

基于过程监督的序列多任务法律判决预测方法



张春云¹ 曲浩² 崔超然¹ 孙皓亮² 尹义龙²

1 山东财经大学计算机科学与技术学院 济南 250014

2 山东大学软件学院 济南 250101

(zhangchunyun1009@126.com)

摘要 法律判决预测是人工智能技术在法律领域的应用,因此对法律判决预测方法的研究对于实现智慧司法具有重要的理论价值和实际意义。传统的法律判决预测方法大都是只进行单一任务的预测或仅基于参数共享的多任务预测,并未考虑各子任务之间的序列依存关系,因此预测性能难以得到进一步的提升。文中提出了一个端到端的基于过程监督的序列多任务法律判决预测模型,在建模各子任务之间的依存关系时,通过引入过程监督来确保依赖信息的准确性,从而提升序列子任务的预测性能。将所提模型应用到 CAIL2018 数据集上,取得了较好的分类效果,平均分类准确率比现有的 state-of-the-art 方法的准确率提升了 2%。

关键词 法律判决预测;多任务学习;过程监督;深度学习

中图分类号 TP391

Process Supervision Based Sequence Multi-task Method for Legal Judgement Prediction

ZHANG Chun-yun¹, QU Hao², CUI Chao-ran¹, SUN Hao-liang² and YIN Yi-long²

1 School of Computer Science and Technology, Shandong University of Finance and Economics, Jinan 250014, China

2 School of Software, Shandong University, Jinan 250101, China

Abstract Legal judgment prediction is an application of artificial intelligence technology in legal field. Hence, the research on the legal judgment prediction method has important theoretical value and practical significance for the realization of intelligent justice. Traditional legal judgment prediction methods only make single task prediction or just use multi-task prediction based on parameter sharing, without considering the sequence dependence among subtasks, so the prediction performance is difficult to be further improved. This paper proposes a process supervision based sequence multi-task framework (PS-SMTL) by encoding sequence dependency of subtasks in legal judgement. It is an end to end legal judgement prediction method without any external features. By introducing process supervision, the proposed model ensures the accuracy of the obtained dependent prior information from advance tasks. The proposed model is applied to CAIL2018 dataset and a good classification result is achieved. The average classification accuracy is 2% higher than that of the existing state-of-the-art method.

Keywords Legal judgement prediction, Multi-task learning, Process supervision, Deep learning

1 引言

法律判决预测 (Legal Judgment Prediction, LJP) 旨在使机器在阅读案件事实描述后能够预测法律案件的判决结果,是人工智能在法律领域的应用。实现法律判决的自动预测不仅能够辅助法官、律师等专业人士进行更加高效的法律预判,还能为缺乏法律专业知识的人群提供法律援助。近年来,随着机器学习方法和自然语言处理 (Natural Language Processing, NLP) 技术的发展,针对法律判决预测的研究引起了越来越多学者的关注。

LJP 的研究早在几十年前就开始了^[1-5]。目前,法律判决

任务被视为文本分类任务,研究者们围绕这一任务展开研究,并提出了许多里程碑式的方法^[6-13],主要包括基于传统机器学习方法和基于深度神经网络的法律判决预测方法。基于传统机器学习方法的法律判决方法^[9-10]大都是基于特定案件类型通过手工方式来标注特征,难以应用到其他类型的案件中,场景适应性较差;基于深度神经网络的法律判决预测方法^[11-12]不再依赖人工精心设计的模板,而是通过卷积神经网络或者循环神经网络来捕获上下文信息,并根据案件事实描述进行特征表示,进而实现判决预测。虽然现有方法取得了不错的预测效果,但是准确率仍然难以得到较大提升,主要原因是法律判决预测任务仍面临如下挑战。

到稿日期:2020-07-09 返修日期:2020-09-18 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家自然科学基金项目(61703234);国家重点研发计划(2018YFC0830102)

This work was supported by the National Natural Science Foundation of China (61703234) and National Key R & D Plan Project (2018YFC0830102).

通信作者:尹义龙(ylyin@su.edu.cn)

(1)法律判决书子任务之间存在拓扑依赖关系。法律判决通常由详细而复杂的条款组成,如涉及的相关法律条款、犯罪人所犯的罪行和处罚程度(罚款、刑期等)。因此,法律判决中的各个子任务预测应该是一个整体预测,并且各个任务之间存在相互的依存关系^[13]。Zhong 等^[13]指出,对于人类法官,法律判决的子任务之间存在严格的顺序。图 1 给出了特定案例的事实描述,法官首先确定与该场景相关的法条,然后根据相关法条的内容指示确定罪行。根据这些结果,法官再进一步确定刑期。因此,充分利用法律判决预测中各子任务之间的拓扑依存关系,对于提升判决预测的性能具有重要作用。然而,传统的法律判决任务大都是基于单一任务的预测^[7-8,21],或者是采用简单的参数共享的多任务框架,未考虑各个子任务之间的拓扑依存关系。

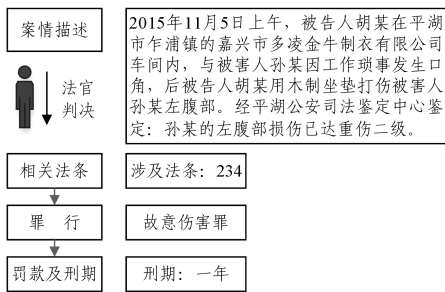


图 1 大陆法体系中法官判决逻辑展示图

Fig. 1 Illustration of judicial logic of human judges in civil law system

(2)仅基于案例事实描述进行的刑期预测缺乏先验知识的指导。相比其他子任务,基于事实描述的刑期预测具有更大的挑战。在现实场景下,在确定一件案例的刑期时,法官会受到很多因素的影响,除了根据案例事实描述中被告的年龄、案件涉案金额等因素进行判决以外,还要考虑其涉及的法律条款内容、所犯罪行等依赖任务中所包含的先验知识等因素。因此,如何在其所依赖的子任务信息的监督下获取刑期预测的先验知识,对提升最终的预测性能具有重要作用。然而,现有的刑期预测任务仅基于案例的事实描述部分进行预测^[7,9,17],或仅考虑训练过程中其依赖的子任务中间特征,将其作为刑期预测的先验知识,但缺乏对这些先验知识的过程监督^[13,18]。

为了解决上述问题,本文提出了一种基于过程监督的序列多任务法律判决预测模型(PS-SMTL),通过考虑各任务之间的序列拓扑依存关系,将法律判决中的相关法条、罪行和刑期的预测作为序列多任务问题。在训练过程中加入各个子任务的过程监督信息,从而确保获取的先行任务信息的准确性,进而实现对后续子任务的有效预测。本文模型是基于端到端的循环神经网络多任务框架,在多任务框架中添加子任务标签的监督信息,通过自注意力机制(self-attention)实现任务相关的特征表示,并将先行任务相关特征作为先验信息,与后续任务相关特征进行融合,将结果作为后续预测任务的输入,从而实现序列多任务的判决预测。

本文的贡献如下:

(1)首次提出基于过程监督的序列多任务预测模型,在建模序列子任务之间的依赖关系时,通过引入过程监督来保证获取的依赖信息的准确性,从而提升最终的预测性能。

(2)模型考虑各子任务之间的序列依赖关系,将法律判决

预测任务定义为序列多任务问题,符合司法判决的基本逻辑,能够通过编码先行任务的依赖特征来提升后续任务的分类性能。

(3)将本文模型应用在 CAIL2018 的法律判决预测任务数据集上,平均分类准确率比目前 state-of-the-art 方法的分类准确率高 2%。

2 相关工作

本文提出的模型涉及法律判决预测和多任务学习两方面的相关工作,下面将进行详细介绍。

2.1 法律判决预测

对于法律判决预测的研究早在几十年前就开始了。早期的研究大都受限于公开案件的局限性,基于统计的方法实现了对小部分判决结果的统计,而非真正意义上的预测^[1-5]。

随着机器学习方法和 NLP 技术的发展,法律判决预测任务被看作文本分类任务。因此,大多数的法律判决预测任务都是针对特定任务,探索如何抽取更有效的文本特征,进而基于机器学习方法实现较好的罪行预测^[9-10,17,19-20]。然而,这些传统的方法依赖于手工制作的浅层文本特征,需耗费大量的人力且领域适应性差,很难迁移到其他场景。

近年来,深度神经网络在 NLP 领域得到了成功应用^[14-16]。受此启发,研究者尝试融合法律知识的深度神经网络框架进行法律判决预测^[7-8,13,18,23-24]。例如,Luo 等^[12]采用基于注意力机制的神经网络模型,来实现罪行预测和相关文档抽取的联合任务。Hu 等^[21]通过定义 10 个有区分性的法律属性来实现对少样本和易混淆罪行的预测。Ye 等^[6]采用一个序列到序列的模型,基于案情描述生成法律文书并实现民事案件的罪行预测。Zhong 等^[13]提出 TOPJUDGE 模型,首次定义法律判决预测为多任务问题,并提出一种拓扑多任务学习框架。但是该方法在考虑任务之间的依赖关系时,未加入过程监督信息,难以保证依赖特征的准确性。Yang 等^[18]提出一种可以编码判决书子任务之间的依赖关系和验证关系的多视角多反馈的多任务学习框架,但是该方法难以实现端到端的预测,需输入外部辅助信息。

2.2 多任务学习

多任务学习(Multi-task Learning, MTL)旨在利用任务之间的共性和差异来同时解决相关任务问题,它可以实现各任务之间的有用信息传递,已被广泛应用于 NLP^[25-26]、语音识别^[27-28]和计算机视觉^[29-30]等领域。

多任务学习在 NLP 中有许多成功的应用^[31,33,35-36,39]。目前,基于 NLP 的多任务框架主要有两种通用方法:参数硬共享和软共享。参数硬共享方法通过在相关任务之间共享表示层或某些编码层来实现参数共享。Collobert 等^[31]通过共享嵌入表示层来解决词性标注和语义角色标注任务问题。Dong 等^[32]和 Minh 等^[33]通过共享编码和解码器实现从一种语言到多种语言的机器翻译问题的改进。Orhan 等^[34]在多语言翻译中采用多种方式的共享注意力机制。基于参数软共享的多任务学习方法软共享为每个任务都学习一个网络,但每个任务的网络都可以访问其他任务对应网络中的信息。例如,Long 等^[35]使用 L2 范数进行正则化,而 Yang 等^[36]使用迹范数进行正则化。Liu 等^[37]在特定任务的 RNN 层之间引入门来控制信息流。Sebastian 等^[38]引入了一个可以决定不

同 NLP 任务之间的共享量的多任务学习模型。

本文提出一种基于过程监督的序列多任务法律判决预测模型,用于实现端到端的法律判决预测。不同于传统的多任务学习研究如何实现参数共享,本文模型在建模各个子任务之间的序列依赖关系的同时,引入了过程监督机制,从而实现了融合先验信息的法律判决预测。

3 基于过程监督的序列多任务法律判决预测模型

本文提出一种基于过程监督的序列法律判决预测模型(PS-SMTL),下面将首先介绍法律预测判别预测任务中序列多任务的定义,然后介绍引入过程监督的循环神经网络框架。

3.1 基于法律判决预测的序列多任务定义

本文研究基于公民法系统的法律判决预测。给定一个案例的事实描述 x ,其可以看成是由词序列组成的, $x = \{x_1, x_2, \dots, x_n\}$,其中 n 为事实描述的长度,每个词 x_i 通过从一个固定的词典 W 中查字典获得。

基于事实描述 x ,法律判决预测任务是预测该案件所涉及的相关法条、罪行和刑期等。对其进行形式化表示,设 T 包含 $|T|$ 个分类子任务,即 $T = \{t_1, t_2, \dots, t_{|T|}\}$,对于第 i 个子任务 t_i ,我们的目标是预测其相应的类别标签 $y_i \in Y_i$,其中 Y_i 是子任务相关的标签集合。以罪行预测为例,相应的罪行预测标签包括欺诈、盗窃、故意杀人等标签。

由于判决预测任务的各个子任务之间存在拓扑依存关系^[13],本文将将其看作序列多任务问题。其中,使用 $t_i \triangleleft t_j$ 来表示第 j 个子任务的预测是依赖于第 i 个子任务的, $D_j = \{t_i \mid t_i \triangleleft t_j\}$ 用于表示第 j 个任务的依存集合。因此,任务向量 T 中的子任务需满足以下约束:

$$i < j, \forall (i, j) \in \{(i, j) \mid t_i \in D_j\} \quad (1)$$

按照上述依存约束进行任务排序,得到序列多任务的向量 $T = \{t_1, \dots, t_i, t_j, \dots, t_{|T|}\}$ 。

3.2 基于过程监督的序列多任务判决预测框架

本文提出的基于过程监督的序列多任务判决预测框架主要包括 4 层:共享事实描述编码层、基于自注意力机制的过程监督层、序列任务依赖信息融合层和输出预测层。下面将根据判决预测的 3 个子任务,即相关法条预测、罪行预测和刑期预测为例,对各个模块进行详细介绍。基于 3 个子任务的 PS-SMTL 模型框架如图 2 所示。

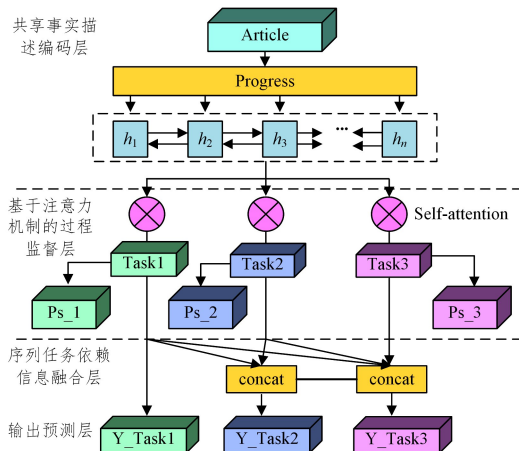


图 2 基于 3 个子任务的 PS-SMTL 模型框架

Fig. 2 Framework of PS-SMTL model based on three subtasks

3.2.1 共享事实描述编码层

在本文采用的多任务框架中,事实描述编码层为各个任务的共享层。对于事实描述,本文采用 BiLSTM^[41] 实现对事实描述的编码。给定输入 x ,LSTM 单元通过输入门、遗忘门、更新门和输出门进行计算,并对输入进行特征编码。

在每个时间步 $t \in [1, n]$,LSTM 单元以 x_t 作为输入,重新计算存储单元 c_t ,并输出新的隐藏状态 h_t ,如式(2)~式(7)所示:

$$f_t = \sigma(W_f[x_t; h_{t-1}] + b_f) \quad (2)$$

$$i_t = \sigma(W_i[x_t; h_{t-1}] + b_i) \quad (3)$$

$$o_t = \sigma(W_o[x_t; h_{t-1}] + b_o) \quad (4)$$

$$\hat{c}_t = \tanh(W_c[x_t; h_{t-1}] + b_c) \quad (5)$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \hat{c}_t \quad (6)$$

$$h_t = o_t \cdot \tanh(c_{t-1}) \quad (7)$$

其中, f_t , i_t 和 o_t 分别表示遗忘门、输入门和输出门, \cdot 表示元素乘, σ 和 \tanh 分别是 sigmoid 和双曲正切激活函数, W 和 U 分别是权重矩阵和偏差向量。在 BiLSTM 编码过程中,前向 LSTM 网络可获得关于事实描述的从左到右的特征表示,后向 LSTM 网络可以编码融合后续特征的语义。因此,可得到两个隐含层特征表示序列 $\{\vec{h}_1, \vec{h}_2, \dots, \vec{h}_n\}$ 和 $\{\overleftarrow{h}_1, \overleftarrow{h}_2, \dots, \overleftarrow{h}_n\}$,将每个单词的两个隐含状态进行拼接,可得到其隐含状态表示 $h_j = \{\vec{h}_j; \overleftarrow{h}_j\}$,最终可得:

$$h = BiLSTM(x, \phi) \quad (8)$$

即,输入向量 x 通过 BiLSTM 网络的编码来获得高层语义输出 h ,这里设置前向和后向隐含状态的维度为 $d/2$,从而拼接后隐含状态的维度为 d , ϕ 为上述编码过程涉及到的参数。

3.2.2 基于自注意力机制的过程监督层

基于自注意力机制的过程监督层主要通过引入过程监督来获取有效的先行任务相关依赖信息,从而为后续序列多任务依赖信息融合层抽取提供重要的特征保障。多任务过程监督层采用基于自注意力机制的 BiLSTM 网络,通过利用子任务标签信息进行监督,来获得面向各子任务的事实描述特征。

将共享事实描述编码层的输出 h 分别输入到 $|T|$ 个子任务各自的 BiLSTM 网络中,得到关于各个任务的高层语义表示:

$$h^i = BiLSTM(h, \phi_i), i = 1, 2, \dots, |T| \quad (9)$$

其中,参数 ϕ_i 为第 i 个子任务对应 BiLSTM 的参数。为了更好地获取各任务的相关特征,我们引入注意力机制,即通过将 h^i 输入到一个全连接层,并采用 softmax 函数将其映射到 $[0, 1]$ 之间。

$$a^i = \text{softmax}(W^i h^i + b_{a^i}) \quad (10)$$

然后,通过基于自注意力权重的加权求和获得面向各个子任务分类的特征表示。

$$\hat{h}^i = h^i \cdot a^i \quad (11)$$

为了保证各个子任务在建模中从事实描述特征中获取其相关的特征表示,在该部分训练时,添加各子任务的类别标签进行过程监督,即:

$$\hat{y}_{pi} = \text{softmax}(W_{pi}^i \hat{h}^i + b_{pi}^i) \quad (12)$$

基于获得的预测结果 \hat{y}_{pi} ,采用最小化 \hat{y}_{pi} 和 y_{pi} 之间的交叉熵:

$$L_{p_i}(\hat{y}_{p_i}, y_{p_i}) = - \sum_{k=1}^{|\mathcal{Y}_i|} y_{p_i,k} \log(\hat{y}_{p_i,k}) \quad (13)$$

通过引入过程监督,来确保其获得有效的面向子任务的特征表示,为后续基于依赖特征的序列多任务预测的准确性提供保障。

3.2.3 序列任务依赖信息融合层

序列任务依赖信息融合层根据序列任务之间的依赖关系,通过从其依赖任务相关的特征中获取先行任务的先验信息,并与当前任务相关的特征进行融合,实现对当前任务预测结果的提升。对于任务 i ,本模型简单采用其依赖任务相关的特征作为先验信息,并与当前任务相关特征进行简单拼接,然后输入到当前任务的分类器中进行分类,即:对于任务 j ,将其依赖任务集合 D_j 中所有任务 $t_i \in D_j$ 的相关特征 $\hat{h}^i, i=1, 2, \dots, k$ 和该任务的相关特征 \hat{h}^j 进行拼接,获得其分类特征,并对其进行仿射变换。

$$\hat{h}^d = \hat{h}^1 \oplus \hat{h}^2 \oplus \dots \oplus \hat{h}^k \quad (14)$$

$$h_{dj} = W_{dj} [\hat{h}^d; \hat{h}^j] + b_j^d \quad (15)$$

其中, $h_{dj} \in R^d, W_{dj}, b_j^d$ 是任务相关参数。对于罪行预测任务,其依赖于法条预测任务,故其分类特征为:

$$h_{dc} = W_c [\hat{h}^r; \hat{h}^c] + b_c \quad (16)$$

刑期预测任务依赖于法条和罪行预测任务,其分类特征为:

$$h_{di} = W_i [\hat{h}^r; \hat{h}^c; \hat{h}^t] + b_i \quad (17)$$

3.2.4 输出预测层

基于融合依赖任务信息的特征表示 h_{dj} ,对其进行线性变换并通过 softmax 函数实现对包含依赖信息任务的最终预测,从而最小化包含依赖任务的交叉熵,如式(16)和式(17)所示:

$$\hat{y}_{dj} = \text{softmax}(W_j^d h^j + b_j^d) \quad (18)$$

$$L_{dj}(\hat{y}_{dj}, y_{dj}) = - \sum_{k=1}^{|\mathcal{Y}_j|} y_{dj,k} \log(\hat{y}_{dj,k}) \quad (19)$$

3.2.5 训练

针对所有任务,我们将各子任务的过程监督交叉熵损失和最终的包含依赖信息任务的交叉熵进行加权求和,从而获得最终的损失函数:

$$L = \sum_{i=1}^{|\mathcal{T}|} \alpha_i \cdot p_i(\hat{y}_{p_i}, y_{p_i}) + \sum_{j=1}^{|\mathcal{T}|} \beta_j \cdot d_j(\hat{y}_{d_j}, y_{d_j}) \quad (20)$$

其中, α_i 为在过程监督过程中各个子任务的权重系数, β_j 为在包含依赖信息的训练过程中各个子任务的权重系数。在实际操作中,我们使过程监督的损失函数的权重保持一致,包含依赖信息的子任务损失权重一致,只需获取两部分的权重占比,这里我们采用 Adam^[42] 进行参数优化。为了防止训练过程过拟合,本文还在事实描述表示部分使用了 dropout 策略^[43]。

4 实验

为了验证本文方法的有效性,我们将其应用到 CAIL2018 数据集上实现对相关法条、罪行和刑期的预测。

4.1 数据集

本文使用了中国司法人工智能挑战赛¹⁾(CAIL2018)发布的训练阶段的公开数据集 CAIL2018-small 进行实验。

数据集中的每一个案例都包含事实描述、相关法条、被判罪行和被判的刑期。其中,有一部分案例涉及到多个相关法条和多个罪行。由于实验部分主要是为了验证在编码任务的序列依赖性时引入过程监督机制的有效性,因此我们过滤掉一些多标签的样本。除此之外,数据集中还有一些出现频率较低的相关法条和罪行,因此我们过滤掉出现频率低于 100 的类别,保留高于 100 的类别作为本文进行分类的目标类别。针对惩罚刑期,我们将其分成不重合的刑期段^[13],并对其进行分类预测。表 1 列出了数据集的统计信息描述。

表 1 数据集的统计信息描述

Table 1 Statistic information of datasets

Datasets	CAIL2018-small
Training Set Case	96925
Validation Set	12127
Test Set Cases	22772
Law Articles	94
Charges	115
Term of Penalty	11

4.2 参数设置

由于案例文件是中文案例,因此我们用 jiebe 进行分词,在过滤掉常见的停用词后,使用 word2vec²⁾ 对输入向量进行初始化。对于所有的模型,我们设定事实描述特征表示和任务相关的特征表示维度为 300 维,设置句子的最大长度为 128,最大文档长度包含 32 个句子。在训练过程中,将 Adam 优化器的学习率设置为 10^{-3} , dropout 参数设置为 0.5。训练的批次为 128,模型训练 20 个周期,并在测试集上进行最终模型的评估。

我们采用精确度(Acc.)、宏平均精确度(MP)、宏平均召回率(MR)和宏平均 F1 值(F1)作为评估指标。其中,宏平均指标指所有类别指标的平均值。

4.3 实验结果及分析

本文对法律判决预测的 3 个子任务的预测效果进行了评估:相关法条预测、罪行预测和刑期预测。为了对比实验结果,本文采用以下 3 个模型作为基线模型(每个基线系统都是在多任务框架下进行训练的,并取最好的实验结果进行对比)。

(1)CNN。文本采用包含多个滤波器长度的 CNN 模型^[15]对事实描述进行特征表示并进行分类。

(2)HSLTM。借鉴 Yang 等^[40]在情感分类任务中采用的层次神经网络结构,本文采用 LSTM 编码句子特征,使用另一个 LSTM 编码事实描述的文档特征。

(3)TOPJUDGE。Zhong 等^[13]考虑各个子任务之间的拓扑依存关系的法律判决预测模型,其中事实描述特征表示采用 LSTM 编码获得。

(4)SMTL。采用本文序列多任务学习框架,但不加入过程监督机制。

对法律判决的相关法条、罪行和刑期 3 个子任务的预测

¹⁾ <http://cail.cipsc.org.cn:2018/>

²⁾ <http://code.google.com/p/word2vec/>

实验结果如表 2—表 4 所列。由实验结果可知,本文采用的 PS-SMTL 模型在 3 个任务的 4 个分类指标上都高于基线系统,这说明了本文模型的有效性。同时,针对有序列依存关系的多任务序列,相关法条(t_1)、罪行(t_2)和刑期预测(t_3)3 个子任务的依赖关系如下:

$$D_1 = \Phi, D_2 = \{t_1\}, D_3 = \{t_1, t_2\} \quad (19)$$

即罪行的预测依赖于相关法条的预测,刑期的预测依赖于相关法条和罪行的预测。针对有依存关系的序列多任务预测,相比传统的仅采用参数共享的多任务学习框架 CNN 和 HLSTM 模型,对序列依存关系进行编码的多任务学习框架 TOPJUDGE, SMTL 和 PS-SMTL 模型能够取得更好的预测效果,说明序列任务之间这种明确的依赖关系符合人类法官进行判决的逻辑,对提升最终的罪行、刑期预测结果有重要作用,这一结论在实验结果中已得到验证。

表 2 相关法条任务预测结果

Table 2 Prediction results of related law tasks

Methods	Task	Law Articles			
		Acc.	MP	MR	F1
Baselines	CNN	0.815	0.771	0.740	0.751
	HLSTM	0.856	0.816	0.790	0.798
	TOPJUDGE	0.863	0.825	0.807	0.813
Ours	SMTL	0.866	0.839	0.812	0.821
	PS-SMTL	0.870	0.847	0.817	0.826

表 3 罪行预测结果

Table 3 Results of charges prediction task

Methods	Task	Charges			
		Acc.	MP	MR	F1
Baselines	CNN	0.812	0.781	0.756	0.764
	HLSTM	0.852	0.817	0.803	0.803
	TOPJUDGE	0.860	0.828	0.824	0.822
Ours	SMTL	0.863	0.845	0.831	0.834
	PS-SMTL	0.868	0.852	0.835	0.836

表 4 刑期子任务预测结果

Table 4 Prediction results of term of penalty sub-tasks

Methods	Task	Term of Penalty			
		Acc.	MP	MR	F1
Baselines	CNN	0.372	0.316	0.277	0.258
	HLSTM	0.413	0.372	—	0.327
	TOPJUDGE	0.428	0.382	0.341	0.332
Ours	SMTL	0.429	0.385	0.334	0.332
	PS-SMTL	0.433	0.403	0.355	0.342

由表 2—表 4 可知,在考虑各任务的序列依赖关系的条件下,相比 TOPJUDGE 模型,本文采用的序列多任务框架 SMTL 和 PS-SMTL 的分类效果优于 TOPJUDGE 模型,尤其是基于各类别的宏平均分类准确率能够达到良好的效果,最大可提升 2%。对于各类别的宏平均召回率和 F1 值也都能提升 1%。这说明本文采用的序列多任务框架对各任务序列依赖关系的编码方式是有效的。

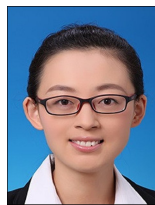
为了进一步验证过程监督的有效性,本文对过程监督机制进行了消融实验。通过比较发现,引入过程监督机制的 PS-SMTL 模型比未引入过程监督机制的 SMTL 模型的效果有进一步的提升,说明了过程监督机制的有效性,即引入过程监督机制能够保证获得有效的先行任务的依赖特征,从而为将其输入到后续任务、提升其分类性能提供了重要保障。

结束语 针对法律判决预测任务中各子任务存在的序列依赖关系,本文提出一种基于过程监督的序列多任务法律判决预测模型。该模型在序列多任务学习框架下,在编码各任务之间的依赖关系时引入了过程监督机制。过程监督机制能保证在序列多任务信息传递中基于先行任务的依赖先验信息的正确性,从而实现后续任务预测性能的提升。实验结果验证了本文模型采用的序列多任务框架和过程监督机制的有效性。然而,针对本文模型对基于先行任务先验信息的特征获取,融合部分仅采用先行任务相关特征和当前任务特征进行拼接来获得,之后可以尝试其他的特征获取和融合方案,以提升最终分类性能提供帮助。

参考文献

- [1] FRED K. Predicting Supreme Court decisions mathematically: A quantitative analysis of the “right to counsel” cases [R]. American Political Science Review, 1957.
- [2] STUART S. Applying correlation analysis to case prediction [R]. Texas Law Review, 1963.
- [3] KEOWN R. Mathematical Models For Legal Prediction [J]. Computer Journal, 1980(1): 829-831.
- [4] JEFFREY A. Predicting Supreme Court cases probabilistically: The search and seizure cases (1962-1981) [R]. American Political Science Review, 1984, 78(4): 891-900.
- [5] BENJAMIN E L, TOM S C. The Supreme Court’s many median justices [J]. American Political Science Review, 2012, 106(4): 847-866.
- [6] CHEN H J, CAI D, DAI W, et al. Charge-Based Prison Term Prediction with Deep Gating Network [C] // Proceedings of EMNLP-IJCNLP, 2019: 6362-6367.
- [7] YE H, JIANG X, LUO Z C, et al. Interpretable charge predictions for criminal cases: Learning to generate court views from fact descriptions [C] // Proceedings of NAACL, 2018: 1854-1864.
- [8] JIANG X, YE H, LUO Z C, et al. Interpretable Rationale Augmented Charge Prediction System [C] // Proceedings of COLING, 2018: 146-151.
- [9] LIU C L, CHENG T C, JIM H H. Case instance generation and refinement for case-based criminal summary judgments in Chinese [J]. Journal of Information Science & Engineering, 2004, 20(4): 783-800.
- [10] LIU C L, CHWEN D H. Exploring phrase-based classification of judicial documents for criminal charges in Chinese [C] // Proceedings of ISMIS, 2006, 681-690.
- [11] DANNIEL M K, MICHAEL J B II, JOSH B. A general approach for predicting the behavior of the supreme court of the united states [J]. Plos One, 2017, 12(4).
- [12] LUO B F, FENG Y S, XU J B, et al. Learning to predict charges for criminal cases with legal basis [C] // Proceedings of EMNLP, 2017: 1-10.
- [13] ZHONG H X, GUO Z P, TU C C, et al. Legal judgment prediction via topological learning [C] // Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2018: 3540-3549.

- [14] BAHARUM B, LAM H L, KHAIRULLAH K. A review of machine learning algorithms for text-documents classification [J]. *Journal of Advances in Information Technology*, 2010, 1(1): 4-20.
- [15] YOON K. Convolutional neural networks for sentence classification [C] // *Proceedings of EMNLP*. 2014.
- [16] TANG D Y, QIN B, LIU T. Document modeling with gated recurrent neural network for sentiment classification [C] // *Proceedings of EMNLP*. 2015:1422-1432.
- [17] LIN W C, TSUNG T K, TUNG J C. Exploiting machine learning models for chinese legal documents labeling, case classification, and sentencing prediction [C] // *Proceedings of ROCLING*. 2012:140.
- [18] YANG W, JIA W, ZHOU X, et al. Legal Judgment Prediction via Multi-Perspective Bi-Feedback Network [C] // *Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI-19)*. 2019.
- [19] NIKOLAOS A, DIMITRIOS T, DANIEL P, et al. Predicting judicial decisions of the European court of human rights: A natural language processing perspective [J]. *PeerJ Computer Science*, 2016(2):e93.
- [20] OCTAVIA M S, MARCOS Z, MIHAELA V, et al. Exploring the use of text classification in the legal domain [C] // *Proceedings of ASAIL Workshop*. 2017.
- [21] HU Z K, LI X, TU C C, et al. Few-shot charge prediction with discriminative legal attributes [C] // *Proceedings of the 27th International Conference on Computational Linguistics*. 2018:487-498.
- [22] RONAN C, JASON W. A unified architecture for natural language processing: Deep neural networks with multitask learning [C] // *Proceedings of ICML*. 2008:160-167.
- [23] LONG S, TU C, LIU Z, et al. Automatic Judgment Prediction via Legal Reading Comprehension [J]. *arXiv:1809.06537v1*, 2018.
- [24] MAX R S M, TOMMASO B, MAXIME R, et al. Machine learning for explaining and ranking the most influential matters of law [C] // *Proceeding of ICAIL*. 2019:239-243.
- [25] SANDEEP S, ADAM T, YOSHUA B, et al. Learning General Purpose Distributed Sentence Representations via Large Scale Multi-task Learning [C] // *Proceeding of ICLR*. 2018.
- [26] WANG A, AMANPREET S, JULIAN M, et al. GLUE: A Multi-Task Benchmark and Analysis Platform for Natural Language Understanding [C] // *Proceeding of ICLR*. 2019.
- [27] WU Z, VALENTINI B C, WATTS O, et al. Deep neural networks employing Multi-Task Learning and stacked bottleneck features for speech synthesis [C] // *IEEE International Conference on Acoustics*. 2015:4460-4464.
- [28] RAVANELLI M, ZHONG J, PASCUAL S, et al. Multi-Task Self-Supervised Learning for Robust Speech Recognition [C] // *Proceeding of ICASSP*. 2020.
- [29] TREVOR S, AMIR R Z, DAWN C, et al. Which Tasks Should Be Learned Together in Multi-task Learning? [C] // *Proceeding of ICCV*. 2019.
- [30] LIU S K, EDWARD J, ANDREW J D. End-to-End Multi-Task Learning with Attention [C] // *Proceeding of CVPR*. 2019.
- [31] RONAN C, JASON W. A unified architecture for natural language processing: Deep neural networks with multitask learning [C] // *Proceedings of ICML*. 2008:160-167.
- [32] DONG D X, WU H, HE W, et al. Multi-task learning for multiple language translation [C] // *Proceedings of ACL*. 2015:1723-1732.
- [33] MINH T L, QUOC V L, ILYA S, et al. Multi-task sequence to sequence learning [C] // *Proceedings of ICLR*. 2016.
- [34] ORHAN F, KYUNGHYUN C, YOSHUA B. Multi-way, multi-lingual neural machine translation with a shared attention mechanism [C] // *Proceedings of NAACL*. 2016:866-875.
- [35] LONG D, TREVOR C, STEVEN B, et al. Low resource dependency parsing: Cross-lingual parameter sharing in a neural network parser [C] // *Proceedings of ACL*. 2015:845-850.
- [36] YANG Y X, TIMOTHY H. Deep multi-task representation learning: A tensor factorization approach [C] // *Proceedings of ICLR*. 2017.
- [37] LIU P F, QIU X P, HUANG X J. Recurrent neural network for text classification with multi-task learning [C] // *Proceedings of IJCAI*. 2016.
- [38] SEBASTIAN R, JOACHIM B, ISABELLE, et al. Learning what to share between loosely related tasks [J]. *arXiv:1705.08142*, 2017.
- [39] POORYA Z, GHOLAMREZA H. Neural Machine Translation for Bilingually Scarce Scenarios: A Deep Multi-task Learning Approach [C] // *Proceeding of NAACL*. 2018.
- [40] YANG Z C, YANG D Y, DYER C, et al. Hierarchical attention networks for document classification [C] // *Proceedings of NAACL-HLT*. 2016:1480-1489.
- [41] ZHANG S, ZHENG D Q, HU X C, et al. Bidirectional long short-term memory networks for relation classification [C] // *Proceeding of PACLIC*. 2015:1480-1489.
- [42] DIEDERIK K, JIMMY B. Adam: A method for stochastic optimization [C] // *Proceedings of ICLR*. 2015.
- [43] NITISH S, GEOFFREY E H, ALEX K, et al. Dropout: a simple way to prevent neural networks from overfitting [J]. *Journal of Machine Learning Research (JMLR)*, 2014, 15(1):1929-1958.



ZHANG Chun-yun, born in 1986, Ph.D, assistant professor, is a member of China Computer Federation. Her main research interests include machine learning, natural language processing and information extraction.



YIN Yi-long, born in 1972, Ph.D, professor, Ph.D supervisor, is a member of China Computer Federation. His main research interests include machine learning, data mining and biometrics.