



天津科技大学学报

Journal of Tianjin University of Science & Technology

ISSN 1672-6510, CN 12-1355/N

《天津科技大学学报》网络首发论文

题目: KaNER:融合 KAN 和动态注意力机制的中文命名实体识别方法
作者: 龙卫平, 宋琛, 孙晗娟, 王丽媛, 张贤坤
DOI: 10.13364/j.issn.1672-6510.20250096
收稿日期: 2025-06-12
网络首发日期: 2026-01-29
引用格式: 龙卫平, 宋琛, 孙晗娟, 王丽媛, 张贤坤. KaNER:融合 KAN 和动态注意力机制的中文命名实体识别方法[J/OL]. 天津科技大学学报.
<https://doi.org/10.13364/j.issn.1672-6510.20250096>



网络首发: 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。



DOI: 10.13364/j.issn.1672-6510.20250096

KaNER:融合 KAN 和动态注意力机制的 中文命名实体识别方法

龙卫平¹, 宋琛^{1,2,3}, 孙晗娟¹, 王丽媛¹, 张贤坤¹

(1. 天津科技大学人工智能学院, 天津 300457; 2. 先进计算与关键软件海河实验室, 天津 300459;

3. 南开大学密码与网络空间安全学院, 天津 300071)

摘要: 命名实体识别(named entity recognition, NER)旨在从非结构化文本中识别出预定义的实体类别。当前中文 NER 方法常依赖静态词汇匹配或外部词典增强, 面临多义词歧义、上下文依赖缺失等问题, 导致实体边界模糊、嵌套和不连续结构识别困难。为此, 本文提出一种融合 KAN(Kolmogorov-Arnold Networks)与动态多头注意力机制的中文命名实体识别方法(KaNER)。设计一种 KAN 驱动的多频基函数动态筛选方法, 协同位置感知样条插值增强字符对特征, 并结合自适应语义门控权重调整字符对关联强度。进一步通过多头注意力整合模块, 结合多子空间特征交互优化全局-局部语义融合。实验结果表明, KaNER 在 OntoNotes、MSRA、Resume 和 Weibo 4 个平面实体数据集上 F_1 值最优达 96.92%, 较最佳基线均有提升; 在 ACE2004 与 ACE2005 嵌套实体数据集上 F_1 值达 88.41% 和 88.94%。证明了 KaNER 的有效性。该模型为复杂中文实体识别提供了高精度、轻量化解决方案, 对知识抽取、智能问答等下游任务具有重要应用价值。

关键词: 中文命名实体识别; Knowledge-Aware Network; 动态注意力机制; 多频特征融合; 位置感知样条

中图分类号: TP391

文献标志码: A

KaNER: A Chinese Named Entity Recognition Method Integrating KAN and Dynamic Attention Mechanism

LONG Weiping¹, SONG Chen^{1,2,3}, SUN Hanjuan¹, WANG Liyuan¹, ZHANG Xiankun¹

(1. College of Artificial Intelligence, Tianjin University of Science and Technology, Tianjin 300457, China;

2. Haihe Laboratory of Information Technology Application Innovation, Tianjin 300459, China;

3. College of Cryptology and Cyber Science, Nankai University, Tianjin 300071, China)

Abstract: Named entity recognition (NER) aims to identify predefined entity categories from unstructured text. Existing Chinese NER methods often rely on static vocabulary matching or external dictionary enhancement, yet they suffer from issues like polysemous ambiguity and insufficient contextual dependency, resulting in blurred entity boundaries and challenges in recognizing nested and discontinuous structures. To address these problems, this paper proposes a Chinese named entity recognition method (KaNER) that integrates KAN (Kolmogorov-Arnold Networks) with a dynamic multi-head attention mechanism. Specifically, a KAN-driven dynamic screening method for multi-frequency basis functions is designed; subsequently, character pair features are enhanced via synergistic position-aware spline interpolation, and the association strength of character pairs is adjusted using adaptive semantic gating weights. Furthermore, a multi-head attention integration module is employed to optimize global-local semantic fusion through multi-subspace feature interaction. Experimental results demonstrate that KaNER achieves an F_1 -score of 96.92% on four flat entity datasets (OntoNotes, MSRA, Resume, and Weibo), outperforming the best baseline. On the nested entity datasets ACE2004 and ACE2005, it achieves an F_1 -score of 88.41% and 88.94%. These results verify the effectiveness of KaNER. The proposed model provides a high-precision and

收稿日期: 2025-06-12; 修回日期: 2025-09-19

基金项目: 国家自然科学基金面上项目(62377036); 天津市科技计划项目(No. 22KJXMRC00210)

作者简介: 龙卫平(2001—), 女, 湖南湘西人, 硕士研究生; 通信作者: 宋琛, 讲师, songchen@tust.edu.cn

lightweight solution for complex Chinese entity recognition, and holds significant application value for downstream tasks such as knowledge extraction and intelligent question answering.

Key words: Chinese named entity recognition; Kolmogorov-Arnold Networks ; dynamic attention mechanism; multi-frequency feature fusion; position-aware spline

命名实体识别(named entity recognition, NER), 是自然语言处理里的一项基础任务, 主要是从非结构化文本当中识别并且分类预先定义好的实体类型^[1-2], 如人名、组织名以及地点等, 可为知识图谱构建^[3]、智能问答^[4]、舆情分析^[5]等后续任务给予结构化语义方面的支持。中文 NER 因为语言本身的特性, 面临一些独特的挑战: 一方面, 汉语没有十分突出的词边界, 字符组合很灵活, 而且多义性比较强^[6], 另一方面; 实体大多时候呈现出嵌套以及不连续的结构, 传统的方法很难在精度和效率之间达成平衡^[7]。

针对中文 NER 的复杂情况, 现有的研究可归纳成 4 类主流方法。一是序列标注方法: 借助 BIO(Begin, Inside, Outside)标注法逐个字符分配标签, 该方法在简单实体识别中表现稳定, 但受到单标签约束的限制, 无法处理嵌套结构, 且在长实体识别的时候容易受到梯度消失的影响^[8]。二是基于跨度分类的方法: 通过罗列出所有可能的子序列进行类型判断, 但计算复杂度会随着序列长度呈平方增长^[9], 同时负样本淹没问题会让模型倾向于短实体预测。三是生成式方法: 这类方法借助序列生成或者扩散模型输出实体, 虽然支持复杂结构, 却因解码效率低下与训练资源消耗大面临应用瓶颈^[10]。四是基于网格标注的方法: 构建字符对关系矩阵统一处理各类实体^[11], 是当前的主流范式, 存在特征融合效率不足、外部词典依赖引发的语境噪声干扰等缺陷。

近年来, 基于网格标注的方法借助显式建模字符对关系成为统一处理平面、嵌套以及不连续实体的有效范式^[12-13]。然而在中文场景下的应用依旧面临着多个挑战, 现有的方法一般依赖静态词表匹配引入外部词汇特征, 却忽略了上下文动态关联特性^[14], 导致非语境化噪声干扰比较大; 同时低频全局依赖与高频局部特征的协同建模不够, 模型对复杂实体的解析能力受到限制^[15]。另外, 句法信息与语义特征的简单拼接策略很难实现跨模态深度交互, 造成句法语义割裂现象, 影响实体边界的判定精度。

为解决以上问题, 本文设计一种融合

Kolmogorov-Arnold Network(KAN)与动态多头注意力机制的中文命名实体识别框架 KaNER, 有效地解决了汉语文本中实体边界模糊、嵌套及不连续结构识别险阻的关键问题, 主要贡献总结为以下几点:

(1)设计一种基于 KAN 的字符对特征增强模块。通过多频基函数动态融合与位置感知样条插值的协同机制, 在统一架构下实现局部非线性特征提取与长距离依赖建模, 解决低频全局依赖与高频局部特征协同建模不足的问题。

(2)提出自适应语义相似性加权方法。结合上下文感知门控机制动态调整字符对关联权重, 有效抑制跨实体噪声干扰, 缓解了外部词典依赖导致的非语境化噪声干扰问题。

(3)构建了多子空间注意力整合策略。利用多头注意力机制融合全局频域特征与局部位置特征, 增强了复杂语义关系的解码能力, 解决了句法语义割裂对边界判定的负面影响。

1 相关工作

1.1 KAN

由于 KAN 基于样条插值动态函数的良好逼近能力, 近些年其在自然语言处理领域得到了快速应用^[16]。传统多层感知机因为固定激活函数与线性权重的分离设计而受到限制, 很难应对高维数据里的维度灾难以及语义耦合问题^[17]。KAN 把激活函数从节点转移到边权, 引入基于样条插值的动态非线性映射, 达成对多变量函数的自适应分解。这样的设计严格遵循 Kolmogorov-Arnold 表示定理的数学框架, 还依靠可解释的样条基函数提高了模型对局部特征的捕捉能力。

2024 年, Yang 等^[18]提出 Kolmogorov-Arnold Transformer(KAT), 首次将 KAN 层系统整合至 Transformer 架构以替代传统多层感知器(MLP)模块。该工作通过自适应基函数优化机制, 动态平衡计算效率与函数表达能力, 在图像识别等任务中持续超越基准模型, 印证了 KAN 通过频域特征解耦增强长

距离依赖建模的普适优势。同一时期, 创新性融合KAN与循环门控机制的TKAN模型^[19], 通过递归Kolmogorov-Arnold层实现多步时序依赖建模, 显著提升长期预测稳定性, 验证了动态记忆管理对复杂模式捕捉的有效性。这些进展说明KAN已成为解决复杂语义建模问题的新范式。

在信息抽取领域, KAN的架构特性正逐渐呈现其应用潜力。卷积KAN研究证实, 把固定权重卷积核替换为基于样条的动态函数后, 模型能以50%的参数量在图像分类任务中保持同等精度^[20]。这一特性对语言处理任务有关键启示, 中文命名实体识别中嵌套结构与不连续实体的本质可抽象为多尺度空间关系建模问题: 实体内部字符的紧邻组合对应高频局部模式, 类似图像边缘; 而跨句子的实体语义关联则对应低频全局模式, 类似图像纹理。图学习领域的最新探索中, KAN驱动的图卷积层通过增强特征聚类能力在低资源场景显著提升节点分类性能^[21], 其同类特征紧聚类与异类特征强分离特性为字符对关系建模提供了跨领域实证。

中文NER面临的核心挑战在于需要一种架构, 该架构可协同对局部非线性与全局语义依赖进行建模, 而KAN的数学特性恰好为此奠定了理想基础。样条基函数有局部紧支性, 这种特性支持对字符间距展开细粒度建模, 可精确区分实体内部的紧密度以及边界噪声。傅里叶基函数拥有频谱解耦能力, 可以分离高频边界特征与低频类型语义。KAN有参数效率优势, 可为引入类似多头注意力的复杂模块预留计算资源。尽管现在尚未发现KAN在NER领域有直接应用, 然而上述跨领域的实证已经充分证实了其在空间关系建模、多频特征分离以及轻量化设计方面所具有的普适价值, 这为本文构建动态多频KAN框架提供了坚实的理论和技術支撑。

1.2 注意力机制

近些年来, 注意力机制在中文命名实体识别任务里呈现出独特的优势, 研究者借助多维度特征交互以及动态权重分配, 使模型对复杂语义关系的建模能力得到了提升^[22], 基于传统多头注意力的Transformer架构凭借全局上下文感知对长距离依赖捕捉的能力进行增强, 然而在中文场景中却遭遇计算复杂度高以及局部细粒度特征提取不足的问题。例如, ChineseBERT借助融合字形与拼音信息构建预训练注意力权重^[23], 尽管在简单实体识别任务中表现较为稳定, 但其静态注意力机制难以动态适应多义词歧义与嵌套结构, 造成微博数据集中低频实体面临挑战; 这类方法依赖单一频域特征融合, 无

法有效区分高频边界信号与低频语义关联, 在复杂实体解析时容易引入噪声干扰。

为了缓解上述问题, 研究者引入了外部知识引导的注意力机制, 借助句法依存树或词典特征提高局部语义感知, W²NER模型采用扩展卷积与注意力联合建模字符对关系, 利用句法距离约束注意力权重^[24], 在嵌套实体识别方面相较于传统序列标注方法, F₁值提升了14.59%, 但该方法依赖静态句法解析器且未考虑上下文动态关联, 导致跨实体噪声传播问题较为突出。同样地, HiNER依靠混合句法-语义注意力融合提升边界定位精度^[25], 但其固定的句法权重分配机制在非规范文本中引发显著误匹配问题, 暴露出外部知识引入与动态语境适配之间的矛盾。

动态注意力机制的最新研究成果为中文NER提供了优化思路。DiffusionNER将命名实体识别建模为边界去噪扩散过程^[26], 采用噪声注入策略, 提高对不连续实体的覆盖能力, 然而其迭代式解码过程致使推理效率下降。Zhang等^[27]提出的MGBERT模型引入全局指针机制, 凭借动态门控筛选有效字符对关联, 在MSRA数据集上达成了96.10%的F₁值, 但其线性注意力核函数难以对字符间距的非线性影响进行建模, 在长实体识别中仍存在边界判断不准的问题。这些方法虽然在一定程度上实现了动态权重调整, 但对多频特征的解耦能力存在局限, 并且未充分融合位置先验信息, 对复杂实体结构的解析精度造成了制约。

现有方法存在的核心局限是静态注意力机制在匹配动态语义需求方面存在险阻, 而且多频特征与位置依赖的协同建模有所缺失。传统的多头注意力依靠固定子空间投影进行特征提取, 很难依据上下文动态地筛选出关键频率成分。位置编码与语义特征采用简单叠加策略, 这导致局部边界感知与全局依赖建模出现割裂。另外外部知识引导的注意力调整缺乏语境自适应能力, 容易引入噪声干扰, 加剧实体边界模糊问题, 针对这些挑战, 本文提出融合KAN的多频动态注意力机制, 借助傅里叶基函数与样条插值的混合架构达成频谱解耦, 再结合位置感知门控动态调整字符对关联强度。

2 模型架构

本文提出的中文命名实体识别KaNER模型的整体框架如图1所示, 它主要由编码层、KAN动态机制融合层和协同预测层组成。编码层采用语义与位

置联合表征模块, 结合预训练模型 BERT 提取子词语义特征效率, 通过最大池化与位置补偿编码构建融合表征, 增强实体边界感知能力。在 KAN 动态机制融合层, KAN 特征增强融合动态频域筛选与位置感知样条建模, 自适应捕捉字符对的全局频率特征与局部间距依赖。注意力层利用多头机制协同增强特征与关系权重, 输出为嵌套感知的网格表示。之后, 协同预测层结合 KAN 和双仿射(biaffine)预测器的双路机制, 被用于多任务联合推理所有可能实体的边界和类型。

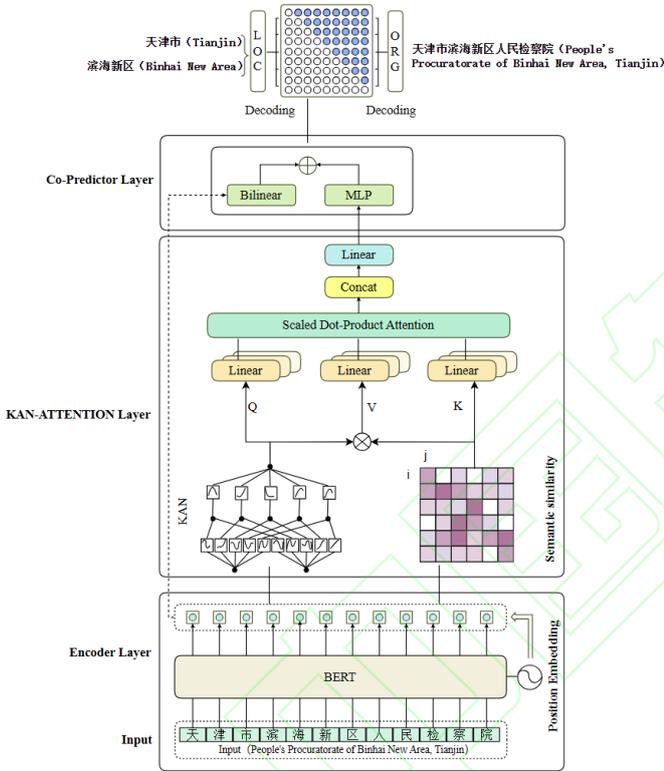


图 1 KaNER 模型的整体框架

Fig. 1 Overall framework of KaNER model

2.1 字符对关系建模

在基于网格的标记任务中, 字符对关系建模是识别平面的、重叠的以及不连续实体的关键。对给定的一个包含 N 个字符的输入句子 $X=\{x_1, x_2, \dots, x_N\}$, 目标是识别每对字符之间的语义关系 $R(x_i, x_j)$ 。KaNER 模型把字符对关系定义为 3 种类型。

无关系(NONE): 表示字符对之间不存在任何语义关系。这种关系在模型中用于排除无关字符对的干扰, 提高模型的识别精度。

相邻字符关系(next-neighbor-word, NNW): 表示字符对属于同一命名实体, 且在文本中是连续的。这种关系有助于模型识别实体的连续部分。

尾首字符关系(tail - head-word, THW-*): 表示字符对分别属于同一命名实体的尾部和头部。这种关系有助于模型识别实体的边界。其中, “*”表示实体类型, 如“地名”“人名”等。

基于网格的字符对关系建模方法的示例如图 2 所示。这种字符对关系建模能够更准确地识别中文中平面的、嵌套的和连续的命名实体。

		天	津	市	滨	海	新	区	人	民	检	察	院
天		NNW											
津			NNW										
市	THW-L		NNW										
滨				NNW									
海					NNW								
新						NNW							
区			THW-L			NNW							
人							NNW						
民								NNW					
检									NNW				
察										NNW			
院	THW-O												

图 2 基于网格的字符对关系建模方法的示例

Fig. 2 Example of a grid-based approach to character pair relationship modeling

2.2 双模态编码层

在编码层中, KaNER 模型通过结合 BERT 子词编码和位置补偿编码, 从语义和位置多个维度获得了中文表示。

预训练 BERT^[28]在命名实体识别任务中表现出色, 因此选择 BERT 模型获得子词级别的语义嵌入。给定一个输入句子 $X=\{x_1, x_2, \dots, x_N\}$, 首先将每个字符 x_i 转换为子词片段, 然后把子词片段输入到预训练的 BERT 模块中, 形成子词嵌入矩阵 E_{sub} 。

$$E_{sub} = \text{BERT}(X) \in \mathbb{R}^{M \times d} \quad (1)$$

式中: M 为子词单元的总数, d 为 BERT 输出的嵌入维度。

经过计算后, 一个字符可能被分割成多个子词, 导致其语义信息分散在多个子词向量中。由此采用最大池化操作, 将属于同一个字符的所有子词向量聚合为一个词级表示, 为

$$e_i = \max_{k \in \text{sub}(x_i)} E_{sub}[k] \in \mathbb{R}^d \quad (2)$$

式中: $\text{sub}(x_i)$ 表示字符 x_i 被分割成的子词单元的

索引集合, 用来保留关键的语义特征。

然而, 在某些情形之下, 预训练模型直接输出的句子表示可能无法充分捕获细粒度位置特征。为增强模型对实体边界位置的感知能力, 本文引入位置补偿编码机制, 该机制需要计算每个字符的累积位置偏移量 p_i 。

$$p_i = \sum_{k=1}^i \text{Length}(\text{sub}(x_k)) \quad (3)$$

式中: $\text{Length}(\text{sub}(x_k))$ 表示字符被分割成的子词单元数量。

之后, 位置偏移量 p_i 被转换为正弦位置编码

h_i^{enc} , 上标 enc 表示编码后的位置特征。为了更好地整合向量特征, 把经过最大池化聚合的 BERT 子词嵌入与位置补偿编码相加, 生成融合语义和位置信息的最终编码表示, 为

$$H^{\text{enc}} = e_i + h_i^{\text{enc}} \in \mathbb{R}^{N \times d} \quad (4)$$

该机制不仅恢复了字符的位置信息, 还增强了模型对实体边界的感知能力, 从而提升处理嵌套和间断实体的性能。

2.3 KAN 动态机制融合层

该层借助多频基函数动态融合以及位置感知样条建模的方式, 可自适应地捕捉字符对的全局长程依赖和局部语义关系, 这一做法相比传统方法有着明显优势。传统方法中固定卷积核或者静态注意力权重难以对字符间距给语义造成的非线性影响进行建模, 而 KaNER 采用 KAN 的样条函数与傅里叶基函数混合机制, 提升了中文命名实体识别的鲁棒性, 实现动态特征筛选与权重调整, 充分融合字符对的语义相似性与位置关联性。

2.3.1 动态相似性加权

在字符对关系建模中, KaNER 计算字符对之间的语义相似性和动态调整关系权重, 以此增强模型的语义理解能力。

语义相似性为

$$\text{Sim}_{\text{sem}}(i, j) = \cos(H_i^{\text{enc}}, H_j^{\text{enc}}) \quad (5)$$

式中: \cos 为余弦相似度, 用于衡量两个字符表示之间的语义相似性。

动态调整关系权重为

$$R_{ij} = \lambda \cdot \text{Sim}_{\text{sem}}(i, j) \quad (6)$$

式中: $\lambda = \sigma(W_\lambda [H_i^{\text{enc}}; H_j^{\text{enc}}])$, σ 是 sigmoid 激活函数, W_λ 是可学习的权重矩阵, 用于动态调整语义相似性的影响权重。

2.3.2 KAN 特征增强

模型的最终目标是预测字符对间的关系, 受 Hou 等^[26]动态关系捕捉机制的启发, 采用 KAN 自身优异的激活函数和样条插值方法获取特征增强的语义表示。构建字符对网格表示 $V \in \mathbb{R}^{N \times N \times 2d}$, 其中各元素

V_{ij} 表示字符对 (x_i, x_j) 的语义相似性和位置关系。具体来说, 通过特征拼接操作生成字符对的表示。

$$V_{ij} = H_i^{\text{enc}} \oplus H_j^{\text{enc}} \quad (7)$$

式中: \oplus 表示特征拼接操作, 将两个特征向量连接形成更高维度的向量。由此, 每个字符 x_i 与序列中的其他字符 x_j 构成的二维字符对网格, 同时保留原始字符特征及其交互信息。

该层通过多频基函数动态融合和位置感知样条建模, 增强词对网格表示中的全局和局部特征。多频基函数动态融合机制通过混合傅里叶基函数与样条激活建模频率成分, 位置感知样条建模直接利用字符间距作为输入扩展, 动态调整语义关系。

(1)多频基函数动态融合: 为捕捉全局和局部特征, 在 KAN 的激活函数中引入了傅里叶函数与样条的混合机制。通过将傅里叶基函数与样条函数结合, 直接在单变量层面建模频率特征。

$$\phi(x) = \sum_{m=0}^3 c_m B_m(x) + \sum_{\omega \in \Omega} (a_\omega \sin(\omega x) + b_\omega \cos(\omega x)) \quad (8)$$

式中: $B_m(x)$ 为 3 次 B 样条基函数, 负责捕捉局部

非线性特征; $\sin(\omega x)$ 和 $\cos(\omega x)$ 为预设频率集的基函数, 低频分量建模长距离依赖, 高频分量提取局部细节; c_m 、 a_ω 和 b_ω 为可学习参数, 动态平衡样条与傅里叶成分的贡献。

为动态筛选关键频率成分, 设计自适应权重机制。

$$V_{\text{freq}} = \sum_{\omega \in \Omega} \text{sigmoid}(w_{\omega}) \odot (V \cdot F_{\omega}) \quad (9)$$

式中： F_{ω} 为频率对应的特征映射矩阵， \odot 表示逐元素乘法；通过 $\text{Sigmoid}(w_{\omega}) \in (0,1)$ 归一化权重对有效频率成分进行筛选；获得全局增强特征 $V_{\text{freq}} \in \mathbb{R}^{N \times N \times 2d}$ ，通过学习不同频率成分的重要性，使KaNER模型能够根据任务需求动态地筛选出更有价值的特征成分。

(2)位置感知样条建模：为捕捉字符间距对语义关系的非线性影响，直接在KAN的边激活函数中引入位置感知。

$$\alpha_{ij} = \text{softmax}(\text{KAN}_{\text{edge}}(V_{ij}, |i-j|)) \quad (10)$$

其中， KAN_{edge} 的边激活函数定义为

$$\phi(x, d) = \sum_{m=0}^3 c_m B_m(x) + \sum_{n=0}^3 k_n B_n(d) \quad (11)$$

式中： x 为字符对的原始特征；定义 $d = |i-j|$ 为字符间距，通过独立样条基函数 $B_n(d)$ 建模位置依赖。

使用可学习参数 c_m 和 k_n 联合优化特征与位置的非线性映射。

通过加权聚合生成局部语义特征。

$$V_{\text{local}} = \sum_{i,j} \alpha_{ij} \cdot V_{ij} \in \mathbb{R}^{N \times N \times 2d} \quad (12)$$

通过可学习权重 β 实现局部和全局语义特征的最终融合。

$$V_{\text{enhanced}} = \beta V_{\text{freq}} + (1-\beta) V_{\text{local}} \quad (13)$$

式中： β 是一个介于0和1之间的自适应标量参数，用于控制全局特征和局部特征之间的相对贡献。

2.4 多头注意力整合

为有效整合KAN特征增强后的全局-局部特征和字符对关系建模的语义关联权重，该方法引入了多头注意力机制，实现对字符对交互特征进行多子空间建模。具体地，查询(Query)矩阵 Q 采用经过增

强特征 V_{enhanced} ，包含字符对的频率成分与间距依赖信息；键(Key)矩阵 K 源自动态语义关系权重 $R \cdot W_k$ ，通过可学习参数 W_k 将关系分数扩展为高维表示；值(Value)矩阵 V 结合增强特征与关系权重的联合信息 $V = V_{\text{enhanced}} + R$ 。

将 Q 、 K 、 V 沿特征维度均匀切分为 h 个注意力头，每个头独立计算注意力分数并聚合特征：

$$\text{Attention}_k = \text{softmax}\left(\frac{Q_k K_k^T}{\sqrt{d_k}}\right) V_k \in \mathbb{R}^{N \times N \times \frac{2d}{h}} \quad (14)$$

式中： $d_k = 2d/h$ 为单头特征维度，通过缩放点积抑制梯度异常。

将所有头的输出拼接在一起，通过线性投影融合多子空间特征：

$$H_{\text{fused}} = W_o \cdot \text{Concat}(A_1, A_2, \dots, A_h) \quad (15)$$

式中： $W_o \in \mathbb{R}^{2d \times d}$ 是一个可学习的权重矩阵，用于将拼接后的特征映射到目标维度并保留核心语义信息； A_k 表示第 k 个头的注意力聚合特征。

为适配嵌套与不连续实体的解码需求，将序列特征 H_{fused} 重构为字符对网格矩阵 Y 。针对每对字符对，通过外积运算显式建模其联合特征。

$$y_{ij} = W_{\text{grid}} \cdot \text{vec}(H_{\text{fused}}[i] \otimes H_{\text{fused}}[j]) \quad (16)$$

式中： \otimes 表示外积操作， W_{grid} 为可学习参数矩阵。

最终的网格矩阵 $Y \in \mathbb{R}^{N \times N \times d_t}$ ，其中 d_t 是融合后的特征维度。这个网格矩阵 Y 集成全局语义关联与局部结构特征，为协同预测层提供细粒度交互表示。

2.5 协调预测层

2.5.1 协调预测器

协同预测器通过融合多层感知机(MLP)和双仿射机制^[29]，动态建模字符对的复杂语义关系。输入为网格矩阵 Y ，其每个元素表征字符对之间的语义、位置及频域交互特征。

MLP捕捉每对字符 (x_i, x_j) 之间的关联，通过多层

非线性变换捕捉字符对的深层交互特征。

$$Y'_{ij} = \text{MLP}(Y_{ij}) \quad (17)$$

同时, 在双仿射预测器中, 对网格矩阵进行残差连接设计, 分别对字符*i*和*j*提取独立特征。

$$s_i = \text{MLP}(Y'_{ij}[i]) \quad (18)$$

$$o_j = \text{MLP}(Y'_{ij}[j]) \quad (19)$$

通过双仿射机制计算关系分数。

$$E'_{ij} = s_i^T U o_j + W[s_i; o_j] + b \quad (20)$$

式中: 其中 U 、 W 、 b 为可训练参数, 编码字符对的联合语义空间映射。

最终将 MLP 和双仿射预测器得出的分数进行融合, 生成归一化关系概率分布。

$$y_{ij} = \text{Softmax}(Y'_{ij} + E'_{ij}) \quad (21)$$

式中: $y_{ij} \in \mathbb{R}^{|R|}$ 为预定义关系集合 R 中的关系分数。

对输入序列, KaNER 模型优化的目的在于最小化结构化负对数似然损失函数。该损失函数衡量字符对关系预测分布与真实标签的差异, 其数学形式为

$$\mathcal{L} = -\frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \sum_{r=1}^{|R|} \hat{y}_{ij}^r \log y_{ij}^r \quad (22)$$

式中: N 为输入序列的字符数量, R 为预定义语义关系集合包含无关联关系类别, \hat{y}_{ij}^r 是字符对的真实标签向量, y_{ij}^r 为模型输出的归一化概率分布向量。

2.5.2 解码层

基于 KaNER 模型关系的定义, 通过 NNW 关系维护实体内部连续性, THW-*关系提供类型化边界约束与消歧信息, 两者协同支持复杂实体结构的自适应解析, 如图 3 所示。

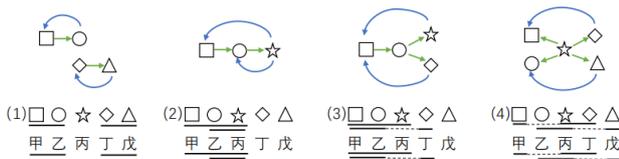


图 3 字符序列“甲乙丙丁戊”的解码情形

Fig. 3 Decoding of the character sequence “甲乙丙丁戊”

平面实体识别: 路径“甲→乙”与“丁→戊”

通过 NNW 关系连接, 对应实体“甲乙”和“丁戊”, THW-*关系显式标记边界与类型, 避免“甲→乙→丁→戊”的误跨距连接。

嵌套实体识别: 仅依赖 NNW 关系时, 路径“甲→乙→丙”解析为单一实体“甲乙丙”, 引入 THW-*关系后, “乙→丙”的局部边界激活, 拆分出嵌套实体“乙丙”。

不连续实体识别: 路径“甲→乙→丙”与“甲→乙→丁”通过 NNW 关系生成, 但实体“甲乙丁”为不连续实体, 则利用 NNW 关系连接不连续的语义片段, 构成“甲→乙”与“乙→丁”的跨距关联。

重叠结构协调筛选: 不连续实体“甲丙丁”与“乙丙戊”需要通过 NNW 与 THW-*关系的有效组合, 协同筛选出路径“甲→丙→丁”与“乙→丙→戊”, 若仅使用 NNW 关系只能识别出“甲乙丙丁”而非正确实体。

3 实验

3.1 实验设置

3.1.1 数据集

为了验证 KaNER 模型的效果, 在 7 个数据集上对模型性能进行全面评估, 包括 4 个中文数据集: Resume 数据集、Weibo 数据集、MSRA 数据集以及 OntoNotes 新闻数据集^[30-34]。Resume 数据集包含 1027 份来自新浪财经的简历文档。Weibo 数据集则采集自社交媒体平台。MSRA 数据集是微软公司发布的新闻语料库, 该数据集未提供验证集, 直接通过测试集对模型性能进行评估。此外, 同时引入了 2 个中文嵌套数据集和 1 个不连续 NER 数据集。针对嵌套的 NER 任务, 采用 ACE2004^[35]、ACE2005^[36], 将训练集、验证集、测试集按照 8: 1: 1 的比例进行划分。针对非连续命名实体识别任务的评估, 选择 ND-NER 语料库^[37]作为基准数据集(包含 7% 不连续实体)。每个数据集的具体情况详见表 1 和表 2。

表 1 中文平面 NER 数据集统计

Tab. 1 Statistics of Chinese flat NER dataset

数据集	句子数			
	训练集	验证集	测试集	总计
OntoNotes	15736	4306	4351	24393
MSRA	46471	—	4376	50847
Weibo	1350	270	270	1890
Resume	3819	463	477	4759

3.1.2 参数设置

采用的超参数配置具体内容见表 3, 在语义编码阶段, 借助 BERT-base 和 BERT-large 架构获取字符级嵌入, 这两种架构所得到的维度分别是 768 和 1024, 位置补偿编码模块会构建 256 维的正弦波位置向量以增强边界感知能力。在 KAN 动态特征提高层, 将隐藏维度设置为 512, 运用 3 组 3 次样条基函数对字符间距非线性依赖关系进行建模。动态相似性加权模块通过可学习参数矩阵优化注意力权重,

网格解码器凭借外积操作把特征投影到 1024 维目标空间, 训练阶段利用随机失活率抑制过拟合, 配置 8 头注意力机制以及 2 层堆叠架构用以实现特征整合。为验证模型鲁棒性, 实验在 PyTorch 框架下^[38]部署 AdamW 优化器, 选用 NVIDIA A100 GPU 进行多种子并行训练, 每个种子独立运行 10 轮次, 最终指标是 30 次实验的平均值, 以此保证结论有统计显著性。

表 2 中文嵌套和不连续 NER 数据集统计

Tab. 2 Statistics of Chinese nested and discontinuous NER datasets

数据集	句子数				实体数		
	训练集	验证集	测试集	总计	嵌套	不连续	总计
ACE2004	5754	721	660	7135	15219	—	15219
ACE2005	5876	660	740	7276	15734	—	15734
ND-NER	6021	2007	2007	10035	21768	3628	25396

表 3 超参数设置

Tab. 3 Hyperparameter settings

超参数	取值
d_w	[768, 1024]
d_p	256
d_s	512
k_s	3
d_t	1024
随机失活率	[0.3, 0.4, 0.5]
多头注意力(头数, 层数)	8 头, 2 层
随机种子	[42, 58, 256]
学习率	[1×10^{-5} , 5×10^{-5}]
批次大小	[8, 12, 16]

3.1.3 评价指标

为对模型性能展开全面评估, 采用了精确率 Precision(P)、召回率 Recall(R)以及 F_1 值 3 种核心评价指标。精确率指的是, 在模型预测出的正类实体当中, 真正属于正类的那部分所占的比例, 其重点在于呈现模型针对实体边界以及类型判定的准确程度。召回率指的是, 模型可正确识别出来的正类实体, 在所有实际存在的正类实体里所占的比例, 它可体现模型对有挑战性的嵌套实体以及不连续实体的识别能力。 F_1 值作为精确率与召回率的调和平均数, 可以平衡两者可能出现的此消彼长的关系, 给模型综合性能提供一个更全面的评估角度。

$$P = \frac{N_{TP}}{N_{TP} + N_{FP}} \times 100\% \quad (23)$$

$$R = \frac{N_{TP}}{N_{TP} + N_{FN}} \times 100\% \quad (24)$$

$$F_1 = \frac{2 \times P \times R}{P + R} \times 100\% \quad (25)$$

式中: N_{TP} 表示被正确识别为正类的样本数量, N_{FP} 表示被错误分类为正类的负样本数量, N_{FN} 表示被错误预测为负类的正样本数量。

3.1.4 基线模型

基于词汇增强的方法, 通过融合词典信息强化字符表示, 例如 SoftLexicon 的词典特征融合^[38], 以及 GS-Lexicon 的自适应动态词向量权重计算词汇融合机制^[39]。

基于多模态预训练的方法联合字形与拼音信息进而增强语义表征, 例如 ChineseBERT^[23]的异构特征编码框架。

基于生成式建模的方法, 采用序列生成或去噪框架生成实体信息, 例如 MGBERT^[27]的全局指针机制, 与 DiffusionNER^[26]的迭代扩散优化。

基于边界优化的方法, 通过自蒸馏或多粒度学习提升实体识别精度, 例如 MSE 的边界感知蒸馏策略^[40]。

基于网格标注的方法, 显式建模字符对关系矩阵, 例如 W^2 NER^[24]的扩展卷积架构, mixed-scale 的粗细粒度融合表征^[41], 以及 HiNER^[25]的句法 - 语义混合注意力融合。

3.2 实验结果

在4个平面NER数据集上进行对比实验, 结果见表4。以SoftLexicon为代表的词汇提高传统方法, 尽管借助融合词典信息强化了字符表示, 但其对静态词表的依赖容易引入噪声干扰。后续提出的GS-Lexicon利用增强的全局注意力计算词-句关联性, 显著优化词典融合质量。然而实验显示, 在

OntoNotes数据集上, KaNER的 F_1 值仍显著优于这类模型, 尤其在噪声敏感的Weibo数据集领先1.68%, 印证了KAN特征增强模块的动态频率筛选机制对非规范文本的适应性优势。

表4 在平面数据集上的评估结果

Tab. 4 Evaluation results on a flat dataset

模型	OntoNotes			MSRA			Resume			Weibo		
	P/%	R/%	F_1 /%	P/%	R/%	F_1 /%	P/%	R/%	F_1 /%	P/%	R/%	F_1 /%
SoftLexicon	83.37	82.20	82.77	95.70	95.12	95.38	96.05	96.11	96.08	70.89	67.05	70.46
GS-Lexicon	83.38	82.55	82.95	95.77	95.28	95.51	96.29	96.36	96.32	72.28	70.13	71.04
ChineseBERT	80.05	83.28	81.62	95.16	94.62	94.88	95.97	94.42	95.70	68.26	69.80	69.17
MGBERT	83.02	83.11	83.05	95.98	96.21	96.08	96.21	96.29	96.24	71.93	72.55	72.22
MSE	82.74	83.20	82.96	96.29	96.14	96.21	96.94	96.75	96.84	72.23	71.92	72.07
W ² NER	82.33	83.38	83.07	96.11	96.07	96.09	96.95	96.37	96.65	70.86	73.89	72.34
mixed-scale	83.15	83.28	83.21	96.45	96.60	96.52	96.75	96.84	96.79	72.18	73.02	72.60
HiNER	82.88	83.46	83.15	96.34	96.74	96.54	96.85	96.83	96.84	71.53	73.86	72.65
KaNER	82.92	83.61	83.26	96.57	96.82	96.69	96.89	96.96	96.92	72.01	73.45	72.72

与MGBERT、ChineseBERT等依赖多模态预训练的模型相比, KaNER模型展现出更优越的性能。这种改进可归因于KaNER中的多头注意力模块, 该模块通过并行捕捉字符对的多维度语义关系, 在减少参数冗余的同时提升特征表达能力。在MSRA数据集上, KaNER以96.69%的 F_1 超越MGBERT和ChineseBERT。此外, 进一步地, 针对简历类文本的密集短实体识别, KaNER通过位置感知样条建模精准定位实体边界, 召回率较MGBERT略有提升, 验证局部非线性映射对边界模糊实体的敏感性。

面对mixed-scale提出的粗细粒度融合表征与HiNER的句法-语义混合注意力机制, KaNER通过动态特征融合与噪声抑制机制, 解决了传统方法在长距离依赖建模与语义冗余引入上的局限性。KaNER引入KAN多频基函数融合机制, 动态平衡低频全局特征与高频局部特征。在MSRA和Resume数据集上, 分别以96.69%和96.92%的 F_1 值超越mixed-scale、HiNER和W²NER。另外两个平面数据集中, KaNER虽因短实体边界清晰提升有限, 但通过多头注意力模块对细粒度语义共现的捕捉, 精准率仍保持领先。

不同方法在ACE2004、ACE2005和ND-NER数据集的评价结果如图4所示。KaNER模型在ACE2004和ACE2005数据集上分别以88.41%和

88.94%的 F_1 值刷新性能记录, 较此前最优模型HiNER有所提升, 显著超越基于标记的TOE^[35]与基于网格的W²NER。这一改进得益于动态频率筛选模块对跨层嵌套结构的建模能力, 通过低频傅里叶基函数捕捉长距离层级依赖, 结合位置补偿编码的样条边界定位从而减少跨层实体混淆误差; 而生成式框架DiffusionNER因迭代去噪机制难以建模深层嵌套实体, 性能受限。

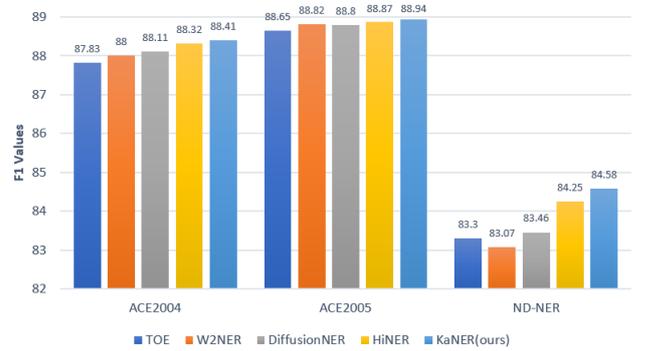


图4 在ACE2004、ACE2005和ND-NER数据集的评价结果

Fig. 4 Evaluation results on ACE2004, ACE2005 and ND-NER datasets

在中文不连续实体识别任务里, KaNER凭借84.58%的 F_1 值超越了现有的方法, 相较于HiNER

提升了 0.33%，相较于生成式 DiffusionNER 提升了 1.12%，这得益于多头注意力整合模块与动态关系权重机制共同发挥作用。与之相比，W²NER 依靠静态卷积，难以捕捉远距离片段之间的交互，TOE 则受限于序列标注框架的连续性假设，这也验证了 KaNER 在不连续场景下有鲁棒性和泛化性。

在表 5 的长句案例分析示例中，KaNER 模型凭借 KAN 特征增强与动态注意力机制的协同优化，提升了对复杂实体结构的识别能力。它精准解析了多层次嵌套结构，可有效区分内外层实体边界，完整抽取实体“天津市滨海新区人民检察院”。相较之下，HiNER 模型存在符合结构丢失与长距离依赖建模不足的问题，验证了本模型在实体感知和边界定位方面的技术优势。

表 5 长句案例分析示例

Tab. 5 Analysis of long sentence cases

项目	内容
示例句	[¹ 天津市] ¹ LOC[² 滨海新区] ² LOC[³ 人民检察院] ³ ORG 近期处理了市民[¹ 李明] ¹ PER 反映的，关于[¹ 某科技公司] ² ORG 推出的[² 智能设备] ² PRODUCT ¹ PRODUCT 相关问题的咨询，李明曾在[⁴ 南京大学] ⁴ ORG[¹ 物理系] ¹ DEPT ³ ORG 学习。
KaNER 结果	[天津市, LOC], [滨海新区, LOC], [人民检察院, ORG], [天津市滨海新区人民检察院, ORG], [某科技公司, ORG], [南京大学, ORG], [李明, PER], [南京大学物理系, ORG], [物理系, DEPT], [智能设备, PRODUCT]
HiNER 结果	[天津市, LOC], [滨海新区, LOC], [人民检察院, ORG], [某科技公司, ORG], [南京大学, ORG], [李明, PER], [物理系, DEPT], [智能设备, PRODUCT]

3.3 消融实验

KaNER 模型通过 KAN 特征增强模块动态融合字符对的局部位置与全局频率特征，并通过字符对关系建模与多头注意力整合实现多粒度语义交互。为验证各模块的有效性，在中文平面实体数据集 Resume 与嵌套实体数据集 ACE 2004 上进行消融实验，了完整模型与移除关键模块后 F₁ 值的差异见表 6。

表 6 消融实验结果对比

Tab. 6 Comparison of ablation experiment results

方法	F ₁ /%	
	Resume	ACE2004
KaNER	96.92	88.41

w/o KAN	95.80	86.59
w/o KAN-多频基函数动态融合	96.11	87.24
w/o KAN-位置感知样条建模	96.36	88.01
w/o 动态相似性加权	96.23	87.35
w/o 多头注意力整合	96.45	87.82

4 结 论

本文提出一种中文命名实体识别框架 KaNER，该框架基于字符对关系分类，借助动态融合多频基函数以及位置感知样条建模来强化字符对表示，还结合动态相似性加权与多头注意力机制达成全局与局部语义协同优化，面对中文文本里实体边界模糊、嵌套以及不连续结构等难题，KaNER 设计了 KAN 驱动的特征提高模块，提高了长距离依赖建模与噪声关联抑制能力。实验结果显示：该模型在 OntoNotes、MSRA、Resume 和 Weibo 这 4 个中文数据集以及 ACE2004、ACE2005 等嵌套实体数据集上均取得当前最优性能，最高 F₁ 提升了 1.82%。消融研究证实了多频基函数与位置感知样条、动态语义权重调整和注意力整合模块的有效性，KAN 特征提高对嵌套实体的解析贡献非常关键。该框架为中文复杂实体识别提供了轻量化且可扩展的解决办法，有望促进中文信息处理技术在知识图谱构建、智能问答等领域的应用。

参考文献:

- [1] NADEAU D, SEKINE S. A survey of named entity recognition and classification[J]. *Linguisticae investigationes*, 2007, 30(1): 3-26.
- [2] LI J, SUN A X, HAN J L, et al. A survey on deep learning for named entity recognition[J]. *IEEE Transactions on knowledge and data engineering*, 2020, 34(1): 50-70.
- [3] ZHANG H, BAI L Y. Few-shot link prediction for temporal knowledge graphs based on time-aware translation and attention mechanism[J]. *Neural networks*, 2023, 161: 371-381.
- [4] DIEFENBACH D, LOPEZ V, SINGH K, et al. Core techniques of question answering systems over knowledge bases: a survey[J]. *Knowledge and information systems*, 2018, 55: 529-569.
- [5] 李轩舟. 舆情背景下的多标签文本分类和通用命名实体

- 识别研究[D]. 上海:华东师范大学,2023.
- [6] SCHICK T, SCHMID H, SCHÜTZE H. Automatically identifying words that can serve as labels for few-shot text classification[EB/OL]. (2020-10-26)[2025-06-12]. <https://arxiv.org/abs/2010.13641>.
- [7] SUNG M, JEON H, LEE J, et al. Biomedical entity representations with synonym marginalization[EB/OL]. (2020-05-01)[2025-06-12]. <https://arxiv.org/abs/2005.00239>.
- [8] REIMERS N, GUREVYCH I. Optimal hyperparameters for deep LSTM-networks for sequence labeling tasks[EB/OL]. (2017-08-16)[2025-06-12]. <https://arxiv.org/abs/1707.06799>.
- [9] PYATKIN V, ROIT P, MICHAEL J, et al. Asking it all: generating contextualized questions for any semantic role[EB/OL]. (2021-09-10)[2025-06-12]. <https://arxiv.org/abs/2109.04832>.
- [10] ZHANG H X, ZHANG X F, HUANG H B, et al. Prompt-based meta-learning for few-shot text classification[C]// Association for Computational Linguistics. Proceedings of the 2022 conference on empirical methods in natural language processing. Abu Dhabi:Association for Computational Linguistics, 2022: 1342-1357.
- [11] WANG S, SUN X, LI X, et al. GPT-NER: named entity recognition via large language models[EB/OL]. (2023-10-07)[2025-06-12]. <https://arxiv.org/abs/2304.10428>.
- [12] MO Y, LI Z J. Incorporating entity type-aware and word-word relation-aware attention in generative named entity recognition[J]. Electronics, 2024, 13(7): 1407.
- [13] CHEN Y, HUANG Y, ZHANG H. Domain-Oriented Chinese Named Entity Recognition Based on Enhanced Word-Word Relation Classification[C] //ZHANG H P, SU J S, SHANG J Y. Intelligent Multilingual Information Processing: First International Conference. Singapore: Springer Nature, 2024: 174-188.
- [14] TIAN X T, BU X X, HE L. Multi-task learning with helpful word selection for lexicon-enhanced Chinese NER[J]. Applied intelligence, 2023, 53(16): 19028-19043.
- [15] ZHU P, CHENG D W, YANG F Z, et al. Improving Chinese named entity recognition by large-scale syntactic dependency graph[J]. IEEE/ACM transactions on audio, speech, and language processing, 2022, 30: 979-991.
- [16] LI H C, ZHENG J M, JIN B H, et al. Adaptive knowledge contrastive learning with dynamic attention for recommender systems[J]. Electronics, 2024, 13(18): 3594.
- [17] TOLSTIKHIN I O, HOULSBY N, KOLESNIKOV A, et al. MLP-Mixer: an all-mlp architecture for vision[J]. Advances in neural information processing systems, 2021, 34: 24261-24272.
- [18] YANG X Y, Wang X C. Kolmogorov-arnold transformer[EB/OL]. (2024-09-16)[2025-06-04]. <https://arxiv.org/abs/2409.10594>.
- [19] GENET R, INZIRILLO H. TKAN: temporal Kolmogorov-Arnold Networks[EB/OL]. (2024-05-12)[2025-06-04]. <https://arxiv.org/abs/2405.07344>.
- [20] BODNER A D, TEPSICH A S, SPOLSKI J N, et al. Convolutional kolmogorov-arnold networks[EB/OL]. (2025-03-31)[2025-06-04]. <https://arxiv.org/abs/2406.13155>.
- [21] ZHANG F, ZHANG X. GraphKAN: enhancing feature extraction with graph kolmogorov arnold networks[EB/OL]. (2024-06-19)[2025-06-04]. <https://arxiv.org/abs/2406.13597>.
- [22] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[EB/OL]. (2023-08-02)[2025-06-04]. <https://arxiv.org/abs/1706.03762>.
- [23] SUN Z J, LI X Y, SUN X F, et al. ChineseBERT: Chinese pretraining enhanced by glyph and pinyin information[EB/OL]. (2021-06-30)[2025-06-04]. <https://arxiv.org/abs/2106.16038>.
- [24] Li J Y, Fei H, Liu J, et al. Unified named entity recognition as word-word relation classification[J]. Proceedings of the AAAI conference on artificial intelligence, 2022, 36(10): 10965-10973.
- [25] Hou S X, Qian Y R, Chen J Y, et al. HiNER: hierarchical feature fusion for Chinese named entity recognition[J]. Neurocomputing, 2025, 611: 128667.
- [26] SHEN Y L, SONG K T, TAN X, et al. DiffusionNER: boundary diffusion for named entity recognition[EB/OL]. (2023-05-22)[2025-06-04]. <https://arxiv.org/abs/2305.13298>.
- [27] ZHANG L, XIA P F, MA X X, et al. Enhanced Chinese named entity recognition with multi-granularity BERT adapter and efficient global pointer[J]. Complex & intelligent systems, 2024, 10(3): 4473-4491.
- [28] DEVLIN J, CHANG M W, LEE K, et al. BERT: Pre-training of deep bidirectional transformers for language

- understanding[C]//BURSTEIN J, DORAN C, SOLORIO T. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Minneapolis :Association for Computational Linguistics, 2019: 4171-4186.
- [29] DOZAT T, MANNING C D. Deep biaffine attention for neural dependency parsing[EB/OL]. (2017-03-10)[2025-06-04]. <https://arxiv.org/abs/1611.01734>.
- [30] LEVOW G A. The third international Chinese language processing bakeoff: Word segmentation and named entity recognition[C]//Association for Computational Linguistics. Proceedings of the Fifth SIGHAN Workshop on Chinese Language Processing. Sydney:Association for Computational Linguistics, 2006: 108-117.
- [31] WEISCHEDEL R, PRADHAN S, RAMSHAW L, et al. Ontonotes release 4. 0[EB/OL]. [2025-06-04]. <https://catalog.ldc.upenn.edu/docs/LDC2011T03/OntoNotes-Release-4.0.pdf>.
- [32] ZHANG Y, YANG J. Chinese NER using lattice LSTM [EB/OL]. (2018-07-05)[2025-06-04]. <https://arxiv.org/abs/1805.02023>.
- [33] PENG N, DREDZE M. Named entity recognition for chinese social media with jointly trained embeddings[C]// Association for Computational Linguistics. Proceedings of the 2015 conference on empirical methods in natural language processing. Lisbon: association for Computational Linguistics , 2015: 548-554.
- [34] DODDINGTON G R, MITCHELL A, PRZYBOCKI M A, et al. The automatic content extraction (ace) program-tasks, data, and evaluation[EB/OL]. [2025-06-04]. <https://course.ccs.neu.edu/csg224/resources/muc/ACE.pdf>.
- [35] WALKER C, STRASSEL S, MEDERO J, et al. ACE 2005 multilingual training corpus[DS/OL]. [2025-06-04]. <https://www.seas.upenn.edu/~why16gZL/>.
- [36] LI X Y, LI D X, YANG Z H, et al. ND-NER: a named entity recognition dataset for osint towards the national defense domain[C]// Tanveer M, Agarwal S, Ozawa S, et al. International Conference on Neural Information Processing. Singapore: Springer Nature, 2022: 361-372.
- [37] LOSHCHILOV I, HUTTER F. Decoupled weight decay regularization [EB/OL]. (2019-01-04)[2025-06-04]. <https://arxiv.org/abs/1711.05101>.
- [38] MA R, PENG M, ZHANG Q, et al. Simplify the usage of lexicon in Chinese NER[EB/OL]. (2020-10-14)[2025-06-04]. <https://arxiv.org/abs/1908.05969>.
- [39] XU Y P, YING M T, FANG K Y, et al. Chinese named entity recognition based on adaptive lexical weights[J]. Computer speech & language, 2025, 90: 101735.
- [40] LIU J, JI D H, LI J Y, et al. TOE: a grid-tagging discontinuous NER model enhanced by embedding tag/word relations and more fine-grained tags[J]. IEEE/ACM transactions on audio, speech, and language processing, 2022, 31: 177-187.
- [41] GUO X, CHEN Y P, TANG R X, et al. Camouflaged named entity recognition in 2D sentence representation[J]. Expert systems with applications, 2024, 257: 125096.