

# 多场景文本的细粒度命名实体识别

向政鹏<sup>1</sup> 盛剑<sup>2</sup> 秦兵<sup>3</sup> 刘铭<sup>4</sup> 王莉峰<sup>5</sup> 王巍<sup>6</sup>

<sup>1,2,3,4,6</sup>哈尔滨工业大学, 黑龙江省哈尔滨市西大直街92号

<sup>5</sup>腾讯科技(深圳)有限公司, 广东省深圳市南山区腾讯滨海大厦S27

<sup>1,2,3,4</sup>{zpxiang, jsheng, bqin, mliu}@ir.hit.edu.cn

<sup>5</sup>fandywang@tencent.com

<sup>6</sup>wangweidmail0809@gmail.com

**摘要.** 命名实体识别一直是数据挖掘领域的经典问题之一, 尤其随着网络数据的剧增, 如果能对多来源的文本数据进行多领域、细粒度的命名实体识别显然能够为很多的数据挖掘应用提供支持。本文即提出一种多领域、细粒度的命名实体识别方法。其利用网络词典回标文本数据获得了大量的粗糙训练文本。为防止训练文本中的噪声干扰命名实体识别的结果, 本算法将命名实体识别的过程划分为两个阶段, 第一个阶段先获得命名实体的领域标签, 之后利用命名实体的上下文确定命名实体的细粒度标签。实验显示, 本文提出的方法在多个领域上均获得了显著的识别准确率。

**关键字:** 命名实体识别, 细粒度类别划分, 语料回标

**Abstract.** Name entity recognition is a classical research issue in data mining community. Especially along with the explosive augment of web data, if we can handle the data generated from multi-sources and recognize the entities in them with multi-domain and fine-grained labels, it obviously can provide support to many upstream data mining applications. In this paper, we propose a way which utilizes web thesaurus to annotate web data automatically to acquire large-scale training corpus with noises. To avoid the noises in training corpus to drop the performance of entity recognition, we separate the process of entity recognition into two phases. At first, the entity's domain label is obtained. After that, the context of each recognized entity is used to determine the fine-grained label for one entity. Experimental results demonstrate that via our proposed method, it can obtain high accuracy on entity recognition in many domains.

**Keywords:** Named entity recognition, fine-grained category annotation, corpus annotation

## 1 引言

命名实体识别(Named Entity Recognition, NER)是自然语言处理领域(Natural Language Processing, NLP)的基础任务之一, 也是信息抽取中最为重要的一个子

任务，并且可以对后续的抽取任务中提供帮助。命名实体识别任务意在识别文本中的事物的名称，例如人名、地名和机构名。本文主要在多场景多领域下研究命名识别实体识别，以LSTM-CRF为基础并引入CNN（卷积神经网络）从文本中进一步提取有用的语义特征。

早期的命名实体识别大多是基于规则的方法，但是由于语言结构本身具有不确定性，制定出统一完整的规则难度较大。基于规则的方法需要构造特定的规则模板，采用的特征包括统计信息、标点符号、关键字、位置词、中心词等，以模式和字符串相匹配为主要手段，尤其依赖于知识库和词典的建立。对于不同领域需要对应的专家重新书写规则，代价较大，存在规则建立周期长、移植性差且需要建立不同领域知识库作为辅助以提高系统识别能力等问题。

传统的命名实体识别方法大多采用有监督的机器学习模型，如隐马尔可夫模型[1]、最大熵[2][3]、支持向量机和条件随机场[4]。最大熵模型具有结构严谨，良好通用性的特点，但同时训练时间复杂性高，需要明确的归一化计算[5]，导致计算上的开销比较大。条件随机场在分词和命名实体识别上表现出色，提供了一个特征灵活、全局最优的标注框架，但同时存在收敛速度慢、训练时间长的的问题。通常来讲，最大熵和支持向量机在正确率上比隐马尔可夫模型略高，但是隐马尔可夫模型在训练和识别时的速度要更快一些，主要是由于Viterbi算法在解码时具有较高的效率。基于统计的方法对特征的选取依赖较高，需要从文本中分析选择对于此项任务影响因子较大的特征，并将这些特征加入到特征模板中。通过对训练语料所包含的语言语义信息进行统计和分析，进行有效的特征选择，从训练语料中不断发现强特征。有关特征可以分为具体的停用词特征、上下文特征、词典及词性特征、单词特征、核心词特征以及语义特征等。与基于规则的方法相似，基于统计的方法对于语料库的依赖也较大，而建立较大的领域语料库又是一大难点。

当前最好的实体识别模型是LSTM-CRF模型[6][7]，该网络由两个长短期记忆网络组成，一个前向记忆网络和一个后向记忆网络，前者用于学习前向的序列信息，后者用于学习后向的序列信息，得到每个隐层的表示，将该隐层映射到所需分类的特征维度，之后选取概率最大的一维作为其实体类别，该方法也称之为softmax[8]。尽管该模型在独立的序列标注任务中取得了成功，例如词性标注，但是该模型忽略了标签间的依赖关系，这一缺点导致了部分精度的损失。实体识别任务存在某些内在限制，例如I-PER 标签并不能接在B-LOC标签的后边。因此，用条件随机场模型(CRF)来学习标签之间的关系，而不是进行独立的标注。

本文提出的细粒度命名实体识别算法以LSTM-CRF模型为基础，并引入CNN从文本中进一步提取有用的语义特征做实体边界划分，之后交给细粒度划分模块做小类别分类，在第一步通过使用RNN对命名实体的上下文进行表示，并使用softmax分类确定该命名实体所属的大类别（即领域）；第二步利用每个领域下的语料构建模型以确定命名实体所属的细粒度类别（小类别）。实验结果显示，命名实体识别的准确率、召回率和F1值能达到80%以上，能在一定程度上说明实现的分阶段方案是有效的。

## 2 方法描述

本文实现的模型共包含三个模块：语料回标模块、命名实体识别模块、命名实体细粒度划分模块。在获取某个词条为命名实体后，则将该命名实体和命名实体的上下文交由细粒度的命名实体类别划分模块，先由模块确定该命名实体的大类别标签（所属领域），然后再确定该实体的小类别标签。对句子的识别提供了有交叉多标签的命名实体识别结果，即有可能出现输入“ABCDE”，识别结果为“AB”和“BC”为不同领域下的命名实体，另外，在对实体打标签时也会出现一个实体有多个标签的情况，类似于“马龙”可能是一个人名，也是一个体育明星，这时该方案会对命名实体提供属于多个领域的细粒度类别标签。

举例来说，输入文本为：

范冰冰,1981年9月16日出生于山东青岛,华语影视女演员、制片人、流行歌手,毕业于上海师范大学谢晋影视艺术学院。

输出文本为：

[范冰冰][影视明星,人名],1981年9月16日出生于[山东][地名][青岛][地名],华语影视女演员、制片人、流行歌手,毕业于[上海师范大学][机构名,教育机构]谢晋影视艺术学院。

本文共识别包括12个领域的命名实体（领域列表见表2），该文本在经过12个领域的命名实体识别模块后，在传统、娱乐和教育领域均有命名实体被识别出来，将结果全部保留送入细粒度类别划分模型得到最终的分类。此例中，范冰冰被识别成影视明星和人名，山东和青岛被识别为地名，上海师范大学被识别成机构名和教育机构名。

模型的整体结构如图1所示：

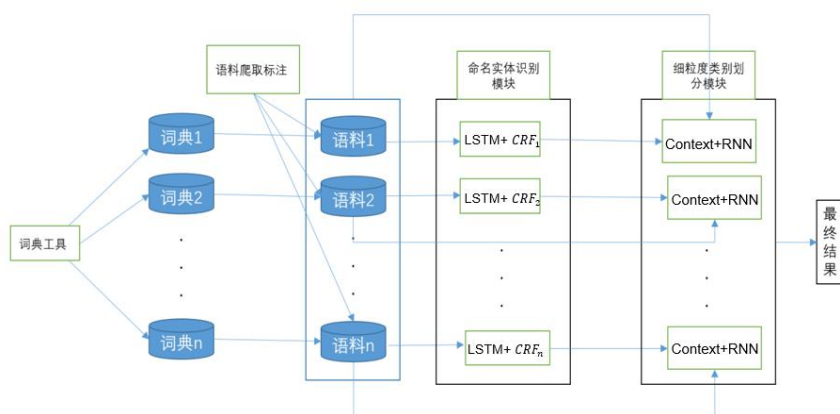


图1. 模型整体结构图

2.1 词典回标模块

在词典标注模块中，从网络中爬取了除人名、地名、机构名外其它类别下的细粒度命名实体词典。并从网络中爬取文本数据，利用词典回标数据形成大规模的训练集、开发集和测试集。实现阶段发现，自动标注的数据里面有很多误标注的数据（噪声）存在，例如把“你的名字是什么”这句话标注为”[你的名字][影视剧]是什么”，即认为“你的名字”是个影视剧，但是显然该词条不是一个命名实体。

通过抽样分析语料回标的质量不是很高，回标结果举例如表1所示：

表 1.回标实体举例

效果	领域	语料	识别结果
效果较好	餐饮	清蒸[黄花鱼][食品名],可以说是“一鱼两吃”吧	正确
效果较差	餐饮	[苹果][食品名]申请智能保护套专利以防ipad跌落	错误
效果较好	旅游	【超值套餐】[千岛湖][景点]梅地亚君澜豪华2天1晚美食套餐自由行	正确
效果较差	旅游	雅马哈鬼火天蝎[黄龙][景点]600摩托车改装配件	错误
效果较好	电商	[七匹狼][品牌]家居裤男士睡裤纯棉长裤单件装纯色款xxl	正确
效果较差	电商	[中华][品牌]成语故事小学生语文新课标必读书系	错误

基于此类的错误，人工对粗糙的机器标注的语料进行二次标注，得到了小规模的较为准确的标注语料。

2.2 命名实体识别模块

本文以LSTM+CRF构建基准的命名实体识别模块[9][10]，引入CNN特征提取模块[11]，用于识别输入文档中的命名实体。

CNN在句子建模上有着广泛的应用，CNN强大的特征捕捉能力使得在句子建模过程中，经过组合底层邻近的词语信息，逐步向上传递，上层又组合新的短语信息，从而使得相距较远的句子也存在联系，这种联系通常是语义上的联系。

本文中，卷积的时候是整行进行的，filter的高为词向量的维度，宽为2,3不等。如图2最下方CNN提取句子特征部分所示，宽为2的filter对整个序列进行卷积得到维度为4的向量。对每个filter卷积得到的向量通过max pooling操作，将结果拼接在一起得到的向量作为整个句子的表示。将此句子表示加入到LSTM获取到的每个词的上下文表示，从而使每个词可以既具有句子表示又可以具有词级别表示。

LSTM+CNN+CRF的基本结构如图2所示：

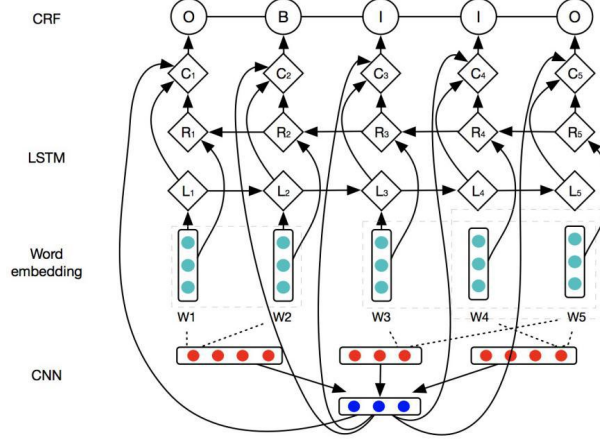


图2. LSTM+CRF+CNN结构图

LSTM+CRF为两层结构，其中输入表示层由双向LSTM组成，通过LSTM中的每个cell单元对输入文档进行编码，将隐层得到的输出与CNN提取的类似N-Gram的特征进行拼接[12]，通过全连接层进行tag分类，得到的矩阵作为CRF层的发射矩阵[13]。

在CRF中，给定一个观察序列 $\vec{x}$ ，目的是希望找到一个概率最大的标记序列 $\vec{y}$ 。在命名实体识别任务中， $\vec{x}$ 表示词序列， $\vec{y}$ 表示命名实体标记。那么给定 $\vec{x}$ ， $\vec{y}$ 的概率可由以下公式计算：

$$p(\vec{y}|\vec{x}) = \frac{1}{Z(\vec{x})} \prod_{j=1}^n \psi_j(\vec{x}, \vec{y}) \quad (1)$$

$$Z(\vec{x}) = \sum_{\vec{y}'} \prod_{j=1}^n \psi_j(\vec{x}, \vec{y}') \quad (2)$$

$$\psi_j(\vec{x}, \vec{y}) = \exp \left( \sum_{i=1}^m \lambda_i f_i(y_{j-1}, y_j, \vec{x}, j) \right) \quad (3)$$

在上述三个公式中， $j$ 代表词序列 $\vec{x}$ 中的第 $j$ 个位置， $n$ 表示 $\vec{x}$ 的长度， $m$ 表示特征的数量， $f_i(y_{j-1}, y_j, \vec{x}, j)$ 是CRF的特征函数， $\lambda_i$ 则是对应特征函数的权重。假设 $\vec{x}$ 中的一个子串 $x_k x_{k+1} \dots x_{k+l}$ 被标注为一个命名实体，记为 $Ne$ ，则 $Ne$ 的标记序列为 $y_k^* y_{k+1}^* \dots y_{k+l}^*$ ，记做 $\vec{y}_{Ne}$ 。使用式(4)计算边缘概率 $\phi(Ne)$ ，作为识别到的命名实体的置信度。计算时，可采用前向后向算法[14]。

$$\phi(N_e) = \frac{1}{Z(\vec{x})} \sum_{\vec{y}: y'_k \dots y'_{k+1} = \vec{y}^*} p(\vec{y} | \vec{x}) \quad (4)$$

根据每个被识别出的命名实体在语料中出现的频率和命名实体的置信度 ( $\phi(N_e)$ ) 设定阈值, 然后选择大于阈值的命名实体作为可信的标注结果。

### 2.3 细粒度类别划分模块

本文将命名实体的细粒度类别划分切分为两个步骤: 1) 根据命名实体及命名实体的上下文进行领域(大类别标签)的划分; 2) 根据第一步得到的大类别标签进行类别的细分以得到命名实体的小类别标签。

我们的模型采用Tang[15]提出的对句子中目标词的建模方法对命名实体的表示进行建模, 这个模型在中文零指代消解任务[16]中也发挥了较大的作用。具体来说, 在第一步通过使用两个单向RNN对命名实体的上文和下文分别进行表示, 再将上文的表示和下文的表示拼接在一起作为命名实体的表示, 最后使用 $softmax$ 分类确定该命名实体所属的大类别(即领域); 第二步利用每个领域下的语料构建模型以确定命名实体所属的细粒度类别(小类别)。在两个步骤中使用同样的模型, 区别在于: 第一个步骤分类的类别为十二个领域, 第二个步骤将训练十二个分类器, 每个分类器只对一个领域下的小类别标签进行分类。

模型结构如图3所示:

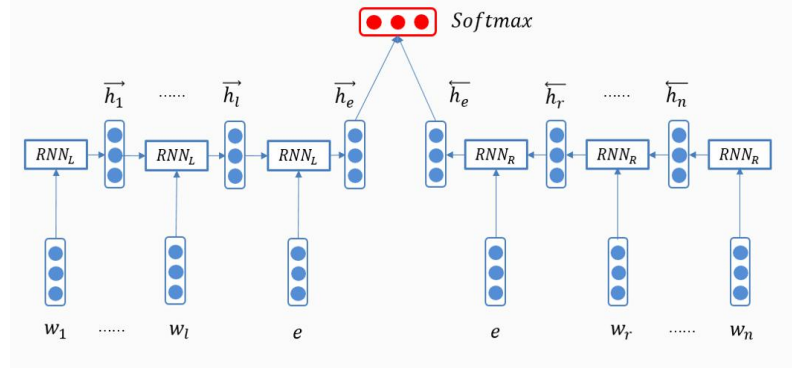


图3. 细粒度类别划分算法模型图

依据模型, 我们对每个识别出的命名实体采用单向RNN表示其上下文, 上下文的表示分为两部分。其中, 中间的两个 $e$ 表示的是同一个命名实体,  $w_1$ 到 $w_l$ 表示句子中 $e$ 实体之前的词,  $w_r$ 到 $w_n$ 表示句子中 $e$ 实体之后的词。从句子中的第一个词 $w_1$ 开始到命名实体 $e$ 为止, 从左到右依次读入 $RNN_L$ 中, 得到的表示记为 $\vec{h}_e$ 。从句子中的最后一词 $w_n$ 开始到实体 $e$ 为止, 从右到左依次读入 $RNN_R$ 中, 得到的表示记为 $\vec{h}_e$ 。将这两个表示拼接在一起之后使用 $softmax$ 进行分类。

在建模命名实体的上下文时, 本文对实体的表示按以下三种情况分别处理:

- (1) 每次仅对一个命名实体进行分类；
- (2) 句中如果有多个命名实体，则每次仅处理一个命名实体，将其余的部分都看作是上下文；
- (3) 命名实体如果由多个词组成，直接将多个词拼在一起看作是一个词。

实验过程中，第一个步骤使用十二个领域的训练集进行训练（在训练的时候会对训练语料随机打乱），然后在十二个领域的开发集上进行模型的选择，最后保留在开发集上效果最好的模型。在第二个步骤中，分别按上述方法利用十二个领域的训练语料依次训练十二个分类器做命名实体的细粒度划分。

### 3 实验分析

本文将命名实体的识别分为了两个阶段。第一个阶段利用序列标注算法确定某个词条是否为命名实体（利用12个领域下的命名实体识别器判断某一词条是否为该领域下的命名实体），第二个阶段根据第一阶段识别出的命名实体进行细粒度类别划分结果。因此实验部分分别对这两个阶段进行检测，并最终检测了两个阶段合二为一后的结果。

#### 3.1 标签体系

表 2. 标签体系

领域（大类别）	细粒度标签（小类别）
传统	人名、地名、机构名
教育	学校、课程名、书籍名、作者、培训机构
游戏	游戏名、虚拟角色、道具、技能
娱乐	明星、影视剧、歌曲、专辑、综艺
金融	股票、股票代码、理财产品
汽车	品牌、车系、车型、零部件
旅游	景点、酒店、航空公司、车站
餐饮	品牌、菜名、食品名、餐馆名、商场/购物中心
房产	小区、开发商、物业公司、家居/家具品牌
电商	品牌、店铺名、快递公司
体育	明星、体育项目、赛事
医疗	医院、疾病、科室、药品、症状

本文从领域角度出发，定义一套命名实体（NE，Named Entity）标签体系，研发面向互联网多场景文本数据的命名实体识别通用解决方案，提升文本词法分析质量。表2中展示了我们所定义的十二个领域以及这十二个领域下的总计四十八个细粒度标签。

### 3.2 数据集

本实验采集了与需求对应的12个类别下的命名实体构建词典。在获取词典后，利用从网络爬取和腾讯方面给出的文本数据进行回标后获取每个类别下的数据集，获得我们的数据集，句子级数据集规模（单位：条）如表3所示：

表 3.回标数据集统计

领域	句子条数	领域	句子条数
传统	14060	旅游	20500
教育	13559	餐饮	14466
游戏	20716	房产	14811
娱乐	9552	电商	23543
金融	5797	体育	61429
汽车	7821	医疗	14605

在语料中，每个领域分别随机抽样100条语料，并计算出准确率（标注正确实体/所有实体）如表4所示：

表 4.回标抽样准确率

领域	准确率	领域	准确率
传统	98.91	旅游	96.67
教育	90.24	餐饮	93.07
游戏	86.96	房产	83.01
娱乐	70.05	电商	62.50
金融	82.79	体育	75.00
汽车	91.35	医疗	65.15

### 3.3 词向量

在模型的训练过程中，使用了预训练的中文词向量。其使用gensim中的Word2Vec的接口训练中文维基百科的语料得到。

我们首先对训练集统计了词表。对于词表中的词 $w$ ，如果有对应的预训练的词向量，那么将使用对应的预训练的词向量作为词 $w$ 的词向量的初始值；如果没有对应的预训练的词向量，那么词 $w$ 的词向量将由随机产生。在模型的训练过程中词向量会随着训练时得到的梯度进行微调。

### 3.4 实验结果

**实体识别模块实验结果.**此模块利用12个领域下的语料分别训练12个命名实体识别器，这些命名实体识别器仅进行0-1识别，即确定某词条是否为某领域下的命名实体。因此，本部分给出了在各个领域下命名实体的0-1识别结果，并对比了与LSTM-CRF模型的结果，效果有所提高。实验结果如表5所示。由表中可见，



加入CNN后，在大部分领域的命名实体识别结果均有所提高，说明CNN能够从文本中获得对命名实体识别有益的特征。

表 5.命名实体识别模块实验结果

领域	准确率		召回率		F值	
	无CNN	加入CNN	无CNN	加入CNN	无CNN	加入CNN
传统	78.35	<b>79.92</b>	<b>87.21</b>	86.62	82.54	<b>83.13</b>
教育	91.21	<b>91.90</b>	89.03	<b>91.01</b>	90.10	<b>91.45</b>
游戏	95.03	<b>97.23</b>	95.33	<b>96.72</b>	95.17	<b>96.98</b>
娱乐	<b>91.32</b>	90.87	89.60	<b>93.67</b>	90.45	<b>92.25</b>
金融	<b>94.78</b>	94.22	92.23	<b>94.59</b>	93.48	<b>94.40</b>
汽车	92.34	<b>93.15</b>	95.23	<b>96.62</b>	93.76	<b>94.86</b>
旅游	<b>89.98</b>	89.67	92.54	<b>94.49</b>	91.24	<b>92.01</b>
餐饮	86.45	<b>89.68</b>	<b>94.22</b>	93.77	90.16	<b>91.68</b>
房产	78.30	<b>81.75</b>	<b>88.93</b>	87.10	83.27	<b>84.34</b>
电商	87.32	<b>91.60</b>	91.47	<b>94.52</b>	89.34	<b>93.04</b>
体育	79.21	<b>80.30</b>	<b>78.17</b>	77.28	78.68	<b>78.75</b>
医疗	92.11	<b>93.86</b>	<b>92.40</b>	91.16	92.25	<b>92.49</b>

**命名实体的细粒度类别划分模块实验结果.**命名实体的细粒度类别划分分为两个阶段，因此实验部分也给出了两个阶段的划分结果。第一个阶段的输入为由命名实体识别模块识别出的命名实体及其上下文。输出为该命名实体的大类别标签。在进行此部分实验时，假设输入的命名实体为正确的命名实体。此部分实验能够检测命名实体被划分到了正确的大类别下的准确程度。本部分的实验数据均为词典回标得到的数据按照7:1:1的比例划分为训练集、开发集和测试集。

实验结果如表6所示。本模块只对命名实体识别模块识别正确的实体进行细粒度类别划分，并不对实体的识别结果进行修改，因此本模块只对命名实体类别划分的准确率进行分析。

表 6.命名实体的大类别划分结果

训练集	开发集	测试集
97.93	95.22	95.32

由表5可见，大类别划分的准确率在95%左右，说明在命名实体识别准确的情况下，依据上下文可以获得较为准确的大类别划分结果。

在第二个阶段，对命名实体在第一个阶段识别出的大类别下进行细粒度划分。表6给出了实验结果。表6中医疗、娱乐、游戏等领域实验中训练集的效果不如开发集和测试集的原因是我们的训练数据中是由于回标的误差有噪声的。表6中同样假设输入为由命名实体识别模块识别出的命名实体及其上下文，输出

则为该命名实体的细粒度类别标签。此部分实验能够检测命名实体被划分到细粒度的类别下的准确程度。

表 7.命名实体的细粒度类别划分结果

领域	训练集	开发集	测试集
金融	96.52	93.13	92.01
医疗	74.30	87.95	85.63
餐饮	65.27	73.29	75.87
教育	76.30	87.68	85.75
体育	95.15	98.58	98.82
旅游	86.96	87.70	88.16
传统	92.59	95.78	94.93
汽车	98.02	95.04	93.12
房产	86.45	93.71	93.62
游戏	79.31	88.68	86.06
电商	96.11	85.21	80.77
娱乐	64.49	69.71	72.05

**命名实体识别整体实验结果.**最后，将命名实体识别和命名实体的细粒度类别划分两个步骤合二为一，测试了本文实现的命名实体细粒度划分的整体准确率，使用的语料是人工标注的语料，结果如表8所示。由于在细粒度类别划分模块中只对命名实体识别模型识别正确的实体进行细粒度类别划分，因此最后结果的准确率等于两个模块准确率的乘积，召回率等于命名实体识别模块的召回率。在此实验中，输入为未经处理的原始文本，而输出为带有细粒度标签的命名实体。

从表7的实验结果可见，大部分领域的效果都在80%左右，少部分领域效果较差，这些领域效果较差的主要原因是数据在标注的过程中出现了错误，而在人工标注的语料中就不包含这些错误的信息，因而模型学到的更多的是错误的信息。

另外，由于训练语料是由词典回标产生的，所以模型往往可以正确识别常见的命名实体，而对于一些出现次数较少的命名实体就难以正确识别。部分实体的类别在自动标注的过程中被误标，原因是在数据回标时将部分非命名实体标注为命名实体，使得在命名实体识别时，识别出错误的命名实体。

例如，对于文本“losea磨砂拼接小方包2017新款斜挎包迷你锁扣包韩版复古单肩包小”中的“迷你”在汽车领域的标签是“汽车一品牌”，但是在当前语境下就是错误的标签。在该领域下实验效果较差的原因就是在数据回标时，将大量的“迷你”标注为“汽车一品牌”，所以在训练的时候，模型学习到的信息是将“迷你”的类别识别为“汽车一品牌”。

表 8.细粒度命名实体识别的整体结果

领域	准确率	召回率	F1
金融	82.28	94.59	88.01
医疗	61.60	91.16	73.52
餐饮	58.48	93.77	72.04
教育	69.95	91.01	79.10
旅游	80.70	94.49	87.05
汽车	87.65	96.62	91.92
游戏	86.77	96.62	91.48
体育	79.62	77.28	78.44
房产	87.12	87.10	79.30
电商	83.19	94.52	88.49
娱乐	89.70	93.67	91.65
传统	61.66	86.62	72.04

部分抽样的错误样例见表9。

表 9.错误示例

序号	例子	类型	实体	真实类别	识别类别
1	[渗出]的血珠,就像一腔热血无处挥洒而倾泻的情感。	实体识别错误	渗出	无	医疗
2	[长安CS75] 车型挡泥板汽车配件用品外饰改装专用挡泥胶带标	细粒度类别划分错误	长安CS75	汽车-车系	汽车-车型
3	张家辉负心汉我与孙瑜共患难张家辉负心汉我我爱杨凯迪爱华晨宇都可以进群欢迎[火星]大家好！	细粒度类别划分错误	火星	娱乐-明星	娱乐-专辑

4 总结

本文设计并实现的分阶段细粒度命名实体识别方案能够将大部分命名实体识别出来，并确定该命名实体的细粒度标签。对于某些细粒度的命名实体，比如课程、影视名、汽车车系等，由于在训练语料中分布稀疏，并且该类命名实体的命名非常随意，比如车系有“唐”、“A7”等，使得针对这类的命名实体的细粒度划分变得不准确，但是如果不强调小类别，而是仅进行大类别的划分，本

文实现的方法能够达到较高的准确率，达到80%以上，其能在一定程度上说明本项目实现的分阶段方案是有效的。

## 5 参考文献

1. Shai Fine, F., Yoram Singer, S., Naftali Tishby, T.: The hierarchical hidden markov model: Analysis and applications. *Machine Learning*, 32:41–62(1998).
2. Andrew Borthwick, F.: A Maximum Entropy Approach to Named Entity Recognition. Ph.D. thesis, New York University (1999).
3. Andrew McCallum, F., Dayne Freitag, S., Fernando Pereira, T.: Maximum entropy Markov models for information extraction and segmentation. In: *ICML-2000*(2000).
4. J. Lafferty, F., A. McCallum, S., F. Pereira.: Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In: *Proceedings of ICML*(2001).
5. Adwait Ratnaparkhi, F.: A maximum entropy model for part-of-speech tagging. In: *EMNLP 1*, pp. 133–142(1996).
6. Tomas Mikolov, F., Ilya Sutskever, S., Kai Chen, T.: Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119(2013).
7. Felix A Gers, F., Nicol N Schraudolph, S., Jurgen Schmidhuber T.: Learning precise “ timing with lstm recurrent networks. *The Journal of Machine Learning Research*, 3:115–143(2003).
8. Zhiheng Huang, F., Wei Xu, S., Kai Yu, T.: Bidirectional lstm-crf models for sequence tagging. *arXiv preprint arXiv:1508.01991*(2015).
9. Rie Kubota Ando, F., Tong Zhang, S.: A framework for learning predictive structures from multiple tasks and unlabeled data. *The Journal of Machine Learning Research*, 6:1817–1853(2005).
10. Hongyan Jing, F., Tong Zhang, S.: Named entity recognition through classifier combination. In: *Proceedings of HLT-NAACL-2003*, pp. 168–171(2003).
11. Kim Y, F.: Convolutional neural networks for sentence classification[J]. *arXiv preprint arXiv:1408.5882*(2014).
12. Zhang Y, F., Wallace B, S.: A Sensitivity Analysis of (and Practitioners’ Guide to) Convolutional Neural Networks for Sentence Classification[J]. *arXiv preprint arXiv:1510.03820*(2015).
13. M. Boden, F.: A Guide to Recurrent Neural Networks and Back-propagation. In the Dallas project(2002).
14. J. Hammerton, F.: Named Entity Recognition with Long Short-Term Memory. *Proceedings of HLT-NAACL*(2003).
15. Tang, D, F., Qin, B, S., Feng, X, T.: Effective LSTMs for Target-Dependent Sentiment Classification. In: *COLING 2016*, pp. 3298–3307.(2016).
16. Yin, Q, F., Zhang, Y, S., Zhang, W., T.: Chinese Zero Pronoun Resolution with Deep Memory Network. In: *Empirical Methods in Natural Language Processing*. pp. 1309–1318(2017).