

基于BERT的机构名命名实体细粒度实体识别研究

朱 森¹, 刘松林^{2*}, 高一品¹

¹曲阜师范大学, 山东 日照

²信息工程大学, 河南 郑州

收稿日期: 2022年8月17日; 录用日期: 2022年9月21日; 发布日期: 2022年9月30日

摘 要

机构名识别是命名实体识别的核心任务之一, 其识别效果的优劣极大地影响着下游知识图谱构建、意图搜索等任务。现有的机构名实体识别在工作中存在着实体标注粒度通常过粗, 识别率低的问题。基于以上, 提出一种不依赖人工特征选择和领域知识的端到端细粒度识别模型(BERT-BiLSTM-CRF)。该模型在传统实体识别模型BiLSTM-CRF的基础上引入BERT预训练模型, 通过引入BERT预训练语言模型获取动态字向量表示, 有效解决了由于缺乏大量训练语料而无法获得丰富语义信息的问题, 然后将BERT和BiLSTM-CRF模型联合起来进行实体识别, 学习语料上下文信息和标签转移关系, 最终获得机构名细粒度类别标签。同时, 根据实际应用需求重新划分了机构名的类别, 对数据集进行了细粒度的语料标注。实验表明, 该模型在使用的训练数据集上, F1值最佳, 达到了92.41%, 相较于主流序列标注模型准确率更高, 具有潜在应用价值。

关键词

BERT模型, 机构名实体识别, 端到端模型, 自然语言处理

Research on Fine-Grained Recognition of Organization Named Entities Based on BERT Model

Miao Zhu¹, Songlin Liu^{2*}, Yipin Gao¹

¹Qufu Normal University, Rizhao Shandong

²Information Engineering University, Zhengzhou Henan

Received: Aug. 17th, 2022; accepted: Sep. 21st, 2022; published: Sep. 30th, 2022

*通讯作者。

Abstract

Organization name recognition is one of the core tasks of named entity recognition, and its recognition effect greatly affects the downstream tasks, like knowledge map construction, intention search and so on. There are many problems in the existing entity recognition research of organization names, such as the granularity of entity annotation is usually too coarse and the recognition rate is low. Based on the above, an end-to-end fine-grained recognition model (BERT-BiLSTM-CRF) independent of artificial feature selection and domain knowledge is proposed. Based on the traditional entity recognition model BiLSTM-CRF, this model introduces the BERT pre-training model. By introducing the BERT pre-training language model to obtain the dynamic word vector representation, it effectively solves the problem that it is unable to obtain rich semantic information due to the lack of a large number of training corpus. Then it combines the BERT and BiLSTM-CRF models for entity recognition to learn the corpus context information and label transfer relationship. Finally, the fine-grained category label of the institution name is obtained. At the same time, according to the needs of practical application, the category of organization name is re-divided, and the data set is marked with fine-grained corpus. Experiments show that the F1 value of this model is the best in the training data set, reaching 92.41%. Compared with the mainstream sequence annotation models, the proposed model in this paper has higher accuracy and great potential for practical application.

Keywords

BERT Model, Organization Name Recognition, End-To-End Model, Natural Language Processing

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

命名实体识别(Named Entity Recognition, NER)是信息抽取的重要环节之一,其目的是从文本中识别出命名实体的边界并判断其所属的实体类型。传统实体识别任务主要集中在人名、地名、机构名的识别上,但是,由于机构名具有罕见词多、结构复杂、名称差异性较大等问题,和其他类别相比,识别准确率较低[1]。

机构名在日常生活中覆盖面极广且划分领域细致,通常分为政府机关、医疗机构、科教服务场所、餐饮等多个类型,因此,对机构名进行细粒度实体识别更符合现实世界的知识体系。通过在已有类别基础上进行进一步的类别划分,使得从文本中抽取出来的实体有更为详细的定义,为下游兴趣点搜索、意图识别、面向导航系统等实际任务提供更有力的支撑。

命名实体识别主要有基于规则的、基于统计的和基于深度学习的方法。基于规则的方法通过分析实体的构成特点和上下文结构,人工构造大量的特定规则,再从语料中匹配符合这些规则的实体,人工成本高,可移植性低。基于机器学习的常见方法有隐马尔可夫模型和条件随机场[2]等,文献[3]采用规则与条件随机场相结合对中文地名、机构名进行识别并取得了不错的效果。但是,这种方法对于语料库的依赖较大,而且需要语料具有一定的特征结构[4]。

近年来,基于神经网络的方法在实体识别任务中取得了较大的成功,文献[5]用双向长短时记忆网络BiLSTM自动学习文本的词级和字符级表示,在CoNLL-2003数据集上取得了F1值90.94%的先进水平成绩,该方法能够自行从数据中学习出文本序列的特征,不需要人工设计规则模板。相比于英文语系,中文

词边界不明显, 由于分词所产生误差将会影响实体识别的效果[6], 因此, 中文命名实体识别可以转化为字符级别的序列标注任务。文献[7]表明, 基于字符级别未登录字识别效果较词级别提升了 11.05%。基于字符的输入虽然避免了引入噪声, 但同时丢失了词汇信息、时序信息等特征, 无法表示中文语境种的一词多义现象。文献[8]将分词模型和实体识别模型进行联合训练, 实验证明, 关注词汇可以有效提高对社交媒体信息这类不规则文本的实体识别效果。随着词嵌入模型迅速发展, ALBERT、ELMO [9]、BERT 等预训练语言模型被运用到序列标注领域, 文献[10]提出了基于 BERT 嵌入的中文实体识别方法, 在人民日报语料上取得了 94.86% 的 F1 值。文献[11]基于自建语料, 对中文专业术语进行识别, 将基于统计文本特征模型、word2vec 字嵌入和 BERT 预训练的字嵌入进行对比, 证实基于 BERT 的字嵌入向量能更好地表征字符的语义和语句特征。此外, BERT 模型也被成功应用到地名[12]、医药[13] [14]、食品[15]等领域的实体识别任务之中。

目前, 考虑到机构名细粒度实体识别存在的问题: ① 部分实体长度较长, 上下文信息存在远距离依赖。② 一词多义现象明显, 罕见词较多。③ 细粒度分类后存在实体嵌套情况, 例如“王府井商场停车场”错误分词会导致实体被切分, 类型判断错误影响模型识别准确率。本文采用一种基于 BERT-BiLSTM-CRF 的实体识别模型, 该模型选择 BERT 作为获取机构名特征向量的预训练方法, 结合 BiLSTM 对上下文信息的记忆能力和 CRF 对标签的约束能力, 在完整保留语义信息的基础上, 提升模型对中文命名实体的识别准确率, 并在完成实体识别任务的同时完成了实体分类问题。

2. 机构名语料集特点

本文实验使用的数据来自北京市高德地图 POI 数据, 关注的是 POI 的名称, 即一句话中是否含有 POI 名称并判断其所属的品类。根据实际应用需求结合 POI 数据中标签划分类型, 本文将传统领域的机构名进一步细粒度划分成了 12 个类别, 分别为交通设施服务、休闲服务、住宿服务、公司企业、医疗服务、政府机构、生活服务、社会团体、科教服务、金融服务、风景名胜和餐饮服务。数据集统计情况见表 1。

Table 1. Statistical table of the number of organizations and entities in the data set

表 1. 数据集组织机构实体数目统计表

标签类型	数量	样例
交通设施服务	7974	正阳北里停车场
休闲服务	2817	进启采摘园
住宿服务	1731	北京欣双环大旅社
公司企业	12,982	东方瑞丰公司
医疗服务	2550	康达门诊
政府机构	3657	吉祥里社区党委
生活服务	13,186	二院快递驿站
社会团体	1978	常营地区温馨家园
科教服务	5930	张坊中心小学
金融服务	2912	平安银行
风景名胜	1791	金星公园
餐饮服务	9716	京食饭店
总计	67,224	

不难发现，现实生活中对机构名的实体类别划分通常是细粒度的。当用户搜索“肯德基(王府井大街店)”通常是对“王府井大街附近的餐饮信息”感兴趣，而在常规 NER 任务中通常粗粒度的划分为“地名类”或者“机构名类”，这并不适用于用户的实际需求，因此，对机构名进行细致的分类可以更贴切的向用户推荐合适的信息。

此外，机构名的语料表述存在着较大的复合性和嵌套性，例如“全聚德朝阳旗舰店”、“肯德基(王府井大街店)”等餐饮服务实体词，可细化为餐饮店名称、餐饮店所在地、餐饮店规模等复合型实体词；或者在表述时具有明显的不同类别实体嵌套特点，如“天安门停车场”、“故宫公交站”等实体前半部分属于风景名胜而后半部分属于交通设施服务，其实体词含义和类别需要结合上下文来分析理解。

同时，机构名表述的结构也多种多样，如“快递驿站”表述为“菜鸟驿站”、“快递点”、“菜鸟”等；部分机构名完全不具有指代作用，如“蚝先生”、“地磅”、“赛科龙”等；同时机构名可能参杂数字、字符及特殊符号，如“colala2 号店”、“.com 服务中心”等。

综上所述可以看出，机构名语料在实际应用中存在细粒度类别划分、复合性和嵌套性强、语言表述不规范等特点，给识别及划分类型过程带来巨大的困难。

3. 基于 BERT-BiLSTM-CRF 的机构名实体识别方法

针对机构名实体对象的特点，本文采用 BERT-BiLSTM-CRF 模型的实体识别方法解决机构名识别与细粒度划分问题。

3.1. BERT-BiLSTM-CRF 模型

本文采用的 BERT-BiLSTM-CRF 模型分为三部分，具体模型结构如图 1 所示。

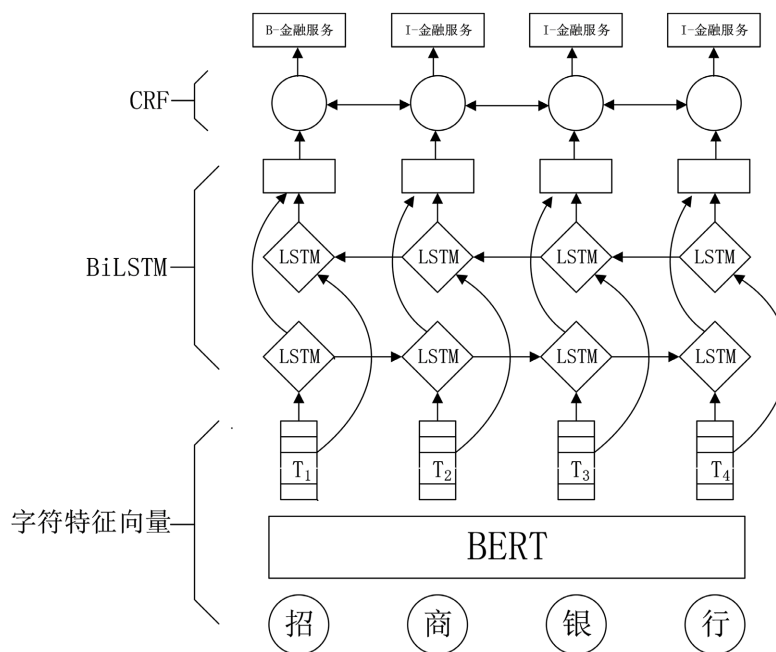


Figure 1. BERT-BiLSTM-CRF model structure diagram

图 1. BERT-BiLSTM-CRF 模型结构图

假定输入语料 $X = (x_1, x_2, \dots, x_n)$ ，其中， x_i 表示语料 X 中第 i 个字。模型第一层利用预训练语言模型 BERT 获取语料 X 的字符级别特征向量序列 $T = (T_1, T_2, \dots, T_n)$ ，其中， T_i 为字符 x_i 所对应的字符特征向量。

在输入阶段,为 X 中的每个字符 x_i 依次叠加句嵌入Eis、字符嵌入Eic和位置嵌入Eip形成字向量 E_i ,将其送入模型得到具有全局语义信息的输出序列 T 。

模型第二层为BiLSTM层,由正向和反向的LSTM组成。第一层获取的 n 维字符向量作为双向LSTM神经网络各个时间步的输入,经正向LSTM后输出的隐状态序列为 \vec{h} ,经反向LSTM后输出的隐状态序列为 \overleftarrow{h} ,然后将前后向信息拼接得到完整的隐状态序列 $h_i = (h_1, h_2, \dots, h_n)$ 。通过线性输出层将完整的隐状态序列映射到 k 维(k 维为标签类别总数),从而得到特征矩阵 $P_{n \times k}$,其中 p_{ij} 表示字符 i 被标注为标签 j 的概率。

模型第三层CRF层接受BiLSTM输出的特征矩阵 $P_{n \times k}$,同时引入转移得分矩阵,通过维特比算法在所有备选标签序列中求得全局最优序列。最终识别“招商银行”应被标注为“金融服务”。

3.2. BERT 模型

在进入神经网络训练之前,需要先将语料中的文字表示成向量的形式作为模型的输入。与one-hot、Skip-gram、Word2Vec等静态词向量表示不同,BERT是一种基于微调的双向多层Transformer编码器,它可以更好地捕捉当前字符在真实语境地上下文信息,模型结构简图如图2所示。

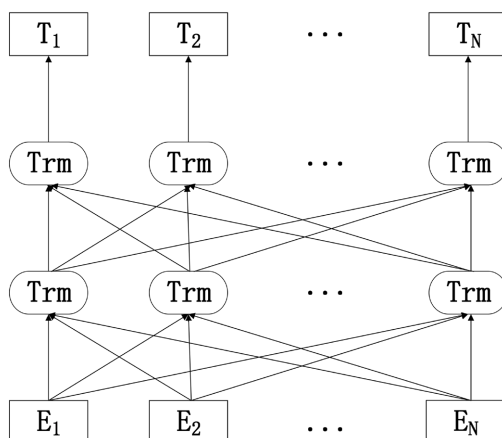


Figure 2. Bert model diagram

图 2. Bert 模型简图

由于中文分词技术在准确率方便存在各种问题,错误的分词会误导模型的训练效果,因此,本文BERT模型使用字符级表示作为输入。对于每一个字 W_i 其输入向量 E_i 是由三个嵌入向量相加组成的:字符级向量(Tokens_tensor)、位置向量(Positin_tensor)和句向量(Segments_tensor),具体如图3所示。

通过查找字符表,将输入中的字符转化为字符向量;位置向量用于标记每个字符在输入序列中所处的位置,为字符添加时序信息;句向量的作用是语料进行分句,用[CLS]和[SEP]表示一个句子的开始位置和结束位置,如不考虑上下句之间的关系,句向量可忽略不计。与单纯的静态向量相比,BERT模型输出融合了全局语义信息的特征向量,从而可以解决中文表达中“一词多义”的问题,这对机构名的实体分类会有很大的帮助。因此,本文采用BERT预处理模型生成字向量,提高字向量的质量,以期获得等好的分类效果。

3.3. BiLSTM 模型

组织机构的命名具有名称多样性、实体长度较长、实体嵌套等问题,在进行序列标注的过程中往往需要利用上下文信息来提高标注效果。LSTM全称是Long Short-Term Memory,即长短时记忆网络,它可以有效学习长期依赖信息。LSTM模型结构如图4所示。

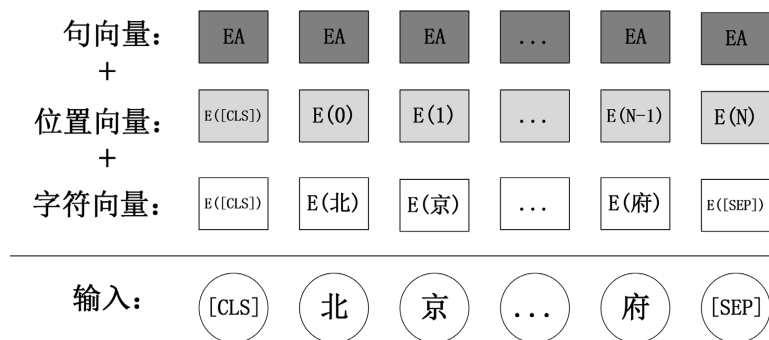


Figure 3. Input vector composition of Bert model

图 3. Bert 模型输入向量构成

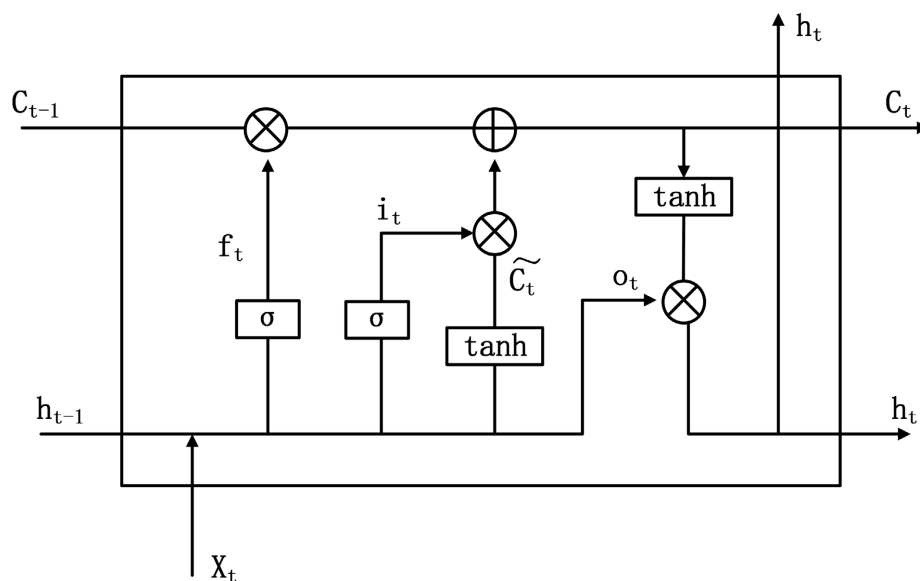


Figure 4. LSTM model structure diagram

图 4. LSTM 模型结构图

相较于传统循环神经网络，LSTM 主要增加了遗忘门 f_t 、记忆门 i_t 和输出门 o_t 三个门控单元来控制输入和输出，各个门之间的计算过程如下：

$$f_t = \sigma(W_f [h_{t-1}, X_t] + b_f) \quad (1)$$

$$i_t = \sigma(W_i [h_{t-1}, X_t] + b_i) \quad (2)$$

$$\tilde{C}_t = \tanh(W_c [h_{t-1}, X_t] + b_c) \quad (3)$$

$$o_t = \sigma(W_o [h_{t-1}, X_t] + b_o) \quad (4)$$

$$h_t = \tanh(C_t) * o_t \quad (5)$$

$$C_t = C_{t-1} * f_t + i_t * \tilde{C}_t \quad (6)$$

其中， σ 代表 sigmoid 函数， \tanh 代表双曲正切激活函数， W_f ， W_i ， W_c ， W_o 是各个门的权重， b_o ， b_i ， b_c ， W_f 是各个门的偏置项。由于 LSTM 只能处理一个方向传来的信息，因此，本文采取的 BiLSTM 模型是一个双向循环神经网络，由前向 LSTM 和后向 LSTM 组合而成，每个时刻 BiLSTM 模型从自左向右

\vec{h} 和自右往左 \overleftarrow{h} 两个方向来记忆上下文信息, 然后将双向网络输出的隐状态拼接得到 t 时刻 BiLSTM 的输出 $h_t = [\vec{h}, \overleftarrow{h}]$, 各时刻输出构成完整的隐状态序列 $\{h_1, h_2, \dots, h_n\}$, 通过线性输出层将完整的隐状态序列映射到 s 维(s 维为标签类别总数), 从而得到特征矩阵 $P_{n \times k}$, 其中, p_{ij} 表示字符 i 被标注为标签 j 的概率。

3.4. CRF 算法

BERT 与 BiLSTM 层学习到的标签序列只受字符本身及其上下文特征的影响, 仅仅依据打分值的高低进行最佳标签的选择, 这个结果并不准确。在 BIO 标注模式中, 标签之间存在着一定的依赖关系: 1) 句子只能以标签 “B-” 或 “O” 开始; 2) 标签 “B-X1、I-X2、I-X3” 中 X1、X2、X3 应该属于统一类别。条件随机场(Conditional Random Field, CRF)可以通过学习标签之间的转移规则获取全局最优解, 进而降低预测结果中出现非法序列的概率。

CRF 层的输入为 BiLSTM 层输出的特征矩阵 $P_{n \times k}$ 和一个随机初始化的转移矩阵 $A_{k \times k}$, 其中 k 表示标签类别的个数, n 表示输入序列的长度。

其可能输出标签序列 Y 为:

$$Y = (y_1, y_2, \dots, y_n) \quad (7)$$

输入序列 X 得到输出序列 Y 的分数计算公式为:

$$score(X, Y) = \sum_{i=1}^n P_{i, y_i} + \sum_{i=1}^n A_{y_i, y_i} \quad (8)$$

式中, P_{i, y_i} 代表第 i 个字符被预测为标签 y_i 的概率, A_{y_{i-1}, y_i} 代表标签 y_{i-1} 转移到标签 y_i 的概率。最终选取全部可能的标签序列 Y 中得分函数值最大的全局最优序列作为输出

$$Y^* = \arg \max score(X, Y) \quad (9)$$

4. 实验结果及分析

4.1. 数据集

本文将细粒度划分后的数据集按照 7:1:2 划分成训练集、验证集和测试集, 数据集统计情况见表 2。

Table 2. Distribution of entities in each category in the data set

表 2. 数据集中各类别实体分布情况

标签	总数	训练集	验证集	测试集
交通设施服务	7974	5516	859	1599
休闲服务	2817	1935	310	572
住宿服务	1731	1194	174	363
公司企业	12,982	9203	1250	2529
医疗服务	2550	1762	268	520
政府机构	3657	2535	372	750
生活服务	13,186	9206	1326	2654
社会团体	1978	1393	203	382
科教服务	5930	4167	531	1232
金融服务	2912	2044	301	567

Continued

风景名胜	1791	1251	171	369
餐饮服务	9716	6850	958	1908
总计	67,224	47,056	6723	13,445

在获取数据后, 需要将数据转化为带标签的格式, 本文采用的是 BIO 标注模式对实体进行标注, 标注示例图如表 3 所示。在该模式中“B”代表实体的开始, I 代表实体内部, O 代表其他, 结合表中的 12 小类, 带识别的小类标签共有 25 个, 例如“B-医疗服务”、“I-医疗服务”、“B-社会团体”、“I-社会团体”、“O”等。

Table 3. Schematic diagram of organization name corpus annotation

表 3. 机构名语料标注示意图

字符	标签	字符	标签
湿	B-科教服务	去	O
地	I-科教服务	益	B-住宿服务
博	I-科教服务	友	I-住宿服务
物	I-科教服务	客	I-住宿服务
馆	I-科教服务	栈	I-住宿服务

4.2. 评价指标

在测试过程中, 只有实体边界与实体类别完全匹配时, 才判断该实体类型预测正确。由于本实验数据集中的标签具有不平衡性, 为了能够客观地衡量模型在各类别上的分类性能, 实验主要用 F1 值作为评价标准, 并以准确率(Precision)和召回率(Recall)作为参考:

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \quad (10)$$

其中, F1 的值越大, 模型的实验效果越好。

4.3. 实验结果

为验证本文构建的 BERT-BiLSTM-CRF 模型在细粒度组织名识别的效果, 分别使用 BERT-CRF、BiLSTM-CRF 以及 CNN-BiLSTM-CRF 的三种命名实体识别模型与本文构建的 BERT-BiLSTM-CRF 模型进行对比实验。实验结果如表 3 所示。

BERT-CRF: BERT 模型在海量文本数据集上进行了预训练, 其动态文本表示方法相较于传统静态文本表示方法在文本数据上表现出优异性能, 加入 BERT 验证其在命名实体识别上的性能。

BiLSTM-CRF: 该模型是经典的命名实体识别模型, 由 BiLSTM 层学习输入序列的上下文信息, 再通过 CRF 层对标签之间的转移规则进行约束。

CNN-BiLSTM-CRF: CNN 可以学习各级别的输入特征, 因此, 加入 CNN 验证其有效性。

通过表 4 可知, 增加了 BiLSTM 的效果比基于 BERT-CRF 的实体识别模型 F1 值略有提升, 这是因为部分组织机构名长度较长或存在实体嵌套现象, 例“中共共青团拱辰街道建人外语艺术幼儿园支部委员会”等, BERT-CRF 会错误的将其标注为“科教服务”而不是“政府机构”。作为经典的实体识别模

型 BiLSTM-CRF 模型的 F1 值为 88.48%，在引入了 BERT 模型之后，模型 F1 值提高了 3.93 个百分点，这说明使用 BERT 对输入语料进行向量化之后，句子中的字、词表达更为准确，语义信息也更加丰富。CNN 可以学习语料间的语义相关性，和 BERT 具有相似的作用，但卷积核池化层可能会丢失某些输入的局部信息，因此，加入 CNN 的效果不如加入 BERT 好。

Table 4. Comparison of experimental results of different models
表 4. 不同模型的实验结果对比

模型	F1 值
BERT-CRF	91.73
BiLSTM-CRF	88.48
BiLSTM-CNN-CRF	91.49
BERT-BiLSTM-CRF	92.41

4.4. 不同实体实验结果对比

表 5 列出了不同模型对机构名数据集上各类实体的识别结果。由实验可知，交通设施服务实体在不同模型上的识别率都较高，因为这类组织名称具有较明显的指向性特征(例如：“富恒停车场”、“中山公园停车场”、“天安门地铁站”等)。而休闲服务、社会团体和风景名胜的分类效果均低于 0.88，部分原因是该类别的实体名称形式多变，神经网络难以捕捉到其特征信息；语料标注存在部分的错误和混乱，例如“故宫博物院观众休息厅”被标注为“生活服务”而“故宫博物院”被标注为“风景名胜”，高度相似的实体名称却被标注为不同实体，影响实验分类效果。由表中可以看出，BERT-BiLSTM-CRF 在各类实体上的识别效果都较好。

Table 5. Recognition effect of different models on various entities (F1 value)
表 5. 不同模型在各类实体上的识别效果(F1 值)

	BiLSTM-CRF	BiLSTM-CNN-CRF	BERT-CRF	BERT-BiLSTM-CRF
交通设施服务	0.9860	0.9897	0.9894	0.9919
休闲服务	0.7490	0.8146	0.8041	0.8288
住宿服务	0.8870	0.9068	0.8916	0.9183
公司企业	0.8752	0.9052	0.9061	0.9099
医疗服务	0.8686	0.8881	0.8925	0.9135
政府机构	0.8148	0.9000	0.8884	0.9021
生活服务	0.8723	0.8987	0.9019	0.9081
社会团体	0.8101	0.8798	0.8557	0.8764
科教服务	0.8590	0.9154	0.9225	0.9280
金融服务	0.9181	0.9466	0.9325	0.9474
风景名胜	0.7790	0.8045	0.8556	0.8591
餐饮服务	0.9443	0.9527	0.9665	0.9682
平均	0.8848	0.9151	0.9173	0.9241

5. 结论

本文构建了机构名数据集, 标注并划分了不同类型的 12 类实体, 共计 67,224 个, 在此基础上完成细粒度机构名的识别分类。本文构建的基于 BERT-BiLSTM-CRF 的神经网络模型, 利用 BERT 解决中文文本特征表示时存在的一词多义问题, 结合 BiLSTM 充分学习语料的上下文信息, 通过 CRF 方法约束标签转移规则, 最终提取全局最优的标注序列, 完成机构名的细粒度实体识别分类任务。最后, 在构建的数据集上对比了其他三种实体识别模型, 进行实验验证。实验结果表明, 本文采用的 BERT-BiLSTM-CRF 模型在不同实体类型上都可以保持较高的准确率。

参考文献

- [1] 丁晟春, 方振, 王楠. 基于 Bi-LSTM-CRF 的商业领域命名实体识别[J]. 现代情报, 2020, 40(3): 103-110.
- [2] Mo, H.M., Nwet, K.T. and Soe, K.M. (2016) CRF-Based Named Entity Recognition for Myanmar Language. In: Pan, J.-S., et al., Eds., *International Conference on Genetic and Evolutionary Computing*, Springer International Publishing, Berlin, 204-211. https://doi.org/10.1007/978-3-319-48490-7_24
- [3] 鞠久朋, 张伟伟, 宁建军, 周国栋. CRF 与规则相结合的地理空间命名实体识别[J]. 计算机工程, 2011, 37(7): 210-212+215.
- [4] 盛剑, 向政鹏, 秦兵, 刘铭, 王莉峰. 多场景文本的细粒度命名实体识别[J]. 中文信息学报, 2019, 33(6): 80-87.
- [5] Lample, G., et al. (2016) Neural Architectures for Named Entity Recognition. *The 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, San Diego, 12-17 June 2016, 260-270. <https://doi.org/10.18653/v1/N16-1030>
- [6] 左亚尧, 陈致然, 洪嘉伟, 陈坤. 融合多语义特征的命名实体识别方法[EB/OL]. 计算机应用: 1-9. <http://kns.cnki.net/kcms/detail/51.1307.tp.20211026.1608.008.html>, 2022-03-20.
- [7] 朱丹浩, 杨蕾, 王东波. 基于深度学习的中文机构名识别研究——一种汉字级别的循环神经网络方法[J]. 现代图书情报技术, 2016(12): 36-43.
- [8] Peng, N. and Dredze, M. (2016) Improving Named Entity Recognition for Chinese Social Media with Word Segmentation Representation Learning. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, Volume 2, 149-155. <https://doi.org/10.18653/v1/P16-2025>
- [9] Peters, M., Neumann, M., Iyyer, M., et al. (2018) Deep Contextualized Word Representations. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Volume 1, 2227-2237. <https://doi.org/10.18653/v1/N18-1202>
- [10] 王子牛, 姜猛, 高建筑, 陈娅先. 基于 BERT 的中文命名实体识别方法[J]. 计算机科学, 2019, 46(S2): 138-142.
- [11] 吴俊, 程垚, 郝瀚, 艾力亚尔·艾则孜, 刘菲雪, 苏亦坡. 基于 BERT 嵌入 BiLSTM-CRF 模型的中文专业术语抽取研究[J]. 情报学报, 2020, 39(4): 409-418.
- [12] 朱鹏, 石丽红, 焦明连, 刘晓东, 孙浩. 混合神经网络的中文地名识别方法[J]. 测绘科学, 2021, 46(11): 159-165. <https://doi.org/10.16251/j.cnki.1009-2307.2021.11.023>
- [13] 杨璐, 张恬, 郑丽敏, 田立军. 兽药致病命名实体 Att-Aux-BERT-BiLSTM-CRF 识别[J]. 农业机械学报, 2022, 53(3): 294-300.
- [14] Wang, Y., Sun, Y., Ma, Z., et al. (2020) Named Entity Recognition in Chinese Medical Literature Using Pretraining Models. *Scientific Programming*, 2020, Article ID: 8812754. <https://doi.org/10.1155/2020/8812754>
- [15] 董哲, 邵若琦, 陈玉梁, 翟维枫. 基于 BERT 和对抗训练的食品领域命名实体识别[J]. 计算机科学, 2021, 48(5): 247-253.