

# 基于注意力机制的 BiGRU 判决结果倾向性分析<sup>①</sup>



王 宁<sup>2</sup>, 李世林<sup>1,2</sup>, 刘堂亮<sup>3</sup>, 赵 伟<sup>4</sup>

<sup>1</sup>(中国科学院大学 北京 100049)

<sup>2</sup>(中国科学院 沈阳计算技术研究所, 沈阳 110168)

<sup>3</sup>(辽宁省人民检察院沈阳铁路运输分院, 沈阳 110001)

<sup>4</sup>(沈阳东软系统集成工程有限公司, 沈阳 110000)

通讯作者: 李世林, E-mail: [lishilin16@mails.ucas.ac.cn](mailto:lishilin16@mails.ucas.ac.cn)

**摘 要:** 对裁判文书中判决结果的倾向性分析是完成律师推荐系统的前提, 如何高效的实现判决结果倾向性分析是本文的重点. 本文提出了基于注意力机制和 BiGRU 的判决结果倾向性分析模型. 首先, 训练词向量, 得到词向量表; 然后, 通过查找词向量表, 将文书数据转化为词向量序列, 将词向量序列作为输入来训练判决结果倾向性分析模型. 实验结果表明: 注意力机制和 BiGRU 算法在判决结果倾向性分析中具有一定的有效性. 该模型能够对裁判文书中判决结果的倾向性做一个合理的判断, 为后期律师推荐系统的实现提供一个合理的评分依据.

**关键词:** 倾向性分析; BiGRU; 注意力机制; 词向量模型; 律师推荐

引用格式: 王宁, 李世林, 刘堂亮, 赵伟. 基于注意力机制的 BiGRU 判决结果倾向性分析. 计算机系统应用, 2019, 28(3): 191-195. <http://www.c-s-a.org.cn/1003-3254/6816.html>

## Attention-Based BiGRU on Tendency Analysis of Judgment Results

WANG Ning<sup>2</sup>, LI Shi-Lin<sup>1,2</sup>, LIU Tang-Liang<sup>3</sup>, ZHAO Wei<sup>4</sup>

<sup>1</sup>(University of Chinese Academy of Sciences, Beijing 100049, China)

<sup>2</sup>(Shenyang Institute of Computing Technology, Chinese Academy of Sciences, Shenyang 110168, China)

<sup>3</sup>(Shenyang Railway Transportation Branch of Liaoning People's Procuratorate, Shenyang 110001, China)

<sup>4</sup>(Shenyang Neusoft System Integration Engineering Co. Ltd., Shenyang 110000, China)

**Abstract:** The tendency analysis of the judgment results in the judgment documents is the premise of completing the lawyer recommendation system. How to effectively realize the tendency analysis of judgment results has been focused on. This study proposes a biased analysis model based on attention mechanism and BiGRU. First, training the word embedding to obtain a word embedding table. Then, by looking up the word embedding table, the document data is transformed into a word embedding sequence, and the word embedding sequence is used as input to train the judgment result orientation analysis model. The experimental results show that the attention mechanism and BiGRU algorithm have certain effectiveness in the analysis of decision results. The model can make a reasonable judgment on the propensity of the judgment results in the judgment documents, and provide a reasonable scoring basis for the realization of the later lawyer recommendation system.

**Key words:** tendency analysis; BiGRU; attention mechanism; word embedding model; lawyers' recommendation

① 基金项目: 水专项课题 (2018ZX07601001); 环境质量预警项目 (2017CXCY-C-06)

Foundation item: Major Project of National Water Pollution Control and Treatment Science and Technology (2018ZX07601001); Environmental Quality Early Warning Project (2017CXCY-C-06)

收稿时间: 2018-09-07; 修改时间: 2018-10-08; 采用时间: 2018-10-19; csa 在线出版时间: 2019-02-22

随着深度学习在机器学习研究领域的表现日益突出,深度学习已经被广泛应用于自然语言处理、语音识别以及图像处理等领域<sup>[1]</sup>.由于深度学习的结构灵活,可以抽象局部特征且拥有记忆功能,这些特点使深度学习在众多算法中脱颖而出.在自然语言处理的领域中,深度学习同样已经表现出其强大的优势.

判决结果倾向性分析的研究可以表述为文本情感分析在裁判文书中的具体应用,文本情感分析又是自然语言处理方法的一种应用.由于深度学习中的 LSTM、GRU 等深度循环神经网络在处理序列时能有效地缓解句子中长距离依赖的问题,有着较好的计算效果,所以深度学习在文本情感分析研究领域中日益流行.

近年来,注意力机制被广泛用于自然语言处理的任务中.最早应用于自然语言处理中的是 Bahdanau 等人<sup>[2]</sup>使用类似注意力机制的算法完成机器翻译任务.之后,类似的基于注意力机制的 RNN 模型扩展并开始应用到各种自然语言处理任务中.

通过研究近几年的文本情感分析现状,本文选取了几种主流的方法进行分析.王业沛等人<sup>[3]</sup>在处理文本情感分析的时候使用了基于三层 LSTM 的神经网络模型,获得了较高的准确率.本文在通过本地数据集实验时,发现由于单纯的使用 LSTM 网络很难将文本中的重要信息凸显出来,这导致实验结果难以达到预期效果.周瑛等人<sup>[4]</sup>将注意力机制加入 LSTM 神经网络实现对微博情感的分析,并取得较好结果.但是单向的 LSTM 很难将下文信息利用起来,在一定程度上降低了使用效果.白静等人<sup>[5]</sup>在处理中文微博立场时使用了基于注意力的 BiLSTM-CNN 模型,由于该模型需要分别通过 BiLSTM, CNN 计算获得中间文本表示和局部卷积特征,最后再由注意力机制融合,增加了计算复杂度.

综合以上的分析,本文提出了一种基于注意力机制的 BiGRU 判决结果倾向性分析模型.由于 GRU 是 LSTM 的一种变体,且在准确率相近的基础上,GRU 具有更高效的计算速度<sup>[6]</sup>.在 BiGRU 层中每一个训练序列向前和向后分别是两个循环神经网络,两个网络都连接着同一个输出,这个结构可以为输出层提供输入序列中每一个点完整的过去和未来的上下文信息.由于注意力机制能选择性地筛选出重要信息并聚焦到这些重要信息上,有效的减少关键信息的丢失.所以注意力机制和 BiGRU 可以有效提升本文中的模型准确性.

## 1 文本预处理

本文以从中国裁判文书网下载的 2000 份裁判文书作为实验数据.通常裁判文书可以分为一审、二审和再审,三种文书中的法律称谓有所区别,比如,二审文书中的法律称谓为上诉人和被上诉人等,会标注出双方在一审中的法律称谓.本文忽略当事人在一审中的法律称谓,未对二审裁判文书与一审裁判文书做关联.由于本文是对判决结果的倾向性分析,将一审、二审的裁判文书做关联是没有必要的工作.因为,本文所做的关于判决结果的倾向性分析是判断判决结果倾向于当事人中的哪一方,而以一审的中当事人称谓计算,根据相应的当事人称谓做出倾向性分析,最终也是达到同样结果,并没有对分类精确度有影响.所以未对二审裁判文书与一审裁判文书做关联.

针对需要分析的文本数据内容,本文先对裁判文书的半结构化文体做分析.通常裁判文书的结构如图 1 所示.

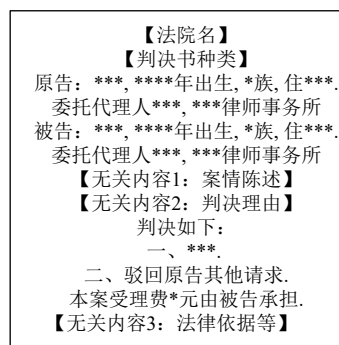


图 1 裁判文书结构图

### 1.1 数据抽取

由于【无关内容 1】、【无关内容 2】和【无关内容 3】在本文中无倾向性分析的任务无关,因此,本文只提取原被告的相关信息以及最终的判决结果.

由于裁判文书具有半结构化特征,故可以通过关键字抽取文本中与判决结果倾向性分析相关的信息.例如:“原告:”、“被告:”、“判决如下:”等关键字.抽出原告、被告信息为下一步处理文本数据做法律称谓归一化使用.

### 1.2 法律称谓标准化

在大部分判决结果中会以原告、被告作为当事人的称呼,但有些情况下,判决结果中会存在非标准法律称谓,比如:“二审案件受理费 2513 元,由张\*负担”,未

明确当事人法律称谓的,难以分析判决结果对于原被告的倾向性,因此需要对文本中的非标准称谓进行标准化.具体是通过将上一步提取出的原被告字符串与抽取出的判决结果进行最长公共子序列匹配,然后替换成标准法律称谓.完成法律称谓标准化.

### 1.3 数据集构建

上步完成之后,对判决结果进行按条拆分,以便细化判决结果的倾向性,然后按条做标注:0表示支持被告;1表示支持原告;对于没有明显胜负倾向的单条判决结果,本文以支持原告来标记.表1是数据集标注样例.

表1 数据集标注样例展示

单条判决结果	标注
二审案件受理费1300元由,由上诉人谢某负担	0
撤销**县人民法院(2012)*民一初字第*号民事判决	1
维持本院(2010)*中法民一终字第*号民事判决	0
本案受理费200元,由原、被告各半负担	1
本判决生效之日起十日内,被告*有限公司给付原告*因迟延履行2010年4月1日至2013年12月31日房屋租金136665元的违约金13666.5元(136665×10%)	1
准许周某某撤回起诉	1
驳回上诉,维持原裁定	0

### 1.4 训练词向量

在训练词向量时,本文使用从中国裁判文书网下载的裁判文书作为语料,对裁判文书去除停用词后做分词,然后使用 Word2Vec 训练得到词向量表. Word2Vec 是一种将词表示为实数值向量的算法模型,使用 Distributed representation 的词向量表示方式.通过词向量表可以将需要处理的文本内容转换成低维向量空间中的向量.文本语义上的相似可以用向量空间上的相似度来表示.

本文使用了 Word2Vec 中 skip-gram 算法,该算法输入特定的词的词向量,输出是特定词对应的上下文词向量.利用分词后的文本数据,通过 skip-gram 算法计算,训练得到本文所使用的词向量表.

### 1.5 词向量表示

先对处理好的训练集和测试集进行去除停用词操作,然后做分词,将分词之后的文本词序列通过词向量表转化为词向量序列,完成文本向量化处理.最后将以词向量序列形式表示的判决结果输入到构建好的深度神经网络中进行模型训练.

## 2 基于注意力机制的 BiGRU 判决结果倾向性分析模型

本文提出了一种基于注意力机制的 BiGRU 判决

结果倾向性分析模型.本模型主要是由注意力机制和 BiGRU 神经网络构成,将以词向量表示的判决结果通过输入层输入到模型,经过 BiGRU 神经网络层特征提取以及注意力层的信息聚焦,最终将计算的结果通过 Softmax 进行分类.其中 BiGRU 层是由多个双向的 GRU 单元相互连通构成,进行双向计算.具体框架如图2所示.

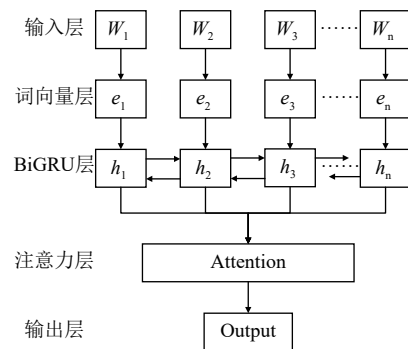


图2 基于注意力机制的 BiGRU 判决结果倾向性分析模型

### 2.1 BiGRU 算法模块

GRU (Gated Recurrent Unit) 是长短时记忆神经网络 (LSTM, Long Short-Term Memory) 的一种变体,和 LSTM 同属于循环神经网络 (RNN, Recurrent Neural Network) 的改进模型.由于 RNN 在处理序列时具有严重的梯度消失问题,即越靠后的节点对于前面的节点感知能力越低.为了解决梯度消失问题, Hinton 等人<sup>[7]</sup>提出了长短时记忆神经网络 (LSTM).而 GRU 作为 LSTM 的变体,对序列数据处理同样非常适合,也是通过“门机制”来记忆前面节点的信息,以此解决梯度消失问题. GRU 运用于文本情感分类时,通常将句子视为序列,词语则是序列的节点,通过 GRU 单元将词语按序进行组合,每个词语会获得其前面词语的信息,最终得到包含整个句子上下文信息的句向量,进而实现文本信息的特征提取.

与 LSTM 不同的是,GRU 模型只有两个门,分别为更新门和重置门,即图3中的  $z_t$  和  $r_t$ .更新门用于控制前一时刻的状态信息被传入到当前状态中的程度,更新门的值越大说明前一时刻的状态信息传入就越多.重置门用于控制遗忘前一时刻的状态信息的程度,重置门的值越小说明遗忘得越多.

由于 GRU 只能得到前向的上下文信息,忽略了后向的上下文信息,因此本文使用了双向 GRU 神经网络,即从前后向上同时获取上下文信息,以提高特征提取的准确率.并且 BiGRU 具有对词向量的依赖性小,复杂度低,响应时间快的优点.



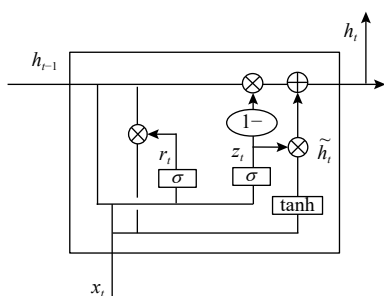


图3 GRU 模块示意图

$x_t$ 是输入数据,  $h_t$ 是 GRU 单元的输出,  $r_t$ 是重置门,  $z_t$ 是更新门,  $z_t$ 和 $r_t$ 共同控制了从 $h_{t-1}$ 隐藏状态到 $h_t$ 隐藏状态的的计算, 更新门同时控制当前输入数据和先前记忆信息 $h_{t-1}$ , 输出一个在 0~1 之间的数值 $z_t$ ,  $z_t$ 决定以多大程度将 $h_{t-1}$ 向下一个状态传递. 具体门单元计算为:

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

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \quad (2)$$

$$\tilde{h}_t = \tanh(W \cdot [r_t \times h_{t-1}, x_t]) \quad (3)$$

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

公式中 $\sigma$ 是 Sigmoid 函数,  $W_z$ ,  $W_r$ ,  $W$  分别为更新门, 重置门以及候选隐含状态的权重矩阵. 重置门控制  $h_{t-1}$ 对结果 $h_t$ 的重要程度, 当先前记忆 $h_{t-1}$ 和新的记忆完全相关性较大时, 可以通过重置门发挥作用, 提升先前记忆的影响. 根据重置门、更新门和隐含状态的计算结果可以通过公式得到当前时刻的输出 $h_t$ .

### 2.2 注意力机制模块

注意力机制的目标是实现从大量信息中忽略不重要的信息, 有选择地筛选出少量重要信息并聚焦. 聚焦表现为权重系数的计算, 权重表示信息的重要程度, 即权重越大表示在对应的文本信息上越聚焦.

输入文本序列较短时, 注意力机制在模型处理的结果数据中没有明显提升, 当输入文本序列比较长时, 文本语义只能表示为一个中间向量, 文本的信息已经丢失, 原句中的关键信息也已经部分丢失. 引入注意力机制的重要原因就是减少文本序列中关键信息的丢失. 由于注意力机制在自然语言处理中对强化关键信息的有效性, 所以本文在双向 GRU 神经网络的基础上添加了注意力机制来提高模型的准确率.

注意力机制可以减少文本序列关键信息丢失的原因是通过强化关键信息的比重, 实现方式是通过通过对不同的关键信息分配不同的权重, 以提升权重的方式来强化关键信息的比重, 以减少关键信息的丢失.

注意力机制权重矩阵的计算是通过由注意力机制分配的不同初始化概率权重与 BiGRU 层各个时间步的输出向量的乘积的相加, 最后经过 Softmax 函数计算得到数值.

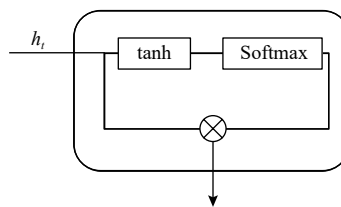


图4 注意力机制示意图

注意力机制的计算公式如下:

$$M = \tanh(h_t) \quad (5)$$

$$\alpha = \text{softmax}(w^T M) \quad (6)$$

$$\gamma = h_t \alpha^T \quad (7)$$

$h_t$ 为 BiGRU 层  $t$  个时间步的输出特征向量,  $w^T$ 为随机初始化并在训练中学习的权重矩阵,  $\gamma$ 为注意力层的输出值, 最后的结果经由激活函数将结果进行映射, 得到本文的分类结果.

## 3 实验分析

### 3.1 模型评价标准

对于模型的评价标准, 本文采用准确率 *Precision*、召回率 *Recall* 和 *F1* 值来进行评价. 计算公式如下:

$$Precision = \frac{N_{right}}{N_{right} + N_{wrong}} \quad (8)$$

$$Recall = \frac{TP}{TP + FN} \quad (9)$$

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

$N_{right}$ 表示分类正确的数量,  $N_{wrong}$ 表示分类错误的数量.  $TP$ 表示样本为正向且被判断为正向的个数,  $FN$ 表示样本为正向被误判为负向的个数. 本文利用前文中构建的测试集, 对以下设计的四种基于深度学习算法的情感分析模型进行实验.

### 3.2 实验设计与结果

实验以本文中构建的数据集按照 4:1 的比例划分为测试集和训练集. 数据集的总数为 2537 条, 其中支持原告的为 1728 条, 支持被告的为 809 条.

本文设计了四组实验, 使用了不同的神经网络算

法, 分别为: LSTM(baseline), GRU, BiGRU, Attention+BiGRU. 实验对比结果如表 2.

图 5 为 Attention+BiGRU 模型和 BiGRU 模型随着迭代次数变化而改变的准确率折线图:

表 2 对比实验结果

	Precision	Recall	F1
LSTM	0.856	0.862	0.859
GRU	0.860	0.864	0.862
BiGRU	0.878	0.881	0.879
Attention+BiGRU	0.900	0.903	0.901

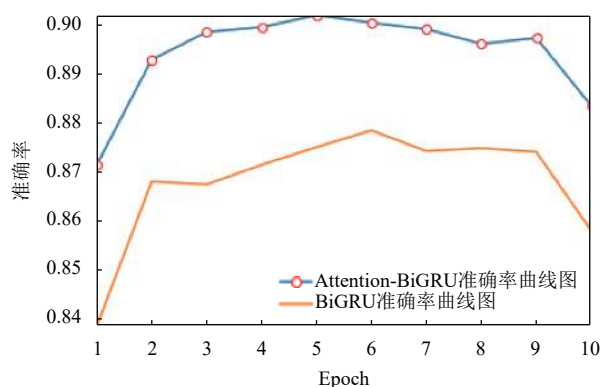


图 5 对比实验准确率折线图

### 3.3 结果分析

通过对比四组实验的数据结果可以发现: 基于 LSTM 的模型与基于 GRU 的模型具有相似的准确率, 而低于基于 BiGRU 的模型准确度. 基于注意力机制的 BiGRU 模型具有更高的准确率.

分析以上结果可以得出以下总结:

1) 在对比基于 GRU 算法的模型和基于 LSTM 算法的模型实验结果的准确率、召回率和 F1 值时, 可以得到结论: 在本文的实验中, GRU 和 LSTM 具有相似的准确性. GRU 在保持了 LSTM 的计算效果的同时使模型结构更加简单, 从而减少了计算量, 因此 GRU 的运行时间更短, 效率更高.

2) 在对比基于 GRU 算法的模型和基于 BiGRU 算法的模型实验结果时, 可以发现 BiGRU 准确率高于 GRU 的准确率, 主要是由于 BiGRU 对输入的序列在前后向上进行计算, 可以对文本信息特征的提取更精确. 所以双向的神经元模型对计算结果具有更高的准确性.

3) 在对比基于注意力机制的 BiGRU 算法的模型和基于 BiGRU 算法的模型时, 可以发现在添加注意力

机制后, 模型的准确率有了明显的提升, 这说明了注意力机制在本模型中对模型准确率的有效性.

## 4 结论与展望

本文通过在使用 BiGRU 神经网络的基础上, 利用注意力机制对输入数据中的关键信息分配不同的影响力权重, 进而凸显出关键特征, 从而提高模型对判决结果倾向性分析的准确性. 本文的贡献在于将注意力机制与 BiGRU 神经网络进行结合, 并成功运用到裁判文书的判决结果倾向性分析任务中, 同时, 通过设置三组对比实验, 验证了该模型具有较好的使用效果.

本模型主要实现了对裁判文书的判决结果倾向性做了较为准确的分类, 通过构建数据集验证, 并对比其他的算法模型, 证明了本模型的有效性. 进而为完成律师推荐的工作奠定基础.

本文的不足之处是在判决结果倾向性不明显时, 将判决结果分到了支持原告的一类, 未将倾向性分析做成支持原告, 支持被告和中立的三分类. 同时, 在处理判决结果是只是简单的按文本语义做倾向性分析, 未将裁判文书中的原被告各自诉求考虑在内. 在裁判文书中的可以提取相应的原被告诉求, 然后对原被告的诉求与判决结果做语义匹配, 可以判断判决结果是否符合原被告的诉求. 比如婚姻纠纷中的子女抚养权的归属, 应该参照原被告的诉求来决定倾向. 因此, 在下一步的研究工作中应该考虑添加分类的种类以及添加分类的影响因素.

### 参考文献

- 1 梁军, 柴玉梅, 原慧斌, 等. 基于深度学习的微博情感分析. 中文信息学报, 2014, 28(5): 155-161. [doi: 10.3969/j.issn.1003-0077.2014.05.019]
- 2 Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. arXiv:1409.0473, 2014.
- 3 王业沛, 宋梦姣, 王讚, 等. 基于深度学习的判决结果倾向性分析. 计算机应用研究, 2019, 36(2). <http://www.aocmag.com/article/02-2019-02-011.html>, 2018-01-19.
- 4 周瑛, 刘越, 蔡俊. 基于注意力机制的微博情感分析. 情报理论与实践, 2018, 41(3): 89-94.
- 5 白静, 李霏, 姬东鸿. 基于注意力的 BiLSTM-CNN 中文微博立场检测模型. 计算机应用与软件, 2018, 35(3): 266-274. [doi: 10.3969/j.issn.1000-386x.2018.03.051]
- 6 Chung J, Gulcehre C, Cho KH, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling. Eprint Arxiv:1412.3555, 2014.
- 7 Hinton GE. Learning distributed representations of concepts. Proceedings of the Eighth Annual Conference of the Cognitive Science Society. Amherst, MA, USA. 1986.