

## 基于DS理论的多模态信息抽取方法

王剑, 王京岭, 张革, 王章全, 郭世远, 庾桂铭

### 引用本文

王剑, 王京岭, 张革, 王章全, 郭世远, 庾桂铭. [基于DS理论的多模态信息抽取方法](#)[J]. 计算机科学, 2025, 52(10): 208-216.

WANG Jian, WANG Jingling, ZHANG Ge, WANG Zhangquan, GUO Shiyuan, YU Guiming. [Multimodal Information Extraction Fusion Method Based on Dempster-Shafer Theory](#) [J]. Computer Science, 2025, 52(10): 208-216.

---

## 相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

### Similar articles recommended (Please use Firefox or IE to view the article)

#### [基于双流深度学习的Dockerfile安全误配置检测方法](#)

Dual-stream Feature Fusion Approach for Dockerfile Security Misconfiguration Detection  
计算机科学, 2025, 52(10): 395-403. <https://doi.org/10.11896/jsjcx.241000014>

#### [一种新的基于凸损失函数的离散扩散文本生成模型](#)

Novel Discrete Diffusion Text Generation Model with Convex Loss Function  
计算机科学, 2025, 52(10): 231-238. <https://doi.org/10.11896/jsjcx.240800147>

#### [基于Retinex理论的低照度图像自适应增强算法](#)

Low Light Image Adaptive Enhancement Algorithm Based on Retinex Theory  
计算机科学, 2025, 52(10): 168-175. <https://doi.org/10.11896/jsjcx.240800057>

#### [基于神经辐射场的即时高保真人脸生成算法](#)

Immediate Generation Algorithm of High-fidelity Head Avatars Based on NeRF  
计算机科学, 2025, 52(10): 159-167. <https://doi.org/10.11896/jsjcx.241000066>

#### [基于分步协作融合表示的情感分类方法](#)

Sentiment Classification Method Based on Stepwise Cooperative Fusion Representation  
计算机科学, 2025, 52(9): 313-319. <https://doi.org/10.11896/jsjcx.240700161>

# 基于 DS 理论的多模态信息抽取方法

王 剑<sup>1</sup> 王京岭<sup>2</sup> 张 革<sup>1</sup> 王章全<sup>1</sup> 郭世远<sup>2</sup> 庾桂铭<sup>1</sup>

1 郑州大学计算机与人工智能学院 郑州 450000

2 郑州大学网络空间安全学院 郑州 450000

**摘 要** 在过去多模态信息抽取(Multimodal Information Extraction, MIE)任务中,研究人员通常使用数据层融合的方式训练用于 MIE 的神经网络模型。但是,由于不同模态间的异构性,这种融合方式容易导致特征冗余、特征不兼容和缺乏解释性等问题,进而影响模型训练的效果。针对此问题,提出了一种基于 Dempster-Shafer(DS)理论的决策层融合方法。该方法通过神经网络和狄利克雷函数处理不同模态特征生成证据,经证据修正和权重分配后,利用 Shafer 融合规则得出最终决策,有效提升了特征处理的准确性和模型的可解释性。采用精确率、召回率和 F1 分数作为评价指标,在公开和私有数据集上的实验结果表明,相较于现有方法,所提方法的总体性能提升了 0.22 个百分点到 1.87 个百分点。

**关键词:** 信息抽取;多模态;Dempster-Shafer 理论;深度学习;证据修正;决策融合

**中图分类号** TP391

## Multimodal Information Extraction Fusion Method Based on Dempster-Shafer Theory

WANG Jian<sup>1</sup>, WANG Jingling<sup>2</sup>, ZHANG Ge<sup>1</sup>, WANG Zhangquan<sup>1</sup>, GUO Shiyuan<sup>2</sup> and YU Guiming<sup>1</sup>

1 School of Computer Science and Artificial Intelligence, Zhengzhou University, Zhengzhou 450000, China

2 School of Cyber Science and Engineering, Zhengzhou University, Zhengzhou 450000, China

**Abstract** In the past MIE tasks, researchers usually use data layer fusion to train neural network models for MIE. However, due to the heterogeneity among different modalities, this fusion approach can lead to issues such as feature redundancy, incompatibility, and lack of interpretability, which in turn affect the effectiveness of model training. In view of this, this paper proposes a decision-level fusion method based on the DS theory to solve the problems of feature redundancy, incompatibility, and lack of interpretability caused by data layer fusion. The evidence is generated by processing different modal features through neural networks and Dirichlet functions, and after evidence correction and weight assignment, the Shafer fusion rule is utilized to arrive at the final decision. This method effectively improves the accuracy of feature processing and the interpretability of the model. Using accuracy, recall, and F1 score as evaluation metrics, experiments on public and private datasets show an overall performance improvement of 0.22 to 1.87 percentage points compared to existing methods.

**Keywords** Key information extraction, Multimodality, Dempster-Shafer theory, Deep learning, Evidentiary amendments, Decision fusion

## 1 引言

信息抽取(Information Extraction, IE)<sup>[1]</sup>是自然语言处理(Natural Language Processing, NLP)领域的一项关键任务,旨在自动地从半结构化文本中提取出结构化、有意义的信息和知识,这些信息在待抽取的文档中被视为关键信息。MIE<sup>[2]</sup>则是在 IE 的基础上,引入了除文本以外的模态,如布局 and 图像,以辅助抽取任务。随着互联网和数字化时代的发展,海量的文档数据不断涌现,其中蕴含着大量有价值的信息,例如新闻、公告、发票等(见图 1)。这些数据的规模和复杂性使得人工处理和分析变得不可行,因此自动化的信息

抽取技术变得尤为重要。

针对多模态信息抽取任务,研究人员通常采用深度学习方法将文本信息、布局信息和视觉信息进行融合,促使模型学习不同模态间的特征。这种融合通常发生在模型的输入层,通过拼接或组合文本、布局和图像特征,引导模型学习模态间的潜在联系。这种方法虽然在提高多模态任务的性能方面已取得显著成果,但也可能引发一些潜在问题。

1) 忽视不同特征之间可能存在的影响。在许多抽取任务中,特征间的关联性可能产生负面影响,从而降低模型的性能。例如,文本信息中的某些关键词可能与图像中的特定区域相对应,而这种关联性可能被简单地拼接或加减操作所

基金项目:国家自然科学基金(61972133);河南省重点研发专项(241111212700);信息网络安全公安部重点实验室开放课题(C23600-04)

This work was supported by the National Natural Science Foundation of China(61972133), Key Research and Development Program of Henan Province(241111212700) and Open Subjects of Key Laboratory of the Ministry of Public Security for Information Network Security(C23600-04).

通信作者:王剑(iejwang@zzu.edu.cn)

忽略,甚至引入不必要的噪声,干扰原本特征<sup>[3-4]</sup>。

2) 缺乏清晰的模态关系建模。输入层融合方法通常将不同模态特征进行简单的拼接后送入模型输入层,这可能导致模型难以准确建模模态间的复杂关系<sup>[5]</sup>。模态之间的相互影响和依赖关系可能被模糊化,使得模型在做出预测时难以解释其背后的决策过程,从而无法对模型做出针对性的优化以提高模型的精确率<sup>[6]</sup>。例如,在一个融合了图像和文本的信息抽取任务中,模型可能会在某个隐藏层上融合两种模态的特征,但由于缺乏明确的关系建模,研究人员无法解释模型是如何根据文本和图像之间的特定关系做出决策的。

3) 加剧硬件资源的消耗。多模态融合方法通常需要处理来自不同模态的大量数据,并且在模型的不同层次上进行复杂的计算和操作。这要求更强大的硬件资源以支持模型训练和推理过程,导致大量的资源消耗。例如,相较于单模态,多模态模型的训练需要更大的内存空间来存储训练数据集,以及更高的计算能力来处理复杂的模态融合过程。此外,在推理阶段,多模态模型需要更快的计算速度和更大的内存容量,以满足实时应用的需求。



图1 公告和发票类文档

Fig.1 Announcement and invoice type documents

为了解决上述 3 个问题,本文提出了一种新的可解释性较强的融合方法用于多模态信息抽取任务。主要研究工作与成果如下:

1) 将 DS 理论引入信息抽取任务中,通过将文本特征和布局特征分别转化为 DS 理论框架下的证据形式进行推理决策,避免模态干扰问题。

2) 提出了一种基于置信度的距离矩阵计算方法,用于矫正 DS 理论中证据过度自信问题。

3) 通过观察决策过程中的信度函数与似然函数来清晰地了解决策的过程,提高任务的可解释性。

4) 在一个公开数据集和一个私有数据集上进行实验,结果显示所提方法在精确率、召回率和 F1 分数 3 项指标上总体优于现有的多模态信息抽取技术。

本文第 2 章回顾了信息抽取领域的最新进展;第 3 章介绍了本文方法的理论基础;第 4 章阐述了本文方法;第 5 章通过在公开数据集和自编标注数据集上的评估,详细验证了本文方法的有效性;最后总结全文。

## 2 相关工作

IE 旨在通过分析文本并识别语义、定义实体及其关系的方式从非结构化文本数据中提取信息并将其转化为结构化数据。通过 IE 方法将结构化的信息从非结构化文档中提取出来,使得原本散乱的文本信息变得有序和易于分析。

1) 基于单模态的方法:在早期的工作中,研究人员主要采用单模态的方法进行信息抽取。这些方法所使用的技术主要包括基于规则的方法<sup>[7-12]</sup>、基于启发式的方法<sup>[13-15]</sup>和基于神经网络的方法<sup>[16-18]</sup>。基于规则的方法依赖于领域专家事先定义的规则和模式;基于启发式的方法通过考虑文本结构、语言模式等启发性信息来提取关键信息;基于神经网络的方法借助神经网络强大的学习和表达能力对文本进行信息提取。以上方法虽然一定程度上可以实现信息抽取任务,但在处理结构复杂的文本或多模态任务时,泛化能力有限,且对数据质量敏感。

2) 基于多模态的方法:多模态方法融合了文本、布局和视觉等多种信息维度,并通过神经网络来实现信息抽取。这类方法自动化程度高、泛化能力强,在多种任务上表现出色。例如,2018 年,Zhang 等<sup>[19]</sup>通过改良的 BiLSTM 模型,在推特数据集上取得了显著效果。文献<sup>[20-21]</sup>采用图神经网络挖掘文本与布局的内在联系,有效解决了半结构化文档中的实体识别与关系提取问题。Yu 等<sup>[22]</sup>在 2020 年提出 PICK 模型,该模型结合了 Transformer、CNN 和图注意力机制,在 SROIE2019<sup>[23]</sup>数据集上获得了最佳性能。与 PICK 不同,LayoutLM<sup>[24]</sup>,LayoutLMv2<sup>[25]</sup>和 LayoutXLM<sup>[26]</sup>模型使用多模态预训练的方法,通过在 Transformer 上使用文本、布局和视觉模态进行预训练,最终在多个信息抽取任务中获得了最佳性能。Sui 等<sup>[27]</sup>将声学信息通过跨模态注意模块集成到文本表示中,使用声音辅助文本进行信息抽取,最终获得了较好的效果。

虽然上述方法在不同信息抽取任务上取得了较好的效果,但都没有考虑到数据层直接融合可能出现的特征不兼容和缺乏解释性等问题,这给后续预测结果的分析 and 模型优化带来了很大的影响。除此之外,LayoutLM 等模型所包含的参数量级较大,因此训练所花费的时间和硬件资源较高。Han 等<sup>[28]</sup>在 2022 年提出了可信多视图分类的方法,将 DS 理论<sup>[29-30]</sup>用于图片分类任务中,在提高模型可解释性的同时验证了该方法在分类精度方面的有效性。然而,对于证据过度自信问题,该模型并没有给出很好的解决办法。

基于此,本文提出一种以 DS 证据理论框架为基础的方法,用于改进信息抽取任务,并通过实验验证了其有效性。

## 3 DS 理论基础

DS 理论是一种用于处理不确定性推理和决策的数学框架,它提供了一种有效的方法来处理不完全或冲突的证据,以推断出关于不确定性的合理结论。例如,将多个主体(如专家的判断、传感器的数据、分类器输出结果等)的证据相融合,从而形成最终的决策结果。具体框架如图 2 所示。

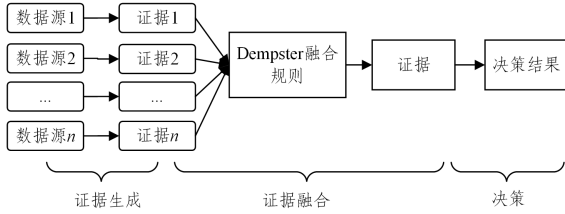


图2 基于DS理论证据的融合方法

Fig 2 DS evidence theory-based evidence fusion approach

### 3.1 识别框架

在DS理论中,将所有命题组成的样本空间称为识别框架,记作 $\theta$ 。其中, $A$ 被称为识别框架的事件,即识别框架的事件之间相互排斥且可以构成完备集的基本假设。识别框架可以表示为:

$$\theta = \{A_1, A_2, A_3, \dots, A_n\} \quad (1)$$

$\theta$ 的所有子集构成的集合称为其幂集 $\Omega$ ,记为 $\Omega = 2^\theta$ 。 $\Omega$ 可以表示为:

$$\Omega = \{\emptyset, \{A_1\}, \{A_2\}, \dots, \{A_n\}, \{A_1, A_2\}, \dots, \theta\} \quad (2)$$

其中, $\emptyset$ 表示空集。

### 3.2 基本概率分配

基本概率分配(Basic Probability Assignment, BPA)是证据的数学表达,是定义在识别框架的幂集上的一个函数,主要用于证据生成阶段。它将每个子集映射到区间 $[0, 1]$ 上的一个值,这个值表示对该子集的信任程度,即分配给该子集的概率。基本概率分配函数用 $m(x)$ 表示。 $m(x)$ 反映了事件置信度,它满足以下条件:

$$m(\emptyset) = 0 \quad (3)$$

$$\sum_{i=1}^n m(A_i) = 1 \quad (4)$$

### 3.3 证据融合规则

证据融合规则用于合并来自不同源的BPA,生成一个综合的概率分配,以更准确地反映整体证据。对于任意 $n$ 个属于同一识别框架时间的基本概率分配函数,可以使用Dempster规则进行融合,如式(5)所示:

$$(m_1 \oplus m_2 \oplus m_3 \oplus \dots \oplus m_n)(A) = \frac{\sum_{A_1 \cap A_2 \cap A_3 \cap \dots \cap A_n = A} m_1(A_1) m_2(A_2) m_3(A_3) \dots m_n(A_n)}{1 - K} \quad (5)$$

其中, $K$ 为归一化因子,计算式如下:

$$K = 1 - \sum_{A_1 \cap A_2 \cap \dots \cap A_n = \emptyset} m_1(A_1) m_2(A_2) \dots m_n(A_n) \quad (6)$$

### 3.4 可解释性

通过DS理论得出决策结果的推理过程清晰合理,具有一定的可解释性。具体示例如下。

假设给出3条证据 $\{E_1, E_2, E_3\}$ ,其识别框架为 $\theta = \{A_1, A_2, A_3\}$ ,则BPA如表1所列。

表1 BPA示例

Table 1 Example of BPA

证据	$A_1$	$A_2$	$A_3$
$E_1$	0.25	0.25	0.50
$E_2$	0.70	0.15	0.15
$E_3$	0.40	0.30	0.30

利用DS理论对上述内容进行合并,过程如下:

$$(m_{E_1} \oplus m_{E_2} \oplus m_{E_3})(A_1) = \frac{1}{K} (0.25 \times 0.7 \times 0.4) = 0.6747 \quad (7)$$

$$(m_{E_1} \oplus m_{E_2} \oplus m_{E_3})(A_2) = \frac{1}{K} (0.25 \times 0.15 \times 0.3) = 0.1084 \quad (8)$$

$$(m_{E_1} \oplus m_{E_2} \oplus m_{E_3})(A_3) = \frac{1}{K} (0.5 \times 0.15 \times 0.3) = 0.2169 \quad (9)$$

其中, $K$ 的计算过程及结果如下:

$$\begin{aligned} K &= (0.25 \times 0.7 \times 0.4) + (0.25 \times 0.15 \times 0.3) + (0.5 \times 0.15 \times 0.3) \\ &= 0.07 + 0.01125 + 0.0225 \\ &= 0.10375 \end{aligned} \quad (10)$$

通过上述融合过程,得到最终的BPA为 $\{0.6747, 0.1084, 0.2169\}$ ,取最大值对应的事件为最终结果,即 $A_1$ 。从表1中可知,证据 $E_1$ 更倾向于 $A_3$ ;证据 $E_2$ 对 $A_1$ 的支持度更高;证据 $E_3$ 虽然较为模糊,但其对 $A_1$ 的支持度更大。可见融合后,最终BPA更倾向于 $A_1$ 为最终结果。因此,通过DS理论框架可以清晰了解结果的产生过程。

## 4 本文方法

本章详细阐述了本文方法。其工作流程如图3所示,主要包括证据生成、证据修正、权重分配、证据融合以及数据后处理5个部分。根据本文信息抽取任务的设定,该任务的识别框架为 $\theta = \{A_1, A_2, A_3, \dots, A_n\}$ ,其中 $A_1$ 到 $A_n$ 分别代表 $n$ 个互斥的分类。

1) 证据生成:结合文本和布局特征,分别通过神经网络和狄利克雷分布生成证据,并形成证据集。其中,文本特征通过SentenceBert<sup>[31]</sup>编码和BiLSTM推理,从而获得不同标签的分类概率,即文本特征对应证据的BPA。布局特征采用狄利克雷分布建模训练数据集中位置信息与分类标签之间的关系,并利用狄利克雷分布的概率密度函数生成布局特征对应证据的BPA。

2) 证据修正:为了解决证据过度自信问题,使用本文提出的基于置信度的距离矩阵计算方法对证据进行修正。

3) 权重分配:使用Li等<sup>[32]</sup>提出的基于关联度的可信度计算方法来计算证据的可信度;使用KL散度计算证据与均匀分布的距离来确定证据的模糊程度;使用Wang等<sup>[33]</sup>提出的基于证据可信度和清晰程度的权重分配方法进行权重分配,确定每个证据的权重,减少冲突和模糊证据的负面影响。

4) 证据融合:首先将上述权重与修正后的证据相乘得到最终用于决策的证据集合,然后使用Dempster规则,将证据集合中的证据进行融合,进而得到最终的决策结果。

5) 证据后处理:将基于DS理论框架下的决策结果转化为信息抽取任务的标准输出。

本文方法综合考虑了不同模态证据的冲突和关联性,通过有效的证据修正与权重分配机制,提高了决策的质量和可信度。同时,在该方法使用证据形成最终决策的过程中,每一个步骤都以白盒的形式进行,推理过程透明。因此本文方法具有较好的可解释性。

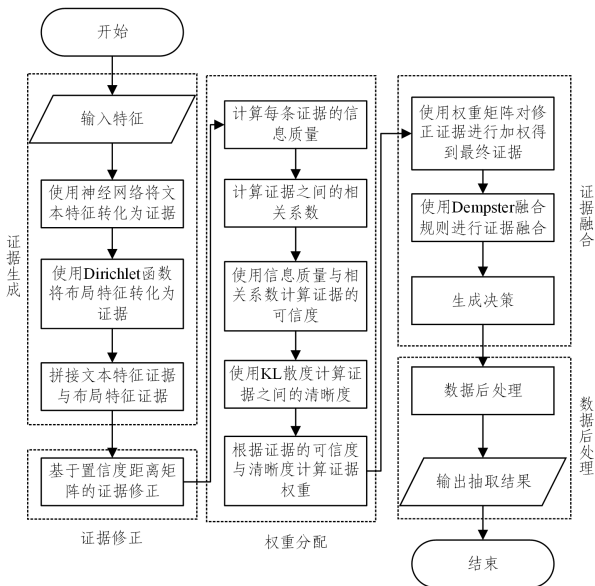


图3 本文方法流程

Fig. 3 Flow of the proposed method

## 4.1 证据生成阶段

### 4.1.1 基于神经网络生成文本模态的证据

如图4所示,本文模型采用预训练 SentenceBert 实现文本嵌入,并通过3层 BiLSTM 进行特征学习。在训练阶段,模型通过最小化 CRF(Conditional Random Field) 损失函数来调整权重;在推理阶段,为了计算每个标签的概率,本文方法不使用 CRF,而是采用 Softmax 模块输出概率向量,以便后续任务的处理。具体流程如下。

首先,从文档  $D$  中提取  $K$  个文本块信息,表示为  $S = \{s_1, s_2, s_3, \dots, s_K\}$ 。其中,  $s_k = \{c_1^k, c_2^k, c_3^k, \dots, c_n^k\}$  表示第  $k$  个文本块,  $c_j^k$  表示第  $k$  个文本块中的第  $j$  个文字。对于给定的文本数据  $S$ ,首先使用词典将文本转化为编码的向量,如式(11)所示,然后使用 SentenceBert 模型完成最终编码。文本嵌入的表示如式(12)所示:

$$ce_k = \text{Encode}(c_j^k) \quad (11)$$

$$se_k = \text{SentenceBertEncode}(ce_{1:n}^k) \quad (12)$$

其中,  $ce_{1:n}^k = \{ce_1^k, ce_2^k, ce_3^k, \dots, ce_n^k\}$  表示第  $k$  个文本块内文字序列的编码;  $se_k$  表示第  $k$  个文本块的句嵌入。因此,将每一个文本块  $s$  经过编码后可以得到整个文档  $D$  的编码  $DE = \{se_1, se_2, se_3, \dots, se_K\}$ 。通过这种方式能够捕捉到文本之间的语义相似性,进而提升下游任务的效果。

然后,将编码  $DE$  输入 BiLSTM 模型,该模型在推理过程中能够从文本数据中提取上下文信息,并输出对各个标签的分类概率。通常 LSTM 模型后会添加 CRF 层以输出预测的标签值,为了获取模型对每一个标签的支持度,将模型的 CRF 层替换为 Softmax 层。这种替换和变换的过程能够从模型中获取更为丰富的信息。函数定义如式(13)所示:

$$P_t = \text{Softmax}(\text{BiLSTM}(DE)) \quad (13)$$

其中,  $P_t = \{y_1, y_2, y_3, \dots, y_K\} \subset R^{K \times 1 \times N}$  表示整个文档的分类情况,也就是证据;  $y_k$  表示第  $k$  个文本块的特征对  $N$  个分类的置信度。

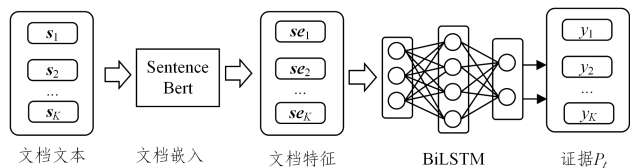


图4 基于神经网络生成文本模态的证据

Fig. 4 Neural network-based generation of evidence for textual modality

### 4.1.2 基于狄利克雷分布生成布局模态的证据

如图5所示,在处理布局模态数据时,首先根据训练数据集中的标注信息提取每个样本的位置信息及其对应的分类标签。为了有效地将位置信息映射为狄利克雷分布的参数,将整个文本的长度和宽度按照坐标等比例分割,从而将位置信息表示为一个连续的数值区间。在此基础上,统计每个位置区间内不同标签出现的频率,并将这些频率转化为相对频率。最后,利用得到的相对频率构建狄利克雷分布,其中每个位置区间的相对频率成为该分布的一个参数。这样的建模方式不仅准确反映了位置信息与分类标签之间的关系,而且能够为每个位置区间内的标签概率提供准确估计。具体流程如下。

首先,使用包含位置信息和分类标签的数据集,该数据集中每个样本都包含一个8维的边界框(Bounding Box, bbox)信息向量和对应的分类标签。为了实现建模和预测,需要从每个样本的边界框信息中提取出位置信息构成位置特征  $P_k = \{x_k, y_k, w_k, h_k, a_k\}$ 。其中,  $x_k$  表示第  $k$  个文本块左上角的  $x$  轴坐标占整个文档宽度的比例,  $y_k$  表示第  $k$  个文本块左上角的  $y$  轴坐标占整个文档高度的比例,  $w_k$  表示第  $k$  个文本块的宽度占整个文档宽度的比例,  $h_k$  表示第  $k$  个文本块的高度占整个文档高度的比例,  $a_k$  表示第  $k$  个文本块的面积占整个文档面积的比例。该特征将作为后续任务的输入。

对于每个不同的分类标签,需要计算其在已有位置信息中的分布情况。具体而言,收集所有包含特定分类标签的样本,并使用区间数的方法获取该分类标签在某一区域的频率。这样做的目的是为了捕捉每个类别在位置空间中的特征分布,从而为狄利克雷分布的参数估计提供依据。

随后,将每个分类标签的位置分布视为狄利克雷分布的参数  $\alpha$ ,并进行参数估计。为了满足概率分布的性质,需要对每个分类的位置分布进行归一化处理,以确保其元素之和为1。这些归一化后的位置分布将作为狄利克雷分布的参数,用于后续生成新的分类标签。基于统计得到的狄利克雷分布参数,为每个分类标签建立对应的狄利克雷分布,其参数由上述步骤中的位置分布确定。这些分布将用于预测新给定的位置信息的分类,计算式如下:

$$\text{Dirichlet}(p | \alpha) = \frac{\prod_{i=1}^5 p_i^{\alpha_i - 1}}{B(\alpha)} \quad (14)$$

其中,  $\alpha_i$  是归一化后的位置分布中第  $i$  个元素。

最后,对于给定的新位置信息,使用已建模的狄利克雷分布来生成分类标签的概率分布。将新的文档位置信息输入相应的狄利克雷分布中,计算每个文本块不同布局参数在  $N$  个分类标签下的概率  $P_t \subset R^{K \times 5 \times N}$ 。

$$\mathbf{P}_s = [\text{Dirichlet}(\mathbf{p}_{\text{new}} | \boldsymbol{\alpha}_s)]_{s=1}^K \quad (15)$$

其中,  $\mathbf{p}_{\text{new}}$  为新位置信息的特征,  $\boldsymbol{\alpha}_s$  为基于历史数据中不同类别的位置分布估计得到狄利克雷分布的参数向量。

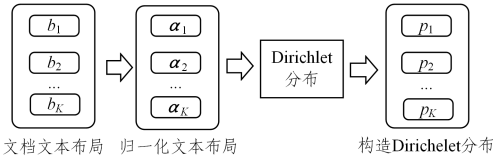


图5 基于狄利克雷分布生成布局模态的证据

Fig. 5 Evidence for generating layout modalities based on the Dirichlet distribution

#### 4.1.3 合并

如图6所示,将4.1.1节与4.1.2节生成的证据 $\mathbf{P}_i$ 与证据 $\mathbf{P}_s$ 按照第二个维度拼接成新的证据 $\mathbf{M}$ ,其中 $\mathbf{M} = \mathbf{P}_i \oplus \mathbf{P}_s = \{\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3, \dots, \mathbf{m}_K\} \subset R^{K \times 6 \times N}$ 。其中,  $K$  表示整个文本包含的  $K$  个文本块;  $6$  表示 1 个文本模态证据和 5 个布局模态特征  $P_k$  对应的证据;  $N$  表示分类标签的数量;  $\mathbf{M}$  整体表示在某一文档中, 每一个文本块的文本特征和布局特征对该文本块属于某一个分类的置信度。

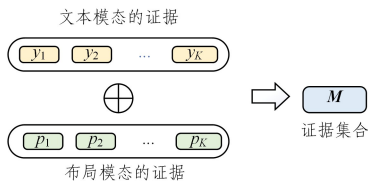


图6 合并证据

Fig. 6 Consolidation of evidence

通过上述方法得到的矩阵  $\mathbf{M}$  中不同模态的证据处于不同维度, 其对分类的概率分配不会互相干扰。因此, 不存在因模态之间异构性产生的模态冗余以及模态互相干扰问题。

#### 4.2 证据修正

本文使用基于置信度的距离矩阵计算方法来得到距离矩阵  $\mathbf{d}_{k,i,j}$ , 并通过该矩阵计算得到修正证据  $\mathbf{F}_k = \{f_{k,1}, f_{k,2}, f_{k,3}, \dots, f_{k,6}\}$ , 其目的是防止由于 DS 理论的特性, 当某条证据过度自信时, 其他证据将无法对最终决策做出贡献。

首先, 使用基于置信度的距离矩阵计算式  $\text{Dis}(\cdot)$  计算当前证据的距离矩阵  $\mathbf{d}_{k,i,j}$ , 计算式如下:

$$\mathbf{d}_{k,i,j} = \text{Dis}(\mathbf{m}_{k,i}) = \frac{\sum_{l=1}^N |\mathbf{m}_{k,i,j} - \mathbf{m}_{k,i,l}|}{N-1} \quad (16)$$

其中,  $\mathbf{m}_{k,i,j}$  和  $\mathbf{m}_{k,i,l}$  表示对于第  $k$  个文本块, 其第  $i$  个证据对第  $j$  个和第  $l$  个分类的支持程度,  $\mathbf{m}_{k,i,j}, \mathbf{m}_{k,i,l} \subset \mathbf{M}$ 。

然后, 证据  $\mathbf{m}_{k,i}$  对各个分类置信度之间的平均距离进行归一化处理, 再与原本置信度相加后求均值, 得到每条证据对各分类置信度的修正值  $d\omega_{k,i,j}$ , 计算式如下:

$$d\omega_{k,i,j} = \frac{1}{2} \left( \frac{\mathbf{d}_{k,i,j}}{\sum_{j=1}^N \mathbf{d}_{k,i,j}} + \mathbf{m}_{k,i,j} \right) \quad (17)$$

最后, 将  $d\omega_{k,i,j}$  拼接得到该条证据的修正值  $f_{k,i}$ , 计算式如下:

$$f_{k,i} = \{d\omega_{k,i,1}, d\omega_{k,i,2}, d\omega_{k,i,3}, \dots, d\omega_{k,i,N}\} \quad (18)$$

如表2所列, 在1中原本过度自信的证据, 经过修正后, 在一定程度上降低了自信水平。而2中完全模糊的证据在修

正后与原本的数值一致。对于其他没有出现过度自信的证据, 经过修正后的结果与原本数值相差不大。

表2 证据修正

Table 2 Amendments to evidence

序号	原始置信度	修正后置信度
1	0.9, 0.01, 0.01, 0.08	0.68, 0.10, 0.10, 0.12
2	0.25, 0.25, 0.25, 0.25	0.25, 0.25, 0.25, 0.25
3	0.35, 0.27, 0.25, 0.13	0.32, 0.23, 0.21, 0.24
4	0.62, 0.21, 0.05, 0.12	0.51, 0.20, 0.14, 0.15

#### 4.3 权重分配

本节将介绍证据权重系数的计算方法。以整个文档的证据  $\mathbf{M}$  中单独一个文本块的证据集合  $\mathbf{m}_k$  为例, 对于第  $k$  个文本块中各个证据的权重系数  $\mathbf{W}_k$ , 本文使用基于证据的清晰程度算法来计算。

##### 4.3.1 权重分配相关公式介绍

1) 可信度 Cred: 本文使用基于关联度的可信度计算方法来计算证据的可信度。步骤如下:

步骤1 计算每条证据的信息质量  $IQ$ :

$$IQ(\mathbf{m}_{k,i}) = \sum_{A_n \in \theta} \left( \frac{m_i(A_n)}{2^{|A_n|} - 1} \right)^2 \quad (19)$$

其中,  $m_i(A_n)$  表示第  $i$  条证据对分类  $A_n$  的支持度,  $\theta$  表示所有可能的文本分类。

步骤2 基于证据的信息质量, 计算证据之间的相关系数  $CIQ$ :

$$CIQ(\mathbf{m}_{k,i}, \mathbf{m}_{k,j}) = \frac{\sum_{A_n \in \theta} \frac{m_{k,i}(A_n)}{2^{|A_n|} - 1} \times \frac{m_{k,j}(A_n)}{2^{|A_n|} - 1}}{\sqrt{IQ(\mathbf{m}_{k,i}) \times IQ(\mathbf{m}_{k,j})}} \quad (20)$$

步骤3 求每条证据与其他证据相关系数之和并取平均值得到该条证据的可信度  $Cred$ :

$$Cred_{k,i} = \frac{1}{5} \sum_{j=1, j \neq i}^6 CIQ(\mathbf{m}_{k,i}, \mathbf{m}_{k,j}) \quad (21)$$

2) 权重系数  $\mathbf{W}_k$ : 本文使用基于证据可信度和清晰程度的权重分配方法进行权重  $\mathbf{W}_k = \{\omega_{k,1}, \omega_{k,2}, \omega_{k,3}, \dots, \omega_{k,6}\}$  的计算, 计算式如下:

$$\omega_{k,i} = Cred(\mathbf{m}_{k,i}) \times \frac{6 \times Clar(\mathbf{m}_{k,i})}{\sum_{j=1}^6 Clar(\mathbf{m}_{k,j})} \quad (22)$$

3) 清晰度 Clar: 本文方法使用3.1节得到的证据集合  $\mathbf{m}_k$  与均匀分布  $\mathbf{U}$  的 KL (Kullback Leibler) 散度<sup>[34]</sup> 来表示证据的清晰程度  $Clar$ 。KL 散度是一种度量两个概率分布之间差异的统计量, 用于衡量在给定一个分布的情况下, 使用另一个分布来近似表示所引入的信息损失。证据  $\mathbf{m}_k$  和均匀分布  $\mathbf{U}$  的清晰度计算式为:

$$KL(\mathbf{m}_{k,i} \parallel \mathbf{U}) = \sum_i m_i \log \frac{m_i}{U} \quad (23)$$

$$Clar(\mathbf{m}_k) = [KL(\mathbf{m}_{k,i} \parallel \mathbf{U})]_{i=1}^6 \quad (24)$$

当 KL 散度较大时, 意味着该条证据的分布与均匀分布之间的差异较大, 表明证据的模糊程度较低。

##### 4.3.2 算法流程

本文方法的权重分配具体过程如算法1所示, 为了方便展示, 算法中使用  $\mathbf{m}$  表示上文中的文本块  $\mathbf{m}_k$ , 用  $\mathbf{W}$  表示表示上文中的  $\mathbf{W}_k$ 。

**算法 1** 权重分配算法输入:证据  $\mathbf{m}$ 输出:权重系数  $\mathbf{W}$ 

```

1. /* 权重分配方案 */
2. /* IQs 为证据信息质量, CIQs 为证据间相关系数, Cred 为证据可
   信度, Clar 为证据清晰度, U 为均匀分布 */
3. 初始化: IQs, CIQs, Cred, Clar,  $\mathbf{W} \leftarrow \mathbf{O}$ , U
4. /* 使用 KL 散度计算证据清晰度 */
5. for  $i \leftarrow 1$  to 6 do
6.    $\text{Clar}[i] \leftarrow \text{KL}(\mathbf{m}[i], \text{U})$ 
7. end for
8. /* 计算每条证据的信息质量 */
9. for  $i \leftarrow 1$  to 6 do
10.   $\text{IQs}[i] \leftarrow \text{IQ}(\mathbf{m}[i])$ 
11. end for
12. /* 计算证据间相关系数 */
13. for  $i \leftarrow 1$  to 6 do
14.   for  $j \leftarrow 1$  to 6 do
15.      $\text{CIQs}[i][j] \leftarrow \text{CIQ}(\text{IQ}[i], \text{IQ}[j], \mathbf{m}[i], \mathbf{m}[j])$ 
16.   end for
17. end for
18. /* 使用证据间的相关系数计算证据的可信度 */
19. for  $i \leftarrow 1$  to 6 do
20.   for  $j \leftarrow 1$  to 6 do
21.     if ( $i \neq j$ ) then
22.        $\text{Cred}[i] \leftarrow \text{Cred}[i] + \text{CIQ}[i][j]$ 
23.     end if
24.   end for
25. end for
26. /* 计算证据权重系数  $\mathbf{W}$  */
27. for  $i \leftarrow 1$  to 6 do
28.    $\mathbf{W}[i] \leftarrow N * \text{Cred}[i] * \text{Cred}[i] * \text{sum}(\text{Clar})$ 
29. return  $\mathbf{W}$ 

```

**4.4 证据融合**

在证据融合阶段,首先将权重系数  $\mathbf{W}$  对修正证据  $F$  进行加权计算得到可以用于融合的证据集合  $\mathbf{m}_{\text{final}} = \{\mathbf{m}f_1, \mathbf{m}f_2, \mathbf{m}f_3, \dots, \mathbf{m}f_K\} \subset R^{K \times 6 \times N}$ , 其中  $\mathbf{m}f_k$  的计算式如下:

$$\mathbf{m}f_k = \mathbf{W}_k \times F_k = \{\mathbf{m}f_{k,1}, \mathbf{m}f_{k,2}, \mathbf{m}f_{k,3}, \dots, \mathbf{m}f_{k,6}\} \quad (25)$$

然后,使用 Dempster 融合公式对加权后的证据进行融合。该计算式可以将多个来源的证据合并,然后计算出  $K$  个文本块下各个分类标签的最终概率  $\mathbf{P} = \{p_1, p_2, p_3, \dots, p_K\} \subset R^{K \times N}$ 。其中  $p_k = \{p_k(A_1), p_k(A_2), p_k(A_3), \dots, p_k(A_N)\}$ , 计算步骤如式(26)所示:

$$p_k(A_i) = \frac{\mathbf{m}f_{k,1} \oplus \mathbf{m}f_{k,2} \oplus \mathbf{m}f_{k,3} \oplus \dots \oplus \mathbf{m}f_{k,6}(A_i)}{1 - \text{Norm}_k} \quad (26)$$

其中,  $A_i$  表示第  $i$  个分类;  $\mathbf{m}f_{k,j}(A_i)$  表示第  $k$  个文本块中第  $j$  条证据对第  $i$  个分类的支持度,  $1 \leq j \leq 6$ 。  $\text{Norm}_k$  表示第  $k$  个文本块证据融合的归一化因子, 计算式为:

$$\text{Norm}_k = 1 - \sum \mathbf{m}f_{k,1}(A_i) \mathbf{m}f_{k,2}(A_i) \mathbf{m}f_{k,3}(A_i) \dots \mathbf{m}f_{k,6}(A_i) \quad (27)$$

最后,根据证据对不同分类的最终支持度,选取概率最大的标签作为文本块的最终分类标签  $L = \{\max(p_1), \max(p_2), \max(p_3), \dots, \max(p_K)\}$ 。

**4.5 数据后处理阶段**

后处理阶段将文本块内容进行清洗后与标签共同组成结构化数据。

根据 4.4 节得到的每个文本块的分类标签,将标签和根据规则处理后的文本块内容拼接为结构化数据。例如,标签为“1”,对应分类为“招标单位名称”,文本块内容为“招标人:盐城市第一人民医院”,则结构化数据为“招标单位名称:盐城市第一人民医院”。在后处理阶段,遍历整个分类标签信息  $L$ ,重复上述操作,将每个文档中的关键信息抽取并组织为结构化信息。

**5 实验****5.1 数据集**

本文实验使用医疗公告数据集(Medical)和公开数据集(SROIE<sup>1)</sup>)。其中 SROIE 数据集用于消融实验和对比实验 1, Medical 数据集用于对比实验 2。下面分别介绍两组数据集。

Medical 数据集是由本文收集并标注的数据集,包含 2 000 条医院公告图片和标注好的 bbox 格式文本,其中 1 500 条用于训练,500 条用于测试。如图 1(a)所示,该数据集有 5 个字段,分别是采购单位名称、采购人姓名、联系方式、地址和邮箱(部分文档会出现字段未出现情况)。该数据集主要由中、英文字符和数字组成。因为公告内容长短和格式不统一,所以不同公告中的关键信息可能出现在不同的位置,且文本中重复出现语义、格式与目标字段相近的内容。因此,该数据集属于复杂的动态结构布局。

SROIE 数据集是公开数据集,包含 626 张训练收据和 347 张测试收据,数据集内的格式同医疗公告数据集一样,包含图片和标注好的 bbox 格式文本。如图 1(b)所示,不同于医疗公告数据集, SROIE 数据集有 4 个字段,分别是公司名称、地址、日期和收据总金额。该数据集的内容仅包含英文字符和数字,且布局相对单一,因此该数据集属于固定布局。

**5.2 实验环境与任务描述**

本实验运行环境为 Windows 系统,代码执行环境为 Python 3.8.6, GPU 型号为 RTX 3060,显存为 12 GB,运行内存为 16 GB。

在本文实验中,待抽取文档中的每个目标字段属于一个独立的文本块内,从文本块内剔除多余的符号(如“:”“\”“\$”等)的后续处理不是本任务的重点,不做讨论。因此,将信息抽取任务的目标设定为从文档中选出标签内容所在的文本块。本文实验使用精确率(P)、召回率(R)、F1 值<sup>[35]</sup>作为评价指标,将本文方法与其他方法进行性能对比。

**5.3 对比实验**

本节进行了对比实验,以 BiLSTM-CRF 为基线方法,对

<sup>1)</sup> <https://rrc.cvc.uab.es/?ch=13>

比并分析了本文方法在信息抽取任务中的性能表现。BiLSTM-CRF 架构作为信息抽取领域广泛应用的基准模型,其有效性已在多个实际场景中得到证明。

除了与基线方法进行比较以外,还与 LayoutLM, PICK 和 TRIE 方法在 SROIE 数据集上进行对比实验 1,与两个基于 DS 理论的决策融合方法在 Medical 数据集上进行对比实验 2。结果如表 3、表 4 所列。BiLSTM-CRF+DS 方法将经典的 BiLSTM-CRF 模型输出结果作为 DS 证据理论框架的输入,并使用 Dempster 组合规则对模型输出的结果进行处理,并生成最终决策结果;BiLSTM-CRF+Deng 方法则是在 BiLSTM-CRF+DS 方法的基础上使用 Deng 熵对证据进行校正后再进行最终决策。其中 LayoutLM, PICK 和 TRIE 的实验结果来源于 SROIE2019 比赛的公开结果。

表 3 在 SROIE 数据集上进行对比实验 1 的结果

Table 3 Result of comparative experiments 1 on SROIE dataset (%)

Method	P	R	F1
Baseline	90.14	91.16	90.19
LayoutLM(Large) <sup>[24]</sup>	96.04	96.04	96.04
LayoutLMv2(Large) <sup>[25]</sup>	99.04	96.61	97.81
PICK <sup>[22]</sup>	96.79	95.46	96.12
TRIE <sup>[36]</sup>	97.72	95.89	96.80
Ours	98.76	97.32	98.03

在 SROIE 数据集上,本文方法相较于其他基于数据层融合的神经网络方法在召回率和 F1 分数上有 0.22 到 1.87 个百分点的提升,在精确率上,本文方法的分数仅低于 LayoutLMv2 模型 0.28 个百分点。总体而言,本文方法能够高效地从收据文档中捕获位置信息与标签之间的关联关系,并以证据形式支持最终决策的形成,同时具有较强的可解释性。值得注意的是,与本文方法相比,LayoutLM 和 LayoutLMv2 两个模型均为大语言模型,模型参数约为  $200 \times 10^6$ ,并使用了其他数据集进行预训练任务。本文方法仅在文本证据生成阶段使用深度学习模型,总参数仅有  $20 \times 10^6$  左右,并且仅在 SROIE 数据集上进行训练而未引入其他知识库。由此可以看出,在更小的参数量及更少的预处理流程下,本文方法展现出较好的学习与决策能力。相对于其他方法,本文方法能够在充分利用有效证据的同时,减弱模糊证据对决策的影响,从而显著提升了最终正确分类的概率。

表 4 在 Medical 数据集上进行对比实验 2 的结果

Table 4 Result of comparative experiments 2 on Medical dataset (%)

Method	P	R	F1
Baseline	93.93	94.34	93.89
BiLSTM-CRF + DS	94.04	94.73	94.08
BiLSTM-CRF+Deng <sup>[35]</sup>	94.49	94.38	94.23
Ours	95.16	95.81	95.21

从表 4 可以看出,相较于其他决策融合方法,本文方法在精确率上提高了 0.67 个百分点到 1.23 个百分点,这一结果表明,本文方法可以处理布局复杂度较高文档的信息抽取任务。

综上所述,本文方法不仅在 SROIE 数据集上取得了较好的效果,而且在 Medical 数据集上也表现出不错的性能。这

进一步验证了本文方法无论是面对复杂的动态结构布局还是较为简单的固定布局,都有高效且稳定的信息提取能力。

#### 5.4 消融实验

为了验证本文方法中每个模块的贡献,本节进行了消融实验,并使用精确率作为评估指标。

如表 5 所列,使用本文方法在 SROIE 数据集上进行了一系列消融实验。在这些实验中,对模型训练数据的输入方式和特征融合方法进行了不同的调整和组合,以评估不同特征、不同融合方式对最终效果的影响。

表 5 在 SROIE 数据集上的消融实验结果

Table 5 Result of ablation experiments on SROIE dataset

实验	method	Fusion Method	Text	Layout	P/%
1	LSTM-CRF	Feature	✓	×	90.14
2	LSTM-CRF	Fusion	✓	✓	86.76
3	LSTM-CRF+DS		✓	✓	90.80
4	LSTM-CRF+Deng <sup>[35]</sup>	Decision	✓	✓	91.31
5	Ours	Fusion	✓	✓	98.76

在实验 1 中,仅将文本特征作为模型的输入。这意味着模型仅能从文本信息中学习特征与标签之间的关联关系。在实验 2 中,将文本特征与布局特征融合后作为模型的输入。然而,从表 5 的结果可以观察到,与实验 1 相比,实验 2 的精确率下降了 3.38 个百分点。原因可能是布局特征包含较多的噪声,如果不进行处理,这种噪声可能会干扰模型学习文本特征与标签之间的关联关系,导致融合后的特征信息变得模糊,使得模型难以准确学习到特征与标签之间的映射关系。

与实验 1 和实验 2 不同,本文在实验 3 中采取了一种不同的特征融合策略。具体来说,将文本特征作为模型的输入,同时在决策层引入基于布局特征构建的狄利克雷分布。通过这种方式,试图在决策层融合文本特征和布局特征的信息。实验 3 相较于实验 2,精确率提升了 4.04 个百分点,比实验 1 提高了 0.66 个百分点。这表明,在决策层融合时,引入布局特征的狄利克雷分布可以有效提升模型的性能,原因是这种方法能够降低布局特征中噪声的干扰,同时更好地捕捉特征之间的复杂关系,从而增强模型对特征与标签关联性的学习能力。

在实验 3 中,面对模糊证据与冲突证据时,Dempster 组合规则存在局限性,会导致结果不够准确和稳定。现有的方法如实验 4,使用 Deng 熵<sup>[35]</sup>来修正原本的证据,从而达到较为准确的预测结果。由表 5 可知,在 SROIE 数据集上,使用 Deng 熵的方法可以在一定程度上提高分类精确率。

实验 4 和实验 5 中对比了使用和不使用证据修正策略的情况。本文方法在实验 4 的基础上,引入了证据修正策略。该策略可以根据证据的可靠性和一致性动态对证据进行修正,从而更好地避免证据过度自信。结果显示,在引入证据修正后,最终得分显著提升,精确率提升了 7.45 个百分点。这一结果表明,本文提出的证据修正策略能够有效地突破 Dempster 组合规则在面对证据过度自信时的局限性。通过动态调整证据,本文方法能够更准确地利用多个证据,降低过度自信证据的影响,从而提高模型的整体性能。

综上所述,消融实验结果不仅显示了特征融合策略对模型性能的影响,还展示了证据修正策略在处理证据过度自信

时的重要性。实验表明,文本特征和布局特征的融合方式对于模型性能具有重要影响。合理的特征融合策略可以提高模型的性能,而不恰当的融合方式可能导致信息损失和性能下降。本文方法通过在决策层使用狄利克雷分布构建的布局特征,同时引入权重分配策略,在 SROIE 数据集上取得了显著的性能提升,为不确定性建模和证据组合在信息抽取领域的应用提供了有价值的思路。

**结束语** 本文提出了一种基于决策层融合的多模态信息抽取方法,该方法将 DS 理论引入信息抽取任务中;为了解决证据过度自信问题,本文提出了一种基于置信度的距离矩阵计算方法对证据进行修正。本文方法有效避免了在数据层融合中常见的特征冗余和不兼容问题,增强了模型的可解释性,同时一定程度上降低了证据过度自信导致的决策偏差问题。在公开数据集上的验证结果表明,该方法表现出良好的性能。此外,本文方法在真实的医疗招标投标公告数据集上进行验证,结果展示出足够高的精确度。因此,其可以用于同类复杂布局的信息抽取任务中,如企业年度财报、人事任免公告等信息抽取任务。

本文方法仅适用于对文档、公告等单主体文档进行信息抽取,因此会出现主体混淆的情况。在下一步工作中将继续探索并改进本文方案,以解决不同主体下同类标签的信息抽取问题。除此以外,本文方法的总体参数较小,对各模态特征提取的能力受限,导致方法泛化能力不如其他大模型。因此,下一步的工作将会考虑在保持模型参数量尽可能小的同时,引入特征提取能力更好的模型。

## 参 考 文 献

- [1] LANDOLSI M Y, HLAOUA L, BEN ROMDHANE L. Information extraction from electronic medical documents: state of the art and future research directions[J]. *Knowledge and Information Systems*, 2023, 65(2): 463-516.
- [2] XU B, HUANG S, DU M, et al. A Unified Visual Prompt Tuning Framework with Mixture-of-Experts for Multimodal Information Extraction[C]// *International Conference on Database Systems for Advanced Applications*. Cham: Springer, 2023: 544-554.
- [3] RAHATE A, WALAMBE R, RAMANNA S, et al. Multimodal co-learning: Challenges, applications with datasets, recent advances and future directions[J]. *Information Fusion*, 2022, 81: 203-239.
- [4] YANG Y, ZHAN D C, JIANG Y, et al. Reliable multi-modal learning: a survey[J]. *Journal of Software*, 2020, 32(4): 1067-1081.
- [5] LI X, ZHAO X, XU J, et al. IMF: Interactive Multimodal Fusion Model for Link Prediction[C]// *Proceedings of the ACM Web Conference 2023*. 2023: 2572-2580.
- [6] DOU H, ZHANG L M, HAN F, et al. Survey on Convolutional Neural Network Interpretability[J]. *Ruan Jian Xue Bao/Journal of Software*, 2023, 35(1): 159-184.
- [7] CUNNINGHAM H, DING Y, KIRYAKOV A. Workshop on Human Language Technology for the Semantic Web and Web Services[EB/OL]. <https://gate.ac.uk/conferences/iswc2003/proceedings/iswc2003-hlt4sw-proceedings.pdf>.
- [8] ETZIONI O, CAFARELLA M, DOWNEY D, et al. Unsupervised named-entity extraction from the web: An experimental study[J]. *Artificial Intelligence*, 2005, 165(1): 91-134.
- [9] SEKINE S, NOBATA C. Definition, Dictionaries and Tagger for Extended Named Entity Hierarchy[C]// *LREC*. 2004: 1977-1980.
- [10] ZHANG S, ELHADAD N. Unsupervised biomedical named entity recognition: experiments with clinical and biological texts. [J]. *Journal of Biomedical Informatics*, 2013, 46(6): 1088-1098.
- [11] HANISCH D, FUNDEL K, MEVISSSEN H T, et al. ProMiner: rule-based protein and gene entity recognition[J]. *BMC Bioinformatics*, 2005, 6(1): 1-9.
- [12] QUIMBAYA A P, MÚNERA A S, RIVERA R A G, et al. Named entity recognition over electronic health records through a combined dictionary-based approach[J]. *Procedia Computer Science*, 2016, 100: 55-61.
- [13] FLESCA S, MANCO G, MASCIARI E, et al. Web wrapper induction: a brief survey[J]. *AI communications*, 2004, 17(2): 57-61.
- [14] NADEAU D, TURNEY P D, MATWIN S. Unsupervised named-entity recognition: Generating gazetteers and resolving ambiguity[C]// *Advances in Artificial Intelligence: 19th Conference of the Canadian Society for Computational Studies of Intelligence*. Berlin: Springer, 2006: 266-277.
- [15] COLLINS M, SINGER Y. Unsupervised models for named entity classification[C]// *1999 Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora*. 1999.
- [16] YAO L, LIU H, LIU Y, et al. Biomedical named entity recognition based on deep neural network[J]. *International Journal of Hybrid Information Technology*, 2015, 8(8): 279-288.
- [17] YANG J, ZHANG Y, DONG F. Neural reranking for named entity recognition[J]. *arXiv:1707.05127*, 2017.
- [18] KURU O, CAN O A, YURET D. Charner: Character-level named entity recognition[C]// *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*. 2016: 911-921.
- [19] ZHANG Q, FU J, LIU X, et al. Adaptive co-attention network for named entity recognition in tweets[C]// *Proceedings of the AAAI Conference on Artificial Intelligence*. 2018.
- [20] CARBONELL M, RIBA P, VILLEGAS M, et al. Named entity recognition and relation extraction with graph neural networks in semi structured documents[C]// *2020 25th International Conference on Pattern Recognition (ICPR)*. 2021: 9622-9627.
- [21] LIU X, GAO F, ZHANG Q, et al. Graph convolution for multimodal information extraction from visually rich documents[J]. *arXiv:1903.11279*, 2019.
- [22] YU W, LU N, QI X, et al. PICK: processing key information extraction from documents using improved graph learning-convolutional networks[C]// *2020 25th International Conference on Pattern Recognition (ICPR)*. 2021: 4363-4370.

- [23] HUANG Z, CHEN K, HE J, et al. Icdar2019 competition on scanned receipt ocr and information extraction[C]//2019 International Conference on Document Analysis and Recognition(ICDAR). 2019;1516-1520.
- [24] XU Y, LI M, CUI L, et al. Layoutlm: Pre-training of text and layout for document image understanding[C]//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2020;1192-1200.
- [25] XU Y, XU Y, LYU T, et al. Layoutlmv2: Multi-modal pre-training for visually-rich document understanding[J]. arXiv; 2012. 14740, 2020.
- [26] XU Y, LYU T, CUI L, et al. Layoutxlm: Multimodal pre-training for multilingual visually-rich document understanding[J]. arXiv; 2104. 08836, 2021.
- [27] SUI D, TIAN Z, CHEN Y, et al. A large-scale chinese multimodal ner dataset with speech clues[C]//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing. 2021;2807-2818.
- [28] HAN Z, ZHANG C, FU H, et al. Trusted multi-view classification with dynamic evidential fusion[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(2): 2551-2566.
- [29] DEMPSTER A P. Upper and Lower Probabilities Induced by a Multivalued Mapping[J]. Annals of Mathematical Statistics, 1967, 38(2): 325-339.
- [30] SHAFER G A. A Mathematical Theory of Evidence [J]. Technometrics, 1978, 20(1): 106-106.
- [31] REIMERS N, GUREVYCH I. Sentence-bert: Sentence embeddings using siamese bert-networks [J]. arXiv; 1908. 10084, 2019.
- [32] LI D, DENG Y, CHEONG K H. Multisource basic probability assignment fusion based on information quality[J]. International Journal of Intelligent Systems, 2021, 36(4): 1851-1875.
- [33] WANG Y C, WANG J, HUANG M J, et al. An evidence combination rule based on a new weight assignment scheme[J]. Soft Computing, 2022, 26(15): 7123-7137.
- [34] ZENG J, XIAO F. A fractal belief KL divergence for decision fusion [J]. Engineering Applications of Artificial Intelligence, 2023, 121: 106027.
- [35] DENG Y. Deng entropy[J]. Chaos, Solitons & Fractals, 2016, 91: 549-553.
- [36] ZHANG P, XU Y, CHENG Z, et al. TRIE: end-to-end text reading and information extraction for document understanding [C]//Proceedings of the 28th ACM International Conference on Multimedia. 2020:1413-1422.



**WANG Jian**, born in 1978, Ph.D, professor, is a member of CCF(No. 28300S). Her main research interests include social computing, cybersecurity, and artificial intelligence.

(责任编辑:何杨)