

交通事故的自动判案研究

尹何举¹ 咎红英¹ 陈俊怡¹ 翟新丽²

(1. 郑州大学 信息工程学院, 河南省 郑州 450000; 2. 郑州大学 法学院, 河南省 郑州 450000)

摘要: 该文针对法律领域民事案件中的“交通事故”类案件进行研究, 期望在该“交通事故”数据集上实现自动判案。该文从“中国裁判文书网”采集数据文本, 按照两种分类模式进行分类, 分别为粗粒度分类和细粒度分类, 粗粒度分类为 4 类, 细粒度分为 8 类, 并对文本进行人工标注。该文使用了三种模型, 分别为: 基于 SVM 的模型、基于 BI-GRU 的模型和基于 Attention+BI-GRU 的模型。三种模型的实验结果表明: 在本数据集上, 对数据进行粗粒度分类时, 基于 Attention+BI-GRU 的模型 F1 值为 80.26%, 基于 SVM 的模型为 77.24%, 基于 BI-GRU 的模型为 72.65%。在细粒度分类时, 基于 BI-GRU 的模型 F1 值为 48.59%, 基于 SVM 的模型的为 38.29%, 基于 Attention+BI-GRU 的模型为 40.87%。

关键词: 自动判案; 神经网络; 支持向量机

中图分类号: TP391

文献标识码: A

Study on Automatic Judgment of Traffic Accidents

YIN Heju¹, ZAN Hongying¹, CHEN Junyi¹, ZHAI Xinli²

(1. School of Information Engineering, Zhengzhou University, Zhengzhou, Henan 450000, China;

2. Law school, Zhengzhou University, Zhengzhou, Henan 450000, China)

Abstract: This article studies the cases of “traffic accidents” in civil cases of the legal field, and expects to achieve automatic judgment on the “traffic accident” data set. This paper collects data texts from the “China Judgment Document Network” and classifies them according to two classification modes, namely coarse-grained classification and fine-grained classification, coarse-grained classification into 4 categories, fine-grained classification into 8 categories, and manually label the texts. Three models are used in this paper: SVM-based model, BI-GRU-based model, and Attention+BI-GRU-based model. The experimental results of the three models show that: In this data set, when the data is coarse-grained, the F1 value of the model based on Attention+BI-GRU is 80.26%, the model based on SVM is 77.24%, and the model based on BI-GRU is 72.65%. In the fine-grained classification, the F1 value based on the BI-GRU model is 48.59%, the SVM-based model is 38.29%, and the Attention+BI-GRU based model is 40.87%.

Key words: Automatic judgment; Neural network; SVM

0 引言

自动判案是对给定一个案件的事实描述给出一个判案结果, 例如杀人、盗窃和交通肇事等。对每一个案件判案而言需要分析案件的事实描述文本, 找出事实描述文本中相同的特征。主流的处理过程为查找相似案件或者律师根据

自己的专业知识对文本中的关键词等分析, 因此可以将让机器对案件的自动判案看作是对法律文本的研究。现阶段关于自动判案的研究也主要被当作自然语言处理中的文本分类任务, 对事实文本进行分析找到文本的特征, 根据提取到的特征找寻特征和判案结果之间的映射关系。

文本分类是自然语言处理中的一项基础研究。近年来, 深度学习技术逐渐取代传统的基

作者简介: 尹何举 (1996—), 男, 本科, 主要研究领域为自然语言处理; 咎红英 (1966—), 女, 教授, 主要研究领域为自然语言处理; 陈俊怡 (1991—), 女, 博士, 主要研究领域为自然语言处理。

于机器学习的方法,成为文本分类领域的主流。深度学习在自然语言处理领域的突破以 Mikolov、Bengio 等人^[1-2]为代表提出的神经语言模型,如 CBOW 和 Skip-gram 等。通过神经语言模型如 Word2Vec 提供词向量让神经网络在文本分类有了更好的表现。

传统机器学习主要使用支持向量机 (Support Vector Machine, SVM)、近邻算法 (k-NearestNeighbor, kNN)、朴素贝叶斯方法,其中 SVM 算法由于核函数的使用在文本分类取得较好的应用。

深度学习中循环神经网络的出现对文本分类问题有较好的帮助,通过循环神经网络去学习词和词之间的上下文关系,词和词之间不再独立。同时近些年来借鉴人类视觉注意力机制 (Attention Mechanism) 将注意力机制引入神经网络中,在文本分类时可以通过学习上下文信息的权重来生成文本表示。这样可以进一步加深特征,提升神经网络学习中的准确率。

本文通过使用传统机器学习 SVM 算法、循环神经网络和基于注意力机制的循环神经网络的模型来进行自动判案的研究,并比较分析三种模型的实验结果。

1. 相关工作

近几年来,随着词向量的引用,文本可以很好的表示成词向量序列,这就为各种神经网络架构提供了良好的基础。例如卷积神经网络 (Convolutional Neural Network, CNN),循环神经网络 (Recurrent Neural Network, RNN) 等。词向量训练工具多种多样,例如 Word2Vec 和等,通过这些工具可以在数据量小时实现较好的结果。

Kim^[3]提出了多尺度并行的单层卷积神经网络结合预训练的词向量在多个句子分类任务上的表现优于传统的机器学习算法以及早期的神经网络方法。得益于 RNN 的出现,分文分类又有了进一步的提升,由于 RNN 更适合学习长文本序列,但是 RNN 在训练时会出现“长期依赖问题”,例如预测“我住在中国,……我说汉语。”中的“汉语”,由于句子程度过长,

无法学习到关键信息。其变体长短时记忆网络^[4] (Long Short-Term Memory Network, LSTM) 引入门机制,通过使用输入门、遗忘门和输出门来控制记忆的长短。随后 LSTM 的变种门控循环网络^[5] (Gated Recurrent Unit, BI-GRU) 出现,GRU 和 LSTM 相似,但 GRU 将遗忘门和输入门合成了一个单一的更新门,同时混合了细胞状态和隐藏状态。

近两年来,注意力机制作为深度学习中的重要研究方向,首先在神经机器翻译领域取得了突破,随后拓展到其他领域,例如语音识别、文本分类等。Zhou^[6]等人提出了基于注意力机制的双向长短时记忆网络 (bidirectional Long Short-Term Memory Network, BI-LSTM) 并且用于文本分类, Yang^[7]等人又提出分层注意力机制的神经网络模型 (Hierarchical Attention Networks, HAN), 并应用在文本分类。

法律判案研究在 NLP 中常被当作一个对文本多分类的问题处理。其中台湾的 Chao-Lin Liu^[8]等人在 2006 提出用 kNN 对 12 种和 6 种犯罪结果进行预测,在 2015 台湾的 Yi-Hung Liu^[9]又提出使用 SVM 来进行结果的分类提高了判案的准确率。另外英国伦敦大学的 Nikolaos Aletras^[10]在 2016 年提出对欧洲人权法庭的预测,考虑人权等其他伦理性因素,使用 SVM 构建模型对判案结果进行预测,准确率达到 79%。美国芝加哥肯特伊利诺伊理工学院法学院的 Daniel Martin Katz^[11]等人在 2016 年使用随机森林构建模型对美国最高法院行为的预测,准确率达到 70%。同时还有日本的 Mi-Young Kim^[12]等人用 SVM 的方法来计算文本的相似度从而对判案结果进行胜负预测。2017 年中国北京大学的 Bing-Feng Luo^[13]等人使用基于神经网络的模型对刑事案件进行预测,准确率达到 98%。

本文使用基于循环神经网络和机器学习的模型,并对其中一个循环神经网络使用注意力机制。模型需要较大数据量,本文对河南省 2014-2018 年部分“交通事故”数据集进行实验,对比基于 SVM 的模型、基于双向门控网络 (Bidirectional Gated Recurrent Unit, BI-GRU) 的模型和基于 Attention+BI-GRU 的模型结果。

2. 交通事故自动判案模型

2.1 基于 SVM 的模型

基于 SVM 的“交通事故”自动判案模型如图 1 所示：将文本分词，将文本通过词向量表映射成词向量组成的序列，将词向量组合成文本特征进入分类器进行训练。

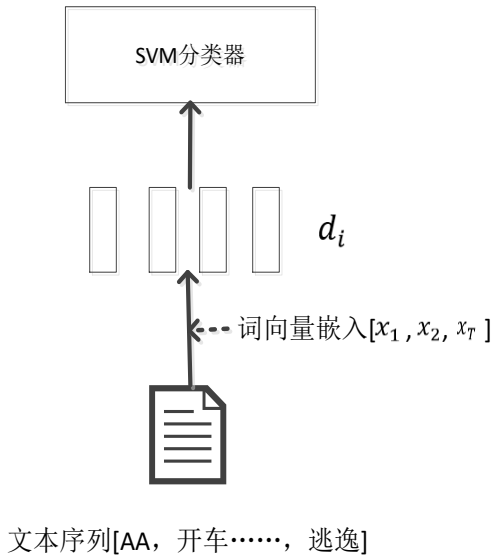


图 1 基于 SVM 的模型

如图 1 所示，文本可以视为是词的表示序列，可以将原始文本序列表示成式 1 所示，通过词向量嵌入将序列文本 S 转换成如图 1 向量序列 $[x_1, x_2, \dots, x_T]$ ，其中 x_t 代表的是将单词 $word_t$ 通过词汇表 V 查到对应词向量。

$$S = [word_1, word_2, \dots, word_T] \quad (1)$$

将词向量序列 $[x_1, x_2, \dots, x_T]$ 转换成图 1 所示的文本向量 d_i ，表示文本的特征。计算方法如式 2 所示：

$$d_i = \frac{\sum_{t=1}^T x_t}{T} \quad (2)$$

其中 T 代表的序列中元素的个数， x_t 代表在 t 时刻的向量。将文本向量 d_i 和标注答案输入到 SVM 中进行有监督训练，SVM 内核函数使用径向基函数。

2.2 基于 BI-GRU 的模型

基于 BI-GRU 的模型结构如图 2 所示，模型主要包括以下几部分：（1）首先将本文需要的事实文本通过词向量嵌入转换成词语向量集合 $[x_1, x_2, \dots, x_T]$ 。（2）将向量集合 $[x_1, x_2, \dots, x_T]$ 传入到事实文本编码器去编码生成文本特征向量 d_f 。（3）随后将文本向量 d_f 通过 Softmax 分类器去为每一个输入事实计算判案结果的概率。下面将会对具体结构进行介绍。

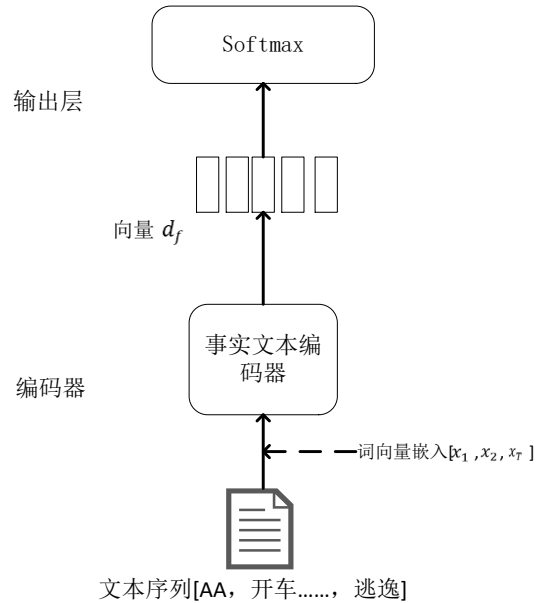


图 2 基于 BI-GRU 的模型

2.2.1 事实文本编码器

Yang^[7]等人提出分层注意机制的神经网络模型，并应用在文本分类。直观上来看，一个句子由多个单词的序列组成，一篇文档由多个句子的序列组成。文档编码的问题的可以转换成两个序列编码的问题，即序列 Encoder 的问题。如图 3 所示，首先在句子序列编码器中对单个句子编码，生成向量 S ，随后在文档序列编码器中，将句子向量 S 作为输入，生成文档向量的向量编码 d 。

¹<http://wenshu.court.gov.cn/>

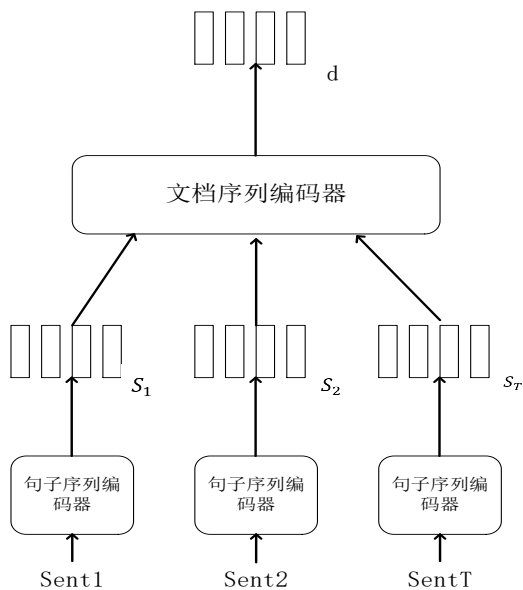


图 3 文档编码器

Sent 代表句子序列，经过句子序列编码器后生成句子向量 S ，将所有句子向量 S 经过文档序列编码器生成文档向量 d 。其中文档序列编码器和句子序列编码器可以使用不同的序列编码器，本文在这里为了方便使用相同的序列编码器。2.2.2 节对序列编码器进行介绍。

2.2.2 BI-GRU 序列编码器

本文使用双向的门控循环网络对序列文本进行编码。如图 4 所示，序列编码器中包括一个前置网络和后置网络，其中每个网络都使用 GRU 实现。对输入序列 $[x_1, x_2, \dots, x_T]$ 在每个时刻 t 产生前置网络中间状态 h_{ft} 和一个后置网络中间状态 h_{bt} 。在每个时刻产生中间状态 h_t 为式 3 所示，对中间状态 h_{bt} 和 h_{ft} 可以使用拼接或者求平均，本文对 h_{bt} 和 h_{ft} 进行拼接：

$$h_t = [h_{ft}, h_{bt}] \quad (3)$$

最后的状态向量 g 可以将中间状态 h_t 进行连接或者求平均，本文将中间状态 h_t 求平均，表示如式 4 所示：

$$g = \frac{\sum_t^T h_t}{T} \quad (4)$$

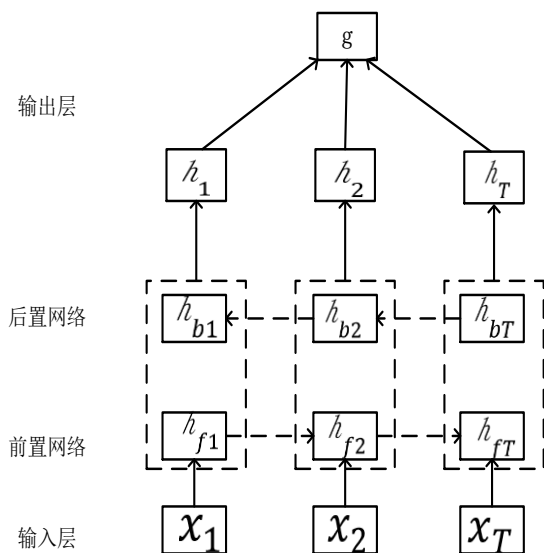


图 4 BI-GRU 序列编码器

2.2.3 输出层

为了更好的得到输出，首先将事实文本向量 d 经过一个全连接层去产生一个新的向量 d' ，全连接层计算如式 5：

$$d' = \sigma(W_f \cdot [d_f] + b_f) \quad (5)$$

其中 $d' \in R^k$ ，答案类别数为 k 。随后将向量 d' 传入一个 Softmax 分类器中去产生可能预测的结果概率向量 p_f ， $p_f \in R^k$ 。其中 p_f 中每个维度数值对应分类标准中每种类别的可能概率值，概率值最高的类别为预测结果。由于本文将自动判案视为单分类问题，即一个文本事实对应一个判案结果，因此不需要设置最后可能概率的阈值。

本文在使用 BI-GRU 模型进行训练时使用交叉熵损失函数，公式如式 6：

$$Loss = -\sum_{t=1}^T y_t \log(p_t) \quad (6)$$

其中 T 是判决的结果的分类数目， y_t 和 p_t 分别是真实的判案结果和预测判案结果概率。其中真实判案结果中如果在 T 时刻为真的判案结果，则标签为 1，否则标签为 0。

2.3 基于 Attention+BI-GRU 的模型

基于 Attention+BI-GRU 的模型和基于 BI-GRU 的模型基本相同，最大的改进是在基于 BI-GRU 的模型中序列编码器直接使用了 BI-GRU，如图 5 所示，其中最终输出 g 如式 7

²<http://ictclas.nlp.ir.org/>

³<https://code.google.com/p/word2vec/>

所示，对中间状态取平均，即所有的信息都统一处理。但在这里使用 attention 机制，对中间状态动态产生权重 α ，即可以动态的学习序列中重要的部分，如图 5 所示：

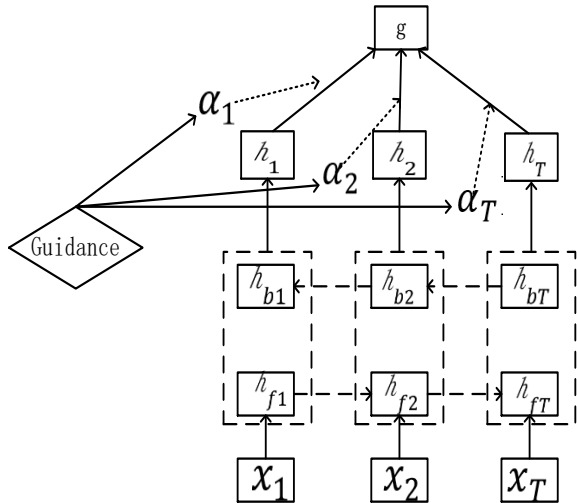


图 5 Attention+BI-GRU 序列编码器

如图 5 中所示，此处所用模型在传统的 BI-GRU 网络上增加了权重机制。对于给定的任意序列 $[h_1, h_2, \dots, h_T]$ ，模型中的 Attention 机制计算出每个元素的权重 $[\alpha_1, \alpha_2, \dots, \alpha_T]$ 。最后的计算公式如式 7：

$$g = \sum_{t=1}^T \alpha_t h_t; \quad \alpha_t = \frac{\exp(\tanh(W h_t)^T u)}{\sum_t \exp(\tanh(W h_t)^T u)} \quad (7)$$

其中 W 是一个权重矩阵，将输入的中间状态 h_t 线性转换， u 是一个内部的向量从没有信息的部分中找出有关信息的部分。通过用这种方法，可以有效地有效地区分出句子和文档中有效的信息。

3 实验设置

3.1 数据集

本文数据集来源为中国裁判文书网¹的公开数据，针对河南省 2014-2018 年的“交通事故”文书，最终获取法律文本数 15000 条。对数据集使用正则表达式抽取事实，规则如“案件事实.(.*?)本院认为”，同时去除数据长度过短的文书，对人名进行匿名化减少噪音。最终获取有效数据集 14000 条。

清洗过的数据集总数 14000 条，选择其中 12000 条作为训练集，2000 条作为测试集，各类别数目分布如表 2 所示。

通过法律判案事实总结规律，对判案结果进行人工分类，分类结果为两类，如表 1 所示，粗分类分为四类，细分类分为 8 类。标准划分后对有效数据集进行人工标注判案结果，分布如表 2 所示。

表 1 分类标准

粗粒度分类	细粒度分类
● 驳 回	● 驳 回
● 危险驾驶罪	● 危险驾驶罪
● 交通肇事罪	● 交通肇事罪
● 罚款不判罪	● 罚款：0-2 万
-	● 罚款：2-5 万
-	● 罚款：5-7 万
-	● 罚款：7-10 万
-	● 罚款：10- 万

表 2 数据集分步统计

样本集		训练集 (条)	测试集 (条)
驳回		324	64
危险驾驶罪		1,879	302
交通肇事罪		1,535	303
罚款不判罪	罚款: 0-2 万	2,039	359
	罚款: 2-5 万	1,539	249
	罚款: 5-7 万	817	141
	罚款: 7-19 万	1,097	162
	罚款: 10- 万	2,658	420
总计		12,000	2,000

3.2 词嵌入设置

本文对数据集的分词使用中科院的分词系统 ICTCLAS², NLPPIR 汉语分词系统, 主要功能包括中文分词; 词性标注; 命名实体识别; 用户词典功能。本文使用 ICTCLAS 进行无词性分词。

在数据集规模较少时使用词向量嵌入可以明显提高网络准确率。本文使用 Word2Vec³ 工具对实验语料进行训练。训练词向量维度为 100 维, 生成有效词向量个数为 63812。对于训练后的词向量表中不存在的词, 随机生成[-1, 1]上的 100 维向量来初始化。

3.3 超参数设置及训练细节

本文模型中使用统一的数据输入格式, 输入文本向量。经过对数据结果进行分析后设置一篇文档中的最大句子个数为 15, 单个句子的最大长度为 40 个单词, 一个单词的维度为 100, 其中达不到句子最大长度时用 100 维的“0”填充。

本文基于 SVM 的模型使用 Sklearn 框架完成, 设置迭代次数为 150, 内核使用径向基函数。

基于 BI-GRU 的模型和基于 Attention+BI-GRU 的模型使用 TensorFlow 框架完成。基于 BI-GRU 的模型批处理大小设置为 100, 基于 Attention+BI-GRU 的模型的批处理大小设置为 7。每个 GRU 单元的隐藏层数目设置为 75, 基于 BI-GRU 的模型和基于 Attention+BI-GRU 的模型都使用随机梯度下降

法作为优化器进行训练, 其中学习速率设置为 0.01。

3.4 评价指标

文本分类的评价标准类似于信息检索的评价标准, 包括准确率、召回率和 F1 值。准确率是指分类正确的文本数和实际分类的文本数的比率, 如式 8:

$$\text{准确率 (Precision)} = \frac{\text{分类正确的文本数}}{\text{实际分类的文本数}} \quad (8)$$

召回率代表正确文本数和分类应有的所用文本的比率, 如式 9:

$$\text{召回率 (Recall)} = \frac{\text{分类正确的文本数}}{\text{应有的所有文本数}} \quad (9)$$

F1 为了综合考虑准确率和召回率不同方面的影响, 表示如式 10:

$$F1 = \frac{\text{准确率} \times \text{召回率} \times 2}{\text{准确率} + \text{召回率}} \quad (10)$$

本文对所有类别的 F1 值进行综合考虑, 计算 $F1_{macro}$ 值, 如式 11 所示, 其中 N 代表结果类别, $f1_i$ 代表类别 i 多对应的 F1 值:

$$F1_{macro} = \frac{\sum_i^N f1_i}{N} \quad (11)$$

4 结果及分析

4.1 粗粒度分类结果

本文首先按照表 1 中粗粒度分类标准对三种模型进行实验, 使用准确率(P)、召回率(R)、F1 值作为评价指标, 实验结果如表 3 所示:

表 3 粗粒度分类结果

模型	分类类别	1. 驳回	2. 危险驾驶罪	3. 交通肇事罪	4. 罚款不判罪

SVM 模型	P (%)	44.18	56.79	87.41	96.33
	R (%)	30.16	91.39	87.13	97.30
	F1 (%)	35.85	89.03	87.27	96.81
	$F1_{macro}$ (%)	77.24			
BI-GRU 模型	P (%)	66.67	88.13	98.02	94.52
	R (%)	6.30	98.34	81.52	99.70
	F1 (%)	11.59	92.96	89.01	97.04
	$F1_{macro}$ (%)	72.65			
Attention+BI-GRU 模型	P (%)	84.21	90.21	64.57	95.08
	R (%)	25.40	97.69	86.74	99.40
	F1 (%)	39.02	93.80	90.48	97.71
	$F1_{macro}$ (%)	80.26			

由表 3 得出, 基于 Attention+BI-GRU 的模型的四个类别 F1 值均高于 SVM、Bi-GRU 两个模型。其他两个模型中基于 SVM 的模型的 $F1_{macro}$ 值高于基于 BI-GRU 的模型的 $F1_{macro}$, 但是从分类的每类结果看基于 SVM 的模型只有“驳回”的 F1 值高于基于 BI-GRU 的模型。

进行纵向的比较发现三组实验中“驳回”的实验结果都相对其他类别较低, 在驳回这一类中 SVM 的 F1 值是 35.85%, BI-GRU 的 F1 值是 11.59%, 基于 Attention+BI-GRU 的模型是 39.02%。主要原因有如下几个方面: (1) 通过观察表 2 所示, 其中“驳回”类别在训练集上数目为 324 条, 数目相对较少。(2) 通过观察数据集发现是在训练和测试数据中“驳回”文书特征差距较大, 即并不严格和“交通案件”相关, 例如“交通案件二审”、“离婚案件”等。

4.2 细粒度分类结果

表 4 所示为细粒度分类实验结果, 对比粗粒度分裂实验结果表 3, 其中明显可以看出基于 SVM 的模型、基于 BI-GRU 的模型、基于 Attention+BI-GRU 的模型的 $F1_{macro}$ 均有大幅度下降。其原因是在罚款的分类处理过程中文本的特征不明显。例如“……被告人 A 在原告骨折, 住院治疗等花费 6000……”和“……被告人 B 造成原告脑震荡, 且进行药物治疗等共计 9000……”文本内容相似, 但是前例判案结果为 2 万以下, 后例的判案结果为罚款 2-5 万。同时一篇文档中平均 500 词, 关键信息在文本

序列中不固定, 造成循环神经网络即使通过上下文学习也不能准确找到。

在表 4 中, 基于 BI-GRU 的模型和基于 Attention+BI-GRU 的模型在罚款 5-7 万、7-10 万两个类别的召回率和精确率都为 0, 基于 SVM 的模型召回率和精确率不为 0, 但是分类效果较差。原因主要有以下几个: (1) 观察表 1 数据后发现其罚款在罚款 5-7 万、7-10 万的文本数较少。(2) 在罚款 5-7 万、7-10 万之间的文本和其他罚款不判刑类别罚款文本之间特征区分小。(3) 神经网络在学习过程中会遗忘信息, SVM 则由于是将文本特征映射到高维空间丢失特征较少, 因此 SVM 准确率和召回率不会为 0。

从表 4 中可以看出, 虽然其在罚款 5-7 万、7-10 万之间两类的 F1 值为 0, 但是基于 BI-GRU 的模型和基于 Attention+BI-GRU 的模型的 $F1_{macro}$ 都相比传统 SVM 的 $F1_{macro}$ 高。其中基于 Attention+BI-GRU 的模型的 $F1_{macro}$ 低于基于 BI-GRU 的模型。分析原因如下: (1) 在细粒度划分时, 部分类别文本之间相似, 特征不明显。带 Attention 的 BI-GRU 模型会放大某些特征, 去掉某些在注意力机制下不重要的特征。但是对于相似文本特征的细微之处就会影响结果, 导致在对罚款这种相似文本的分类时反而结果不显著。(2) 本文数据量不大, 对于注意力机制来说可能没有更好的发挥作用。

表 4 细粒度分类结果

分类类别 模型		驳回	危险 驾驶 罪	交通 肇事 罪	罚款 0-2 万	罚款 2-5 万	罚款 5-7 万	罚款 7-10 万	罚款 10- 万
		SVM 模型	P (%)	48.00	88.49	85.71	36.53	20.00	14.10
R (%)	18.75		89.07	89.10	97.97	2.81	39.00	13.58	10.24
F1 (%)	26.97		87.78	87.28	47.52	4.92	20.72	13.25	13.76
$F1_{macro}$ (%)	38.29								
BI-GRU 模型	P (%)	71.88	93.43	97.49	51.25	31.30	0.00	0.00	62.10
	R (%)	35.95	99.01	89.77	79.67	14.46	0.00	0.00	77.62
	F1 (%)	47.93	96.14	93.47	62.37	19.78	0.00	0.00	69.00
	$F1_{macro}$ (%)	48.59							
Attention+BI-GRU 模型	P (%)	66.67	94.08	89.97	45.91	18.75	0.00	0.00	69.00
	R (%)	18.75	94.70	94.72	62.12	2.41	0.00	0.00	80.96
	F1 (%)	29.27	94.39	92.28	58.79	4.27	0.00	0.00	53.96
	$F1_{macro}$ (%)	40.87							

表 5 验证实验结果

分类类别 模型		罚款：0-2 万	罚款：2-10 万	罚款：10- 万
		SVM 模型	P (%)	22.40
R (%)	65.07		33.48	4.55
F1 (%)	33.33		39.52	8.34
$F1_{macro}$ (%)	27.06			
BI-GRU 模型	P (%)	54.13	57.97	62.92
	R (%)	77.29	56.79	45.90
	F1 (%)	63.67	57.37	53.07
	$F1_{macro}$ (%)	58.04		
Attention+BI-GRU 模 型	P (%)	39.32	45.62	50.00
	R (%)	20.08	83.71	10.94
	F1 (%)	26.58	59.06	17.95
	$F1_{macro}$ (%)	34.53		

4.3 相似文书对注意力机制的影响

为了验证在本数据集上带 Attention 的 BI-GRU 模型适合特征明显的文本，本文又做了补充实验以验证该结论，如表 5 所示。

从表 5 可以看出基于 BI-GRU 的模型在对金额分成三类的分类实验中各项指标均处于领先地位。在对罚款这种相似案件分类时，从 $F1_{macro}$ 看基于 SVM 的模型的结果最差，基于 Attention+BI-GRU 的模型相对基于 BI-GRU 的模型 $F1_{macro}$ 相差 23.51%，在 4.1 节的粗粒度实验中，文本特征明显，基于 Attention+BI-GRU 的模型实验结果最好。得出结论在本数据集上基于 Attention+BI-GRU 的模型对特征不明显的文本进行学习时会抛弃掉某些特征，放大某

些特征导致分类不明显，反而没有经过 Attention 的 BI-GRU 模型没有丢失掉相似文本中的细微特征，分类结果更好。

5 结论

本文主要实现从中国裁判文书网上获取公开文书数据，抽取关键信息，并对文书数据按照两种分类标准进行人工标注。通过三种模型：基于 SVM 的模型、基于 BI-GRU 的模型和基于 Attention+BI-GRU 的模型在一定程度上实现“交通事故”自动判案。结果发现在粗粒度分类结果上基于 Attention+BI-GRU 的模型最好，细粒度分类上基于 BI-GRU 的模型实验结

果最好。实验总体基于 BI-GRU 的模型和基于 Attention+BI-GRU 的模型好用基于 SVM 的模型,同时发现基于 Attention+BI-GRU 的模型相对基于 BI-GRU 更适合于数据集中特征明显的文本。

本文的目的是实现“交通事故”的自动判案,未来工作中,进一步研究注意力机制怎么去更有效的注意到关键信息,进而提高自动判案的准确率。数据集要进一步扩大,对更多的数据进行分析。同时要进一步研究法律事实描述文本中的特殊性,总结法律文本的规则特征添加到深度学习中。

参考文献

- [1] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality[C]// Advances in neural information processing systems. 2013: 3111-3119.
- [2] Bengio Y, Ducharme R, Vincent P, et al. A neural probabilistic language model[J]. Journal of machine learning research, 2003, 3(Feb): 1137-1155.
- [3] Kim Y. Convolutional neural networks for sentence classification[J]. arXiv preprint arXiv:1408.5882, 2014.
- [4] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural computation, 1997, 9(8): 1735-1780.
- [5] Cho K, Merriënboer B V, Gulcehre C, et al. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation[J]. Computer Science, 2014.
- [6] Zhou P, Shi W, Tian J, et al. Attention-based bidirectional long short-term memory networks for relation classification[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). 2016, 2: 207-212.
- [7] Yang Z, Yang D, Dyer C, et al. Hierarchical attention networks for document classification[C]//Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2016:1480-1489.
- [8] Liu C L, Hsieh C D. Exploring phrase-based classification of judicial documents for criminal charges in chinese[C]//International Symposium on Methodologies for Intelligent Systems. Springer, Berlin, Heidelberg, 2006: 681-690.
- [9] Liu Y H, Chen Y L, Ho W L. Predicting associated statutes for legal problems[J]. Information Processing & Management, 2015 51(1): 194-211.
- [10] Aletras N, Tsarapatsanis D, Preotiuc-Pietro D, et al. Predicting judicial decisions of the European Court of Human Rights: A natural language processing perspective[J]. PeerJ Computer Science, 2016, 2: e93.
- [11] Katz D M, Bommarito II M J, Blackman J. A general approach for predicting the behavior of the Supreme Court of the United States[J]. PloS one, 2017, 12(4): e0174698.
- [12] Kim M Y, Xu Y, Goebel R. Legal question answering using ranking svm and syntactic/semantic similarity[C]//JSAI International Symposium on Artificial Intelligence. Springer, Berlin, Heidelberg, 2014: 244-258.
- [13] Luo B, Feng Y, Xu J, et al. Learning to Predict Charges for Criminal Cases with Legal Basis[J]. 2017.