Transformer) 等技术^[10-25] 的发展为可拓学的理论方法实现提供了技术支撑,人工智能技术与可拓学的交叉研究还在不断深入^[26-29]。现有人工智能技术还不能单独地直接应用于潜部特征元的挖掘。虽然 ChatGPT 能根据语义给出综合的结果,但回答内容的真实性具有不确定性。ChatGPT 回答内容中包含推断结果,不能给出来源网址或者给出的网址是错误的,而推断结果的真实性又难以直接考证,这容易误导使用者。潜在信息具有隐蔽性,难以通过直观感受或主观经验进行有效评估,其揭示往往依赖于考证与分析。同样,大数据背景下可拓学需要人工智能技术来实现获取潜部特征元的理论方法。为此,本文通过研究可拓学理论与人工智能技术的结合,提出在互联网大数据背景下高效可靠地挖掘基元潜部特征元的智能方法,以进一步促进可拓学与人工智能领域的交叉研究,助力矛盾问题求解和创新思维拓展。

1 挖掘潜部特征元的理论方法基础

根据可拓学的基元理论和共轭分析原理,基元的特征元可以划分为显部特征元和潜部特征元,显部特征元是容易观察到的特征元;潜部特征元是不易观察到的特征元^[1-3]。因此,可以运用发散树、相关网、蕴含系及分合链等方法拓展潜在信息。以发散分析为例,根据发散分析原理得到目标基元 B_g 的同对象的特征元集合,再通过筛选得出 B_g 的潜部特征元集合:

$$B_g = (O, c, v) \cup \{(O, c_i, v_i), i = 1, 2, \dots, n\} \quad (1)$$

目标基元 B_g 的同对象的特征元集合记作 C_g :

$$C_g = \{(c_i, v_i), i = 1, 2, \dots, n\} \quad (2)$$

对 C_g 进行筛选,得出 B_g 的潜部特征元集合,

记作 $C_{g,lt}$:

$$C_{g,lt} = \{(c_i, v_i), i = 1, 2, \dots, n\}, C_{g,lt} \subset C_g \quad (3)$$

挖掘潜部特征元的过程涉及大量的数据处理和分析工作,对于人工而言是巨大的挑战。实现大数据背景下的可拓智能创新^[9],从互联网大数据中挖掘基元的潜部特征元的方法基础如下:爬虫技术能快速并且自动化地从互联网上抓取描述事物的各种信息,包括潜在信息^[10-11];数据预处理技术能对包含潜在信息的各种原始信息进行清洗、转换和整理,以便更好地用于分析和挖掘^[12-14];数据挖掘技术能通过统计、事例推理、规则推理等方法从语料获得潜在信息识别的规则^[15-17];自然语言处理技术能比较好地理解人类语言,识别语句中的特定部分,从而能从语句中抽取并组织符合规则的潜部特征元^[18-21];ChatGPT 技术依托庞大的语料库以及对人类语言的理解能力,能辅助生成基元模型构建的语料及数据挖掘的匹配规则^[22-25]。

2 挖掘潜部特征元的智能方法

基于上述理论方法基础,制定挖掘潜部特征元的工作流程,如图1,总共有6个步骤。

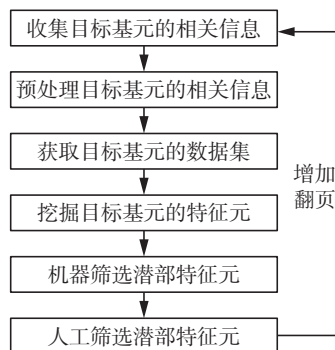


图1 潜部特征元挖掘的工作流程

Fig. 1 Workflow of latent features mining

本文2.1~2.6节依次论述了该流程方法的实现细节并展示核心代码,主要代码的实现基于式(1)~(3)。通过式(1)和(2)拓展出尽可能多的特征元,再通过式(3)筛选潜部特征元。

2.1 收集目标基元的相关信息

应用爬虫技术从互联网收集目标基元 B_g 的相关信息,把 B_g 的名称作为关键词在搜索引擎中搜索并翻页,解析结果页面,获得包含 B_g 相关信息的网页链接,再进一步访问这些链接获得包含 B_g 的网页内容,算法1是Python代码的关键部分。

算法1 收集目标基元的相关信息代码的关键部分

1) for page in range(1, search_count):

```

2) response = requests.get(url, headers=headers)
3) response.encoding = 'utf-8'
4) soup = BeautifulSoup(response.text, 'html.
parser')
5) site_tags_on_page = soup.find_all('cite')
6) site_tags.extend([tag.text for tag in tags_on_
page])

```

2.2 预处理目标基元的相关信息

通过上述步骤得到的网页内容包含 HTML 标签等非目标数据,需要进行预处理,包括正则替换干扰字符,清除脚本、样式、HTML 等标记,转换标点符号等,算法 2 是 Python 关键代码的实现。

算法 2 预处理目标基元的相关信息代码的关键部分

```

1) for script in body.find_all('script'):
2) script.extract()
3) for style in body.find_all('style'):
4) style.extract()
5) for html_tag in body.find_all(True):
6) html_tag.unwrap()
7) cleaned_html = re.sub(r'<body.*?>|</body>', "",
str(body), flags=re.DOTALL)
8) cleaned_html = re.sub(r'\r\n|\n\r', "", cleaned_
html)
9) cleaned_html = re.sub(r'\s+', "", cleaned_html)
10) p_period1 = r'(?<=[\u4e00-\u9fa5])\.(?=[
\u4e00-\u9fa5])'
11) text = re.sub(p_period1, '。', text)
12) p_period2 = r'(?<=[\u4e00-\u9fa5])\.(?=[
\s\n])$)'
13) text = re.sub(p_period2, '。', text)
14) p_comma1 = r'(?<=[\u4e00-\u9fa5]),(?=[
\u4e00-\u9fa5])'
15) text = re.sub(p_comma1, ',', text)
16) p_comma2 = r'(?<=[\u4e00-\u9fa5]),(?=[
\s\n])$)'
17) text = re.sub(p_comma2, ',', text)

```

2.3 获取目标基元的数据集

上述步骤得到的已清洗数据是大段的互联网文本,按句子进行切分处理,形成适合后续操作的数据集,算法 3 是 Python 代码的实现。

算法 3 获取目标基元的数据集代码的关键部分

```

1) sentences = re.split("[。！？；]", text)
2) f_sentences = [s.strip() for s in sentences if
topic in s]

```

2.4 挖掘目标基元的特征元

B_g 的数据集是由包含 B_g 的句子组成,1 个句子可能描述 B_g 的 1 个特征,也可能描述多个特征。

示例 1 B_{g1} =马铃薯,句子“马铃薯最常用的繁殖方式是无性块茎繁殖”,分析该句子,可得到 B_g 的 1 个特征 C_{g1} :

$$C_{g1} = \{(\text{繁殖方式,无性块茎繁殖})\}$$

示例 2 B_{g2} =马铃薯,句子“马铃薯块茎含有多种维生素和无机盐,可防止坏血病”,分析该句子,可得到 B_{g2} 的 2 个特征 C_{g2} :

$$C_{g2} = \{(\text{营养成分,维生素v无机盐}), (\text{功效,防止,支配对象,坏血病})\}$$

通过观察分析 B_g 的数据集,总结其中的模式和规律:句子中的动宾结构常常描述 B_g 的特征。因此,应用自然语言处理技术抽取句子中的动宾结构作为 B_g 的特征描述。本文使用了结巴分词工具来进行中文文本的分词处理^[20],使用了斯坦福自然语言处理工具包进行语言分析^[18]。算法 4 是 Python 代码的关键部分。

算法 4 通过自然语言处理技术挖掘目标基元的特征元代码的关键部分

```

1) seg_list = jieba.cut(sentence, cut_all=False,
HMM=True)
2) seg_str = ''.join(seg_list)
3) parsed_sentence = list(parser.raw_arse(seg_
str))[0]
4) vps = []
5) for subtree in parsed_sentence.subtrees(filter=
lambda t: t.label() == 'VP'):
6) if not any(child.label() == 'VP' for child in sub-
tree):
7) a_vp = ".join(subtree.leaves())

```

如上述示例 1 和示例 2,特征名称(比如“繁殖方式”“营养成分”“功效”)是基于领域知识对特征描述内容的概括总结。因此,可通过匹配规则获得特征名称或特征分类的名称。使用 ChatGPT 迅速检索与特定主题或概念相关的匹配规则和关键词。例如,收集描述用途的关键词和匹配规则。按照这种方式基于领域知识整合出多种匹配规则,包括颜色、形状、质地、营养成分、烹饪方法、生长条件、功能等。为了提高匹配的查全率,使用一些宽松的规则。那么,可能存在相同的匹配关键词,比如“温暖的黄色”,既有温度也有颜色的关键词,这容易产生误判,因此通过匹配次数提高准确率,即在同样的文本中,推荐匹配次数多的特征名称或特征分类的名称,算法 5 是 Py-

thon 代码的关键部分。

算法 5 通过规则匹配挖掘目标基元的特征元代码的关键部分

```
1) # 此处示例, 实际匹配规则有多个且内容更丰富
2) pattern_purpose = r"(用于|用来|适用于|专为.+?设计)"
3) keys=['用途']
4) name_counts = {}
5) patterns=[pattern_purpose]
6) for index, p in enumerate(patterns):
7) matches = re.findall(p,sentence)
8) m_len = len(matches)
9) if m_len>0:
10) name_counts[keys[index]] = m_len
11) if name_counts:
12) return max(name_counts, key=name_counts.get)
```

2.5 机器筛选潜部特征元

潜部特征元具有不易被发现的特征, 因而通过统计分析划分显部特征和潜部特征, 即统计不同特征描述的关键词 (假定名词这类实词) 在 B_g 数据集中出现的频次, 设定阈值 t , 通过多次更改阈值并观察划分结果以便选取恰当的阈值。当特征描述的关键词在 B_g 数据集中出现的频次小于阈值 t , 即认为该特征描述为待选潜部特征。算法 6 为计算机筛选潜部特征元的 Python 代码。

算法 6 机器筛选潜部特征元代码的关键部分

```
1) real_words = []
2) words = pseg.cut(features)
3) for word, flag in words:
4) if flag in ["n", "v"]:
5) real_words.append(word)
6) ws_count=0
7) for sentence in sentences:
8) w_count=0
9) for w in real_words:
10) if w in sentence:
11) w_count+=1
12) if w_count== len(real_words):
13) ws_count+=1
```

2.6 人工筛选潜部特征元

潜部特征元的识别具有一定的主观性, 有时需要通过人在待选潜部特征集合中进行甄别。待选潜部特征元列表包含“频次”“潜部判定”“特征名称或特征分类名称”“特征描述”“来源语料”和“来源网址”。当人工筛选之后不能获得足够的潜

部特征元, 可以增加翻页数量, 访问更多网页, 获得更多目标基元的相关信息。

3 案例对比分析

3.1 基于一般方法获得潜部特征元

假设需要获得“马铃薯”的潜部特征元。在大多数人的普遍观念中, 马铃薯是一种食物, 其一般显部特征元为

$$C_{ap} = \{(\text{颜色, 黄色}), (\text{形状, 椭圆形}), (\text{用途, 食品}), (\text{食用方式, \{烤, 煮, 炒\}}), (\text{主要成分, 淀粉}), (\text{生长特点, 高产})\}$$

进一步分析, 根据生活经验可知发芽的马铃薯有毒性 (显部), 深入了解得到潜部特征元的集合为

$$C_{it} = \{(\text{毒性来源, \{土豆芽\}}), (\text{中毒症状, \{恶心、呕吐、腹泻、发热等\}}), (\text{中毒处理方法, \{催吐、洗胃、药物治疗等\}}), (\text{预防措施, \{识别、避免食用土豆芽\}})\}$$

3.2 挖掘潜部特征元的智能方法

目标基元 B_g = 马铃薯, 按照本文“2 挖掘潜部特征元的智能方法”的论述挖掘马铃薯的潜部特征元。

3.2.1 收集目标基元的相关信息

利用 2.1 节的方法通过网页获取马铃薯的相关信息: “<div class="para MARK_MODULE" label-module="para" data-uuid="txYAXMQaE6xt" data-pid="1">马铃薯 (*Solanum tuberosum* L.) 是茄科茄属的一年生草本植物。地上茎呈菱形, 有毛。</div><div class="para MARK_MODULE" label-module="para" data-uuid="tOqYFzeoW4jU" data-pid="2">转基因土豆: 表面光滑。<sup class="sup--normal" data-sup="18" data-ctrmap=":18,">”。

3.2.2 预处理目标基元的相关信息

利用 2.2 节的技术方法来预处理马铃薯的相关信息: “马铃薯因酷似马铃铛而得名。马铃薯原产热带美洲的山地, 16 世纪传到印度, 继而传到中国, 现广泛种植于全球温带地区。马铃薯喜冷凉干燥气候, 适应性较强, 以疏松肥沃沙质土为宜, 生长周期短而产量高。因种子繁殖会导致性状分离, 所以马铃薯最常用的繁殖方式是无性

块茎繁殖。马铃薯味甘,性平。归胃、大肠经。有益气、健脾、和胃、解毒、消肿等功效。皮色发青或发芽的马铃薯因含过量龙葵素,有毒而不能食用。《本草纲目》中记载了马铃薯可以治疗病后脾胃虚寒,气短乏力。马铃薯块茎含有多种维生素和无机盐,可防止坏血病,刺激造血机能;无机盐对人的健康和幼儿发育成长都是不可缺少的。”

3.2.3 获取目标基元的数据集

通过2.3节的技术方法获取马铃薯的数据集,表1给出了部分数据。

表1 马铃薯的部分数据集
Table 1 Part of the potato dataset

序号	语料
1	马铃薯因酷似马铃薯而得名。
2	《本草纲目》中记载了马铃薯可以治疗病后脾胃虚寒,气短乏力。
3	马铃薯块茎含有多种维生素和无机盐,可防止坏血病,刺激造血机能。
4	马铃薯块茎含有多种维生素和无机盐,可防止坏血病,刺激造血机能。
5	随着西班牙征服印加帝国,马铃薯在16世纪下半叶被企喇承西班牙人带回到欧洲传播开来。
6	马铃薯含钾量丰富,100克中含502毫克,有高钾蔬菜之称。
7	胡萝卜素含量也较高,约为30 mg,此外马铃薯块茎还含有丰富的B族维生素。
8	目前,紫色马铃薯品种有紫洋、紫玫瑰等,红色马铃薯品种有红美、红云等。
9	定西马铃薯是甘肃省定西市特产,中国国家地理标志产品。
10	这个地区生产的马铃薯主要用于鲜食消费。

3.2.4 挖掘目标基元的特征元

通过2.4节的技术方法挖掘马铃薯的特征描述,表2给出了部分数据。

表2 马铃薯的部分特征描述
Table 2 Part of the features of the potato

序号	特征描述
1	酷似马铃薯
2	可以治疗病后脾胃虚寒
3	含有多种维生素和无机盐
4	防止坏血病
5	被企喇承西班牙人带回到欧洲传播开来
6	有高钾蔬菜之称
7	含有丰富B族维生素
8	紫色马铃薯品种
9	是甘肃省定西市特产
10	用于鲜食消费

通过2.4节的技术方法从马铃薯的特征描述

中挖掘特征名称或特征分类的名称,如表3所示。

表3 马铃薯的部分特征描述及其名称
Table 3 Part of the features with names of the potato

序号	名称	特征描述
1	得名原因	酷似马铃薯
2	功效	可以治疗病后脾胃虚寒
3	营养成分	含有多种维生素和无机盐
4	功效	防止坏血病
5	传播历程	被企喇承西班牙人带回到欧洲传播开来
6	营养成分	有高钾蔬菜之称
7	营养成分	含有丰富B族维生素
8	颜色	紫色马铃薯品种
9	文化	是甘肃省定西市特产
10	用途	用于鲜食消费

3.2.5 机器筛选潜部特征元

通过2.5节的技术方法挖掘马铃薯的特征描述在数据集中的频次,设定阈值(比如 $t=3$)机器判定待选潜部特征元。表4是马铃薯的部分特征元标记表。

表4 马铃薯的部分特征元标记
Table 4 Part of the feature elements of the potato

序号	频次	潜部	名称	特征描述
1	1	是	得名原因	酷似马铃薯
2	1	是	功效	可以治疗病后脾胃虚寒
3	2	是	营养成分	含有多种维生素和无机盐
4	2	是	功效	防止坏血病
5	1	是	传播历程	企喇承西班牙人带到欧洲
6	3	否	营养成分	高钾蔬菜之称
7	7	否	营养成分	含有丰富B族维生素
8	3	否	颜色	紫色马铃薯品种
9	1	是	文化	是甘肃省定西市特产
10	4	否	用途	用于鲜食消费

3.2.6 人工筛选潜部特征元

提供马铃薯的待选潜部特征元列表以供人工筛选,实验中访问了8个网页,获得超过100个潜部特征描述,表5给出了部分数据。

实验中该方法找到了325条关于“马铃薯”的语料,得到102条潜部特征元(定义为 s)。代码运行时长13 min(定义为 t_1),人工筛选时长60 min(定义为 t_2)。因此挖掘潜部特征元的效率 v (单位:条/min)的计算公式为

$$v = \frac{s}{(t_1 + t_2)} = 1.4$$

表5 马铃薯的部分待选潜部特征元
Table 5 Part of the feature elements to be selected for potato

序号	频次	潜部	名称	特征描述	语料	来源网址
1	1	是	得名原因	酷似马铃薯	马铃薯因酷似马铃薯而得名。	https://baike.baidu.com/item/马铃薯/416928
2	1	是	功效	可以治疗病后脾胃虚寒	《本草纲目》中记载了马铃薯可以治疗病后脾胃虚寒, 气短乏力。	https://baike.baidu.com/item/马铃薯/416928
3	2	是	营养成分	含有多种维生素和无机盐	马铃薯块茎含有多种维生素和无机盐, 可防止坏血病, 刺激造血机能。	https://baike.baidu.com/item/马铃薯/416928
4	2	是	功效	防止坏血病	马铃薯块茎含有多种维生素和无机盐, 可防止坏血病, 刺激造血机能。	https://baike.baidu.com/item/马铃薯/416928
5	1	是	传播历程	被企喇承西班牙人带回到欧洲传播开来	随着西班牙征服印加帝国, 马铃薯在16世纪下半叶被企喇承西班牙人带回到欧洲传播开来。	https://baike.baidu.com/item/马铃薯/416928
6	3	否	营养成分	有高钾蔬菜之称	马铃薯含钾量丰富, 100克中含502毫克, 有高钾蔬菜之称。	https://www.zhihu.com/question/622443212
7	7	否	营养成分	含有丰富B族维生素	胡萝卜素含量也较高, 约为30mg, 此外马铃薯块茎还含有丰富的B族维生素。	https://zhuanlan.zhihu.com/p/20667272
8	3	否	颜色	紫色马铃薯品种	目前, 紫色马铃薯品种有紫洋、紫玫瑰等, 红色马铃薯品种有红美、红云等。	https://zhuanlan.zhihu.com/p/20667272
9	1	是	文化	是甘肃省定西市特产	定西马铃薯是甘肃省定西市特产, 中国国家地理标志产品。	https://baike.baidu.com/item/中国薯都/10385149
10	4	否	用途	用于鲜食消费	这个地区生产的马铃薯主要用于鲜食消费。	https://zhuanlan.zhihu.com/p/426871804

每一条语料都有来源网址, 即百度百科、知乎、中国科学院官网等。因此, 内容切实可靠, 准确性比较高。

现有方法主要是人工识别并辅助搜索引擎手动检索等。因为人工识别过程通常依赖于人的主观判断和认知能力, 这导致了识别效率的局限性。人的认知能力有限, 容易受到疲劳、注意力分散等因素的影响, 从而降低识别速度和准确性。因此, 该研究的方法在效率、数量以及准确性上均优于现有方法。

上述案例用一般方法获得目标基元的4个潜部特征元, 智能方法访问8个网页获得超过100个潜部描述。不难推测, 如果访问更多网页, 能获得更多潜部特征元。本智能方法能获取更多的潜部特征元, 效率更高, 系统性更好。可拓学与人工智能两者的交叉研究是一个非常具有前景的领域^[30-31]。

4 结束语

论文扩充了可拓学与人工智能的交叉研究范围, 形成一套挖掘基元潜部特征元的流程化智能化方法, 并编码实现相应功能。该算法具有速度快、准确性高的优势, 可高效并大量地挖掘事物背后的潜在信息, 辅助解决问题或者激发创新思

维。语料句法的多变性与复杂性以及语义概括的难点, 导致该研究具有一定局限性, 建议可以进一步结合知识图谱、大语言模型技术和可拓集合等理论, 深入研究精确提取特征名称及其量值的方法。

参考文献:

- [1] 杨春燕, 蔡文, 汤龙. 可拓学[M]. 北京: 科学出版社, 2024.
YANG Chunyan, CAI Wen, TANG Long. Extension[M]. Beijing: Science Press, 2024.
- [2] 杨春燕. 可拓创新方法[M]. 北京: 科学出版社, 2018.
YANG Chunyan. Extension innovation method[M]. Beijing: Science Press, 2018.
- [3] 蔡文, 杨春燕. 可拓学的基础理论与方法体系[J]. 科学通报, 2013, 58(13): 1190-1199.
CAI Wen, YANG Chunyan. Basic theory and methodology on extenics[J]. Chinese science bulletin, 2013, 58(13): 1190-1199.
- [4] 杨春燕, 蔡文, 涂序彦. 可拓学的研究、应用与发展[J]. 系统科学与数学, 2016, 36(9): 1507-1512.
YANG Chunyan, CAI Wen, TU Xuyan. Research, application and development on extenics[J]. Journal of systems science and mathematical sciences, 2016, 36(9): 1507-1512.

- [5] 杨春燕, 蔡文. 可拓学与矛盾问题智能化处理[J]. 科技导报, 2014, 32(36): 15–20.
YANG Chunyan, CAI Wen. Extenics and intelligent processing of contradictory problems[J]. Science & technology review, 2014, 32(36): 15–20.
- [6] 杨春燕, 李兴森. 可拓创新方法及其应用研究进展[J]. 工业工程, 2012, 15(1): 131–137.
YANG Chunyan, LI Xingsen. Research progress in extension innovation method and its applications[J]. Industrial engineering journal, 2012, 15(1): 131–137.
- [7] 李兴森, 洪振挺, 王昊, 等. 可拓学与知识管理交叉研究视角下的问题智能化处理[J]. 包装工程, 2021, 42(12): 51–58, 10.
LI Xingsen, HONG Zhenting, WANG Hao, et al. Intelligent problem solving from the perspective of extenics and knowledge management[J]. Packaging engineering, 2021, 42(12): 51–58, 10.
- [8] 余志伟, 李兴森. 基元库构建模型及其应用研究[J]. 广东工业大学学报, 2015, 32(3): 5–9.
YU Zhiwei, LI Xingsen. Modeling of basic-element and its application[J]. Journal of Guangdong University of Technology, 2015, 32(3): 5–9.
- [9] 李兴森, 李爱红, 施美. 大数据背景下的可拓智能创新[J]. 中国质量, 2018(09): 42–46.
LI Xingsen, LI Aihong, SHI Mei. Extension intelligent innovation in the context of big data[J]. China quality, 2018(09): 42–46.
- [10] 蔡迪阳. 基于 Python 的网页信息爬取技术分析[J]. 科技资讯, 2023, 21(13): 31–34.
CAI Diyang. Analysis of web information crawling technology based on Python[J]. Science & technology information, 2023, 21(13): 31–34.
- [11] 曾健荣, 张仰森, 郑佳, 等. 面向多数据源的网络爬虫实现技术及应用[J]. 计算机科学, 2019, 46(5): 304–309.
ZENG Jianrong, ZHANG Yangsen, ZHENG Jia, et al. Implementation technology and application of web crawler for multi-data sources[J]. Computer science, 2019, 46(5): 304–309.
- [12] 赵海燕, 李欣歌, 陈庆奎, 等. 面向业务过程挖掘和分析的事件日志预处理技术[J]. 小型微型计算机系统, 2022, 43(1): 1–9.
ZHAO Haiyan, LI Xinge, CHEN Qingkui, et al. Event log preprocessing technology for business process mining and analysis[J]. Journal of Chinese computer systems, 2022, 43(1): 1–9.
- [13] 李洪烈, 夏栋, 王倩. 基于回归模型的采集数据清洗技术[J]. 电光与控制, 2022, 29(4): 117–120.
LI Honglie, XIA Dong, WANG Qian. A sampled data cleaning technology based on regression model[J]. Electronics optics & control, 2022, 29(4): 117–120.
- [14] ARAVIND PRAKASH M, INDRA GANDHI K, SRIRAM R, et al. An effective comparative analysis of data preprocessing techniques in network intrusion detection system using deep neural networks[M]//Advances in Parallel Computing. Amsterdam: IOS Press, 2021: 14–19.
- [15] LI Xingsen, ZHANG H, ZHU Zhengxiang, et al. An intelligent transformation knowledge mining method based on extenics[J]. Journal of Internet technology, 2013, 14: 315–325.
- [16] SHI Yong, ZHANG Lingling, TIAN Yingjie, et al. Foundations of intelligent knowledge management[M]//SpringerBriefs in Business. Berlin: Springer, 2015: 13–30.
- [17] SHAH K, SHAH N, SAWANT V, et al. Practical data mining techniques and applications[M]. New York: Auerbach Publications, 2023.
- [18] MANNING C, SURDEANU M, BAUER J, et al. The stanford CoreNLP natural language processing toolkit[C]//Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations. Baltimore: Association for Computational Linguistics, 2014: 55–60.
- [19] 倪渊, 廖世豪, 张健. 基于 Wobert 与对抗学习的中文命名实体识别[J]. 计算机工程, 2024, 50(11): 119–129.
NI Yuan, LIAO Shihao, ZHANG Jian. Chinese named entity recognition based on wobert and adversarial learning[J]. Computer engineering, 2024, 50(11): 119–129.
- [20] 结巴分词研发团队. “结巴”中文分词: 做最好的 Python 中文分词组件[EB/OL]. (2020–02–15)[2023–10–25]. <https://github.com/fxsjy/jieba>.
JieBa Participle R&D Team. "JieBa" Chinese segmentation: being the best Python Chinese segmentation component[EB/OL]. (2020–02–15)[2023–10–25]. <https://github.com/fxsjy/jieba>.
- [21] 赵怡博, 蒋峰, 李培峰. 一种基于 BERT 的多级连贯性文本分割方法[J]. 计算机应用与软件, 2024, 41(10): 262–268, 324.
ZHAO Yibo, JIANG Feng, LI Peifeng. A BERT-based hierarchical adjacent coherence text segmentation method[J]. Computer applications and software, 2024, 41(10): 262–268, 324.
- [22] 李景玉. 预训练语言模型探究[J]. 科技资讯, 2022, 20(19): 5–9, 18.
LI Jingyu. On the pre training language models[J]. Science & technology information, 2022, 20(19): 5–9, 18.
- [23] 江怡, 董化文. 论人工智能与人类智能的双向互动[J]. 自然辩证法通讯, 2023, 45(11): 14–25.
JIANG Yi, DONG Huawen. On the bi-directional interac-

- tion of artificial intelligence and human intelligence[J]. Journal of dialectics of nature, 2023, 45(11): 14–25.
- [24] 朱光辉, 王喜文. ChatGPT 的运行模式、关键技术及未来图景[J]. 新疆师范大学学报 (哲学社会科学版), 2023, 44(4): 113–122.
- ZHU Guanghui, WANG Xiwen. ChatGPT: operation mode, key technology and future prospects[J]. Journal of Xinjiang Normal University (edition of philosophy and social sciences), 2023, 44(4): 113–122.
- [25] 钱力, 刘熠, 张智雄, 等. ChatGPT 的技术基础分析[J]. 数据分析与知识发现, 2023, 7(3): 6–15.
- QIAN Li, LIU Yi, ZHANG Zhixiong, et al. An analysis on the basic technologies of ChatGPT[J]. Data analysis and knowledge discovery, 2023, 7(3): 6–15.
- [26] 蔡国梁, 罗伟, 康丙欣. 可拓学与人工神经网络[J]. 天中学刊, 1998, 13(5): 9–12.
- CAI Guoliang, LUO Wei, KANG Bingxin. Extensics and artificial neural network[J]. Journal of Tianzhong, 1998, 13(5): 9–12.
- [27] 何斌, 张应利. 可拓学在人工智能中的应用初探[J]. 华南理工大学学报 (自然科学版), 1999, 27(6): 88–92.
- HE Bin, ZHANG Yingli. Primary research for extenics' application in artificial intelligence[J]. Journal of South China University of Technology (natural science edition), 1999, 27(6): 88–92.
- [28] 杨春燕, 蔡文. 可拓数据挖掘研究进展[J]. 数学的实践与认识, 2009, 39(4): 134–141.
- YANG Chunyan, CAI Wen. Recent progress in extension data mining[J]. Mathematics in practice and theory, 2009, 39(4): 134–141.
- [29] 李兴森, 石勇, 李爱华. 基于可拓集的企业数据挖掘应用方案初探[J]. 哈尔滨工业大学学报, 2006, 38(7): 1124–1128.
- LI Xingsen, SHI Yong, LI Aihua. Study on enterprise data mining solution based on extension set[J]. Journal of Harbin Institute of Technology, 2006, 38(7): 1124–1128.
- [30] 蔡文, 石勇. 可拓学的科学意义与未来发展[J]. 哈尔滨工业大学学报, 2006, 38(7): 1079–1086.
- CAI Wen, SHI Yong. Extenics: its significance in science and prospects in application[J]. Journal of Harbin Institute of Technology, 2006, 38(7): 1079–1086.
- [31] 李兴森, 许立波, 刘海涛, 等. 因素空间与可拓学的互补性分析及问题处理融合模型[J]. 智能系统学报, 2022, 17(5): 990–998.
- LI Xingsen, XU Libo, LIU Haitao, et al. A complementary analysis of factor space and extenics and a fusion model for solving problems[J]. CAAI transactions on intelligent systems, 2022, 17(5): 990–998.

作者简介:



张丽芳, 软件研发工程师, 主要研究方向为信息系统与电子商务。
E-mail: f200553042@126.com。



李兴森, 教授, 博士, 中国人工智能学会理事, 中国创造学会理事, 中国人工智能学会可拓学专业委员会主任, 主要研究方向为可拓学、知识管理与可拓智能创新。发表学术论文 80 余篇。E-mail: lixingsen@126.com。