**文章编号：**1003-0077（2018）00-0000-00

# 融合依存信息 Attention 机制的药物关系抽取研究*

李丽双，钱爽，周安桥，刘阳，郭元凯

（大连理工大学 计算机科学与技术学院，辽宁 大连 116024）

**摘要：**药物关系（Drug-Drug interaction, DDI）抽取是生物医学关系抽取领域的重要分支，现有方法主要强调实体、位置等信息对关系抽取的影响。相关研究表明，依存信息对于关系抽取具有重要的作用，如何合理地利用依存信息是关系抽取研究中需要解决的问题。本文提出一种融合依存信息 Attention 机制的药物关系抽取模型，衡量最短依存路径与句子的相关性，捕捉对实体间关系有用的信息。首先使用双向 GRU（BiGRU）网络分别学习原句子和最短依存路径（Shortest Dependency Path，SDP）的语义信息和上下文信息，然后通过 Attention 机制将 SDP 信息与原句子信息融合，最后利用融合依存信息之后的句子表示进行分类预测。本文在 DDIExtraction2013 语料上进行了实验评估，模型 F 值为 73.72%。

**关键词：**生物医学关系抽取；药物关系抽取；依存信息；Attention；

**中图分类号：**TP391　　　　**文献标识码：**A

# Research on Drug-Drug Interaction Extraction with the Attention Mechanism Based on the Fusion of Dependency Information

LI Lishuang, QIAN Shuang，ZHOU Anqiao, LIU Yang, GUO Yuankai

(School of Computer Science and Technology, Dalian University of Technology, Dalian, Liaoning 116024, China)

**Abstract:** Drug-Drug interaction (DDI) extraction is an important branch of the biomedical relationship extraction field. Most of existing methods emphasize the key information such as entities and positions in the sentences. Related research shows that dependency information is very significant for relation extraction. There is a problem to be solved that how to make good use of dependency information. This paper proposes a Drug-Drug interaction extraction model based on the Attention mechanism incorporating dependency information. The correlation between the shortest dependency path and the sentence is measured to capture the useful information in the sentence. Firstly, this model uses BiGRU network to learn the semantic information and context information of the original sentence and the Shortest Dependency Path (SDP) respectively. Secondly, fuse the SDP information with the original sentence information through the Attention mechanism. Finally, the final sentence representation is used to classify and predict DDI. This approach is evaluated on DDIExtraction2013 corpus. The experimental results show that our system achieves a micro F-scores of 73.72%.

**Keywords:**Biomedical Relation Extraction;Drug-Drug Interaction Extraction;Dependency Information;Attention;

## 1 引言

生物医学实体关系抽取，旨在抽取生物医学文献中两个实体之间的关系并识别出其关系类型，例如基因、疾病、蛋白质及药物等实体之间的关系。药物相互作用（Drug-Drug interaction, DDI）关系抽取是一种多分类且不区分关系方向的实体间关系抽取任务。研究 DDI 能够为生物医学数据库的构建和维护提供更深层的信息，为疾病治疗、药物开发、生命科学研究提供更为重要的参考。

---

***收稿日期：**　　　　**定稿日期：**

**基金项目：** 国家自然科学基金项目(no.61672126)。

**作者简介：**李丽双（1967—），女，教授，博士生导师，本文通信作者，主要研究领域为自然语言处理、信息抽取与文本挖掘。

近年来，药物-药物关系抽取模型一般分为基于浅层机器学习和深度学习两种方法。例如，Chowdhury [1]等使用两阶段系统来实现 DDI 任务，第一阶段使用特征核、浅层语言学核和闭包树核相结合的组合核分类器进行二分类，第二阶段根据四种不同的 DDI 类型构建组合核分类器进行多分类，在 DDIExtraction2013 评测[2]中获得最好的成绩，F 值达到 65%。Kim[3]等使用词特征、词对特征、解析树特征和名词短语一致性等特征，构建基于特征的核函数，在 DDIExtraction2013 语料上 F 值达到 67.0%。浅层机器学习方法主要依赖于大量复杂的特征集来提高模型性能，但设计和抽取特征需要耗费大量的人力物力。相比之下，深度学习方法能够自动组合低层特征形成更加抽象的高层特征表示，避免了大量人工特征的使用，在关系抽取等领域应用广泛。例如，Liu[4]等仅使用词向量和位置向量作为输入，未添加任何人工特征，结合卷积神经网络实现分类，在 DDIExtraction2013 语料上 F 值达到 69.75%。Quan[5]等提出了多通道卷积神经网络（MCCNN）获取多种语义表示，在 DDIExtraction2013 语料上 F 值达到 70.2%。

注意力 Attention[6]机制最初只应用于计算机视觉中的图片识别任务，现已经被成功应用于问答系统、机器翻译等多个自然语言处理领域。在 DDI 关系抽取任务中，Attention 机制也展现出了明显的优势。例如，Miwa [7]等使用 Attention 机制强调实体信息在句子中的重要性，在 DDIExtraction2013 语料上的 F 值达到 69.12%。Sahu[8]等构建池化 Attention 机制判断句子中各部分的重要性，在 DDIExtraction2013 语料上的 F 值为 71.48%。Zhou[9]等使用 Attention 机制突出句子中距离实体近的单词对关系的影响，在 DDIExtraction2013 语料上的 F 值达到 72.99%。以上 Attention 机制主要突出实体信息、距离信息以及池化作用的影响，但没有从句法结构的角度判断句子各部分的重要性。也就是说，如何构建句法结构的注意力机制从而有效地提高关系抽取的性能还有待研究。

针对 DDI 任务而言，语料中存在着大量复杂句子，例如具有多个从句、具有较复杂的并列关系、具有最长句子长度超过 150 词等特点的复杂句子，仅从原句子层面很难学习到语义和句法结构等依存信息。因此，引入句法信息对于 DDI 抽取至关重要。在一般领域关系抽取中，相关研究[10-12]表明，句法信息例如依存解析树、最短依存路径等，能够获得更多有效信息，去除多余的噪声，从而提高复杂关系抽取性能。其中，利用句法信息的方式主要为 Liu[10]等将句法信息和原句信息词向量拼接作为模型输入、Yan[11]等将句法信息和原句信息的神经网络输出拼接以及 Miwa [12]等树型网络结构等。在 DDI 任务中，主要通过拼接方式使用依存信息提高模型性能，例如，Zhang[13]等使用多层 RNN 网络集成最短依存路径 SDP 信息和原句子信息，首先输入原始句子和对原句进行依存分析得到的最短依存路径 SDP 序列，分别经过双向 RNN（BiRNN）学习，然后将 BiRNN 的输出拼接，最后使用一个顶层 RNN 结合原句子整体信息和 SDP 信息，在 DDIExtraction2013 语料上 F 值达到 72.9%。虽然上述方法使用 SDP 信息丰富了模型的输入信息，但是这种结合原句信息和 SDP 信息的方式存在一定的问题。其一，直接拼接原句子序列和 SDP 序列的特征，不能很好地判断序列中每个单词的重要性，无法准确地获得 SDP 序列包含的依存信息。其二，顶层 BiRNN 能学习输入序列的时序信息，而原句子序列和 SDP 序列之间并不存在时序关系，可能会导致模型引入过量信息。综上所述，如何更合理地利用依存信息是当前关系抽取中需要解决的问题。

本文提出了一种融合依存信息 Attention 机制的药物关系抽取模型，使用注意力机制对原句子信息和最短依存路径 SDP 信息进行融合，该方法能够从句法结构的角度衡量原句子各部分对于实体间关系的重要性。首先，输入原句子序列和最短依存路径 SDP 序列；然后，将两个序列的词向量分别送入两个独立的 BiGRU 中学习其语义信息；接下来，使用最短依存路径 SDP 信息计算原句子序列的 Attention 权重；最后，使用加权后的句子表示进行分类预测。在 DDIExtraction2013 语料上的实验结果表明，该模型在未使用任何人工特征的情况下，有效地提高了模型性能，F 值为 73.7%，达到了目前的先进水平。

## 2 方法

本文提出了融合依存信息 Attention 机制的深度神经网络模型，模型架构如图 1 所示：

（1）输入层：该模型为多输入模型，分别为原句子序列（Sentence sequence）和最短依存路径序列（SDP sequence）。

（2）嵌入层：通过查询训练好的词向量表将输入的语句转换为相应的词向量序列。

（3）编码层：使用两个独立的 BiGRU，分别学习原句子和 SDP 序列的抽象语义表示。

（4）Attention 层：计算原句子和 SDP 的相似度矩阵，融合 SDP 信息得到 Attention 权重，对原句子加权求和得到最终的句子表示。

（5）输出层：利用 Attention 层的输出，即融合过 SDP 信息的句子表示进行分类预测。

下面将详细介绍最短依存路径和模型每层的设计细节。

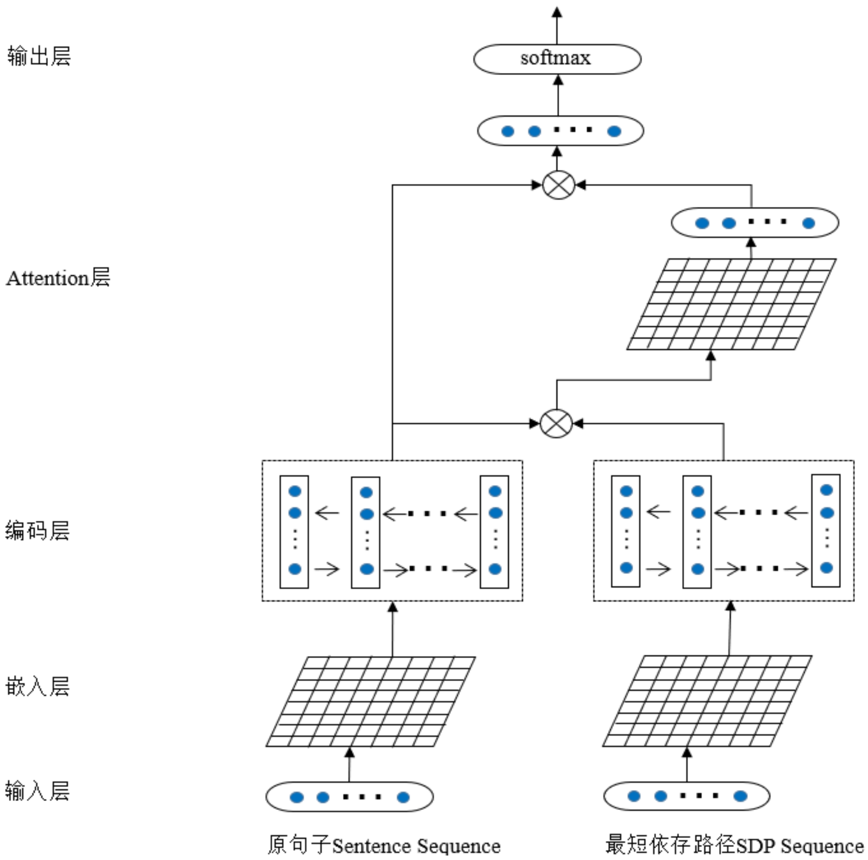



图 1 基于融合 SDP 信息 Attention 机制的 DDI 抽取模型结构图

### 2.1 最短依存路径 SDP

最短依存路径 SDP 信息表征两个实体之间的依存关系，能够包含句子表达的主要信息，同时摒弃冗余的噪声信息。GDep 依存句法分析器由日本东京大学 Tsujii 实验室研发，利用生物医学领域资源 GENIA Treebank 分析器对其训练，使得 GDep 分析器能够更好适用于生物医学文本挖掘领域。本文不采用传统的依存树状结构，而是将两个实体间的最短路径作为词序列输入进行训练。

如图 2 所示，原句子“*Patients treated with acebutolol plus catecholamine-depletors should, therefore, be observed closely for evidence of marked bradycardia or hypotension which may present as vertigo, syncope/presyncope, or orthostatic changes in blood pressure without compensatory tachycardia.*”，“*acebutolol*”为第一个目标实体，“*catecholamine-depletors*”为第二个目标实体。两个实体之间的最短依存路径为“*acebutolol catecholamine-depletors*”，两个实体之间的依存关系为“*amod*”，两个实体之间的 DDI 类型为“*advice*”。

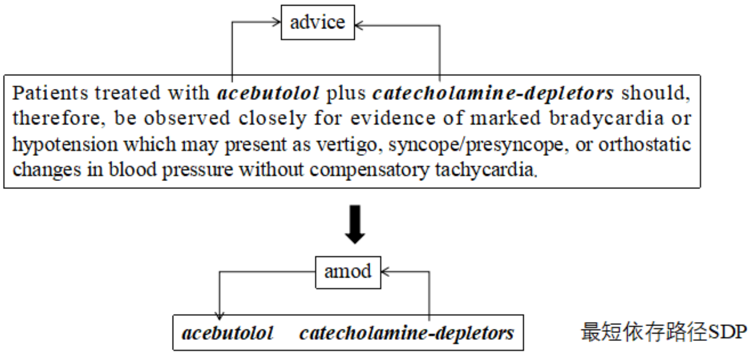



图 2 最短依存路径实例图

## 2.2 嵌入层

该模型有两个输入，即原句子序列（Sentence Sequence）和最短依存路径序列（SDP Sequence）。为了获得词语的句法和语义信息，需要将输入序列的词语转化为词向量。原句子序列的词向量输入表示分别为$X=\{x_1,x_2,\dots,x_i,\dots,x_n\}$，$x_i\in R^d$为句子中第$i$个单词的词向量表示（$i\epsilon[1,n]$），$n$为句子长度，$d$为词向量维度。最短依存路径 SDP 序列的词向量表示为$S=\{s_1,s_2,\dots,s_j,\dots,s_m\}$，$s_j\in R^d$为最短路径中第$j$个单词的词向量表示（$j\epsilon[1,m]$），$m$为 SDP 序列长度。最后，将句子词向量表示$X$和最短依存路径词向量表示$S$，作为模型编码层的输入。

## 2.3 编码层

递归神经网络 RNN 是自然语言处理常用的神经网络模型，对处理时间序列和语言文本序列问题有独特的优势。然而，传统的 RNN 模型由于输入序列过长可能产生梯度消失或者梯度爆炸现象。为了解决这一问题，Chung[14]等提出了 Gated Recurrent Unit（GRU）模型，通过特殊设计的门结构使得模型有选择的保存上下文信息。

在 GRU 中，新记忆$\widetilde{h_t}$由过去的隐含状态$h_{t-1}$和新输入$x_t$共同决定，如公式（1）所示：

$$\tilde{h}_t=\tanh(W\cdot[g_t*h_{t-1},x_t]) \tag{1}$$

其中，$g_t$为重置门，用来判定$h_{t-1}$对$\widetilde{h_t}$的重要程度，如公式（2）所示：

$$g_t=\sigma(W_g\cdot[h_{t-1},x_t]) \tag{2}$$

$z_t$为更新门，决定了过去隐含状态$h_{t-1}$向下一个状态传递的程度，如公式（3）所示：

$$z_t=\sigma(W_z\cdot[h_{t-1},x_t]) \tag{3}$$

隐含状态$h_t$由过去的隐含状态$h_{t-1}$和新记忆$\widetilde{h_t}$共同决定，并通过更新门$z_t$来判断其重要程度，如公式（4）所示：

$$h_t=(1-z_t)*h_{t-1}+z_t*\tilde{h}_t \tag{4}$$

其中，σ是激活函数 *sigmoid*，*tanh* 表示双曲正切激活函数。$x_t$是 GRU 单元输入，$W_z,W_g,W$代表学习得到的参数。*t* 表示第 *t* 时刻，$h_t$表示 GRU 第 *t* 时刻的输出。

为了得到序列的上下文信息，本文采用双向 GRU（BiGRU）得到抽象语义表示。如公式（5）所示，每个句子的最终隐层表示$h_t(x)$由正向和逆向网络的输出拼接得到：

$$h_t(x)=[\overrightarrow{h_t}(x),\overleftarrow{h_t}(x)] \tag{5}$$

本文对原句子输入序列 *X* 和 SDP 输入序列 *S* 分别经过两个独立的 BiGRU 模型学习其语义信息和上下文信息，如公式（6）所示：

$$h_{sen,t} = BiGRU_{sen}(h_{sen,t-1}, x_t)$$
$$h_{sdp,t} = BiGRU_{sdp}(h_{sdp,t-1}, x_t) \tag{6}$$

原句子的 BiGRU 输出序列为 $H_{sen} = \{h_{sen,1}, h_{sen,2}, ..., h_{sen,l}\}(H_{sen} \in R^{n\times 2l})$，SDP 的 BiGRU 输出序列为 $H_{sdp} = \{h_{sdp,1}, h_{sdp,2}, ..., h_{sdp,l}\}(H_{sdp} \in R^{m\times 2l})$，$l$为 BiGRU 隐层单元数。

## 2.4 Attention 层

本文提出了一种融合依存信息 Attention 机制，图 3 给出了信息融合的过程。其中，$H_{sen}$和γ的每个单元表示一个向量，其他单元表示一个值。

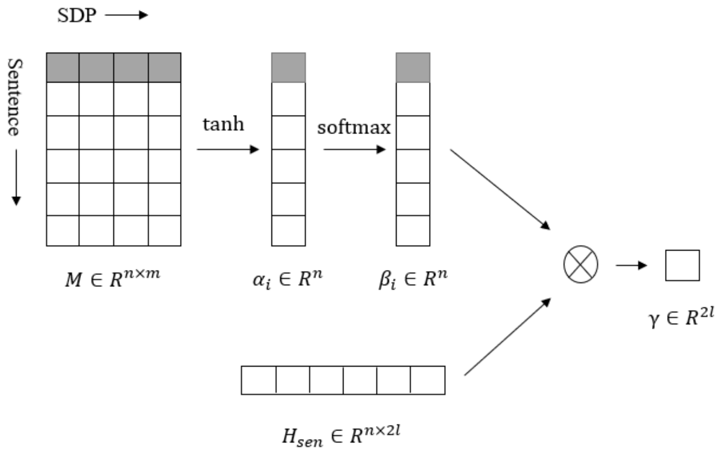



图 3 融合依存信息 Attention 机制原理图

首先，将编码层的输出序列$H_{sen} \in R^{n\times 2l}$和$H_{sdp} \in R^{m\times 2l}$作为输入，计算最短路径 SDP 与原句子 Sentence 的相似矩阵$\boldsymbol{M}$，如公式（7）所示：

$$M(i,j) = H_{sen}(i)^T \cdot H_{sdp}(j) \tag{7}$$

其中，$\boldsymbol{M}(i,j) \in R^{n\times m}$为矩阵中第 $i$ 行第 $j$ 列的分数值，矩阵 $\boldsymbol{M}$ 中的每个分值代表对应的最短路径 SDP 中一个单词与原句子 Sentence 中一个单词的相关性。

然后，使用 SDP 信息计算原句子各部分的 Attention 权重，具体如公式（8-9）所示。其中，公式（8）使用 *tanh* 函数得到考虑 SDP 信息的分值$\alpha_i \in R^n$，$W_\alpha \epsilon R^{1\times n}$为权重矩阵。公式（9）使用 *softmax* 函数计算原句子中每个单词的 Attention 权重$\beta_i \in R^n$。

$$\alpha_i = \tanh(W_\alpha \cdot M + b) \tag{8}$$

$$\beta_i = \frac{\exp(\alpha_i)}{\sum_j \exp(\alpha_i)} \tag{9}$$

$$\gamma = \sum_i^n \beta_i \cdot H_{sen}(i) \tag{10}$$

最后，使用得到的 Attention 权值$\beta_i$与原句子$H_{sen}$进行加权求和，如公式（10）所示，得到融合过 SDP 信息的句子表示$\gamma \in R^{2l}$。

## 2.5 输出层

模型输出层将 Attention 层的输出γ作为最后的分类特征送入全连接层进行分类，候选药物-药物关系对 $y$ 属于第 $c$（$c \in C$）种 DDI 类型的概率$P(y=c)$如公式（11）所示：

$$P(y=c) = soft\max(W_\gamma \cdot \gamma + b) \tag{11}$$

$$\hat{y} = \arg\max_{y \in C}(P(y = c)) \quad (12)$$

其中，W和 $b$ 为权重矩阵和偏置，全连接层的激活函数为 *softmax*，$C$为 DDI 类型标签的集合$C = \{negative，effect，mechanism，advice，int\}$。最后，使用公式（12）计算概率最大的类别标签$\hat{y}$，即为候选药物-药物对的关系类型。

# 3 实验

## 3.1 实验数据与设计

### 3.1.1 语料介绍与预处理

本实验使用 DDIExtraction2013[15,16]语料库，该语料库的文本来自 DrugBank 数据库和以药物关系为主题的 MEDLINE 文献摘要，共计 1017 篇文档。该语料库包含 3 个数据库，分别为 DrugBank、MEDLINE 和包含前两个数据库的 OverAll 数据库，本实验使用 OverAll 数据库。DDIExtraction 任务是多分类任务，共包含五种标注类型，分别为 negative、effect、mechanism、advice、int。语料的具体情况统计如下表 1 所示。

表 1 DDIExtraction2013 语料介绍

| | 训练集 | | | 测试集 | | |
|---|---|---|---|---|---|---|
| | DrugBank | MEDLINE | OverAll | DrugBank | MEDLINE | OverAll |
| negative | 22118 | 1547 | 23772 | 4367 | 345 | 4712 |
| effect | 1535 | 152 | 1687 | 298 | 62 | 360 |
| mechanism | 1257 | 62 | 1319 | 278 | 24 | 302 |
| advice | 818 | 8 | 826 | 214 | 7 | 221 |
| int | 178 | 10 | 188 | 94 | 2 | 96 |

在进行实体关系抽取之前进行了语料预处理，主要进行了以下操作：

（1）由于语料中的句子可能存在两个或者两个以上的实体，本文针对每对实体生成一个句子实例。如果句中有 $n$ 个实体，则生成$C_n^2$个句子实例。

（2）对语料集利用 Stanford Tokenizer 进行分词处理，并且把数据集中的字符全部转化为小写。

（3）根据相关研究[4,7,9]替换句中药物实体，具体将目标药物实体替换成“drug1”和“drug2”，将句中的其他实体替换成“drug0”，使模型更好的学习句子中能表征实体关系的语义信息而非实体本身的语义信息。

### 3.1.2 实验设置与评价标准

本文实验使用 Keras 深度神经网络框架来实现 DDI 抽取任务。结合 DDI 语料和 Pubmed 下载的背景语料，使用 Glove 工具训练得到词向量，词向量维度为 200 维。实验使用的超参数如表 2 所示:

表 2 实验涉及的超参数设置

| 参数名 | 值 |
|---|---|
| 词向量维度 $d$ | 200 |
| 批大小 Batchsize | 256 |
| Sentence BiGRU 隐层单元数 | 300 |
| SDP BiGRU 隐层单元数 | 300 |
| Dropout | 0.5 |
| 学习率 | 0.0001 |

本文使用的评测标准是准确率（*Precision*，*P*）、召回率（*Recall*，*R*）和 F 值（*F-score*，

$F$），如公式（13-15）所示：

$$P = \frac{TP}{TP + FP} \tag{13}$$

$$R = \frac{TP}{TP + FN} \tag{14}$$

$$F = \frac{2PR}{P + R} \tag{15}$$

其中，$TP$ 表示正例中判断正确的样本数，$FP$ 表示负例中判断错误的样本数，$FN$ 表示正例中判断错误的样本数。$F$ 值能够评价准确率和召回率的综合性能。

### 3.2 实验结果与分析

为了验证融合依存信息 Attention 机制的有效性，本文做了多个对比实验，实验结果如表 3 所示。

表 3　融合依存信息 Attention 模型有效性对比实验

| 方法 | P | R | F |
|---|---|---|---|
| Sentence + BiGRU | 70.84% | 73.95% | 72.36% |
| SDP + BiGRU | 55.08% | 49.23% | 51.9% |
| Sentence +SDP + concatenate | 72.73% | 69.46% | 71.06% |
| **Sentence +SDP + 融合依存信息 Attention** | **73.38%** | **74.06%** | **73.72%** |

从表 3 可以看出：

（1）只分别使用原句子、SDP 信息作为模型输入时，F 值为 72.36%和 51.9%。实验结果表明，当单独使用 SDP 信息作为输入时，模型性能较低。主要原因是 SDP 信息可以提取出两实体之间的关键词，摒弃冗余的噪声信息，适用于具有复杂结构的句子，但对于结构简单的句子存在的噪声较少，SDP 可能会丢失更多有用的信息，导致识别精度降低。然而，语料中结构简单的句子数大于复杂句子数，所以只使用 SDP 作为输入时模型整体性能较低。

（2）在同时使用原句子信息和 SDP 信息作为模型输入时，将常用拼接方式（concatenate）和本文的 Attention 融合方式进行了对比实验。拼接方式指将输入的原句子序列和 SDP 序列分别送入两个独立的 BiGRU 模型中学习，然后拼接两个 BiGRU 的输出，最后对拼接后的序列进行学习分类，该方法 F 值为 71.06%。

从上面实验结果可以看出，直接拼接 SDP 信息后模型性能相比于单独用 SDP 作为输入时 F 值升高了 19.16%，比单独用原句子信息作为输入时降低了 1.3%。主要原因是采用两种输入序列经 BiGRU 的结果直接拼接的方式，会导致模型同等程度地学习两种信息。然而对于复杂程度不同的实例，原句子信息和 SDP 信息对关系抽取的影响程度不同，直接拼接不但没有很好地利用 SDP 信息反而影响简单句子的判断，导致拼接后模型整体性能降低。与之相比，本文使用融合依存信息的 Attention 机制，可以在充分学习原句子信息的基础上利用 Attention 机制合理地融合依存信息，相比于单独使用原句子的模型 F 值提高了 1.36%，相比于单独使用 SDP 的模型性能提高了 21.82%，相比常用拼接方式性能提高了 2.66%。综上所述，本文提出的 Attention 方式可以更好的融合 SDP 信息，从而有效地提高关系抽取的精度。

### 3.3 与现有其他工作的对比

#### 3.3.1 采用 Attention 的模型对比

本节对于在 DDI 任务中采用不同 Attention 机制的模型性能进行了对比分析，如表 4 所示。

表 4　采用不同 Attention 机制的模型对比

| 模型 | P | R | F |
|---|---|---|---|
| 强调实体信息的 Attention [7] | 76.30% | 63.25% | 69.12% |
| 池化 Attention[8] | 73.41% | 69.66% | 71.48% |
| 多层 Attention [17] | 73.7% | 70.8% | 72.2% |
| 强调位置信息的 Attention [9] | 75.8% | 70.38% | 72.99% |
| **本文融合依存信息的 Attention 机制** | **73.38%** | **74.06%** | **73.72%** |

Miwa [7]等采用实体级 Attention 机制，分别计算两个实体向量与句子中的每个单词的相似度后求平均，强调实体信息在句子中的重要性，F 值为 69.12%。Yi[17]等使用多层 Attention 网络，使用词级 Attention 机制和句子级 Attention 机制，词级 Attention 机制是指对当前句子做 Attention 池化操作强调句中每个单词的重要性，句子级 Attention 机制对于具有相同实体对的句子集，强调句子间其相关性，F 值达到 72.2%。Zhou[9]等使用关注位置的 Attention 机制，使用位置向量（每个单词和实体之间的距离）计算 Attention 权值，强调句子中距离实体近的单词的重要性，F 值达到 72.99%。从以上结果分析可以看出，不同的 Attention 机制都在一定程度上提高了识别精度。本文使用融合依存信息的 Attention 机制，从句法结构的角度构建 Attention，突出依存路径上的单词对两个实体间的关系的影响，F 值达到 73.72%，高于其他 Attention 模型。从以上实验结果可以看出，本文提出的融合 SDP 信息的 Attention 机制能够有效提高模型性能。

### 3.3.2 模型总体性能对比

本文与目前 DDIExtraction2013 语料上取得较好结果的方法做了比较，如表 5 所示。

表 5　在 DDIExtraction2013 语料上与其他文献的比较

| | 模型 | P | R | F |
|---|---|---|---|---|
| 浅层机器学习 | Kim et al, 2015[3] | ------- | ------ | 67% |
| | Zheng et al, 2016 [18] | ------- | ------ | 68.4% |
| | Raihani et al, 2016[19] | 73.7% | 68.7% | 71.1% |
| 深度学习 | Zhao et al, 2016[20] | 72.5% | 65.1% | 68.6% |
| | Miwa et al, 2017[7] | **76.3%** | 63.25% | 69.12% |
| | Liu et al, 2016[4] | 75.7% | 64.7% | 69.75% |
| | Quan et al, 2016[5] | 76% | 65.3% | 70.2% |
| | Sahu et al, 2017[8] | 73.4% | 69.7% | 71.48% |
| | Yi et al, 2017[17] | 73.7% | 70.8% | 72.2% |
| | Zhang et al, 2018[13] | 74.1% | **71.8%** | 72.9% |
| | Zhou et al, 2018[9] | 75.8% | 70.38% | **72.99%** |
| | **本文模型** | **73.38%** | **74.06%** | **73.72%** |

从表 5 可以看出：

（1）深度学习模型在未使用任何人工特征的情况下，取得了较好的效果。浅层机器学习方法主要依赖于大量复杂的人工特征集来提高模型性能。例如，Kim[3]等结合词特征、n 元组特征、词对特征、依存解析图特征、解析树特征等大量的特征信息，使用基于线性核函数的 SVM 分类器，F 值达到 67.00%。Raihani[19]等在传统特征集的基础上添加触发词类型、模式信息等新的人工特征，使用两阶段策略分类，F 值达到 71.14%。然而，深度学习方法在不使用任何人工特征的情况下，能够在本任务中取得较好的效果。例如，Quan[5]等使用卷

积神经网络模型，F 值达到 70.2%。Sahu[8]等使用长短期记忆网络（LSTM）模型，F 值达到 71.48%。Yi[17]等使用 GRU 神经网络模型，F 值达到 72.2%。本文使用 GRU 神经网络模型，不仅避免了大量人工特征的使用，而且提升了系统性能，F 值达到 73.72%。

（2）本文融合依存信息的 Attention 机制取得了最好的效果。在表 5 先进的系统中，Zhang[13]等考虑了最短依存路径信息，通过构建多层 RNN 网络并使用拼接方法结合最短依存路径和原句子信息，实验得到 F 值为 72.9%。相比于表中其他没有考虑依存信息的模型，Zhang[13]等的模型 F 值明显较高，因此依存信息对于关系抽取有重要的作用。本文同样考虑依存信息，使用融合依存信息的 Attention 机制将最短依存路径信息与原句子相结合，实验得到 F 值为 73.72%，相比 Zhang[13]等的模型 F 值提高 0.82%。

（3）本文的模型总体 F 值最高，且准确率和召回率比较均衡。大多数现有模型准确率偏高，召回率偏低，导致模型 F 值较低。例如，Miwa[7]等模型的准确率最高为 76.3%，但其召回率最低为 63.25%，模型 F 值为 69.12%，比本文模型准确率高 2.92%，召回率低 10.81%，F 值低 4.6%。Zhang[13]等模型的准确率为 74.1%，召回率最高为 71.8%，F 值为 72.9%，比本文模型性能准确率高 0.72%，召回率低 2.26%，F 值低 0.82%。Zhou[9]等模型的 F 值最高为 72.99%，准确率为 75.8%，召回率为 70.38%，比本文模型准确率高 2.42%，召回率低 3.68%，F 值低 0.73%。综合对比以上分析可见，本文的模型 F 值最高，准确率、召回率和 F 值分别为 73.38%、74.06%和 73.72%，准确率和召回率比较均衡。

## 4 结论

生物医学关系抽取是自然语言处理任务中的重要课题，本文针对药物-药物关系抽取进行了研究。提出了一种融合依存信息 Attention 机制的深度学习模型，在 DDIExtraction2013 语料上进行了实验，F 值为 73.72%，该模型性能达到目前的先进水平。主要结论如下：

（1）浅层机器学习方法需要构建合适的人工特征集，进行大量的特征选择实验，导致系统成本较高，泛化能力较低。因而本文使用深层神经网络 GRU 模型，只使用原句子词序列和最短依存路径词序列，未使用任何人工特征，F 值达到 73.72%，性能高于浅层机器学习的模型。

（2）神经网络通过自动学习句子的高层特征表示进而避免了大量人工特征的使用，但无法获得句法结构等对于关系抽取十分重要的依存信息。因此，本文使用最短依存路径信息，相比于单独使用原句子信息的模型，性能提高了 1.36%。

（3）为了更合理地利用依存路径信息，本文摒弃了传统的拼接方式，提出了一种新的融合依存信息 Attention 机制，能够从句法结构的角度衡量原句子各部分对于实体间关系的重要性，该方法相比拼接方式性能提高了 2.66%。

综上所述，本文提出的模型有效地提高了药物关系抽取的识别性能。在未来的工作中，将尝试多层 Attention 机制，进一步提高模型的精度。

## 参考文献

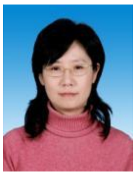

[1] Chowdhury M, Faisal M, Lavelli A, et al. FBK-irst: A multi-phase kernel based approach for Drug-Drug interaction detection and classification that exploits linguistic information[C]. In: 7th International Workshop on Semantic Evaluation, Atlanta, Georgia, USA, 2013:351–355.

[2] Segura Bedmar I, Martinez P, Herrero Zazo M. Semeval-2013 task 9: Extraction of Drug-Drug interactions from biomedical texts (DDIextraction 2013)[J]. Proceeding of Semeval, 2013 ,2(SemEval):341-350.

[3] Kim S, Liu H, Yeganova L, et al. Extracting Drug-Drug interactions from literature using a rich feature-based linear kernel approach[J]. Journal of Biomedical Informatics, 2015(55):23-30.

[4] Liu S, Tang B, Chen Q, et al. Drug-Drug interaction extraction via convolutional neural networks[J]. Comput Math Methods Med, 2016, 2016:6918381.

[5] Quan C, Lei H, Sun X, et al. Multichannel convolutional neural network for biological relation extraction[J]. BioMed Research International, 2016(2-1):1-10.
[6] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate[J]. Computer Science. 2014, arXiv:1409.0473.
[7] Asada M, Miwa M, Sasaki Y. Extracting Drug-Drug interactions with attention CNNs[C]. Proceedings of the BioNLP 2017 workshop, Vancouver, Canada, 4 August ,2017:9-18.
[8] Sahu SK, Anand A. Drug–Drug interaction extraction from biomedical text using long short term memory network[J]. 2017, arXiv:1701.08303.
[9] Zhou D, Miao L, He Y. Position-aware deep multi-task learning for Drug-Drug interaction extraction[J]. Artificial Intelligence in Medicine, 2018.
[10] Liu Y, Wei F, Li S, et al. A dependency-based neural network for relation classification[J]. 2015, arXiv: 1507.04646.
[11] Yan X, Mou L, Li G, et al. Classifying relations via long short term memory networks along shortest dependency path[J]. Computer Science, 2015(42):56-61.
[12] Miwa M, Bansal M. End-to-end relation extraction using lstms on sequences and tree structures[J]. 2016, arXiv: 1601.00770.
[13] Zhang Y, Zheng W, Lin H, et al. Drug-Drug interaction extraction via hierarchical RNNs on sequence and shortest dependency paths[J]. Bioinformatics, 2017(34):828-835.
[14] Kyunghyun Cho, Bart van Merrienboer, Caglar Gulcehre, et al. Learning phrase representations using RNN encoder–decoder for statistical machine translation[C]. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing(EMNLP), Doha, Qatar, 2014:1724–1734.
[15] Herrerozazo M, Segurabedmar I, Declerck T. The DDI corpus: an annotated corpus with pharmacological substances and Drug-Drug interactions[J]. Journal of Biomedical Informatics, 2013(46):914–920
[16] Segurabedmar I, Herrerozazo M. Lessons learnt from the DDIExtraction-2013 Shared Task[J]. Journal of Biomedical Informatics, 2014(51):152-164.
[17] Yi Z, Li S, Yu J, et al. Drug-Drug interaction extraction via recurrent neural network with multiple attention layers[C]. 2017, arXiv:1705.03261.
[18] Zheng W, Lin H, Zhao Z, et al. A graph kernel based on context vectors for extracting Drug-Drug interactions[J]. Journal of Biomedical Informatics, 2016(61):34-43.
[19] Raihani A, Laachfoubi N. Extracting Drug-Drug interactions from biomedical text using a feature-based kernel approach[J]. Journal of Theoretical and Applied Information Technology, 2016(92):109-120.
[20] Zhao Z, Yang Z, Luo L, et al. Drug drug interaction extraction from biomedical literature using syntax convolutional neural network[J]. Bioinformatics, 2016, 32(22):3444-3453.



李丽双（1967—），女，教授，博士生导师，主要研究领域为自然语言处理、信息抽取与文本挖掘。Email: lilishuang314@163.com。

钱爽（1994—）,硕士，主要研究方向为自然语言处理。Email: 18345969582@163.com