面向法律文本的实体关系联合抽取算法

宋文辉[1],周翔[2], 杨萍[2], 孙媛媛[1]†, 杨亮[1], 林鸿飞[1]
[1]大连理工大学/辽宁省大连市
[2]航天科工智慧产业发展有限公司/北京市
songwenhui@mail.dlut.edu.cn, zhouxiang@aiidc.com.cn
yangping@aiidc.com.cn, syuan@dlut.edu.cn
liang@dlut.edu.cn,hflin@dlut.edu.cn

摘要

法律文本中包含的丰富信息可以通过结构化的实体关系三元组进行表示，便于法律知识的存储和查询。传统的流水线方法在自动抽取三元组时执行了大量冗余计算，造成了误差传播。而现有的联合学习方法无法适用于有大量重叠关系的法律文本，也并未关注语法结构信息对文本表示的增强，因此本文提出一种面向法律文本的实体关系联合抽取模型。该模型首先通过ON-LSTM注入语法信息，然后引入多头注意力机制分解重叠关系。相较于流水线和其他联合学习方法本文模型抽取效果最佳，在涉毒类法律文本数据集上抽取结果的F1值达到78.7%。

关键词： 联合学习 ； 智慧司法 ； 信息抽取 ； 关系抽取

Joint Entity and Relation Extraction for Legal Texts

Wenhui Song[1],Xiang Zhou[2], Ping Yang[2],
Yuanyuan Sun[1]†,Liang Yang[1],Hongfei Lin[1]
[1]Dalian University of Technology / Dalian City , Liaoning Province
[2]CASIC Intelligence Industry Development Co.,Ltd/ Beijing City
songwenhui@mail.dlut.edu.cn, zhouxiang@aiidc.com.cn
yangping@aiidc.com.cn, syuan@dlut.edu.cn
liang@dlut.edu.cn,hflin@dlut.edu.cn

Abstract

The abundant information which embraced in legal texts is generally represented by structured triplets composed of entities and relations, which is convenient for the storage and query of legal knowledge. The traditional pipeline method performs a lot of redundant calculations when automatically extracting triples, which causes error propagation. However, the existing joint learning methods cannot be applied to legal texts with a large number of overlapping relationships, and they do not pay attention to the enhancement of text representation by syntactic structure information either. Therefore, we present a model for joint entity and relation extraction to handle legal texts. First, ON-LSTM is used to inject grammatical information into the model, and then a multi-head attention mechanism is introduced to decompose overlapping relations. Compared with the pipeline method and other joint method, the experiments show that our model achieves the best extraction effect, and the F1 score of triplet extraction reaches 78.7% on the drug-related legal dataset.

Keywords: Joint Learning , Intelligent justice , Information Extraction , Relation Extraction

第二十届中国计算语言学大会论文集，第589页-第599页，呼和浩特，中国，2021年8月13日至15日。

1 引言

随着中国司法领域信息公开化的程度不断提高，网络上可获得的法律文书逐渐增多。截止到2021年3月，中国裁判文书网公开法律文本数量达到10亿级别，且每日新增数以万计。公开的海量文书中蕴含着丰富的法律资源，但这些文本的格式内容差异较大，不利于机器进行大规模知识获取，因此将非结构化的文本信息转化为机器可以理解的结构化实体关系三元组（以下简称“三元组”）成为亟需解决的问题。

信息抽取(Information Extraction, IE)是自然语言处理(Natural Language Processing, NLP)领域的基础任务，其目标是从非结构化的文本中抽取结构化的三元组$\langle h,r,t\rangle$，其中h表示三元组的头实体，t表示三元组的尾实体，r表示两个实体之间的语义关系。传统上信息抽取分为两步，首先进行命名实体识别(Named Entity Recognition,NER)获得相应实体，然后进行关系分类(Relation Classification,RC)确定所选择的两实体之间的关系。传统的流水线方法无法获得两阶段任务之间的交互特征，产生了太多冗余信息，容易造成误差传播，导致三元组抽取效果降低，因此研究者开始关注联合抽取方法。实体和关系的联合抽取是指通过端到端模型同时完成实体和关系的提取。早期的联合学习方法依赖于复杂的特征工程，模型泛化性较差。后来研究者通过深度神经网络利用参数共享或多任务学习完成联合抽取。这些方法总是先识别出实体，然后将实体向量融入下一阶段任务，完成关系抽取。

但是上述联合学习方法不适用于存在重叠关系的句子。本文关系重叠遵循该文章的定义(Yuan et al., 2020)，指一个句子中存在的多个关系发生交叉，共享句中的某一实体。如果句子中存在这种重叠问题，那么三元组的抽取结果就无法达到预期。而在司法领域，由于文书所描述事实的特殊性，这种关系重叠的内容范式大量存在。图1表示法律文本中三元组的抽取过程，其输入序列为一段犯罪事实描述，输出结果为该句涉及的实体关系三元组。其中实体“苏某”涉及两个关系，若将该共享实体向量化后进行关系抽取，则容易造成关系混淆，无法准确判断“苏某”在不同关系条件下的尾实体。法律文本总是包含多个关系，关系之间经常发生实体共享，导致关系重叠。若研究者将公开领域的联合学习方法直接迁移到司法领域，模型效果将会显著降低。

Figure 1: 联合抽取过程

在多数深度学习模型中，输入句子的语法树可以为模型提供有效的信息，从而提高模型性能。目前模型主要通过语法树中的词连接构建相邻矩阵，获取词与词之间的依赖关系或者重构网络结构，以多任务学习的方式融入语法信息。该方法容易导致模型过度拟合训练集文本的语法结构，无法泛化到语法变化较大的新文本中。本文首次将有序神经元-长短期记忆神经网络(Ordered Neurons-Long Short Term Memory, ON-LSTM)(Shen et al., 2019) 应用到实体关系联合抽取任务。ON-LSTM在隐藏向量计算中通过主遗忘门和主输入门扩展流行的长短期记忆网络(Long Short Term Memory, LSTM) (Hochreiter and Schmidhuber, 1997)，这两个新门控制隐藏向量的每个神经元在句子的不同单词中的激活时间。基于这样的受控神经元，某个单词对于三元组抽取任务的重要度可以由该词在ON-LSTM 的计算过程中拥有的活跃神经元的数量确定。在模型中加入ON-LSTM 既融入了语法信息，又解决了模型对于训练集文本的过拟合

因此本文面向法律文本提出一种新的联合学习模型，改进了现有的联合学习范式。模型采用编解码器框架进行学习，首先进行关系表示，然后在相应关系的指导下识别头实体和尾实体，完成实体关系三元组的抽取。模型在编码阶段利用ON-LSTM建模输入序列语法树的层级结构，引入语法特征，提升识别效果；然后模型在编解码器之间通过多头注意力机制将关系特定的句子表示连接到原始向量中，生成多个关系解码器，解决关系重叠问题；最后模型在每个解码器上使用双向LSTM(Bi-directional Long Short Term Memory, BiLSTM))进行序列标注，获得每个字的相应标签，确定实体边界和实体类型，实现信息抽取。模型在涉毒类刑事判决书

第二十届中国计算语言学大会论文集，第589页–第599页，呼和浩特，中国，2021年8月13日至15日。

构成的数据集上进行了实验，三元组的抽取效果达到78.7%。

2 相关工作

随着司法领域信息的公开化与透明化，自然语言处理领域的技术开始被迁移到司法领域。Xiao et al. (2017)建立层级卷积神经网络，以多任务学习的方式进行粗细两种粒度的文本建模，完成法律问题分类。Luo et al. (2017)通过基于注意力机制的神经网络同时对罪名预测与法条抽取联合建模，实验结果表明这种多任务学习的方法可以更加准确地识别罪名与相关法条。Zhong et al. (2018)通过拓扑学习的方式构建一个有向无环图，将犯罪事实编码为相应向量，经过不同解码器之间的信息交互，完成判决预测。Ye et al. (2018)以BiLSTM为编码器，编码法律文本中的犯罪事实，再通过注意力机制和LSTM解码器，生成法院观点。Yang et al. (2019)设计了一个多视角的前向预测和后向验证框架，通过注意力机制将事实描述和序列特征整合到网络中，利用子任务之间的信息交互提高了判决预测的准确性。

同时研究者们越来越关注实体和关系的联合抽取，利用端到端模型抽取的结构化三元组信息方便机器获得文本知识。Miwa and Sasaki (2014)提出一种新颖的实体和关系的表格表示，利用束搜索填充表格，完成实体和关系的联合抽取。Zheng et al. (2017)提出一种新的标注策略，将实体和关系的联合抽取转化为序列标注问题，在解决句子级单关系抽取问题方面取得了较好效果。Zeng et al. (2018)提出CopyRE，该模型将相关实体通过复制机制和解码器不断生成三元组，提高了重叠关系三元组提取的准确率。Zeng et al. (2020)等人分析了CopyRE模型不稳定的原因，提出CopyMTL解决该问题，将复制机制与多任务学习结合起来，提高了多词实体的识别率。Dai et al. (2019)提出的模型为包含n个词的句子生成n个标注序列，根据给定的查询词位置生成不同的句子表达，同时抽取全部实体和涉及的关系。Fu et al. (2019)提出GraphRel，该模型分别通过线性结构和依存树结构获得文本的顺序特征和区域特征，利用两阶段的图卷积神经网络预测实体对以及它们之间的关系。Wei et al. (2020)提出CASREL框架，首先识别头实体，而后识别关系特定的尾实体，并通过严格的公式推导证明了方法的正确性。Yuan et al. (2020)提出RSAN框架，先获得句子中的关系，再利用关系注意力机制为每一个关系生成不同的句子向量，从而获得相关实体，生成实体关系三元组。

3 模型

实体关系联合抽取的目的在于通过端到端模型同时获得实体关系三元组即$\langle h,r,t\rangle$，h 和t 分别表示头实体和尾实体，r 表示实体间的语义关系。如果一个句子中蕴含多个关系，则模型会输出多个三元组。此时三元组之间可能共享相同的实体，造成关系重叠。本文提出了简单高效的联合抽取模型，解决了文本中存在的关系重叠问题。

假设将句子描述为$S=(w_1,w_2,\ldots,w_n)$，三元组描述为$triplet=\langle h,r,t\rangle$。很多联合学习的方法专注于先获得头实体$h$，而后在解码端通过序列标注等方法获得尾实体$t$和实体之间的关系$r$。这种"$h\rightarrow r\rightarrow t$" 的抽取模式容易造成头实体和尾实体的混淆，导致抽取出的三元组中头尾实体位置交换，反转关系方向。而关系方向在法律文本中尤为重要，其决定所关注实体是法律动作的施行者或接受者，错误的关系方向严重影响法律从业者对犯罪事实的认识。本文提出的方法改进了上述抽取模式，按照"$r\rightarrow h\rightarrow t$" 的顺序完成相关要素的识别，该过程中关系不再是离散数值，而是作为参数向量参与训练，实现了关系特征的融合。模型首先通过ON-LSTM编码文本，再通过多头注意力机制获得关系向量，最后在关系向量的指导下识别实体h和实体t，获得三元组$\langle h,r,t\rangle$，具体过程见图2。

3.1 文本编码

给定长度为n的句子$S=(w_1,w_2,\ldots,w_n)$，模型将句子编码为$X=(x_1,x_2,...,x_n)$，其中$x_i=(w_i^w;w_i^p;w_i^s)$。其中w_i^w 表示词向量，该向量通过Word2Vec(Mikolov et al., 2013)训练得到；w_i^p 是词性的向量表示，词性特征由此融入相应词的表达向量中；w_i^s 表示中文的笔画向量，将汉字进行拆分可获得其基本单元，将基本单元的表示向量序列通过卷积神经网络(Convolutional Neural Networks, CNN)进行最大池化，生成的新向量可以实现对汉字笔画的表征，参与模型训练从而缓解未登录词的问题。拼接上述三个向量，从而形成神经网络的输入

[1]图2虚线框住的内容为序列的隐层表达h

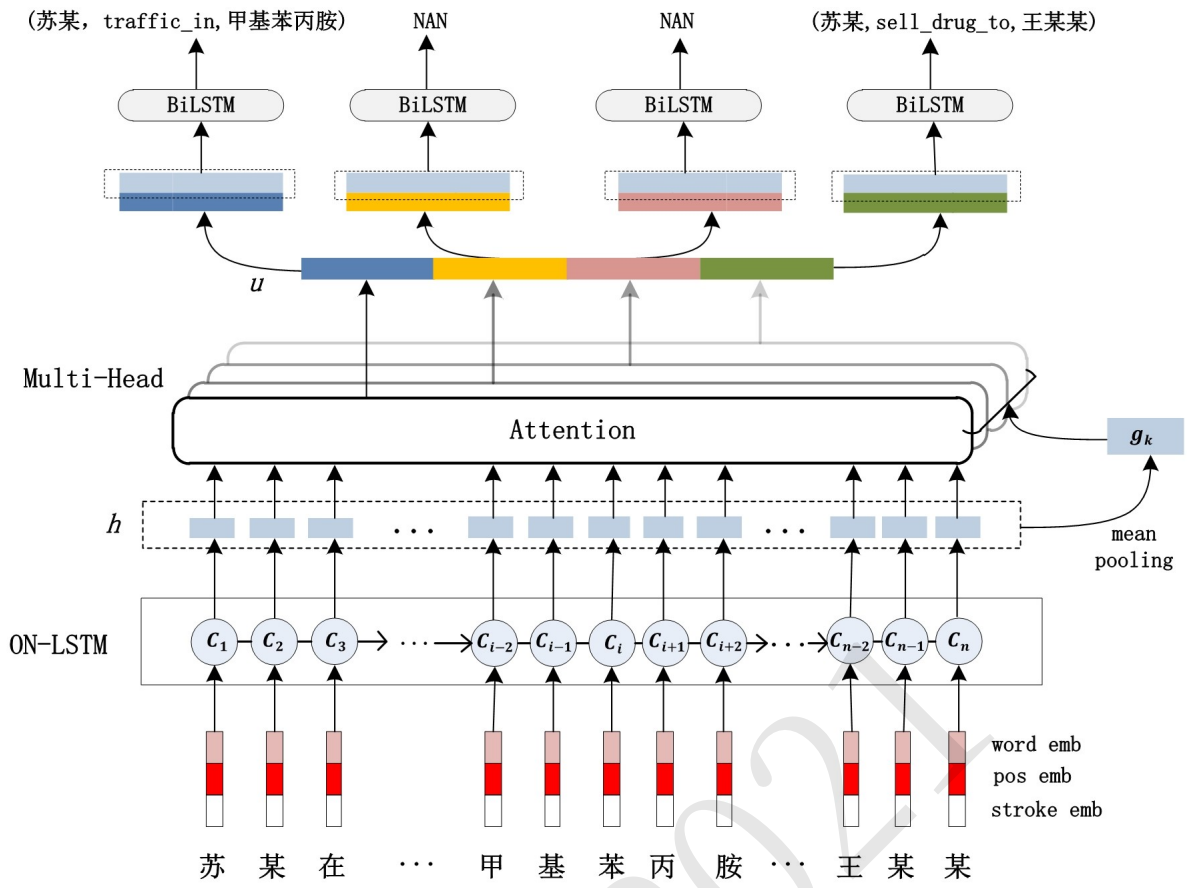

Figure 2: 模型结构图[1]

表示。文本序列第i个词编码如下：

$$x_i = Embedding(w_i), i \in [1, n] \tag{1}$$

3.2 ON-LSTM

LSTM已经广泛应用于自然语言处理领域的许多任务，其神经网络结构可以获得输入序列中每一个词的隐层向量表达。LSTM的三个控制门协同决定信息流的传递，从而在编码文本时缓解长距离依赖的问题。遗忘门f_t决定上一时刻信息的保留度，输入门i_t决定当前输入的重要度，输出门o_t决定当前状态的输出度。其公式如下：

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \tag{2}$$

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \tag{3}$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \tag{4}$$

$$\hat{c}_t = tanh(W_c x_t + U_c h_{t-1} + b_c) \tag{5}$$

$$c_t = f_t \circ c_{t-1} + i_t \circ \hat{c}_t \tag{6}$$

$$h_t = o_t \circ tanh(c_t) \tag{7}$$

为了表示每一个词对所属句子的贡献度，研究者提出了ON-LSTM。该模型优化了传统的LSTM的建模方式，通过神经元的有序排列建模了文本的语法树结构。ON-LSTM将神经元按照语法树的层级结构进行排序，信息的语法层级越高，其对应神经元的所属维度就越高，神经

元状态参数更新越慢，反之亦然。较高语法层次神经元参数更新较慢，保证了语法树的高层次单元可以保留更多的历史信息；较低语法层次神经元参数更新较快，保证了语法树的低层次单元更加关注序列的当前输入。为了完成上述目标，ON-LSTM引入主遗忘门$\hat{f}_t$和主输入门$\hat{i_t}$。其计算公式如下：

$$\hat{f}_t = cumax(W_{\hat{f}}x_t + U_{\hat{f}}h_{t-1} + b_{\hat{f}}) \tag{8}$$

$$\hat{i_t} = 1 - cumax(W_{\hat{i}}x_t + U_{\hat{i}}h_{t-1} + b_{\hat{i}}) \tag{9}$$

$$\bar{f}_t = \hat{f}_t \circ (f_t\hat{i_t} + 1 - \hat{i_t}) \tag{10}$$

$$\bar{i_t} = \hat{i_t} \circ (i_t\hat{f}_t + 1 - \hat{f}_t) \tag{11}$$

$$c_t = \bar{f}_t \circ c_{t-1} + \bar{i_t} \circ \hat{c}_t \tag{12}$$

其中$cumax$为激活函数，$cumax(x) = cumsum(softmax(x))$ 。其中第i个输入对应的隐藏向量表示为:

$$h_i = ONLSTM(x_i), i \in [1, n] \tag{13}$$

3.3 多头注意力机制

实体关系联合抽取方法面对的一个巨大挑战是关系重叠问题，即多个关系之间共享某一实体。而针对句子中存在的某一种特定关系，不同词对该关系识别的贡献度也不同。因此本文模型引入多头注意力机制解决这两个问题，每一个注意力头代表一种关系可以解决不同关系之间的实体共享问题，同时注意力自有的加权机制可以区别句子中每一个词的重要度。假设句子S中存在N种关系，则构建N个关系头，计算出N个关系特定的句子向量。其中第k种关系对应的注意力头计算过程如下：

$$g_k = avg\{h_1, h_2, \ldots, h_n\} \tag{14}$$

$$e_{ik} = v^T tanh(W_g g_k + W_h h_i) \tag{15}$$

$$\alpha_{ik} = \frac{exp(e_{ik})}{\sum_{j=1}^{n} exp(e_{jk})} \tag{16}$$

其中g_k 表示将隐层向量平均池化后获得的句子表示，v, W_g, W_h 是参与模型训练的参数。通过多个关系头的并行计算，模型完成了每个词对每个关系表达的重要度测定。u_k 表示在第k种关系指导下生成的特定句子表达向量，其计算公式如下。

$$u_k = \sum_{i=1}^{n} \alpha_{ik} h_i \tag{17}$$

多头注意力机制为每一个关系r构建了一个表达向量u_r，将其与原始的隐层向量进行拼接可作为下一部分的输入。

3.4 实体抽取

在解码阶段，模型进行序列标注从而抽取相关实体。本文将分别代表头实体和尾实体的标签H,T与代表实体位置的标签B(Begin), I(Inside), E(End), S(Single),O(Outside)结合起来，形成标签集合。模型根据预定义的关系集合长度，为每一个输入序列生成同样数量的标注序列。在某种关系对应的标签序列中，模型仅标注该关系相关的头实体和尾实体，而忽略序列中的其他词，如图3所示。若某一关系中仍存在简单的关系重叠问题，如某个头实体与多个尾实体存在关系，则按照就近原则匹配所有头尾实体对，形成实体关系三元组。

在序列标注过程中，本文建立BiLSTM处理上述过程的输入序列，将输入映射到标签空间。

$$o_i^k = BiLSTM([u_k; h_i]) \tag{18}$$

$$P(y_i^k) = softmax(W_o o_i^k + b_o) \tag{19}$$

其中W_o, b_o 是参与训练的参数，$P(y_i^k)$ 表示在第k种关系下序列中第i个词的标签概率分布。

	苏	某	在	*	*	小	区	甲	基	苯	丙	胺	贩	卖	给	王	某	某
traffic_in	B-H	E-H	0	0	0	0	0	B-T	I-T	I-T	I-T	E-T	0	0	0	0	0	0
sell_drug_to	B-H	E-H	0	0	0	0	0	0	0	0	0	0	0	0	0	B-T	I-T	E-T
possess	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
provide_shelter_for	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 3: 标注策略

4 实验

4.1 数据集

本文数据集的原始语料是非法贩卖毒品罪、非法持有毒品罪和非法容留他人吸毒罪三种涉毒类法律文本的刑事判决书。首先通过关键字匹配的规则抽取出文本中的犯罪事实表达部分，然后在此基础上进行程序的粗粒度标注，粗标结束后人员进行二次标注，再提交第三方进行核验。这种将机器粗标与人工标注相结合的模式既减少了人工参与度，也保证了标注的准确性。

参考《中华人民共和国刑法》中的规定以及三种涉毒类案件的判决依据，本文预定义了4种关系类型，分别为非法持有毒品(*possess*)，贩卖（给某人）(*sell_drug_to*)，贩卖（某种毒品）(*traffic_in*)，非法容留（某人吸毒）(*provide_shelter_for*)，这4种关系涵盖了三类涉毒案件中的各种犯罪行为。其关系数量分布如下:

关系	数量
sell_drug_to	427
traffic_in	400
possess	450
provide_shelter_for	821

Table 1: 关系种类和数量

4.2 实验设置

本文实验在表2所述的硬件环境中完成。

名称	配置
CPU	Intel(R) Xeon(R) CPU E5-2678 v3 @ 2.50GHz
GPU	NVIDIA Tesla K80@11GB
cuda	10.1
Python	3.7.6
Pytorch	1.3.1

Table 2: 实验环境

本文按照8:1:1的比例划分训练集、验证集和测试集，同时在模型的文本编码阶段使用三种表示向量完成词嵌入，分别是词向量(word_embbeding)、词性向量(pos_embedding)和笔画向量(stroke_embedding)，分别将其维度设置为$d_w = 256, d_{pos} = 15, d_{stk} = 50$，后两种向量表示通过随机初始化完成构造。本文设置一个汉字拆分后的笔画序列最大长度为5即stroke_max_len=5，而CNN的两个重要参数分别被设置为window_size=3，filter_num=50。

文本编码将机器不可计算的自然语言映射到机器可以计算的向量空间中。该阶段完成后，模型利用ON-LSTM建模输入序列的语法结构，再通过BiLSTM进行序列标注，其参数如下:

模型	in_size	hidden_size	level_size	num_layers	bidirectional
ON-LSTM	500	300	6	1	False
BiLSTM	500	300	-	1	True

Table 3: 模型参数

在编解码阶段模型均引入Dropout层防止过拟合，其中dropout_rate=0.5。在训练过程中，本文通过多次对比实验选择了模型结果最好的一组参数，其中训练轮数epoch=100，批次大小batch_size=16，学习率learning_rate=0.001，优化器optimizer=Adam。

4.3 评价指标

本文通过三元组最终提取结果的精确率（Precision, P）、召回率（Recall, R）以及F1值（F1-score, F1）评价模型性能，评价指标计算方式如下所示：

$$P = \frac{cor\_num}{pre\_num} \tag{20}$$

$$R = \frac{cor\_num}{true\_num} \tag{21}$$

$$F1 = \frac{2 * P * R}{P + R} \tag{22}$$

其中pre_num表示模型预测出的所有三元组个数，cor_num表示三元组中预测正确的实例个数，$true_num$表示文本中真实存在的三元组个数。

4.4 实验结果

本文将所设计的模型与四种基线模型进行对比，其详细阐述如下：

流水线模型1(Pipeline_nn)：在命名实体识别和关系抽取两阶段任务中均使用神经网络模型为核心组件。

流水线模型2(Pipeline_bert)：在两阶段任务中均使用BERT(Bidirectional Encoder Representations from Transformers)(Devlin et al., 2019)微调实现三元组抽取。

新型标注框架模型(NovelTagging)(Zheng et al., 2017)：提出一种新颖的标签策略，在标签中融入实体和关系信息，通过一次序列标注提取出句中所有的实体关系三元组。

复制机制模型(CopyMTL)(Zeng et al., 2020):改进CopyRE的复制机制，能够确定多字实体的边界和完成头尾实体的差异化识别，最后通过填充表格实现信息抽取。

Model	P(%)	R(%)	F1(%)
Pipeline_nn	43.2	60.1	50.2
NovelTagging	60.9	71.9	65.9
CopyMTL	70.2	74.3	72.2
Pipeline_bert	70.7	**85.6**	77.5
Our Model	**77.4**	80.1	**78.7**

Table 4: 实验结果

表4展示了全部模型结果。整体而言，本文模型优于其他基线模型。该模型所处理的语料来源于司法领域，目标实体对之间可能存在较长的字符距离，其中包含大量无关系交互的实体，编码时容易引入噪声特征。模型优点在于引入ON-LSTM建模语法树结构，避免丢失长距离实体对之间的信息；同时为每一种关系生成一个注意力头的机制，使得序列标注可以仅关注特定关系相关的实体，避免冗余实体对生成的无效信息影响模型结果。

本文模型与流水线模型的对比实验证明了联合学习方法的优越性。流水线方法建立两个独立的学习模型，其中关系抽取模型是对实体识别的结果进行关系分类，这会产生误差传递

和冗余实体对的问题，因此Pipeline_nn模型抽取效果低于其他神经网络为核心的联合抽取方法。Pipeline_bert模型使用BERT为编解码框架的核心，依赖预训练语言模型强大的文本表征能力，虽然优于部分联合学习模型，但并未解决两者之间信息交互的问题，因此效果低于本文提出的模型。

与其他联合学习方法相比，本文模型仍有其优越性。NovelTagging方法提出的标签框架无法处理关系重叠问题，本文方法显著高于该方法证明模型可以识别涉及多个关系的共享实体；而可以处理关系重叠问题的CopyMTL方法效果同样低于本文模型结果，证明模型中引入语法信息与多头注意力机制的必要性。多头注意力机制为预定义的每一个关系确定了一个关系头，每个关系头指导下的序列标注仅识别涉及当前关系的实体，忽略其他实体。在多个关系头相互独立的识别模式下，共享实体在一次识别中仅承担相应关系的角色，忽略是否作用于其他关系，从而实现了重叠关系的分解。模型在每种关系下均执行一次实体识别，将识别出的头尾实体与对应的关系组合在一起，完成实体关系三元组的抽取。

5 结果分析

5.1 消融实验

本文进行了更细致的实验证明模型每个成分的贡献度。在本文提出的模型中，ON-LSTM结构对语法树建模，引入语法特征；词性向量引入词性特征，笔画向量可以缓解未登录词的问题；同时实验采用随机生成的关系向量连接隐层向量指导实体识别，证明多头注意力机制的作用。

Model	P(%)	R(%)	F1(%)
Our Model	**77.4**	**80.1**	**78.7**
–ON-LSTM	67.7	77.9	72.4(-6.3)
–MultiHead attetion	70.4	77.2	73.6(-5.1)
–POS embedding	76.9	79.2	78.0(-0.7)
–Stroke embedding	77.3	78.6	77.9 (-0.8)

Table 5: 模型成分贡献表

通过表5的实验结果可以得出，在模型中使用BiLSTM取代ON-LSTM导致了模型效果的最大幅度下降，证明语法信息的引入尤为重要。法律文本中头尾实体之间一般会有较长的过程描述，所以两者之间具有较长的字符距离。ON-LSTM建模语法树结构可以保证长距离后的高维度神经元仍然保持较多的历史信息，帮助模型有效地过滤实体对之间的无关信息，缓解长距离依赖问题。利用随机生成的关系向量取代多头注意力机制，虽然可以实现关系条件下的实体识别过程从而分解重叠关系，但是随机初始化的状态不稳定，生成的关系向量也并不能体现出原序列中每个字对三元组抽取的贡献度，因此模型效果快速下降。从模型中移除词性向量和笔画向量，同样弱化了模型效果。词性向量可以增强文本序列中每一个字的表示，词性特征可以辅助ON-LSTM建模语法结构；笔画向量引入了字形特征，实现了相似字词向量之间的互相补充，缓解了未登录词的问题；实验结果证明在模型中加入这两个特征可以提高三元组识别效果。

5.2 不同解码器的实验结果

本文亦探索实体识别阶段使用不同解码器对模型的影响，实验证明使用BiLSTM进行解码效果最佳。输入序列经过模型计算获得的隐层向量蕴含着复杂的特征，利用BiLSTM在高维向量空间中捕获关键特征用于预测三元组，提高了三元组的识别率。具体实验结果如图4所示。

与BiLSTM相比，其他神经网络结构在解码阶段表现效果较差，三元组抽取结果较低。在解码阶段模型在关系特定的句子表示中进行序列标注，需要同时解码出关系信息和相关的实体信息，并忽略与该关系不相关却与其他关系相关的易混淆实体，因此该阶段神经网络应具有解码复杂特征的能力。单向循环神经网络(RNN, GRU, LSTM)中当前状态仅与之前的状态相关，并不关注后续文本序列，不利于捕获全局特征，获得关系信息。BiLSTM与BiGRU均获得了序

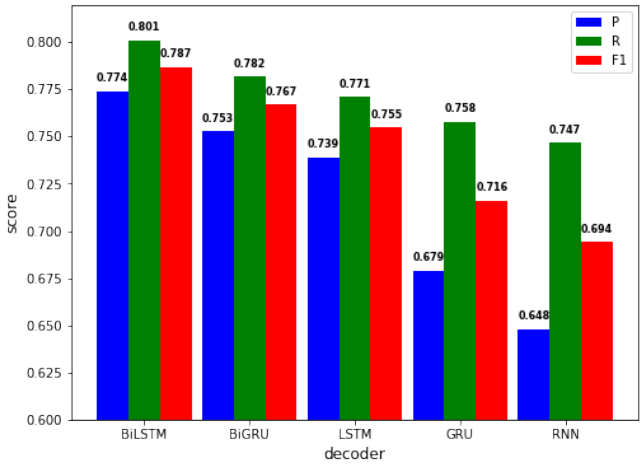

Figure 4: 不同解码器效果

列的双向特征，但前者网络结构中的门机制更加复杂，更适合于处理文本序列较长的法律文本。

5.3 模型案例分析

图5所示为Pipeline_bert，NovelTagging，CopyMTL和本文模型对三种法律文本的预测结果。第一种法律文本中不存在关系重叠问题，第二种和第三种法律文本存在重叠关系，即句子中某一实体与其他多个实体发生关系。同时第三种法律文本中包含字符距离相对较长的实体对，以此评判模型解决长距离依赖问题的能力。

	经审理查明：2013 年 12 月 30 日晚，被告人万某因吸食毒品被正安县公安局瑞溪派出所民警抓获，并从其身上收缴毒品甲基苯丙胺 12.8 克 。	经审理查明，1、贩卖毒品的事实：2015 年 4 月 20 日，被告人邓某在从化区城郊街旺城北路 48 号旺兴家园附近，卖给吸毒人员梁某玲毒品冰毒一小包，并收取 200 元。	2016 年 8 月 18 日 13 时许，被告人黄顺聪在昆明市官渡区官渡镇西庄村 710 号门口将毒品可疑物净重 0.5 克卖给胡某某……从各处查获的毒品可疑物提取的送检检材中均检出海洛因成分。
Golden Label	万某，possess，甲基苯丙胺(span=33)	邓某，sell_drug_to，梁某玲(span=28) 邓某，traffic_in，冰毒(span=31)	黄顺聪，sell_drug_to，胡某某(span=32) 黄顺聪，traffic_in，海洛因(span=182)
Novel-Tagging	万某，possess，甲基苯丙胺	邓某，sell_drug_to，梁某玲	黄顺聪，sell_drug_to，胡某某
CopyMTL	万某，possess，甲基苯丙胺	邓某，sell_drug_to，梁某玲 邓某，traffic_in，冰毒	黄顺聪，sell_drug_to，胡某某
Pipeline_bert	万某，possess，甲基苯丙胺	邓某，sell_drug_to，梁某玲 邓某，traffic_in，冰毒	黄顺聪，sell_drug_to，胡某某
Our Model	万某，possess，甲基苯丙胺	邓某，sell_drug_to，梁某玲 邓某，traffic_in，冰毒	黄顺聪，sell_drug_to，胡某某 黄顺聪，traffic_in，海洛因

Figure 5: 不同模型预测结果示例

由图5的结果可以看出，当文本中不存在重叠关系时，实体关系三元组易于准确抽取；当存在重叠关系时，NovelTagging方法中的序列标注无法解码这类信息，只能按照就近原则输出一组关系。而CopyMTL可以利用特殊的复制机制分解重叠关系，Pipeline_bert依赖于BERT蕴含

的丰富特征减小流水线方法带来的传递误差，完成实体关系三元组的抽取。但当文本长度增加时，CopyMTL所创造出的复制矩阵过大，不利于模型捕捉有效信息；而BERT表征长文本时，其所蕴含的特征不利于在长距离实体之间进行交互，因此降低了长文本的预测能力。在图5所示的第三种文本中CopyMTL和Pipeline_bert仅解析出距离较近的实体对之间的关系即“(黄顺聪,sell_drug_to,胡某某)”，并未抽取出长距离实体对形成的三元组“(黄顺聪,traffic_in,海洛因)”。本文模型利用ON-LSTM建模长文本的层级结构，实现了“毒品可疑物”和“送检检材”这两个长距离关键词之间的链接，抽取出了文本中的所有三元组。

6 结论

本文面向法律文本提出了一个实体关系三元组联合抽取模型，该模型可将非结构化的文本转换为结构化的三元组信息。本文模型引入了ON-LSTM对输入句子的语法结构建模，并通过多头注意力机制结合BiLSTM解决三元组抽取中的关系重叠问题。实验结果证明，本文提出的方法优于传统的流水线方法，同时也比其他的联合学习方法更加精确。本文进行了细粒度的消融实验，证明了模型中每一个成分的重要性。在未来的工作中，将探索更加高效的语法信息引入模式和重叠关系分解方法，进一步完善三元组的抽取模型，形成面向司法领域的知识图谱，构建结构化的法律信息知识库。

致谢

本论文工作受国家重点研发计划项目（2018YFC0831403）资助

参考文献

Dai Dai, Xinyan Xiao, Yajuan Lyu, Shan Dou, Qiaoqiao She, and Haifeng Wang. 2019. Joint extraction of entities and overlapping relations using position-attentive sequence labeling. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*, pages 6300–6308. AAAI Press.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: pre-training of deep bidirectional transformers for language understanding. In Jill Burstein, Christy Doran, and Thamar Solorio, editors, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pages 4171–4186. Association for Computational Linguistics.

Tsu-Jui Fu, Peng-Hsuan Li, and Wei-Yun Ma. 2019. Graphrel: Modeling text as relational graphs for joint entity and relation extraction. In Anna Korhonen, David R. Traum, and Lluís Màrquez, editors, *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*, pages 1409–1418. Association for Computational Linguistics.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural Comput.*, 9(8):1735–1780.

Bingfeng Luo, Yansong Feng, Jianbo Xu, Xiang Zhang, and Dongyan Zhao. 2017. Learning to predict charges for criminal cases with legal basis. In Martha Palmer, Rebecca Hwa, and Sebastian Riedel, editors, *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017, Copenhagen, Denmark, September 9-11, 2017*, pages 2727–2736. Association for Computational Linguistics.

Tomás Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. In Yoshua Bengio and Yann LeCun, editors, *1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA, May 2-4, 2013, Workshop Track Proceedings*.

Makoto Miwa and Yutaka Sasaki. 2014. Modeling joint entity and relation extraction with table representation. In Alessandro Moschitti, Bo Pang, and Walter Daelemans, editors, *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29,*

2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL, pages 1858–1869. ACL.

Yikang Shen, Shawn Tan, Alessandro Sordoni, and Aaron C. Courville. 2019. Ordered neurons: Integrating tree structures into recurrent neural networks. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net.

Zhepei Wei, Jianlin Su, Yue Wang, Yuan Tian, and Yi Chang. 2020. A novel cascade binary tagging framework for relational triple extraction. In Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel R. Tetreault, editors, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, pages 1476–1488. Association for Computational Linguistics.

Guangyi Xiao, Jiqian Mo, Even Chow, Hao Chen, Jingzhi Guo, and Zhiguo Gong. 2017. Multi-task CNN for classification of chinese legal questions. In Omar Hussain, Lihong Jiang, Xiang Fei, Ci-Wei Lan, and Kuo-Ming Chao, editors, *14th IEEE International Conference on e-Business Engineering, ICEBE 2017, Shanghai, China, November 4-6, 2017*, pages 84–90. IEEE Computer Society.

Wenmian Yang, Weijia Jia, Xiaojie Zhou, and Yutao Luo. 2019. Legal judgment prediction via multi-perspective bi-feedback network. In Sarit Kraus, editor, *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, pages 4085–4091. ijcai.org.

Hai Ye, Xin Jiang, Zhunchen Luo, and Wenhan Chao. 2018. Interpretable charge predictions for criminal cases: Learning to generate court views from fact descriptions. In Marilyn A. Walker, Heng Ji, and Amanda Stent, editors, *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2018, New Orleans, Louisiana, USA, June 1-6, 2018, Volume 1 (Long Papers)*, pages 1854–1864. Association for Computational Linguistics.

Yue Yuan, Xiaofei Zhou, Shirui Pan, Qiannan Zhu, Zeliang Song, and Li Guo. 2020. A relation-specific attention network for joint entity and relation extraction. In Christian Bessiere, editor, *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020*, pages 4054–4060. ijcai.org.

Xiangrong Zeng, Daojian Zeng, Shizhu He, Kang Liu, and Jun Zhao. 2018. Extracting relational facts by an end-to-end neural model with copy mechanism. In Iryna Gurevych and Yusuke Miyao, editors, *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 1: Long Papers*, pages 506–514. Association for Computational Linguistics.

Daojian Zeng, Haoran Zhang, and Qianying Liu. 2020. Copymtl: Copy mechanism for joint extraction of entities and relations with multi-task learning. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*, pages 9507–9514. AAAI Press.

Suncong Zheng, Feng Wang, Hongyun Bao, Yuexing Hao, Peng Zhou, and Bo Xu. 2017. Joint extraction of entities and relations based on a novel tagging scheme. In Regina Barzilay and Min-Yen Kan, editors, *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 1: Long Papers*, pages 1227–1236. Association for Computational Linguistics.

Haoxi Zhong, Zhipeng Guo, Cunchao Tu, Chaojun Xiao, Zhiyuan Liu, and Maosong Sun. 2018. Legal judgment prediction via topological learning. In Ellen Riloff, David Chiang, Julia Hockenmaier, and Jun'ichi Tsujii, editors, *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018*, pages 3540–3549. Association for Computational Linguistics.