

JurisCTC: 通过跨领域转移学习和对比学习增强法律判决预测

Zhaolu Kang*

School of Software & Microelectronics
Peking University
Beijing, China
kangzl9966@gmail.com

Hongtian Cai*

School of Electrical & Electronic Engineering
Nanyang Technological University
Singapore
cht7janus@gmail.com

Xiangyang Ji

College of Software
Jilin University
Changchun, Jilin, China
jixy5523@mails.jlu.edu.cn

Jinze Li

College of Computer Science and Technology
Jilin University
Changchun, Jilin, China
lijz2128563286@outlook.com

Nanfei Gu[†]

KoGuan School of Law
Shanghai Jiao Tong University
Shanghai, China
gunanfei@126.com

Abstract—近年来,无监督领域适应(UDA)因其在增强模型跨不同领域的泛化能力上的表现而在自然语言处理(NLP)领域获得了显著关注。然而,它在不同法律领域之间的知识转移的应用却仍然未被广泛探索。为应对冗长和复杂的法律文本以及大规模标注数据集有限的挑战,我们提出了一种新模型JurisCTC,旨在提高法律判决预测(LJP)任务的准确性。与现有方法不同,JurisCTC促进了跨各种法律领域的有效知识转移,并采用对比学习来区分来自不同领域的样本。具体来说,对于LJP任务,我们在民法和刑法领域之间实现知识转移。与其他模型和特定的大型语言模型(LLMs)相比,JurisCTC表现出显著进步,分别实现了76.59%和78.83%的峰值准确率。¹

Index Terms—Unsupervised domain Adaptation, Legal Judgment Prediction, Transfer Learning, Contrastive Learning, Large Language Models

I. 介绍

法律判决预测(LJP)是指根据法律案件的事实以及案件描述中提出的其他相关信息(如论点和主张)来预测法院裁决结果的任务。该领域旨在利用计算技术来预测司法裁决,以在各种法律背景下提供显著的益处。自动化的LJP系统具有相当的实用价值:它们可以协助法律专业人士分析案件并向公众提供咨询服务,从而降低法律成本并改善司法获取。

尽管有潜在的好处,最近尝试主要集中在判决的文本分析和特定法律领域的预测。这些方法常常忽略了司法结果与法律适用的逻辑一致性之间的关系。大多数这些研究从文本(例如,N-grams)或案件档案(例如,日期、术语、地点和类型)中提取有效特征。这些方法需要大量的人力来进行特征工程和案件注释,这既耗时又耗费资源。此外,当这些方法应用于不同的法律场景时,常常面临泛化挑战,限制了它们在各种法律背景下的适用性。这凸显了需要更为强大和适应性强的模型来处理多样的法律数据,并尽量减少手动干预。

* These authors contribute equally to this work. They are co-first authors.

[†] Corresponding author.

¹ JurisCTC is available at: <https://github.com/Zhaolu-K/JurisCTC>

在法律应用的复杂领域中,刑法和民法已成为法律判决预测(LJP)中被广泛探索的领域。然而,它们彼此之间并不具有互操作性。在刑法中,LJP模型旨在预测结果,如适用的法律条款、指控和刑期[?],[?],[?],[?],[?],[?]。在民法中,LJP模型则专注于确定原告的诉求是否能够成立[?],[?]。

LJP模型的有效性在很大程度上取决于用于训练的判决书的数量和质量。然而,由于国家和社会稳定的考虑,中国裁判文书网上公开的刑事判决数目显著减少,数据的短缺直接限制了刑法领域LJP研究的发展。这种数据稀缺性对旨在开发强大LJP模型的研究人员构成了重大挑战。

为了解决这个问题,我们建议利用迁移学习技术。将迁移学习整合到LJP中不仅能缓解训练数据有限的问题,还能增强预测模型的鲁棒性和适应性。通过利用民法中丰富的数据,我们可以创建更通用的模型,即使在刑法的受限环境中也能表现良好。迁移学习能够从民事判决中提取知识,利用法官的集体智慧来改善刑法预测。

此外,我们将无监督领域适应(UDA)融入刑法结果预测中,探索不同法律领域之间的互操作性。这种方法展示了跨领域学习的潜力,来自一个法律领域的见解可以为另一个领域提供信息并加以增强。这种方法也为跨学科研究打开了新的途径,来自一个法律领域的见解可以通知并改善另一个领域的实践。鉴于中国披露司法文件政策的调整,这种研究方法具有更显著的实际价值和理论意义。

我们的贡献总结如下:

- 我们提出了一种在不同法律领域中使用迁移学习的方法,以改善培训数据不足的问题,为跨部门法律研究提供了新的思路,并为其他法律领域的后续知识转移铺平了道路。
- 我们引入了一种跨域迁移和对比学习的方法,以提高LJP的准确性。

II. 相关工作

近年来,法律判决预测(LJP)引起了显著关注并取得了实质性进展。大量法律判决数据的可用性催生了一个不断

增长的研究领域，专注于这一主题。自然语言处理 (NLP) 的最新进展显著推动了 LJP 模型的发展。这些模型利用大规模公共数据集和复杂的算法，以惊人的准确性预测司法结果。然而，法律推理的复杂性和法律论证的主观性带来了持续的挑战。为了应对这些问题，研究人员开始将论证分析纳入 LJP 模型，从而增强它们评估相关方所提出法律论证质量的能力。

当前 LJP 研究主要集中在根据事实信息 [?], [?], [?]、原告叙述 [?], [?] 和其他相关的法庭陈述信息 [?], [?]，预测案件结果 [?], [?], [?], [?]，如适用的法律条款、指控和刑期。

尽管这些任务内容全面，但预测结果的相互依赖性通常导致缺乏直观清晰性。这个问题在刑法中尤其明显，因为各种法律结果的复杂交互可能会掩盖直接的有罪或无罪判决。相反，民法中的预测往往更直接，提供了更清晰的结果描述 [?], [?], [?]。这一区别对于我们的研究至关重要，我们的研究重点是中国法律的逻辑应用。鉴于刑法数据可用性的逐步减少，我们识别出利用民法数据集来填补这一空缺的宝贵机会。我们的研究旨在通过专注于中国民法背景下的法律判定预测应用，利用民法数据集的独特特性来增强预测的准确性和清晰度，从而在这些发展基础上取得进展。

A. 无监督领域自适应

深度前馈架构在各种机器学习任务和应用上带来了令人印象深刻的先进成果。然而，这些性能上的飞跃通常依赖于大量标记的训练数据的可用性。对于标记数据稀缺的问题，通常可以获得足够大的训练集，但这些训练集可能在数据分布上与测试时遇到的实际数据有所不同。一个显著的例子是合成或半合成训练数据，这些数据可能是丰富且完全标记的，但其分布不可避免地与实际世界数据不同。

机器学习已广泛应用于多个领域 [?], [?], [?], [?], [?], [?], [?]。无监督域适应 (UDA) 已被证明是一种有效的策略，可以将知识从标注良好的源域转移到未见过的、多样的和未标注的目标域。通过利用标注的源域和未标注的目标域的数据，UDA 促进了目标域内各种任务的表现。这种方法已成功应用于多个领域，包括自然图像处理、视频分析、自然语言处理、时间序列数据分析和医学图像分析。

在自然语言处理 (NLP) 领域，开发无监督域自适应 (UDA) 方法变得越来越重要，特别是由于为大型语言数据集注释所需的高昂成本。UDA 技术已被用于多种 NLP 任务，如情感分析 [?], [?]、关系抽取 [?], [?] 和语言识别 [?]。NLP UDA 领域的开创性工作包括 Ganin 等人提出的领域对抗神经网络 (DANN) [?]，其通过在特征提取器中整合域分类器并采用梯度反转层实现 UDA。

UDA 研究的主要焦点在于学习域不变特征。这可以通过以下两种方式实现：一是使用一些分布差异度量显式地减小源域和目标域特征空间之间的距离，二是通过对抗训练，其中特征提取器被训练来欺骗域分类器。两种方法都被联合优化，以实现对齐的特征空间。我们的研究重点是将后一种方法应用于基于 transformer 的模型，如 BERT [?]，用于文本任务。

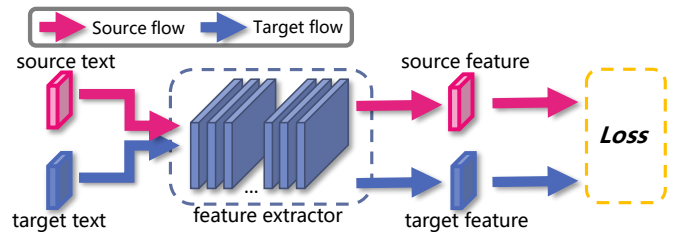
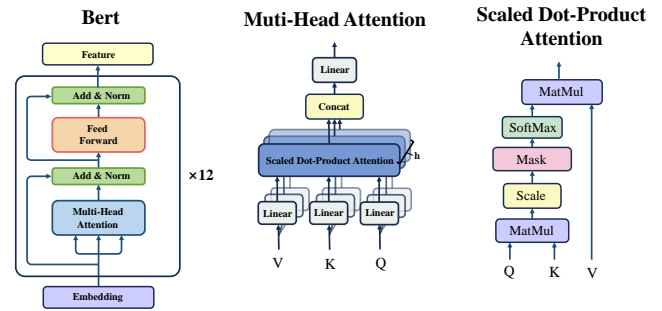
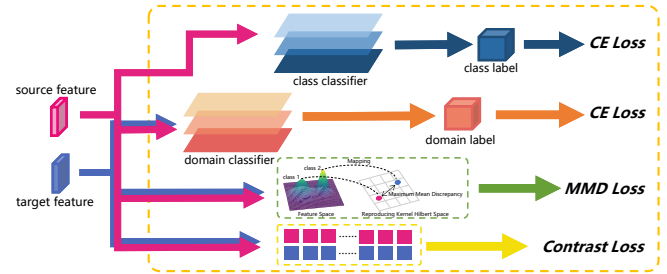


Fig. 1: 该图展示了一个集成 BERT 进行特征提取的高级领域适应模型。该架构并行处理源数据和目标数据，利用 BERT 提取特征，从而为标签预测和领域分类的损失计算提供信息。



(a) 子图 1: 该图描绘了模型错综复杂的结构，展示了将“Bert”模块作为多方面特征提取器，并在其中反复进行“Add & Norm”和“Multi-Head Attention”处理。“Multi-Head Attention”模块通过“Concat”和“Linear”操作详细展示，并引入“Scaled Dot-Product Attention”机制。



(b) 子图 2: 该图表说明了域适配模型中的损失组件。交叉熵 (CE) 损失用于类别和域分类。最大平均差异 (MMD) 损失用于对齐源和目标特征分布。对比损失通过将相似实例聚集在一起并将不相似实例分离开来增强特征学习。

III. 方法

A. 概述

我们的模型由三个关键组件组成：一个 BERT 特征提取器、一个类别分类器和一个领域分类器。图 1 展示了我们模型的整体架构，突出显示了 BERT 特征提取器、类别分类器和领域分类器之间的相互作用。

最初，BERT 特征提取器对输入文本进行解释，将其转换为捕捉上下文和语义复杂性的丰富特征集。这个阶段利用强大的预训练 BERT 模型生成嵌入，以封装输入文本的细微含义。

在第二阶段，复杂的神经网络分类器被应用于源领域。这个全连接网络将提取的特征映射到源领域内的对应标签，微调预训练的 BERT 表示以适应我们的特定分类要

求。此步骤确保模型能够很好地适应特定任务，提高其在正确标签预测方面的准确性。

第三阶段涉及将领域分类器战略性地应用于源领域和目标领域。领域分类器被训练来区分源领域和目标领域的特征。通过迭代地应用领域分类器，我们的模型逐步学习通过对抗性训练来减少领域差异。这个过程包括一个梯度反转层，该层鼓励特征提取器生成领域不变的特征，从而提高领域适应场景中的预测准确性和鲁棒性。

B. 前向传播

BERT 的架构巧妙地设计用以利用双向 Transformer 编码器的力量来捕捉语言的复杂细微差别。过程始于将输入句子分词成离散的标记，然后将这些标记嵌入到向量中。这些向量通过多层 Transformer 编码器进行处理，每一层通过一系列操作来细化标记的表示。

在 BERT 中的每个编码器层对输入嵌入执行一系列操作。每层的第一个操作是多头自注意机制，它允许模型在整个句子的上下文中考虑每个词。这是通过为每个标记生成查询 (Q)、键 (K) 和值 (V) 向量，并计算注意力分数来决定其他标记的影响来实现的。

多头注意力计算如下：

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W^O \quad (1)$$

每个头 head_i 的计算方式如下：

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \quad (2)$$

自注意力机制的数学表达式如下：

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (3)$$

其中 d_k 是键的维度。对于多头注意力中的每个“头”来说，这是并行执行多次的，使模型能够捕捉数据的不同方面。

在注意力机制之后，每一层独立地对每个标记应用一个逐位置的前馈网络。这个网络由两个线性变换组成，中间有一个 ReLU 激活函数。

位置感知前馈网络可以用数学形式表示为：

$$\text{FFN}(x) = \max(0, xW_1 + b_1)W_2 + b_2 \quad (4)$$

其中 W_1 、 W_2 、 b_1 和 b_2 是线性变换的参数。

在预训练期间，BERT 使用 MLM 目标，该目标旨在根据序列中其他未被遮挡的词提供的上下文来预测被遮挡词的原始词汇 ID。

MLM 的损失函数是词汇表上的交叉熵损失：

$$L_{\text{MLM}} = -\sum_{i=1}^N \log p(w_i | w_{\text{context}}) \quad (5)$$

其中 N 是被掩盖的标记数量， w_i 是真实标记， w_{context} 代表周围未被掩盖的标记。

BERT 特征提取器是我们方法的基石，它勤奋地处理输入文本以生成一个特征向量 f ，包含着语言上下文和语义丰富性。这个向量存在于高维空间 $\mathbb{R}^{m \times d}$ 中，其中 m 和 d 分别表示最大输入序列长度和隐藏状态维度，形成了后续预测任务的基础。

通过利用预训练的 BERT 模型并在我们的特定领域语料库上进行微调，我们将其广泛的语言知识适应于我们的特定用例。该模型在识别上下文依赖关系方面的能力不仅局限于相邻的词元，还扩展到包括远距离依赖关系，从而减轻了传统上与顺序处理模型相关的限制。这使得能够提取出对我们关注结果具有高度预测性的特征，从而提高我们预测模型的性能。

类别分类器被设计成具有多层结构的神经网络，然后开始工作。它处理特征向量 f ，应用一系列转换，最终预测出适当的标签。这是通过函数 $g: \mathbb{R}^{m \times d} \rightarrow \mathbb{R}^l$ 实现的，其中 l 是可能的标签数量。

函数 g 定义如下：

$$g(f) = \sigma(W_l f + b_l) \quad (6)$$

其中 W_l 是权重矩阵， b_l 是偏置向量， σ 是引入非线性以使网络能够学习复杂模式的激活函数。

同时，领域分类器，即另一个神经网络，参与一个并行的过程。它评估特征向量 f 以确定其来源领域，使用一个函数 $h: \mathbb{R}^{m \times d} \rightarrow \mathbb{R}^k$ ，其中 k 表示领域的数量。

函数 h 被表示为：

$$h(f) = \sigma(W_d f + b_d) \quad (7)$$

在这个方程中， W_d 是领域特定的权重矩阵， b_d 是对应的偏置向量。激活函数 σ 再次在区分领域的的能力中起到了重要作用。

C. 误差反向传播

为了实现领域对抗训练，我们实施了一个梯度反转层 (GRL)。GRL 是一个独特的组件，它促进了源域和目标域之间的特征分布对齐。除了一个超参数 λ 外，它在运行时并不包含任何可训练的参数，该超参数不会通过反向传播进行更新。在前向传播时，GRL 作为恒等函数运作，允许数据不变地通过。然而，在反向传播期间，GRL 将梯度乘以 λ 并反转其方向，从而有效地鼓励特征提取器生成域不变特征。

从数学上讲，GRL 可以表示为：

域分类器的损失函数 L_d 的定义使用二元交叉熵损失。该损失衡量预测的域标签和真实域标签之间的差异，引导模型准确地区分源域和目标域。

二元交叉熵损失定义为：

$$L_d = -\frac{1}{N} \sum_{i=1}^N [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)] \quad (8)$$

其中 N 是样本数量， y_i 是真实领域标签， p_i 是样本属于源领域的预测概率。

分类损失 L_y 是在源域上计算的，使用预测标签和真实标签之间的交叉熵损失，确保在源域任务上的预测准确性。用于分类的交叉熵损失定义为：

$$L_y = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C y_{i,c} \log(p_{i,c}) \quad (9)$$

其中 C 是类别数量，当类标签 c 是样本 i 的正确分类时， $y_{i,c}$ 是一个二进制指示符 (0 或 1)，而 $p_{i,c}$ 是样本 i 属于类别 c 的预测概率。

为了进一步对齐源域和目标域的特征分布，我们采用了最大均值差异 (MMD) 方法。MMD 是一种非参数度量，用于量化两个域的特征分布之间的距离。通过将 MMD 损失融入到我们的训练过程中，我们鼓励模型最小化域差异。

MMD 基于高斯核，定义为：

$$k(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right) \quad (10)$$

这里 $k(x, y)$ 是样本 x 和 y 之间的高斯核， $\|x - y\|^2$ 是样本之间的平方欧几里得距离， σ 是控制核扩展的带宽参数。

MMD 损失被表述为：

$$L_{\text{MMD}} = \left\| \frac{1}{n_s} \sum_{i=1}^{n_s} \phi(\mathbf{f}_i^s) - \frac{1}{n_t} \sum_{j=1}^{n_t} \phi(\mathbf{f}_j^t) \right\|^2 \quad (11)$$

其中， ϕ 表示一个特征图，将特征投射到再生核希尔伯特空间， \mathbf{f}_i^s 和 \mathbf{f}_j^t 分别是源域和目标域的特征表示， n_s 和 n_t 是每个域中的样本数量。此外，我们通过结合对比学习来增强两个域之间的特征分布对齐。该方法通过最小化相似样本之间的距离，同时最大化不相似样本之间的距离。我们利用源域和目标域样本，通过提取它们的特征并计算一个相似度矩阵，来捕获两个域中所有样本对之间的关系。

对比学习函数可以定义如下。

连接源域和目标域特征：

$$\mathbf{z} = [\mathbf{z}_s; \mathbf{z}_t] \quad (12)$$

其中 \mathbf{z}_s 代表源域特征， \mathbf{z}_t 代表目标域特征。

对特征进行标准化：

$$\mathbf{z}_{\text{norm}} = \frac{\mathbf{z}}{\|\mathbf{z}\|} \quad (13)$$

其中 \mathbf{z}_{norm} 是归一化的特征向量。

计算相似矩阵：

$$\mathbf{S} = \mathbf{z}_{\text{norm}} \mathbf{z}_{\text{norm}}^\top \quad (14)$$

其中 $\mathbf{S}_{i,j}$ 表示样本 i 和样本 j 之间的相似性。

构建对比学习标签矩阵：

$$\mathbf{M}_{i,j} = \begin{cases} 1, & \text{if } y_i = y_j \\ 0, & \text{otherwise} \end{cases} \quad (15)$$

其中 y_i 和 y_j 分别是第 i 和 j 个样本的标签。

对比学习损失定义为：

$$L_{\text{contrast}} = -\frac{1}{N} \sum_{i=1}^N \log \frac{\exp\left(\frac{\text{sim}(z_i, z_j)}{\tau}\right)}{\sum_{k=1}^N \exp\left(\frac{\text{sim}(z_i, z_k)}{\tau}\right)} \quad (16)$$

其中， N 是批大小。 z_i 和 z_j 代表正样本对的特征向量， $\text{sim}(z_i, z_j)$ 是特征向量 z_i 和 z_j 之间的相似性，通常计算为余弦相似性。 τ 是一个温度参数。

总损失 (L_{total}) 被表达为各个损失的加权和：

$$L_{\text{total}} = L_y + \lambda_d L_d + \lambda_{\text{MMD}} L_{\text{MMD}} + L_{\text{contrast}} \quad (17)$$

其中， λ_d 和 λ_{MMD} 是超参数，用于控制领域分类准确性与特征的领域不变性之间的权衡。

通过将 MMD 损失和对比学习整合到整体损失函数中，我们为模型提供了一个稳健的统计基础，以减少域方差，从而提高其在目标域上的适应性和性能。

IV. 实验

A. 数据集

在我们的实验中，我们通过整合两个著名的法律数据集的数据构建了一个综合数据集：LJP-MSJudge [?] 民法数据集和 CAIL-2018 [?] 刑法数据集。这个综合数据集作为我们实验分析的基础，涵盖了来自民法和刑法领域的多样化判决。LJP-MSJudge 数据集包括民事案件的详细集合，每个案件包含原告的诉求、法庭辩论记录和判决结果。CAIL-2018 数据集包括对刑事案件的事实描述、适用法律条款、指控以及处罚条款。

中国的民事案件往往经历多次审理，例如一审、二审和再审。然而，一审判决中的事实认定对于案件结果的确定是至关重要的。鉴于中国民事案件的改判率相对较低，我们主要参考 LJP-MSJudge 数据集中一审判决的案件事实和判决结果。

我们的主要任务是预测判决结果。因此，我们将与判决结果相关的文本分为两类：对于民事案件，结果要么是支持，要么是不支持原告的上诉；对于刑事案件，结果要么是有罪，要么是无罪。数据集的详细统计信息如表 ?? 所示。

对于训练，我们将 Adam 优化器的学习率设置为 10^{-3} ，批量大小设置为 128。每个模型训练 16 个 epoch 后，我们在验证集上选择最佳模型进行测试。

为了比较基线方法和我们方法的性能，我们选择了四个广泛用于多分类任务的指标，包括准确率 (Acc.)、宏观精确率 (MP)、宏观召回率 (MR) 和宏观 F1 值 (F1)。

为了广泛验证所提出模型的有效性，采用以下基线进行比较。

- TextCNN [?] 是一种卷积神经网络，基于预训练的词向量用于句子级分类任务进行训练。
- BERT [?] 是一种微调表示模型，已被应用于学习输入事实摘要的良好表示以进行判决预测。
- TOPJUDGE [?] 是一种用于 LJP 的拓扑多任务学习框架，它在有向无环图中形式化了子任务之间的显式依赖关系。
- MPBFN [?] 是一种用于 LJP 的多任务学习框架，具有多视角前向预测和后向验证。

然后我们比较了几个表现优异的 LLM 模型。这些模型目前是最受欢迎的，并且在理解和推理方面表现出最强的能力。

- GPT-4o 是一个最先进的语言模型，设计用于各种自然语言处理任务。
- Gemini-1.5-Flash 是一个高性能语言模型，专为快速准确的文本生成和理解而定制。
- DeepSeek-V3-Chat 是一个高级对话 AI 模型，针对对话和信息检索任务进行了优化。

TABLE I: 刑法任务的性能比较

Model	Acc.	MP	MR	F1
GPT-4o (Sep. 2024)	66.46	83.00	51.98	63.93
DeepSeek-V3-Chat	67.23	77.75	53.29	63.24
Gemini-1.5-Flash	68.77	78.87	55.59	65.22
BERT	43.51	50.00	6.30	11.14
TextCNN	43.47	41.74	49.92	45.47
TOPJUDGE	44.80	72.04	51.14	59.82
MPBFN	43.51	46.75	49.96	48.30
JurisCTC	76.59	75.92	85.75	80.54

TABLE II: 民法任务的性能比较

Model	Acc.	MP	MR	F1
GPT-4o (Sep. 2024)	52.04	51.80	51.83	51.81
DeepSeek-V3-Chat	46.65	49.09	49.40	49.25
Gemini-1.5-Flash	59.65	58.56	55.24	56.85
BERT	64.37	63.77	86.97	73.58
TextCNN	57.74	56.10	55.59	55.84
TOPJUDGE	67.31	75.07	62.51	68.22
MPBFN	62.89	64.33	58.49	61.27
JurisCTC	78.83	76.59	90.61	83.01

这些基准为评估所提出模型的性能提供了全面的基础。每个基准代表了不同的自然语言处理任务方法，提供了一组多样化的比较方法。

为了评估所提出模型的性能，我们从以下几个方面导出结果：

我们进行了一系列实验，以评估从刑法到民法的迁移学习性能，反之亦然。表 I 和表 II 显示了实验结果。

实验结果表明，所提出的模型 JurisCTC 在刑法和民法任务中均优于所有基准模型。具体而言，JurisCTC 在这两个领域都达到了最高的准确率、宏召回率和宏 F1 分数。

在刑法任务中，JurisCTC 的准确率达到 76.59%，显著高于表现最好的基线模型 Gemini-1.5-Flash，其准确率为 68.77%。同样，在民法任务中，JurisCTC 的准确率达到 78.83%，优于最佳基线模型 TOPJUDGE，其准确率为 67.31%。

JurisCTC 的优异表现可以归因于其有效利用从一个法律领域到另一个法律领域的迁移学习的能力，捕捉法律文本中复杂的关系和依赖性。这从高宏观召回率和宏观 F1 分数中可以看出，这表明 JurisCTC 不仅准确，而且在不同类别的预测中也保持一致。

在刑法任务中，GPT-4o、DeepSeek-V3-Chat 和 Gemini-1.5-Flash 的高 MP 分数表明，这些模型在做出正预测时非常有效地正确识别正实例。高 MP 表明这些模型在正预测方面比较保守，更注重精确性而非召回率。然而，它们在民法任务中的表现则不那么一致，表明这些模型可能更专注或更适合于刑法任务。

虽然 JurisCTC 在刑法和民法任务中表现出色，但仍有一些显著的改进空间。一个主要限制是其在刑法任务中的准确率与如 GPT-4o 等模型相比有所不及，其中 JurisCTC 的准确率为 75.92%，而 GPT-4o 为 83.00%。这表明 JurisCTC 尽管召回率较高，但可能会产生更多的错误阳性，这在需要高精度的场景中可能会有问题。其原因可能是像 GPT-4o 这样的模型从正面的民法数据中获得了更多的知识，但这些模型的准确率并不高，这表明本文提出的模型在学习正面数据和负面数据之间达到了平衡。

TABLE III: 对民法到刑法任务的性能消融

Model	Acc.	MP	MR	F1
BERT	43.51	50.00	6.30	11.14
BERT-UDA	65.64	71.26	65.64	68.33
BERT-CL	60.61	65.47	63.58	65.03
JurisCTC	76.59	75.92	85.75	80.54

TABLE IV: 刑法到民法任务的性能消融

Model	Acc.	MP	MR	F1
BERT	64.37	63.77	86.97	73.58
BERT-UDA	57.74	56.10	55.59	55.84
BERT-CL	65.87	66.43	85.48	74.37
JurisCTC	78.83	76.59	90.61	83.01

总体而言，结果验证了所提出模型在处理复杂法律判决预测任务中的有效性，展示了其在法律领域实际应用中的潜力。

1) 消融研究：在本节中，我们进行了一项消融研究，以评估不同模型在民法和刑法任务之间迁移时的性能。我们比较了基础的 BERT 模型及其变体，包括带有无监督数据增强的 BERT (BERT-UDA)、带有对比学习的 BERT (BERT-CL) 和 JurisCTC。结果如表 III 和 IV 所示。

表 III 展示了在民法数据上训练并在刑法任务上测试的模型性能。基线 BERT 模型的准确率为 43.51%，突显了由于民事和刑事法律语言之间存在显著差异而导致跨领域迁移的挑战。BERT-UDA 将性能提高到 65.64%，表明数据增强在适应新的领域中起到了帮助作用。BERT-CL 的准确率为 60.61%，显示对比学习在跨领域适应中提供了一些益处，尽管比 UDA 效果稍差。JurisCTC 以 76.59% 的准确率超越所有模型，并展现了卓越的精度、召回率和 F1 分数，证明了其在跨领域适应中的稳健性。

表 IV 展示了在刑法数据上训练并在民法任务上测试的模型性能。基线 BERT 模型的准确率为 64.37%，明显优于逆向学习，表明刑法特征可能更容易推广到民法环境。在这种情况下，BERT-UDA 的表现下降，准确率为 57.74%，这表明 UDA 的有效性具有上下文依赖。BERT-CL 的准确率提高到 65.87%，这表明对比学习在转移到民法时可能更有利。JurisCTC 再次以 78.83% 的准确率先，同时具有较高的精准率、召回率和 F1 分数，突显其在跨法律领域的优越适应性。

总体而言，这项消融研究的结果证明了我们提出的领域自适应模型 JurisCTC 在法律判决预测 (LJP) 领域的有效性。该模型始终优于其他变体，展示了其在不同法律领域中的稳健性和适应性。这验证了 JurisCTC 所采用的设计选择和领域适应策略，突出了其解决法律任务中跨领域转移独特挑战的潜力。这些发现强调了模型提高预测准确性和可靠性的能力，确认了其在 LJP 领域实际应用中的适用性。

我们提出了 JurisCTC，这是一种处理法律领域注释文本的知识迁移模型，可以实现从不同部门法中迁移知识。在公开判决数量显著减少和中文领域缺乏大规模注释法律数据集的背景下，我们证明了 UDA 可以从民法中学习法律应用的逻辑，并将其应用于新的刑法 LJP 任务中，显著提高了目标领域的预测准确性。同时，为了测试模型的泛化能力，我们还进行了从刑法领域学习法律应用逻辑并应

用于民法领域的实验，模型的性能显著提升。简言之，与传统模型相比，JurisCTC 有效解决了冗长复杂法律文本的挑战，显著提高了 LJP 任务的预测准确性，并增强了模型的泛化能力。

未来的研究将调查法律语言中有助于提升 JurisCTC 有效性的特定属性。我们计划研究对模型性能影响最大的语言特征，并探索不同的领域适应策略如何增强我们模型的能力。这一分析不仅将深化我们对领域特定适应的理解，还将改善人工智能系统在法律环境中的预测能力。