

基于Ernie胶囊网络的多标签 依赖法条推荐

刘德彬, 施运梅

北京信息科技大学计算机学院, 北京

收稿日期: 2024年5月23日; 录用日期: 2024年6月22日; 发布日期: 2024年6月30日

摘要

针对法律文书中存在多标签法条预测的问题, 为了提高法条推荐任务的准确性, 本文提出了一种基于标签依赖关系的法条推荐模型。该模型利用Ernie中的词感知预训练任务、句子感知预训练任务、语义感知预训练任务增强文本语义表示能力, 利用胶囊网络将低层胶囊的特征传递到高层胶囊, 利用动态路由更新获取层次丰富的语义特征, 再将文本向量输入到BiLSTM获取上下文特征信息。最后, 在法条标签预测层, 通过加权求和每一个时间步的法条标签向量, 捕获法条标签之间的依赖关系, 解决多标签法条推荐问题。实验结果证明, 在CAIL公开数据集上, 本文取得了82.3%的F1值, 有效地提高了法条推荐的准确性。

关键词

法律文书, 胶囊网络, 多标签类别

Multi-Label Dependency Statute Recommendation Based on Ernie Capsule Network

Debin Liu, Yunmei Shi

School of Computer Science, Beijing Information Science and Technology University, Beijing

Received: May 23rd, 2024; accepted: Jun. 22nd, 2024; published: Jun. 30th, 2024

Abstract

In order to improve the accuracy of law recommendation task, this paper proposes a law recom-

mentation model based on label dependency. The model uses the word perception pre-training task, sentence perception pre-training task and semantic perception pre-training task in Ernie to enhance the semantic representation ability of text. Capsule network is used to transfer the features of low-level capsules to high-level capsules, dynamic routing update is used to obtain hierarchical semantic features, and text vector is input into BiLSTM to obtain contextual feature information. Finally, in the law label prediction layer, the dependence relationship between law labels is captured by weighted summing the law label vector of each time step, and the multi-label law recommendation problem is solved. The experimental results show that 82.3% F1 value is obtained on the CAIL open data set, which effectively improves the accuracy of the recommendation.

Keywords

Legal Documents, Capsule Networks, Multi-Label Categories

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着我国法治建设的持续推进,越来越多的人有了法律维权意识,并通过法律手段合理维护社会公平正义,但随之而来的,是司法工作人员所需要受理的案件大幅增加,据中央纪委国家监委网报道,近年来,全国各级人民法院受理案件数量显著增加,其中,在2023年,全国各级人民法院共受理案件高达4557万件。因此,利用计算机及法律相关知识辅助司法工作人员办公,提高案件审理的效率,便极具意义。法律案件的审理过程,其每个环节都离不开相应的法律文书,包括检察文书、公证文书、仲裁文书等。这类法律文书是详尽地记录着案件的信息,因此面向法律文书的自然语言处理研究便拥有着广阔的前景。法条推荐是面向法律文书的自然语言处理任务中的重要组成部分,在传统的人工司法审理过程,法官需要自行翻阅大量法律法规,并从中挑选适用的法律条文作为依据,以此确定审理的结果。考虑到我国的成文法种类复杂繁多,在一个案件的审判过程,可能涉及多个不同的法律法规,司法工作人员需要阅读并熟记大量法律法规,并结合法律案件作出判断,引出相关的法律条文,这个过程不仅耗费了大量的时间和精力,还可能导致“同案不同判,法条适用不统一”的情况,这严重影响了司法的公平正义。针对上述情况,本文提出了基于深度学习的法条推荐方法,旨在利用人工智能技术,给出智能法条推荐方法,优化司法决策过程,提升法律的质量和效率。

法条推荐的主要目的是让机器可以在阅读相关文书的案情描述后,预测本案所涉及的法律法规。根据采用方法的不同,当前方法分为基于分类任务的法条推荐和基于语义匹配的法条推荐。

基于分类任务的法条推荐方法是将相关法律文书转换为文本,法律条文转换为法条标签,利用特征抽取的方式,构建法条分类器,实现法条推荐。从特征抽取的角度看,又可分为基于机器学习的特征抽取及基于深度学习的特征抽取。

机器学习方法需要通过人工方式,提取字、词、短语的浅层语义特征。在早些时候,文本分类研究大多数是基于统计分析的方法[1][2],该方法需要专业人员标注,无法适应不同背景下的法条推荐。随着机器学习的不断研究与深入,支持向量机、逻辑回归、朴素贝叶斯等分类逐渐应用于不同领域,如 Liu [3]

使用基于 SVM 的方法实现初步法条文本分类。但由于从字、词、短语级别提取出的浅层语义特征, 难以区分内容相似、易混淆的法律条文, 因此基于机器学习的特征提取方式, 存在领域复用性差, 模型泛化能力低, 特征信息提取不够充分的问题。

近年来, 受到深度学习领域在图像音频领域取得成功所带来的影响, 基于深度学习的方法也逐渐应用于 NLP 中的分类任务, 如 Luo [4] 和 Jiang [5] 使用双向门控循环单元的方法对法条进行分类。在基于深度学习的方法上, 研究者通过利用 CNN、RNN 等神经网络自动获取文本特征表示, 从而减少了复杂的人工特征工程, 提高了法条推荐的准确率以及效率。如张虎[6]提出基于卷积神经网络的模型, 解决法律判决文书中的法条推荐问题。将 CNN 神经网络应用于法条推荐分类任务, 能从全局信息中提取局部特征, 但无法解决法律文书中存在的长距离依赖关系问题, Zhang [7] 针对案情描述中存在的长距离依赖关系特征, 使用等长卷积和最大池化的 DPCNN 模型, 通过重复执行卷积循环块, 捕获法律文书中存在的长距离依赖关系问题。程豪[8]使用两个 CNN 分类器实现分层预测, 通过卷积层抽取文本特征, 最大池化层选择主要特征, 提高神经网络并行能力, 捕获长距离特征。

基于分类的法条推荐方法利用深度学习进行特征提取, 解决了使用机器学习进行特征提取时, 语义特征提取不充分、领域复用性差、泛化能力低的问题, 但基于文本分类模式的深度学习方法, 仅考虑了案情描述的文本特征提取, 并未充分考虑法条的语义信息。针对上述问题, 基于匹配的方法将法条分类任务转换为语义匹配任务, 通过计算案情描述和法律条文的匹配程度, 实现法律条文的推荐。殷玥[9]提出基于提出一个新的融合记忆网络和双层注意力机制的法条推荐模型。2018 年后, 随着预训练模型 BERT 的大规模应用, 唐光远[10]等人通过融入预训练模型, 提出基于 BERT 预训练模型对法条知识和案件描述分别进行表征, 并针对 RNN 中存在的梯度消失问题, 使用双向 LSTM 对案件描述文本进行特征提取, 然后基于注意力机制提取融合法条知识的案件描述文本特征, 最终实现智能法条推荐, 该方法在法研杯的相关数据集上, 法条推荐的 F1 值有了显著的提升。李琳[11]等人提出了基于深度语义匹配的法条推荐方法, 该方法包含局部语义匹配模块和全局语义推荐模块, 分别设计双向 Transformer 卷积网络模型和基于回归树的推荐模型, 在理解文本序列的同时, 关注与法条匹配学习相关的局部语义特征, 增强法条推荐的准确率和可解释性。丁娜[12]提出一种双向注意力文本关键词匹配的法条推荐模型, 该模型通过提取案情描述和法条两者的关键词序列, 融合关键词序列的匹配特征, 进行法条推荐, 在裁判文书数据集上, 与 BERT 模型相比, 该法条推荐模型在 F1 评价指标上, 平均有了 3.58% 的提升。

然而, 基于匹配的法条推荐任务虽然构建了案情描述与法条语义之间的联系, 提高了信息的利用率, 但针对多法律条文的多标签任务, 忽略了标签之间的依赖关系, 无法找准并找全和案情描述相关的所有法律条文。基于此, 本文提出了一种基于匹配并通过标签之间的依赖关系进行法条推荐的模型。本文方法的主要贡献如下:

- 1) 针对法律文书语义复杂的问题, 本文利用预训练模型 Ernie 中的词感知、结构感知、语义感知预训练任务, 提高模型针对法律文书的语义理解及表示能力。
- 2) 针对法律文书文本较长且存在不同层次结构特征的情况, 本文利用胶囊网络(Capsule-Network, CapsNet)和 BiLSTM, 提高模型的上下文及层次结构特征捕获能力。
- 3) 针对法条推荐中存在多标签问题, 本文对不同时间步的标签向量进行加权求和, 综合考虑标签的整体信息, 建立起标签依赖关系, 避免误差传递问题, 提高法条推荐效果。

2. 模型方法

本文提出基于 Ernie 胶囊网络的多标签依赖法条推荐模型(Multi-label Dependency Statute Recommendation Based on Ernie Capsule Network, MDSR-ECN), 其模型结构如图 1 所示。

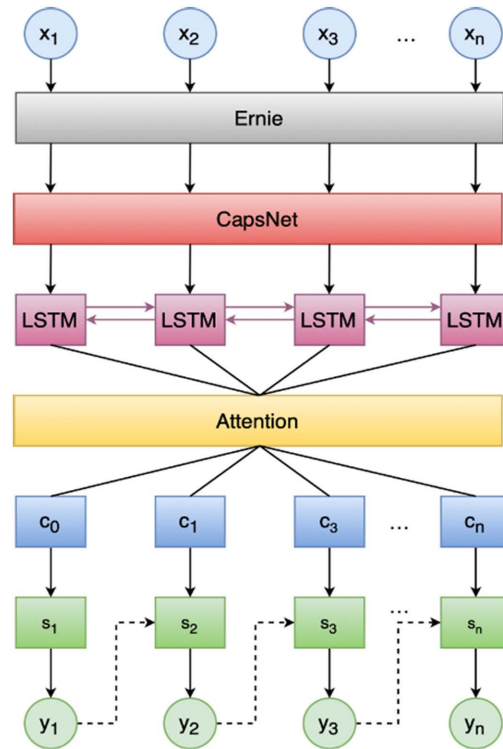


Figure 1. MDSR-ECN model structure
图 1. MDSR-ECN 模型结构图

基于 Ernie 胶囊网络的多标签依赖法条推荐模型 MDSR-ECN 主要分为基于 Ernie 的语义编码层，基于 CapsNet-BiLSTM 的特征提取层以及基于标签依赖关系的法条预测层。

2.1. 基于 Ernie 的语义编码

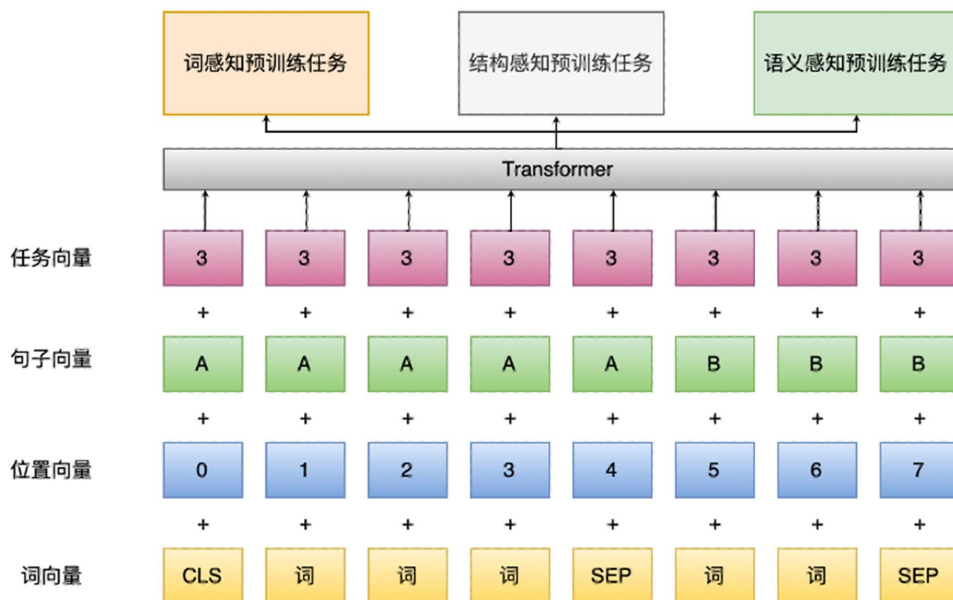


Figure 2. Ernie structure
图 2. Ernie 结构图

Ernie 是百度提出的一种基于 Transformer 框架的预训练模型。在 Ernie 预训练任务中, 主要分为词感知预训练任务、结构感知预训练任务和语义感知预训练任务。在词感知预训练任务中, 引入短语掩码和命名实体掩码来预测序列的信息, 通过标记和文档关系的预测, 使得模型拥有复杂文本关键词提取的语义理解能力。在结构感知预训练任务中, 通过句子重排和句子距离计算的方式预测句子之间的关系, 所谓句子重排和句子距离计算, 即将排列好的片段, 随机拆分成多个片段, 再重新排序, 在此过程中, 模型学习到句子是否相邻的结构信息。在语义感知预训练任务中, 根据文章上下文的关联词, 来提取句子之间的语义关系, Ernie 结构中 Transformer 编码器由注意力层和前馈神经网络组成, 其结构如图 2 所示。

2.2. 基于 CapsNet-BiLSTM 的特征提取

CapsNet 是一种神经网络结构, 最早由 Geoffrey Hinton [13] 等人提出, CapsNet 旨在克服传统卷积神经网络在处理图像数据时, 池化过程无法保留原始信息的问题。Zhao [14] 提出模型 Capsnet-A 和 Capsnet-B 并将模型首次应用于文本分类领域, 实验结果证明, 针对多标签文本任务, 胶囊网络的效果要优于 CNN 和 LSTM 模型。

在法律文书法条推荐任务中, 文本数据通常具有层次结构和语义关系, CapsNet 通过动态路由机制可以学习到不同层次之间的关系, 有助于更好地理解文本中的层次信息, 并且 CapsNet 通过胶囊实现信息传递, 可以减少信息丢失, 提高文本的语义特征提取能力。

在结构上, CapsNet 主要由低层次胶囊网络层、动态路由层及高层次胶囊网络层组成。CapsNet 结构如图 3 所示。

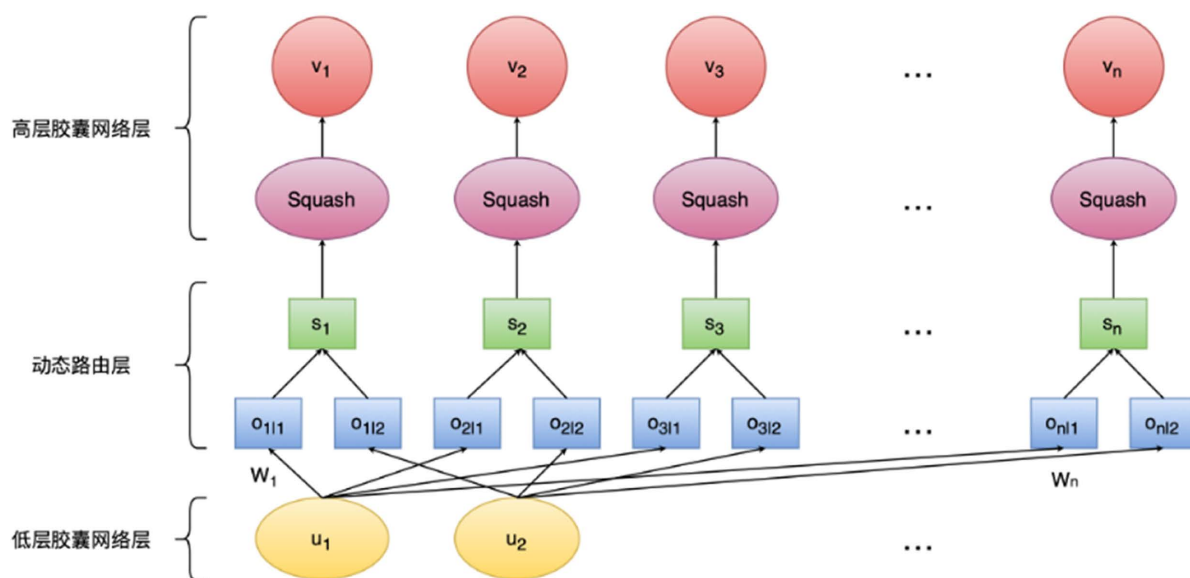


Figure 3. CapsNet structure

图 3. CapsNet 结构图

将 Ernie 编码后的向量作为低层胶囊输入, 经过矩阵 W 的线性变换后得到向量 o_{ji} , 公式如下所示。

$$o_{ji} = W_{ij} u_i \quad (1)$$

为了将低层胶囊信息传入到高层胶囊当中, 需要判断低层胶囊与高层胶囊的耦合程度, 通过计算低层 i 胶囊和高层 j 胶囊的相似度的方式得到胶囊之间的耦合程度, 公式如下所示。

$$c_{ij} = \frac{\exp(b_{ij})}{\sum_K \exp(b_{ij})} \quad (2)$$

将耦合程度与经过 W 矩阵线性变换的向量 o_{ji} 相乘, 经过加权求和后得到高层胶囊向量输出, 计算公式如下所示。

$$s_{ij} = \sum_i c_{ij} o_{ji} \quad (3)$$

为了得到向量归一化表示, 同时保留向量方向信息, 对高层胶囊进行压缩, 计算公式如下所示。

$$v_j = \frac{\|s_j\|^2 s_j}{1 + \|s_j\|^2 \|s_j\|^2} \quad (4)$$

在胶囊网络动态路由机制中, 需要动态更新迭代胶囊特征信息, 计算公式如下所示。

$$b_{ji} \leftarrow b_{ji} + o_{ji} \cdot v_j \quad (5)$$

经过胶囊网络, 得到了局部特征与层次特征的向量表示, 将向量输入到 BiLSTM 层, 进一步学习文本的上下文特征。BiLSTM 层由两个 LSTM 网络组成。计算公式如下所示:

$$\bar{h}_t = \overline{\text{LSTM}}(\bar{h}_{t-1}, x_t) \quad (6)$$

$$\bar{h}_t = \overline{\text{LSTM}}(\bar{h}_{t-1}, x_t) \quad (7)$$

BiLSTM 分别计算前项隐向量 h_{t-1} 产生的新隐向量 h 和后项隐向量 h_{t-1} 产生的新隐向量 h , 并将两个新的隐向量拼接得到向量 h 。

2.3. 基于标签依赖关系的法条预测

由于法律文书中文本的不同部分对不同标签的重要性是不同的, 同时多标签之间也具有相关性[15]。为了解决这些问题, 本文使用 MDSR-ECN 模型进行法条预测, 该模型通过捕捉标签之间的相关性, 提高多标签法条推荐的效果。由于法律文书法条推荐任务是多标签任务, 即每个法律文书的案情描述可以有多个标签, 因此模型需要找到最大化条件概率 $p(y)$ 的最佳标签 $(y|x)$ 序列[16] [17], 其计算公式如下所示。

$$p(y|x) = \prod_{i=1}^n p(y_i | y_1, y_2, \dots, y_{i-1}, x) \quad (8)$$

为了使得模型在预测不同法条时聚焦案情描述的不同部分, 首先计算案情描述向量的权重, 计算公式如下所示。

$$e_{ii} = v_a^T \tanh(W_a s_t + U_a h_t) \quad (9)$$

$$n_{ii} = \frac{\exp(e_{ii})}{\sum_{j=1}^m \exp(e_{ij})} \quad (10)$$

$$c_t = \sum_{i=1}^m s_{ii} h_t \quad (11)$$

其中, W_a 和 U_a 是参数矩阵, s_t 是时间步 t 下的隐藏状态, c_t 是权重分配后的案情描述向量表示。

在得到权重分配的案情描述向量后, 需要对向量进行解码, 计算公式如下所示。

$$s_t = \text{LSTM}(s_{t-1}, [g(y_{t-1}); c_{t-1}]) \quad (12)$$

其中, y_{t-1} 是时间步 $t-1$ 下, 法条标签的概率分布, $g(y_{t-1})$ 表示 y_{t-1} 时概率最高的法条标签向量表示与每一个时间步的法条标签向量表示的加权求和, 计算公式如下所示。

$$\bar{e} = \sum_{i=1}^L y_{t-1}^{(i)} e_i \quad (13)$$

$$H = W_1 e + W_2 \bar{e} \quad (14)$$

$$g(y_{t-1}) = (1-H) \odot e + H \odot \bar{e} \quad (15)$$

其中, y_{t-1} 表示时间步 $t-1$ 时的概率分布, e_i 表示第 i 个标签的向量表示, \bar{e} 表示时间步 t 时法条标签向量的求和, W_1 和 W_2 是参数矩阵, H 表示控制加权比例的门控单元。

使用概率最高的法条标签向量表示与每一个时间步的法条标签向量表示的加权求和的原因是, 当仅通过使用概率最高的法条标签向量时, 若时间步 t 的预测出现了错误, 则接下来的时间步也会得到一系列错误的标签分类, 这种情况, 称之为误差传递, 为了缓解此类问题, 通过两种情况下向量的加权求和, 综合考虑标签的整体信息, 减少由先前时间步中的错误预测造成的偏差, 使得模型能够更准确地预测标签序列。

利用解码后的隐藏状态 s_t 及案情描述向量 c_t 计算法条标签的概率分布, 计算公式如下所示。

$$o_t = W_o f(W_d s_t + V_d c_t) \quad (16)$$

$$y_t = \text{softmax}(o_t + I_t) \quad (17)$$

其中, W_o 、 W_d 、 V_d 表示参数矩阵, I_t 表示掩码向量, 表示若上一时间步输出了第 i 条法条标签, 则当前时间步 I_t 的第 i 个元素为无穷, 避免输出相同的法条标签。

本文将使用交叉熵函数作为损失函数, 计算公式如下所示。

$$\text{Loss} = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M (y_{ij} \cdot \log(p_{ij}) + (1-y_{ij}) \cdot \log(1-p_{ij})) \quad (18)$$

其中, N 表示样本数量, M 表示标签数量, y_{ij} 是第 i 个样本的第 j 个标签真实值, p_{ij} 是模型预测的第 i 个样本的第 j 个标签为正确的概率。

3. 实验

3.1. 实验数据集

本文使用中国法律智能技术评测 2018 法律文书作为数据集, 该数据集由北京大学、清华大学、中国科学院软件研究所和中国司法大数据研究院联合发布, 是第一个用于司法预测的大规模中国法律数据集。本文将该数据集按 7:2:1 的比例划分为训练集 11.2 万条, 测试集 3.2 万条, 验证集 1.6 万条, 作为本实验所用的数据集。数据集涉及法条共计 183 条。

3.2. 实验设计及结果分析

由于目前 CNN、RNN 及注意力机制为多标签文本分类的主流方向, 因此本文首先使用经典 CNN 模型 TextCNN [18], 将 CNN 与 RNN 进行融合的混合模型 TextRCNN [19], 使用多级扩张卷积和混合注意力机制的 SUBDC [20] 作为基线模型。

TextCNN 是 Yoon Kim 利用不同尺寸的卷积核提取文本中的特征信息, 捕获文本的局部相关性, 从而实现文本分类任务。TextRCNN 通过使用双向循环结构获取上下文信息, 同时结合最大池化层获取文本的关键特征信息, 实现了 CNN 和 RNN 的优势融合。

SUBDC 利用扩大感受野的方式, 捕获更大范围的语义信息, 同时通过跨语义单元的连接, 实现语义单元的信息传递, 从而提高多标签文本分类的效果。

基线模型和本文提出的 MDSR-ECN 模型的对比实验结果如表 1 所示。

Table 1. Comparison of experimental results table

表 1. 对比实验结果表

方法	Precision	Recall	F1
TextCNN	0.723	0.539	0.618
TextRCNN	0.719	0.687	0.703
SUBDC	0.813	0.778	0.795
MDSR-ECN	0.835	0.811	0.823

从实验结果看, TextCNN 通过卷积和池化操作, 能够有效地捕获文本中的局部特征和语义信息, 但由于卷积和池化操作会将每个词语特征视为相等重要, 而忽略了文本中的位置信息, 因此无法捕捉文本顺序特征, 同时由于使用固定的卷积核进行卷积操作, 对变长文本进行卷积时, 需要填充或截断, 这将导致部分信息的丢失。

TextRCNN 通过引入 RNN, 用于捕捉文本序列的上下文信息, 使得模型既能获得词语之间的局部信息, 又能获得文本的全局信息, 并且通过动态 RNN 的方式适应不同长度的文本输入, 避免了填充或截断带来的信息损失, 因此在实验结果上, 召回值 R 有了明显的提升, 这意味着减少信息损失后, TextRCNN 找到了更为齐全的正确标签。

在 SUBDC 中, 则使用插空跳跃膨胀卷积的方式, 避免池化操作导致的信息缺失, 同时, 膨胀卷积还可以解决长文本中的长距离依赖关系, 通过调整卷积中的感受野大小, 处理不同长度的文本分类, 另一方面, SUBDC 通过混合注意力的方式, 捕获不同层次的信息, 更好地提取文本词语级, 段落级信息, 从结果上看, 相较于 TextRCNN, 精确率 P 和召回值 R 都有了接近 10% 的提升。

针对法条推荐任务, 上述方法利用主流的 CNN 及 RNN 方法, 结合注意力机制, 在卷积操作上进行了优化, 在文本特征提取上效果显著。

本文提出的模型不仅考虑了文本特征提取的问题, 同时针对多标签分类问题, 进一步考虑标签之间的依赖关系, 提升文本和标签之间的语义关联及信息互补。同时, 一般文本分类方法, 忽略标签之间的依赖关系, 则会导致误差的传播, 即如果一个标签被错误地预测, 那么与它相关的标签也可能会受到影响。

针对误差传播的问题, 本文提出的模型 MDSR-ECN 首先利用 Ernie 编码, 提取丰富的文本语义信息, 再通过 CapsNet, 利用动态路由机制, 实现低层特征与高层特征的集合, 得到分层特征信息, BiLSTM 通过时间步信息传递实现上下文特征融合。标签预测的过程中, 对概率最高的法条标签向量表示与每一个时间步的法条标签向量表示的进行加权求和, 综合考虑标签的整体信息, 建立起标签依赖关系, 避免误差传递问题, 提高法条推荐效果。从实验结果上看, 准确率 P、召回值 R 和 F1 均优于基线模型, 证明了本文模型在法条推荐任务中的有效性及可行性。

考虑到本任务为多标签文本分类, 为进一步探究本文模型在不同标签数量下的效果差异, 首先统计数据集中不同标签数量的样本条数, 根据标签数量的不同, 将训练集、验证集、测试集划分为 3 个类别, 依次表示标签数量为 1、标签数量为 2、标签数量大于等于 3, 其结果如表 2 所示。

Table 2. Data set
表 2. 数据集

标签数量	训练集	验证集	测试集
1	74,863	14,147	27,484
2	23,537	2009	3672
≥3	13,598	975	1352

下面给出不同标签数量下的案情描述样例，案情描述一如图 4 所示，该案情对应法条标签为刑法第 234 条，即标签数量为 1。

昌宁县人民检察院指控，2014 年 4 月 19 日下午 16 时许，被告人段某驾拖车经过鸡飞乡澡塘街子，时逢堵车，段某将车停在“冰凉一夏”冷饮店门口，被害人王某的侄子王 2 某示意段某靠边未果，后上前敲打车门让段某离开，段某遂驾车离开，但对此心生怨愤。同年 4 月 21 日 22 时许，被告人段某酒后与其妻子王 1 某一起准备回家，走到鸡飞乡澡塘街富达通讯手机店门口时停下，段某进入手机店内对被害人王某进行吼骂，紧接着从手机店出来拿得一个石头又冲进手机店内朝王某头部打去，致王某右额部粉碎性骨折、右眼眶骨骨折。经鉴定，被害人王某此次损伤程度为轻伤一级

Figure 4. Case description figure I

图 4. 案情描述图一

案情描述二如图 5 所示，该案情对应法条标签为刑法第 348 条，刑法第 356 条，即标签数量为 2。

公诉机关指控，2016 年 6 月 22 日 14 时许，民警在被告人冯某租赁的重庆市南岸区南坪东路“辰光华府”小区 1 栋 28-3 房间内抓获冯某，并当场查获甲基苯丙胺（冰毒）疑似物一袋、K 粉疑似物一袋、甲基苯丙胺片剂（俗称“麻古”）疑似物一袋。经称重、检验，甲基苯丙胺疑似物净重 25.85 克，从中检出甲基苯丙胺成分；K 粉疑似物净重 1.46 克，从中检出氯胺酮成分；甲基苯丙胺片剂疑似物净重 23.10 克，从中检出甲基苯丙胺和咖啡因成分。

Figure 5. Case description figure II

图 5. 案情描述图二

案情描述三如图 6 所示，该案情对应法条标签为刑法第 383 条、第 385 条、第 386 条，即标签数量为 3。

2008 年 1 月至 2008 年 11 月，被告人姜某在担任某公司副总经理及腾退拆迁安置领导小组副组长期间，利用负责该公司拆迁腾退工作的职务便利，收受被拆迁人某甲仓储有限公司、某物流有限公司、某乙仓储有限公司及杨某（均另案处理）等单位、个人所送贿赂款共计人民币 8 万元，为他人谋取利益。具体犯罪事实如下：
1、2008 年 1 月至 2 月期间，被告人姜某在担任某公司副总经理及腾退拆迁安置领导小组副组长期间，利用负责该公司拆迁腾退工作的职务便利，收受被拆迁单位某甲仓储有限公司由肖某经手的贿赂款人民币 2 万元，为该公司谋取利益。
2、2008 年 8 月，被告人姜某在担任某公司副总经理及腾退拆迁安置领导小组副组长期间，利用负责该公司拆迁腾退工作的职务便利，收受被拆迁单位原名某物流有限公司由李某乙经手的贿赂款人民币 2 万元，为该公司谋取利益。
3、2008 年 9 月，被告人姜某在担任某公司副总经理及腾退拆迁安置领导小组副组长期间，利用负责该公司拆迁腾退工作的职务便利，收受被拆迁单位某乙仓储有限公司由吴某经手的贿赂款人民币 2 万元，为该公司谋取利益。
4、2008 年 10 月至 11 月期间，被告人姜某在担任某公司副总经理及腾退拆迁安置领导小组副组长期间，利用负责该公司拆迁腾退工作的职务便利，收受被拆迁人杨某所送的贿赂款人民币 2 万元，为其谋取利益。

Figure 6. Case description figure III

图 6. 案情描述图三

不同标签数量的对比试验结果如表 3 所示。

Table 3. Table of experimental results for comparison of different label quantities
表 3. 不同标签数量对比实验结果表

标签数量	方法	Precision	Recall	F1
1	TextCNN	0.723	0.539	0.618
	TextRCNN	0.719	0.687	0.703
	SUBDC	0.813	0.778	0.795
	MDSR-ECN	0.835	0.811	0.823
2	TextCNN	0.674	0.603	0.637
	TextRCNN	0.689	0.652	0.670
	SUBDC	0.756	0.676	0.714
	MDSR-ECN	0.782	0.701	0.739
≥3	TextCNN	0.523	0.387	0.445
	TextRCNN	0.546	0.429	0.475
	SUBDC	0.580	0.443	0.502
	MDSR-ECN	0.632	0.511	0.565

通过上述三组实验可以得知, 当数据集中的标签数量较少时, MDSR-ECN 通过编码层, 提取出丰富的文本特征信息, 提高模型的精确率 Precision, 当数据集中的标签数增多时, MDSR-ECN 通过标签依赖关系的解码层, 提高模型的召回值 Recall, 综合提升模型的 F1。因此, 上述三组实验进一步说明, MDSR-ECN 可以很好地解决法律文书中法条推荐的多标签文本分类问题。

4. 结束语

针对分析法律文书中法条推荐任务存在多标签类别的特点, 本文提出了基于预训练模型 Ernie, 融合 CapsNet-BiLSTM 的标签依赖法条推荐模型 MDSR-ECN。针对法律文书篇幅可能较长, 存在长距离文本依赖的问题, Ernie 通过词感知预训练任务、结构感知预训练任务和语义感知预训练任务, 提高模型的文本理解能力。同时, 由于法律文书篇幅较长, 存在词语层次、段落层次、文档层次的结构, 利用 CapsNet, 通过聚类的思想, 代替池化操作, 实现不同层次的特征整合, 利用 BiLSTM, 提高模型的上下文信息提取能力。针对不同法条标签之间存在依赖的问题, 在标签预测的过程中, 对概率最高的法条标签向量表示与每一个时间步的法条标签向量表示的进行加权求和, 综合考虑标签的整体信息, 建立起标签依赖关系, 避免误差传递问题, 提高法条推荐效果。实验结果表明, 该模型在法律文书的法条推荐任务上, F1 值达到了 82.3%, 证明了模型的有效性和可行性。

参考文献

- [1] Lauderdale, B. and Clark, S. (2012) The Supreme Court's Many Median Justices. *American Political Science Review*, **106**, 847-866.
- [2] Segal, J.A. (1984) Predicting Supreme Court Cases Probabilistically: The Search and Seizure Cases, 1962-1981. *American Political Science Review*, **78**, 891-900. <https://doi.org/10.2307/1955796>
- [3] Liu, Y., Chen, Y. and Ho, W. (2015) Predicting Associated Statutes for Legal Problems. *Information Processing & Management*, **51**, 194-211. <https://doi.org/10.1016/j.ipm.2014.07.003>

- [4] Luo, B., Feng, Y., Xu, J., Zhang, X. and Zhao, D. (2017). Learning to Predict Charges for Criminal Cases with Legal Basis. *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, Copenhagen, 7-11 September 2017, 2727-2736. <https://doi.org/10.18653/v1/d17-1289>
- [5] Jiang, X., Ye, H., Luo, Z.C., et al. (2018) Interpretable Rationale Augmented Charge Prediction System. *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstration*, Santa Fe, 20-26 August 2018, 146-151.
- [6] 张虎, 王鑫, 王冲, 等. 面向法律裁判文书的法条推荐方法[J]. 计算机科学, 2019, 46(9): 211-215.
- [7] Zhang, H., Wang, X., Tan, H., et al. (2019) Applying Data Discretization to DPCNN for Law Article Prediction. *Natural Language Processing and Chinese Computing: 8th CCF International Conference*, Dunhuang, 9-14 October 2019, 459-470.
- [8] 程豪, 张虎, 崔军, 等. 基于分层学习的易混淆法条预测[J]. 计算机工程与设计, 2020, 41(1): 278-282.
- [9] 殷玥. 基于神经网络的法条推荐算法研究[D]: [硕士学位论文]. 南京: 南京师范大学, 2019.
- [10] 唐光远, 郭军军, 余正涛, 等. 基于 BERT 与法条知识驱动的法条推荐方法[J]. 山东大学学报(理学版), 2021, 56(11): 24-30.
- [11] 李琳, 段围, 周栋, 等. 基于深度语义匹配的法律条文推荐方法[J]. 软件学报, 2022, 33(7): 2618-2632.
- [12] 丁娜, 刘鹏, 邵惠鹏, 等. 双向注意力文本关键词匹配法条推荐[J]. 北京大学学报(自然科学版), 2024, 60(1): 79-88.
- [13] Sabour, S., Frosst, N. and Hinton, G.E. (2017) Dynamic Routing between Capsules. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, 4-9 December 2017, 3859-3869.
- [14] Zhao, W., Ye, J., Yang, M., et al. (2018) Investigating Capsule Networks with Dynamic Routing for Text Classification.
- [15] 王婉臻, 饶元, 吴连伟, 等. 基于人工智能的司法判决预测研究与进展[J]. 中文信息学报, 2021, 35(9): 1-14.
- [16] Bao, Q., Zan, H., Gong, P., et al. (2019) Charge Prediction with Legal Attention. *Natural Language Processing and Chinese Computing: 8th CCF International Conference*, Dunhuang, 9-14 October 2019, 447-458.
- [17] Yang, P., Sun, X., Li, W., et al. (2018) SGM: Sequence Generation Model for Multi-Label Classification. *Proceedings of the 27th International Conference on Computational Linguistics*, Santa Fe, 20-26 August 2018, 3915-3926.
- [18] Kim, Y. (2014) Convolutional Neural Networks for Sentence Classification. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, 25-29 October 2014, 1746-1751. <https://doi.org/10.3115/v1/d14-1181>
- [19] Lai, S., Xu, L., Liu, K. and Zhao, J. (2015) Recurrent Convolutional Neural Networks for Text Classification. *Proceedings of the AAAI Conference on Artificial Intelligence*, **29**, 2267-2273. <https://doi.org/10.1609/aaai.v29i1.9513>
- [20] Lin, J., Su, Q., Yang, P., Ma, S. and Sun, X. (2018) Semantic-Unit-Based Dilated Convolution for Multi-Label Text Classification. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, Brussels, 31 October-4 November 2018, 4554-4564. <https://doi.org/10.18653/v1/d18-1485>