

基于深度语义匹配的法律条文推荐方法*

李琳¹, 段围¹, 周栋², 袁景凌¹

¹(武汉理工大学 计算机科学与技术学院, 湖北 武汉 430070)

²(湖南科技大学 计算机科学与工程学院, 湖南 湘潭 411201)

通信作者: 李琳, E-mail: cathylin@whut.edu.cn



摘要: 法律条文(简称法条)是司法量刑的主要依据, 法律条文的精准推荐, 能够辅助提高法律智能判决的质量. 目前, 主流的法条推荐模型是将有限数量的法条当作类别标签, 采用分类的思想, 根据法律文书的案例描述将其归类到相关的法条. 但是法条作为法律规范的文字表述形式, 现有的分类方法简单将其作为类别标签的索引编号, 导致对其语义信息利用不足, 影响了推荐质量. 针对此问题, 研究将主流的法条推荐方法从分类模型转化为语义匹配模型, 提出了基于深度语义匹配的法条推荐方法(DeepLawRec). 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 卷积神经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. 在公开数据集上的实验结果表明, DeepLawRec 方法在推荐质量上优于传统的文本分类以及经典的语义匹配方法, 并进一步探讨了如何分析和判决推荐结果.

关键词: 法律条文推荐; 语义匹配; 卷积神经网络; 可解释性

中图法分类号: TP18

中文引用格式: 李琳, 段围, 周栋, 袁景凌. 基于深度语义匹配的法律条文推荐方法. 软件学报, 2022, 33(7): 2618–2632. <http://www.jos.org.cn/1000-9825/6287.htm>

英文引用格式: Li L, Duan W, Zhou D, Yuan JL. Law Article Recommendation Approach Based on Deep Semantic Matching. Ruan Jian Xue Bao/Journal of Software, 2022, 33(7): 2618–2632 (in Chinese). <http://www.jos.org.cn/1000-9825/6287.htm>

Law Article Recommendation Approach Based on Deep Semantic Matching

LI Lin¹, DUAN Wei¹, ZHOU Dong², YUAN Jing-Ling¹

¹(School of Computer Science and Technology, Wuhan University of Technology, Wuhan 430070, China)

²(School of Computer Science and Engineering, Hunan University of Science and Technology, Xiangtan 411201, China)

Abstract: Law articles are the main basis of legal judgment and related law article recommendation can improve the quality of legal judgment prediction (LJP). Currently, the state-of-the-art methods belong to supervised classification model with the finite law articles as discrete class labels, which, however, has the downside that the semantic information of law articles may be underused. This observation implies that the quality of law article recommendation can be further enhanced. To solve this problem, this study proposes a deep semantic matching based law article recommendation approach (DeepLawRec) by transforming the traditional classification solution to pairwise matching learning. The proposed DeepLawRec includes local semantic matching module and global semantic recommendation module with a bi-transformer convolutional neural network and regression tree based recommendation, respectively. It can not only extract the key semantic features from the fact descriptions of cases, but also learn their related and local semantic features with a given law article. Moreover, with the help of regression tree, the recommendation results could be interpreted. The experimental results on a public dataset show the proposed DeepLawRec approach can improve the quality of recommendation and outperform the state-of-the-art techniques in terms of precision and interpretability.

Key words: legal article recommendation; semantic matching; convolution neural network; interpretability

* 基金项目: 国家自然科学基金(61876062)

收稿时间: 2020-07-27; 修改时间: 2020-09-30; 采用时间: 2020-12-18

1996年6月17日,《全国法院计算机信息网络建设规划》由中国最高人民法院发布,信息化自动化在法律领域得到建设。到目前为止,法院审判文书和相关的证据信息等已被有效地保存。2016年11月16日,第3届世界互联网大会“智慧法院暨网络法制”论坛在浙江乌镇举办,“数据法院”系统的建设能够服务法官审判,主要通过“类案推送”,为法官提供类案参考,减轻工作量,避免同案不同判的情况发生;同时,也能从大量数据中发现逻辑错误、遗漏诉讼请求等问题,以提高法律文书的文本质量。

互联网上,中国裁判文书网公开了由法官编写的裁判文书,数量达到千万且表述规范,属于质量高的领域文本数据。此外,每个文书有较为详细的标注,比如,裁判文书中会记录时间、发生的地域、法官姓名、代理律师姓名,涉及的法条内容、罪名和量刑结果等。这些规范的标注信息为NLP技术(natural language processing)在法律领域中的利用提供了便利条件。虽然网络上的法律文本质量高,但是单篇内容的文字长度偏长,是典型的长序列文本。而且由于不同的控诉罪名所对应的案例描述内容在文字上有可能十分类似,不易区分,因此一般需要法律专业人士的解读。如何从长序列的法律文本数据中实现高质量的法律判决预测(legal judgment prediction, LJP),在审判前提供法律判决结果,对辅助法官判案和协助律师维护当事人的合法权益具有重要的研究价值和实际应用意义。

在LJP任务中,相关法律条文推荐任务为法官的判决提供法律依据,同时也支撑智慧法院的犯罪预测等关键问题。智能化法条推荐的研究从2005年开始^[1-3],近年来,主流的推荐方法是建立文本分类模型,将法条看成类别标签,进而预测法律诉讼文书被划分到多个法条的概率^[4-6]。同时,也有研究者关注将推荐的相关法条用于增强罪名预测任务的准确率和可解释性^[2,3,7]。从内容上来看,法条作为法律规范的文字表述形式,是在诉讼辩论中直接引用的法律法规的原文,但是目前的分类模型大多将其作为类别标签的索引编号,未能充分挖掘法条所蕴含的语义信息,影响了推荐质量。

如图1所示,左图给出了对于一篇案例描述(相关联的法律条文为“盗窃”)而言关键词的分布情况,颜色越深代表重要程度越大。可以看到,该文书的案例描述内容高度依赖于词“人财物”,与法条“盗窃”虽有关系,但是语义上也与其他法条相关,比如“故意毁坏财物”。当采用分类模型时,由于大多方法仅将不同的法条当作类别,未在分类过程中考虑法条本身的内容,因此更容易将该案例描述判断成属于其他法条,而不是“盗窃”。本文针对左图的问题,将引入“盗窃”法条的文本描述来约束案件描述和法条的匹配关系学习过程(如图1右图所示),即对法条的文本内容进行语义特征表示学习。通过该法条案例描述相对于此法条的关键词分布情况,可以看到,案例描述内容同时高度依赖于词“非法占用”和“扒窃”等。由此可见,加入法律条文内容有助于对案例描述的语义理解,从而提升法条推荐的精度。

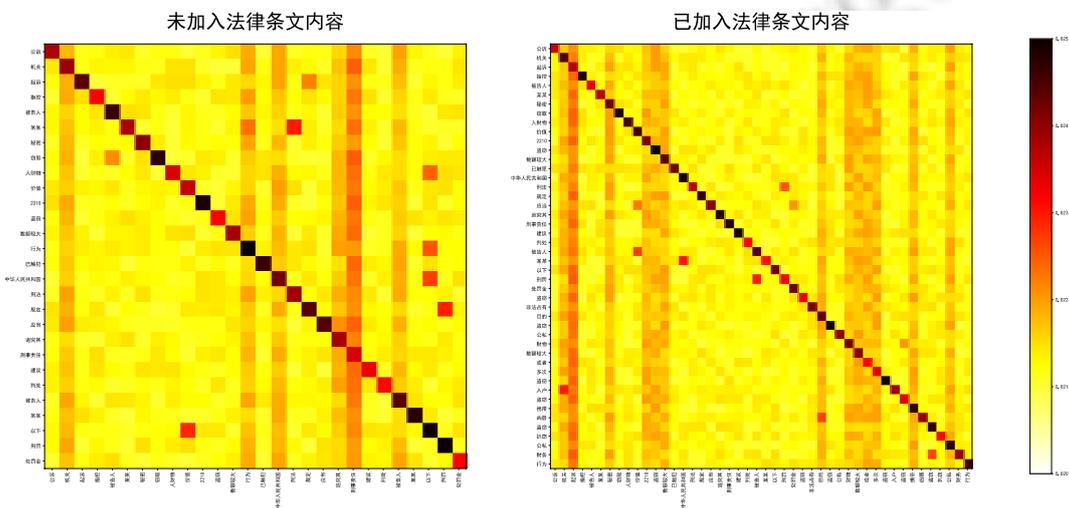


图1 案例描述和法律条文的语义关联情况示例

基于对图 1 例子的观察, 本文研究如何将主流的法条推荐方法从分类模型转化为语义匹配模型, 提出了基于深度语义匹配的法条推荐方法(deep semantic matching based approach for law article recommendation, DeepLawRec). 该方法包含局部语义匹配和全局语义推荐两个模块, 同时考虑法条和案例描述的文本内容, 通过语义特征学习计算两者的匹配度, 进一步以匹配度为引导实现推荐, 并提高结果的可解释性.

DeepLawRec 方法主要包括以下 4 个步骤: (1) 首先, 在匹配模块中使用双向 Transformer 模型, 从相对稀疏的文本向量中提取出语义特征表示, 对序列文本的相关性进行特征学习; (2) 再借由卷积层从语义相关的特征表示中选择性提取重要局部特征, 再经过池化层和激活层, 压缩所提取得到的特征, 得到相对简练的高维语义特征; (3) 接下来通过全连接层计算语义匹配度, 得到输入的案例描述与若干法条的匹配概率分布; (4) 最后, 在推荐模块中, 联合案例描述的特征表示和法条的匹配概率分布, 通过回归树学习最终的推荐概率分布, 增强法条推荐的准确率和可解释性. 在公开数据集上的实验结果表明, 本文提出的 DeepLawRec 模型的推荐质量优于传统的文本分类和经典的语义匹配模型, 同时有助于分析和判读推荐结果.

本文的主要贡献总结如下:

- (1) 提出将法律条文推荐任务的解决思路从传统的分类方法变化成语义匹配方法, 从而便于更好地引入法条文本自身的语义信息.
- (2) 提出了基于深度语义匹配的法条推荐方法(DeepLawRec). 该方法从案例描述中提取与法条语义匹配的重要局部特征, 再联合案例描述的句子级语义特征和匹配概率分布, 通过回归树输出最终推荐结果, 并能对推荐过程给予一定的解释.
- (3) 在公开的中国刑事法律文书数据集上对比了主流的分类和语义匹配模型, 实验结果表明, 本文提出的方法提升了法律条文推荐任务的准确率, 并能够给出如何从若干个法条中进行选择的过程, 达到对推荐结果的判读.

本文首先讨论法律判决预测和语义匹配的国内外研究现状. 接下来介绍本文所提出的方法以及各个模块的技术细节. 然后给出实验数据集、评价指标和实验环境, 报告和讨论实验的结果并分析了样例. 最后对全文工作进行总结, 并探讨未来研究的方向.

1 相关工作

法律判决预测(LJP)是将人工智能技术应用在法律审判领域的典型任务, 主要是训练一种能根据法律文本自动预测判决结果(如相关法律条文、控诉罪名、刑期等)的机器法官. 一个良好运行的 LJP 系统不仅可以为普通民众提供便利, 也可以为专业人士, 如律师和法官提供参考. 接下来先介绍以分类为主流的 LJP 研究现状, 然后再讨论语义匹配模型的特点.

1.1 法律判决预测

早期 LJP 的研究通常集中于如何利用数学和统计方法在某些特定的场景中构建 LJP 系统^[8-13]. 机器学习技术在社会各个领域的应用随着数据和硬件条件的发展, 也为提高 LJP 质量提供了实施途径. 已有的研究工作通常将 LJP 任务, 比如罪名预测和刑期预测, 形式化表示为文本分类问题, 并关注如何从法律文本和额外的信息中计算有效的浅层特征及其组合^[12-15]. 机器学习与 LJP 任务的结合, 虽然能够在一定的指标计算上提升性能, 但是传统机器学习在特征提取阶段人工选择和参与度比较高, 提取的特征属于浅层性质, 在获取更为抽象的高层特征方面能力偏弱. 近年来, 深度学习技术在自然语言处理和计算机视觉等领域有着突出表现, 研究者已经开始尝试将深度学习特别是深度 NLP 技术用于法律领域^[14-18]. 研究者从注意力机制、外部知识、属性增强、多任务框架和多模态数据等不同角度提高 LJP 不同任务的质量^[19-24]. Zhong 等人^[25]提出了一个拓扑关系, 将 LJP 任务中的不同子任务关联起来, 相互辅助做出决策.

前期相关法律条文推荐任务的研究一直受限于数据集的规模, 在研究方法上以传统机器学习为主. 2005 年开始, Liu 等人^[3]尝试使用法条组合的形式将多标签法条推荐问题转换为多类别法条推荐问题. 有研究人员采用分步处理的方式, 使用 SVM(特征选择)进行初分类后, 再使用特征以及共现倾向对结果重排序, 对推荐

结果进行一定程度的修正^[7]。近年来, 为了进一步促进法律智能化的发展, 中华人民共和国最高人民法院与清华大学于 2018–2020 年连续举办了“中国法研杯(CAIL)”比赛。CAIL2018 比赛发布的数据集推动了相关法律条文推荐方法的研究进展, 其中取得优秀成绩的参赛者基本都使用了基于神经网络的文本分类模型, 通过提取深层的文本特征提升预测质量^[26,27], 其中最常用的模型是 CNN^[28], TextCNN^[29], LSTM^[30]和 GRU^[31]等。随后的相关工作在 CAIL2018 数据集上使用多个分类模型融合(CNN 和 RNN)提高法条预测精度^[27]。考虑到法条与罪名和刑期预测两个任务的关联, 研究者从任务之间的相互影响和增强预测结果可判读性这两个方面开展了深入研究。Yang 等人^[32]和 Xu 等人^[33]延续分类思想, 在多任务框架内建立子任务之间的关系, 在完成法条推荐任务的同时, 可以实现罪名和刑期预测。文献[33]中采用基于图卷积的网络结构, 以便增强推荐算法区分容易混淆的法条的能力。将法条推荐和犯罪预测在统一的框架内处理, 从一定程度上可以对结果进行判读, 是法条推荐在 LJP 领域的重要作用。

1.2 语义匹配模型

本文提出采用语义匹配学习的解决方案, 计算案例描述与法条的匹配度, 按匹配程度的大小进行推荐。语义匹配的研究大多集中在 Question & Answer 和句子匹配领域, 例如, Wang 等人^[5]基于 Compare-Aggregate 框架设计了一种 Seq2Seq 模型, 解决句子匹配任务; Feng 等人^[6]应用了一个通用的深度学习框架实现 QACNN, 解决非事实性的问题回答任务。Seq2Seq 和 QACNN 这两种模型是语义匹配任务的经典模型, 近几年, 随着图形处理器(graphics processing unit, GPU)的高速发展, 训练深度神经网络模型的速度增长以及容纳模型参数的显存增加, 预训练模型 GPT^[34]以及 BERT (bidirectional encoder representation from transformers)^[35]迅速发展, 在句子匹配、分类和实体识别等任务中表现突出。而它们的共性是都使用了 Transformer^[36], 其具有良好的并行性以及适合捕捉长距离特征。本文提出的 DeepLawRec 方法利用先进的预训练 BERT 模型获取文本的稠密空间表示, 主要贡献是在文本编码表示之后加入卷积层, 提取与法条相关的局部语义特征, 并结合回归树模型形成一个再选择机制, 提升推荐结果的准确率和可解释性。本文在实验中采用 BERT 编码, 其他编码模型也同样可用于本文提出的方法中。

2 基于深度语义匹配的法条推荐方法

2.1 问题描述

法律条文推荐任务是从法律案件描述(定义为事实)中预测所适用的法律条文, 作为法院判决的依据。如表 1 所示, 对于第 1 条案例描述, 法院依据法律条文“失火罪”, 可以判断嫌疑人的罪名; 对于第 2 条案例描述, 文字上与和第 1 条类似, 都与火灾发生有关, 但是通过深度语义理解, 发现其更加适用法律条文“放火罪”, 存在“用随身携带的打火机点燃”这种故意的行为。在此情况下, 法条推荐任务需要着重关注于违法行为的描述, 从中识别出两种事实描述之间的语义差异, 并最终将其划分到对应的法律条文。

表 1 案例描述以及法律条文示例

	案例描述	法律条文
1. 失火罪	...期间,被告人王某在野外没有保证火灾的安全, 在焚烧纸钱时引起森林火灾, 并在火灾后逃离现场...	由于行为人的过失而引起火灾, 造成严重后果, 危害公共安全的行为
2. 放火罪	...然后用随身携带的打火机点燃房间内的衣服, 烧毁房间内的空调、电视机、电风扇等物品...	故意放火焚烧公私财物, 危害公共安全的行为

相关法条推荐任务在现有的研究中经常将其当作分类问题^[26,27,32,33], 主要预测案例描述所属的类别标签, 即法条编号。通过表 1 的例子可以看出, 分类任务的难点是内容混淆问题。两个案件文字上有相似的描述和用词表达, 但是对应不同的法条, 例如表 1 中的放火罪与失火罪。因此, 本文认为引入法律条文自身文本内容有助于缓解该问题, 从而提出了 DeepLawRec 方法, 将法条文本和案例描述均看成由字词组成的自然语言句子, 计算和学习它们两者之间的匹配度, 将匹配度高的法条作为推荐结果。

DeepLawRec 通过将每一个案例 Q 与法条集合 A 中的每个法条构造造成若干个二元组 $\{Q, A_i\}$, 即可构造成一

个适合语义匹配问题的数据集, 输出则为两者是否匹配. 由于相关法条不再只是一个编号标签, 而考虑其实际本文内容, 通过将分类问题转换成匹配问题, 有利于推荐模型关注更多的语义信息, 提高相关法条推荐任务的质量.

2.2 总体框架

本文提出的基于深度语义匹配的法条推荐方法(DeepLawRec)依据案例描述与法条之间的文本语义关系, 学习和计算两者的语义信息匹配度, 并结合回归树增强推荐结果的可解释性. 总体框架如图2所示, 包含局部语义匹配模块(上部分)和全局语义推荐模块(下部分).

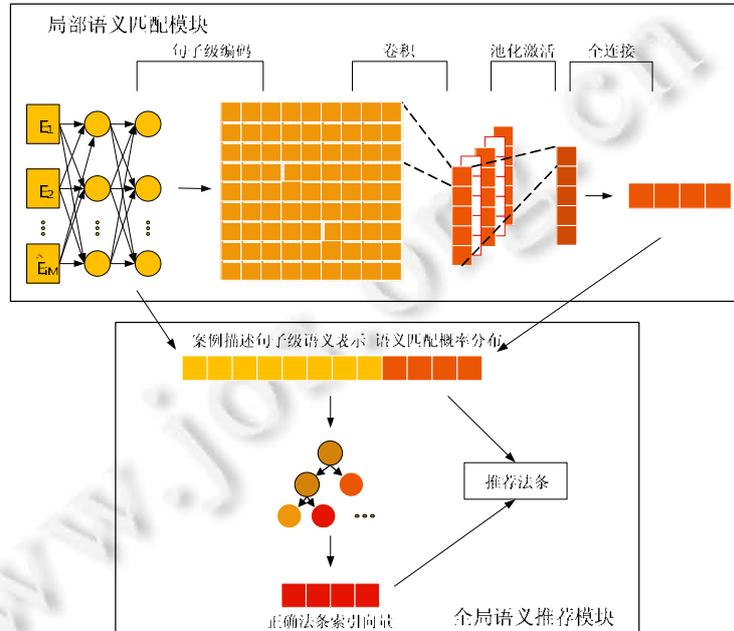


图2 基于深度语义匹配的法条推荐方法(DeepLawRec)

- 局部语义匹配模块

图2上半部分主要给出了本模块中的双向Transformer卷积网络模型, 由句子级编码网络和卷积网络组成. 案例描述的词向量和第 i 个法律条文的词向量为分别为 $E = \{E_1, E_2, E_3, \dots, E_n, \dots, E_N\}$ 和 $\hat{E} = \{\hat{E}_{i1}, \hat{E}_{i2}, \hat{E}_{i3}, \dots, \hat{E}_{iM}\}$, 其中, N 表示案例描述被切割后的词数, M 表示第 i 个法律条文被切割后的词数. 该模块采用句子级编码表示如BERT, 实现文本向量化, 再用CNN卷积网络提取局部特征, 增强与法条内容的匹配计算, 若干个 \hat{E} 形成法条候选集.

在NLP领域多个下游任务表现突出的预训练模型, 以BERT为代表, 产生的文本向量化表示是对于整个句子而言自身的词语与词语的依赖程度. 换句话说, 相对于整个句子而言, 词语的被依赖程度越大越被突出, 从而获得更大的注意力权重. 本文在预训练模型的文本表示之后加入CNN层, 提取文本的局部特征. 这个过程类似于 N -gram, 考虑了相邻的 n 个词 (n 为卷积核宽度) 之间关系的局部特征. 将CNN引入到语义匹配计算, 则是为了考虑案例描述关键词的上下文与法条的匹配, 而不仅仅是句子级文本向量化中关键词对案例句子的重要性. 最后, 池化激活层和全连接层得到语义匹配模型的输出 R_i , 每一个 R_i 都是第 i 个相关法条对于该案例描述的匹配概率.

- 全局语义推荐模块

图2下半部分是基于回归树的推荐模型. 将案例描述的全局句子级语义向量与法条的匹配概率向量(由局部语义匹配模块输出产生)联合表示, 采用回归树形成最终推荐结果, 从树根部到叶子节点的树结构可以提升

推荐的可解释性, 探索从匹配概率分布到最终推荐结果的推荐过程。

本模块中, 输入为案例描述句子级语义向量以及匹配概率分布分别为 S 和 $R=\{R_1, R_2, R_3, \dots, R_l\}$, 输出是推荐匹配度最大的前 k 个法条中, 比如作为正确的法条被推荐的索引概率 $P=\{P_1, P_2, P_3, P_4, \dots, P_k\}$ 。为了便于理解, 通过举例说明。比如, 当得到了匹配概率分布后, 可以从中选出匹配度最大的法条索引为 [35, 63, 70, 102, 161], 进一步通过回归树形成一个再选择机制, 得到的索引概率中最大的是 P_4 , 最后被推荐的法条就是第 102 个法律条文。

2.3 局部语义匹配模块

局部语义匹配模块基于双向 Transformer 网络模型, 由输入层、句子级编码层、卷积层、池化激活层和全连接层构成。计算输入的案例描述和法律条文的匹配程度。

- 输入层: 对案例描述和法律条文进行文本预处理后得到相应的词向量, 形成句子对向量矩阵 $\{E; \hat{E}\}$, 并将其作为句子级编码模型的输入向量。
- 句子级编码层: 主要作用是学习案例描述与法条之间的相关性, 并将更相关的词语赋予更大的权重, 同时也能够从稀疏的长文本向量中获得相应的高维语义信息。它的主要过程就是将输入案例描述和法条句子对组成向量矩阵, 使用双向的 Transformer 模型提取出相应语义表征, 再映射到相应的向量矩阵 W 以及句子向量 S 中。
- 卷积层: 卷积层主要着重提取语义表征中的局部特征, 由于句子级编码层将长文本序列语义关系词语相关性等信息压缩到向量矩阵 W 以及句子向量 S 中, 所以卷积层再抽取高维语义向量表征中重要的语义关联特征。
- 全连接层: 当通过句子级编码层和卷积层提取到不同粒度的重要语义表征后, 通过全连接层提供更丰富的非线性表达, 使用全连接层作为激活层以及输出层之间的桥梁, 提供整合后的表征给输出层, 得到匹配概率。

2.3.1 句子级编码层

NLP 领域的预训练一般是使用大量数据训练神经网络, 得到词嵌入向量表示。当面临一个新任务时, 可以先迁移预训练好的词嵌入向量来初始化神经网络输入, 再在训练过程中不断调整网络参数。预训练充分利用了大规模的语料信息, 既能有效地加快任务训练收敛速度, 也能解决小数据训练复杂网络任务的问题。

BERT 作为一个双向 Transformer 句子级编码器, 是目前在 NLP 领域表现突出的预训练模型, 可以捕捉双向文本信息的预训练模型。同时, 它相较于 RNN, LSTM 和 GRU 等传统的文本序列模型而言, 由于可以减少对距离的依赖, 能够更加高效地捕捉更长距离的逻辑关系。BERT 的词嵌入方法如公式(1)所示, 它由 3 种类型的词嵌入表征求和得到的, 句子级编码表示方法如图 3 所示。

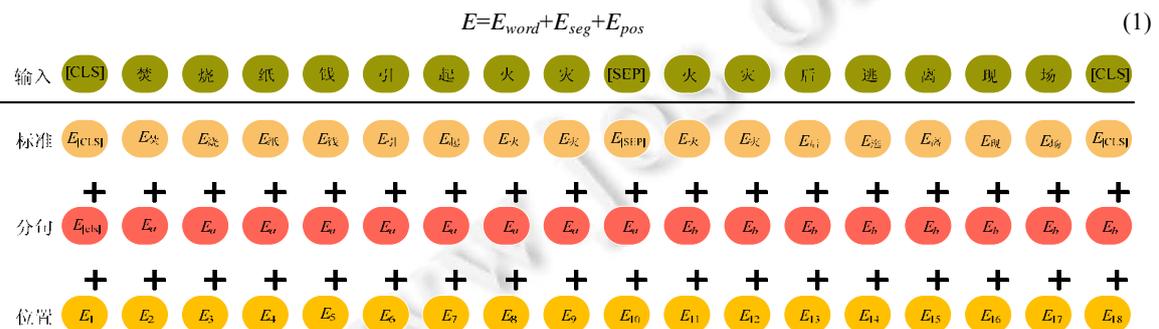


图 3 法律文书和法条的句子级编码表示方法

公式(1)中, 词嵌入表征 E_{word} 为标准的词嵌入向量如 Word2Vec 等; E_{seg} 为分句的词嵌入向量, 表明输入的两个句子的关系, 比如第 1 句和第 2 句, 其中, 输入中的 [CLS] 和 [SEP] 的分句词嵌入与其前一个词的分句词嵌

入相同; E_{pos} 位置表征指明该词语在文章中的位置向量. 在本文中, 分词操作后的每个词先分别求得其对应的标准词嵌入向量, 分句词嵌入向量以及位置表征, 再按照公式(1)求和作为最终的词嵌入表征. 对于本文的匹配任务, 除了开头和分隔符号[CLS]和[SEP]以外, 输入的两个段落中第 1 个段落为案例描述 Q , 另一个段落是第 i 个相关法条 A_i .

由于 BERT 模型是一个预训练模型, 已经在海量的语料信息上完成训练, 所以只需要将其训练好的参数迁移过来, 再使用法条推荐数据集实现微调 Fine-tuning. 主要过程就是冻结预训练的大部分神经网络层, 只训练剩下靠近输出的部分神经网络层. 通过这样的方式, 不仅能保持 BERT 模型原本的特征提取能力, 同时也能使模型胜任法条推荐任务的数据集.

2.3.2 卷积层

卷积层是可以进行卷积操作的神经网络结构, 最初应用于图像领域. 它可以从原始特征中依靠卷积核从中抽取最主要的抽象特征, 适用于端对端网络. 本文使用卷积层接收 BERT 层抽取到的序列向量矩阵, 该向量矩阵是一个二维张量. 卷积层对输入的张量使用自定义的卷积核进行相应的卷积操作, 一般是一个宽度与词嵌入向量长度一致的卷积核. 该卷积核对输入的张量从头到尾、从上到下进行平行移动, 每一次平移后, 卷积核内的每一个参数都会与对应位置的输入相乘再相加作为输出. 具体过程如图 4 所示, 输入窗口中的内容为“盗窃”和“非法”, 特征表示为 w_0, w_1 , 卷积核为 a_0, b_0 . 一次卷积计算如公式(2)所示.

$$[w_0, w_1] \cdot [a_0, b_0] = a_0 w_0 + b_0 w_1 \quad (2)$$

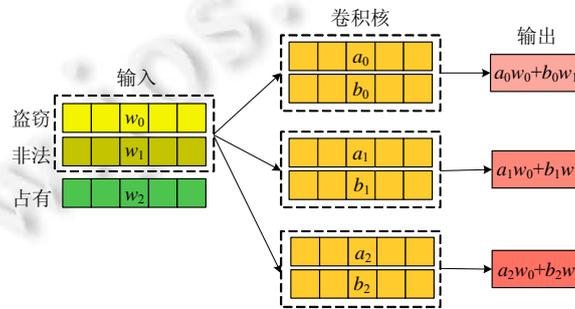


图 4 卷积操作示意图

2.3.3 池化激活层

池化层一般是在卷积层后的操作, 可以从卷积层抽取到的抽象特征中提取最重要的特征作为输出, 起到一个降低维度和压缩特征的作用. 在池化操作中, 有一个需要人为设定的窗口, 窗口的移动与卷积核类似. 在每一次移动后, 最大池化操作就从窗口中的矩阵中找出最大值作为该位置的输出, 而平均池化操作是对窗口中的矩阵中的所有值进行平均值计算, 得到的结果作为输出. 本文使用了最大池化操作(max pooling), 这样就可以从文本特征中提取更为重要的特征, 以此缩减计算量, 减少训练模型的时间; 另外, 也可以防止过拟合.

激活层一般情况下与池化层同时出现, 接收池化层输出的数据. 由于在神经网络中的神经元都是输入的线性组合, 为了使这个神经网络逼近任意函数, 需要引入非线性函数作为激励函数丰富网络的表达. 本文在激活层中引入非线性函数 Rectified Linear Unit (ReLU)函数作为激活层的激励函数, 该函数如公式(3)所示.

$$Relu(x) = \max(0, x) \quad (3)$$

2.3.4 全连接层

全连接层将模型学习得到的特征表示映射到样本标记空间, 其本质是将一个特征空间映射到另一个特征空间. 对于目标空间的每一个数值而言, 它都会受到源特征空间整体的影响. 本文使用全连接层可以将抽取到的特征映射到样本标记空间中, 再拟合成最终的输出, 即输入与法律条文的匹配概率 R_i .

2.4 全局语义推荐模块

在实际情况下, 案例描述与推荐法律条文之间的语义匹配是存在一定的误差的, 会导致如下情况的发生: 在图 2 上半部分的局部语义匹配模型中, 虽然正确的相关法条的匹配概率数值较高, 但可能存在某些法条的匹配概率比其更高, 却不是正确的相关法条. 其原因是在于法律文书内容在文字表达上的多样性. 为了缓解该情况发生对推荐质量的影响, 本文提出了一个基于回归树的推荐模型, 对匹配度较高的相关法条进行再选择, 如图 2 的下半部分所示.

再选择机制的输入是将匹配度概率分布 R 与案例描述的句子级向量 S 的拼接向量, 输出的是该案例描述对应匹配度最大的前 n 个相关法条的独热编码 P . 这里的输入再次引入了案例描述的句子级语义向量, 是一种全局语义信息. 最后, 联合局部语义匹配模块的输出(即匹配概率分布向量)重新构建训练数据, 使用基于回归树的算法作为再选择机制模型, 利于对推荐结果的解释.

本文选取性能表现突出的 XGBoost (extreme gradient boosting), 它改进了传统的 GBDT (gradient boost decision tree) 算法, 将树的复杂度考虑到目标函数中, 并且在迭代优化的过程中对目标函数进行泰勒展开, 求解二阶近似解, 这样可以加快迭代的过程. 本文采用的目标函数定义如公式(4)所示.

$$L = \sum_{i=1}^n \text{loss}_{\text{XG}}(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (4)$$

公式(4)的第 1 部分用来衡量预测结果和真实结果的差距, 第 2 部分则是树的复杂度的正则化项. 进一步可将目标函数改写成公式(5).

$$L^{(t)} = \sum_{i=1}^n \text{loss}_{\text{XG}}(y_i, \hat{y}_i^{t-1} + f_i(x_i)) + \Omega(f_i) \cong \sum_{i=1}^n \left[\text{loss}_{\text{XG}}(y_i, \hat{y}_i^{t-1}) + g_i f_i(x_i) + \frac{1}{2} h_i f_i^2(x_i) \right] + \Omega(f_i) \quad (5)$$

其中, g 是一阶导数, h 是二阶导数, 如公式(6)、公式(7)所示.

$$g_i = \partial_{\hat{y}^{(t-1)}} l(y_i, \hat{y}_i^{t-1}) \quad (6)$$

$$h_i = \partial_{\hat{y}^{(t-1)}}^2 l(y_i, \hat{y}_i^{t-1}) \quad (7)$$

2.5 模型训练和测试算法描述

DeepLawRec 方法由两个串行的模块组成: 局部语义匹配模块采用传统的交叉熵作为损失函数, 在预训练句子级编码模型上, 用法律领域数据进行微调(fine-tuning); 全局语义推荐模型是基于回归树做预测, 损失函数如公式(5)所示. 算法 1 和算法 2 分别给出了整体训练和测试的流程.

DeepLawRec 可以看成语义匹配串联基于回归树的再选择机制, 它是基于预训练模型, 卷积神经网络模型和回归树模型各自不同的特点, 根据智能司法大数据中的相关法律条文推荐存在内容混淆这一问题提出的一个新的组合方法. 算法 1 的 DeepLawRec 训练过程中首先训练局部语义匹配模块, Step 1 和 Step 2 的文本句子级特征表示在本文的实验中来自基于 Transformer 的 BERT 预训练模型, 其他先进的句子级编码模型也可以用于此. 由于在自然语言处理领域中, 以 Transformer 为代表的句子级编码网络擅长捕捉数据或特征的内部相关性, 以及提取案件描述中内容的上下文之间丰富的语义信息, 在 Step 5 中使用卷积神经网络提取最重要的局部特征, 并进行对数据进行相应的压缩.

算法 1 的 DeepLawRec 训练过程中的全局语义推荐模型主要使用 XGBoost 模型作为回归树(Step 10 和 Step 11), 在训练过程中, 它可以降低被正确分类的训练样本的权重, 同时也增大分类错误的样本的权重. 这样被错误分类的训练样本会获得更多关注, 可以有效地对全局语义匹配模块输出的匹配概率做一次再选择的过程, 降低推荐错误法条的机率, 进一步提升推荐方法的效果.

算法 1. DeepLawRec 训练过程.

输入: 案例描述句子序列 $x = \{x^1, \dots, x^n\}$,

法律条文句子序列 $z = \{z^1, \dots, z^n\}$.

输出: 相关法条索引概率 $P = \{P_1, P_2, P_3, P_4, \dots, P_k\}$.

局部语义匹配

repeat:

for $i \leftarrow 1$ to N :

1. 案例描述文本特征表示: $E = \{E_1, E_2, E_3, \dots, E_n, \dots, E_N\}$;
2. 第 i 个法律条文文本特征表示: $\hat{E} = \{\hat{E}_{i1}, \hat{E}_{i2}, \hat{E}_{i3}, \dots, \hat{E}_{im}\}$;
3. 句子对向量矩阵 $\{E; \hat{E}\}$ 输入到句子级编码层;
4. 句子级编码器输出 $T = \{T^1, \dots, T^n\}$;
5. 卷积池化激活 $Relu(x) = \max(0, x)$;
6. 预测与法律条文的匹配概率 R_i 及 $R = \{R_1, R_2, R_3, \dots, R_l\}$;
7. 计算交叉熵损失函数 $Loss$ 并反向传播更新参数;

until: $Loss$ 最小;

全局语义推荐

repeat:

for $i \leftarrow 1$ to N :

8. 案例描述文本的句子级特征表示: $X_i = \{X_i^1, \dots, X_i^n\}$;
9. 法律条文的匹配概率分布 $R = \{R_1, R_2, R_3, \dots, R_l\}$;
10. Step 8 和 Step 9 的向量拼接输入到回归树;
11. 回归树输出推荐匹配度最大的前 k 个法条中正确的法条被推荐的索引概率 $P = \{P_1, P_2, P_3, P_4, \dots, P_k\}$;
12. 计算损失函数 $Loss$ (公式(5)), 并反向传播更新参数;

until: $Loss$ 最小;

算法 2 的 DeepLawRec 测试过程中, 案例描述和法律条文两者的文本在 Step 1 中获得句子级编码, 再送到 Step 2 的局部语义匹配模块获得两者的匹配概率, 从而由若干个法条组成候选集. Step 3 对法条组成的匹配度概率分布与案例描述的句子级向量进行拼接操作, 使用基于回归树的算法作为全局语义推荐模型. 在 Step 4 中输出被推荐的法条; 同时, 树状的推荐路径增强了对推荐结果的解释.

算法 2. DeepLawRec 测试过程.

输入: 案例描述和法律条文句子序列.

输出: 推荐的法条、*Accuracy* 和 *Precision* 分值.

初始化: $MAX_LEN = N$; $i = 1$.

while $i < MAX_LEN$:

1. 计算句子级编码器输出;
2. 在局部语义匹配模块中得到匹配概率分布;
3. 在全局语义推荐模块中得到索引概率分布;
4. 输出概率值最高的法条作为最终推荐结果;
5. $i++$;

end while

6. 计算 *Accuracy* 和 *Precision* 分值;

3 实验结果与分析

这一部分首先介绍实验数据和评价指标, 然后对比本文的 DeepLawRec 方法与基线方法在法条推荐质量上的差异, 对实验结果进行分析和讨论.

3.1 数据集与预处理

为了客观描述 DeepLawRec 方法的有效性, 本文在公开的中文法律数据集 CAIL2018 上进行了一系列的实验对比和分析(<http://cail.cipsc.org.cn:2018/>). 该实验数据来源于中国法研杯竞赛数据集, 是由中国最高人民法院裁判文书网中的案件组成. 比赛中, 训练集的刑事案例数目为 114 824 条, 验证集 14 293 条, 测试集 23 593 条, 一共有 183 个法条名称. 为了利用法条的文本内容, 在数据准备阶段, 本文首先根据法条名称补充其对应的法律条文内容, 见表 2.

表 2 法律条文示例

法律条文	具体内容
聚众斗殴	为了报复他人、争霸一方或者其他不正当目的, 纠集众人成帮结伙地互相进行殴斗, 破坏公共秩序的行为. 聚众斗殴罪侵犯的客体是所谓公共秩序, 不应简单地理解为公共场所的秩序
侵犯著作权	以营利为目的, 未经著作权人许可复制发行其文字、音像、计算机软件等作品, 出版他人享有独占出版权的图书, 未经制作者许可复制发行其制作的音像制品, 制作、展览假冒他人署名的美术作品, 违法所得数额较大或者有其他严重情节的行为
挪用资金	公司、企业或者其他单位的工作人员利用职务上的便利, 挪用本单位资金归个人使用或者借贷给他人, 数额较大、超过 3 个月未还的, 或者虽未超过 3 个月, 但数额较大、进行营利活动的, 或者进行非法活动的行为

接下来对案例描述和法条描述文本进行数据清理, 删去异常数据、无意义的暂停词和时间数值等对法条推荐任务影响不大的信息. 再使用分词工具(结巴中文分词组件)将一整段的案例描述以词为单位分割成片段, 接下来再将某一案例描述与每个相关法条的是否匹配(0 或 1), 按照顺序组合后得到匹配概率分布. 最后, 从中按照顺序构造匹配度最大的前 k 个法条的独热编码, 其中, 如果是输入案例对应的推荐法条, 其编码为 1; 否则为 0.

3.2 实验设置与评价指标

3.2.1 实验方案

实验采用精度和准确率指标, 从以下 3 个方面评估本文提出的法条推荐模型 DeepLawRec 的有效性.

- (1) 与经典语义匹配模型的精度相比, 效果如何?
- (2) 与法律智能预测领域常用的文本分类模型相比, 准确率效果如何?
- (3) 如何对推荐结果做可解释性分析?

对于第 1 个问题, 本文将对对比 QACNN, Seq2Seq 以及 BERT 模型, 采用匹配计算中常用的 *Precision* 指标, 即命中的法条在被推荐法条中的占比; 对于第 2 个问题, 将 CNN, TextCNN, LSTM 和 GRU 作为分类的基线模型, 采用 *Accuracy* 指标, 即被推荐的法条是否正确; 第 3 个问题将通过举例分析验证实验结果.

3.2.2 评价指标

法条推荐结果分为匹配/不匹配两种情况, 在分类模型上是一个二分类问题. 因此, 根据样本的真实结果以及模型对于样本的预测结果, 将按以下规则分成 4 种不同的类别: 真正例(true positive, TP)、假正例(false positive, FP)、假负例(false negative, FN)、真负例(true negative, TN). 精度和准确率的公式如公式(8)和公式(9)所示.

$$accuracy_i = \frac{TP + TN}{TP + FN + FP + TN} \quad (8)$$

$$precision_i = \frac{TP}{TP + FP} \quad (9)$$

3.2.3 实验设置

首先, 将案例描述以及相关法条的词个数分别设置为 270, 30, 将两种文本拼接总词数为 300 个. 在实验中所使用的 Word2Vec 词向量是由文献[32]提供的, 混合了百度百科、中文维基百科、人民日报等语料库训练所得的 300 维词向量.

语义匹配模型中, QACNN 的卷积核宽度数分别为 2, 3, 4, 5, 7, 9, 全连接层第 1 层节点维度设置为 1024, 采用自适应学习率调整算法(AdaDelta)更新模型训练参数, 学习率设置为 $1e-5$, 学习率的衰退系数设置为 0.95, 常数 ϵ 设置为 610, 并且使用 *sigmoid* 函数用于计算匹配概率. 在 XGBoost 中用于控制是否后剪枝的参数 *gamma* 设置为 0.1, 用于控制树的深度的 *max_depth* 设置为 8, 使用 L_2 正则化系数, 设置为 10, 最小叶子节点样本权重和 *min_child_weight* 设置为 3.

对于分类模型, 在实验中设置卷积核宽度为 2, 3, 4, 5, LSTM 以及 GRU 隐藏层的节点维度为 1 000, 采用自适应学习率调整算法(Adam)更新模型训练参数, 学习率设置为 $1e-3$, 学习率的衰退系数设置为 0.95, 损失函数为 *Cross-entropy*, 并且使用 *softmax* 分类器用于分类.

本文的实验环境配置如下: Intel(R) Xeon(R) CPU E5-2650 v4@2.20 GHz; 128 GB DDR4 内存; TITAN Xp 型号 GPU; 10.1 版本的 CUDA. 实验代码则采用 3.6 版本的 Python, Keras 框架和多个机器学习第三方库实现, Anaconda3 环境测试和运行.

3.3 实验结果与分析

3.3.1 与经典语义匹配模型的对比

由于本文首次提出将法条推荐任务转换成语义匹配任务, 所以首先对比了 QACNN 和 Seq2Seq 两个经典语义匹配模型, 以及在 DeepLawRec 中采用 BERT 和 RoBERTa 句子级编码网络(不含卷积层), 以匹配问题常用的 *Precision* 为评价指标. 如表 3 所示, *Precision@1*, *Precision@5* 和 *Precision@10* 这 3 个指标的结果表明, 本文提出的 DeepLawRec 具有优势. *Precision@1* 是 3 个指标中最为严格的, 我们的 DeepLawRec-(不含全局语义推荐模块, 具有句子级编码和卷积层)达到了 0.625, 比 BERT^[35]提升了 20.89%. 将 BERT 替换成 RoBERTa, 可以看到 DeepLawRec-达到了 0.638, 比 RoBERTa^[37]提升 22.2%.

基于 Transformer 结构的 BERT 和 RoBERTa 通过所有层的上下文联合调节来预训练深层双向表示, 在上游任务自然语言处理任务中获得了领先的结果. 本文的 DeepLawRec 借由卷积计算从 BERT 句子级别的表征中选择性提取重要的局部特征. 在数据经过池化层和激活层时, 可以压缩所提取的特征得到相对而言简练但更为重要的特征, 最后通过全连接层给提取到的特征提供更丰富的特征表达, 从而达到了最好的推荐质量.

2015 年, IBM Watson 团队提出的 QACNN^[6]是 QA 方面第一个采取 CNN 实现问答中的答案选择, 所提出的神经网络结构较为简单, 没有加入 RNN 的部分, 没有去刻画自然语言的序列特征, 因此在法条推荐任务上整体表现偏弱, *Precision@1* 为 0.314, 远低于 DeepLawRec 的 0.638. 2017 年的 Seq2Seq^[5]模型是在 Google 的 Parikh 等人提出的 Compare-Aggregate 框架上设计的句子匹配模型, 表 3 的实验结果表明, 该模型比传统的 Siamese model(QACNN)要高出不少, *Precision@1* 达到了 0.511. 可以看到, Siamese model 虽然学习的是句子的整体表示, 但一定程度上还是丢失了一些信息.

表 3 与语义匹配模型对比的实验结果(*precision*)

模型	QACNN ^[6]	Seq2Seq ^[5]	DeepLawRec-(Ours, BERT)	DeepLawRec-(Ours, RoBERTa)	BERT ^[35] (DeepLawRec--)	RoBERTa ^[37] (DeepLawRec--)
<i>Precision@1</i>	0.314	0.511	0.625	0.638	0.517	0.522
<i>Precision@5</i>	0.633	0.948	0.951	0.956	0.847	0.865
<i>Precision@10</i>	0.768	0.974	0.964	0.978	0.922	0.929

图 2 下半部分是本文在语义匹配计算上, 为增强推荐结果的质量而设计的全局语义推荐模块, 可以作为一个独立的功能联合现有的语义匹配模型. 具体而言, 语义匹配计算的输出层是不同法条与法律文书的匹配概率分布, 本文提出的全局语义推荐模块将此匹配度概率分布与案例描述的句子级特征表示组合, 使用回归树增强推荐的精度和可解释性, 实验结果见表 4.

表 4 有/无全局语义推荐模块的对比实验结果(*precision@1*)

有/无全局语义推荐模块	QACNN	Seq2Seq	RoBERTa (DeepLawRec--) ^[37]	DeepLawRec (Ours)
无	0.314	0.511	0.522	0.638
有	0.582	0.809	0.853	0.913

可以看到, 本文提出的全局语义推荐模块对语义匹配模型是有明显的改善. 对于 QACNN 而言, 通过添加该模块, 推荐精度从 0.314 提高到 0.582; Seq2Seq 和 RoBERTa 也同样借助该模块, 使得法条推荐的精度得到较大的提升; 我们的 DeepLawRec 也从 0.638 提高到了 0.913. 这是由于本文中通过回归树实现的再选择机制能够对语义匹配度较高的法条进行重新排序, 将原有的不精准预测通过重新选择后, 给每个案例描述推荐正确的相关法律条文, 从而使预测精度有了明显的提升.

3.3.2 与法律智能预测常用的文本分类模型对比

本文提出的 DeepLawRec 进一步与法律智能预测领域流的文本分类模型 CNN^[28], TextCNN^[29], LSTM^[30], GRU^[31]和 MPBFN^[32]对比, 采用分类中常用的 *Accuracy* 准确率为评价指标, 其实验结果如图 5 所示.

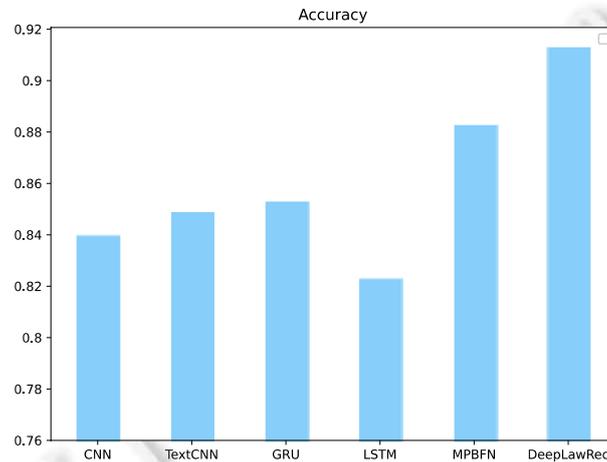


图 5 DeepLawRec 和分类模型的对比实验结果(*accuracy*)

本文提出的方法准确率都高于传统的分类模型, 在法条推荐集上表现最好的分类模型为 GRU, 其推荐准确率为 0.853, 本文的 DeepLawRec 达到了 0.913. 由于本文所使用的法律数据集 CAIL2018 中, 文本长度平均值都是 300 以上, 且具有强烈的上下文关系, 所以分类算法和传统的语义匹配系统算法对长文本处理不够深入, 虽然能从短文本中提取重要关键词, 但是模型效果仍有待提升.

相较而言, 本文提出的 DeepLawRec 将句子级编码的预训练模型在数据集上进行微调, 首先, 预训练模型本身的能力能缓解词语之间的距离进行特征学习, 擅长于理解文本序列; 另外, 由于预训练模型已经通过大量语料进行训练, 再通过 Fine-tuning 的作用, 能够较好地使用本法律领域数据集, 增强模型在领域任务上的能力. 另一方面, DeepLawRec 使用了卷积操作, 能够从长序列的句子级表征中提取重要特征, 从而进一步提升匹配模型效果. 最后, 再结合基于回归树的再选择机制, 可以缓解推荐不相关法条情况的发生. 在这 4 个特点的共同作用下, 本文提出的方法提升了推荐的质量.

在相同阶段的数据集上, 本文从实验结果对比分析了最近发表在 ACL2020 上的 LADAN^[33]模型, 其在法条推荐任务上最高 *Accuracy* 值为 0.823 4. 该模型整体是一个多任务的分类框架, 通过图计算和 Attention 机制增强特征学习能力; 同时, 借助法条推荐、罪名预测和刑期预测这 3 个分类子任务的相关性, 提高分类准确率. 本文提出的 DeepLawRec 关注法条推荐问题, 通过在语义匹配计算框架引入回归树, 提升推荐在准确率和可解释性两方面的质量, 最高 *Accuracy* 值为 0.913. 接下来的案例分析给出了推荐的过程.

3.3.3 案例分析

本文为了提高法条推荐结果的可解释性, 提出了基于回归树的全局语义推荐模块, 使用 XGBoost 作为模块的选择模型, 这里通过例子说明此模块是如何对案例描述的理解和产生法条推荐结果.

对于案例描述“公诉机关起诉指控, 被告人张某某秘密窃取他人财物, 价值 2 210 元, 盗窃数额较大, 其行为已触犯《中华人民共和国刑法》XX 之规定, 应当以 XX 追究其刑事责任. 建议判处被告人张某某 XX 以下

刑罚,并处罚金”。本文得到其对应的 163 个法条的匹配概率分布后,与该案例描述自身的 768 维句向量拼接输入到 XGBoost,对应的输出 2(法条索引)。由于该案例描述对应的匹配度最大的前 5 个推荐指数的法条索引是[49,43,37,78,5](以匹配度数值大小排序),所以在该索引为 2 的法条即为第 43 条法条。由图 6 中可以看到,对于该案例描述而言,由于它第 121 个特征小于 10.84,所以进入左子树;又由于第 543 个特征小于 0.68,所以进入右子树。最后一直到叶子节点输出 0.124。其中,特征 802, 814 以及 808 都与盗窃罪有关。由于大多数误差都发生在法条相似且关联较大的情况,其推荐指数偏高且容易误判。通过回归树的再次选择,较好地这类情况识别处理,降低其错分率并提高可解释性。

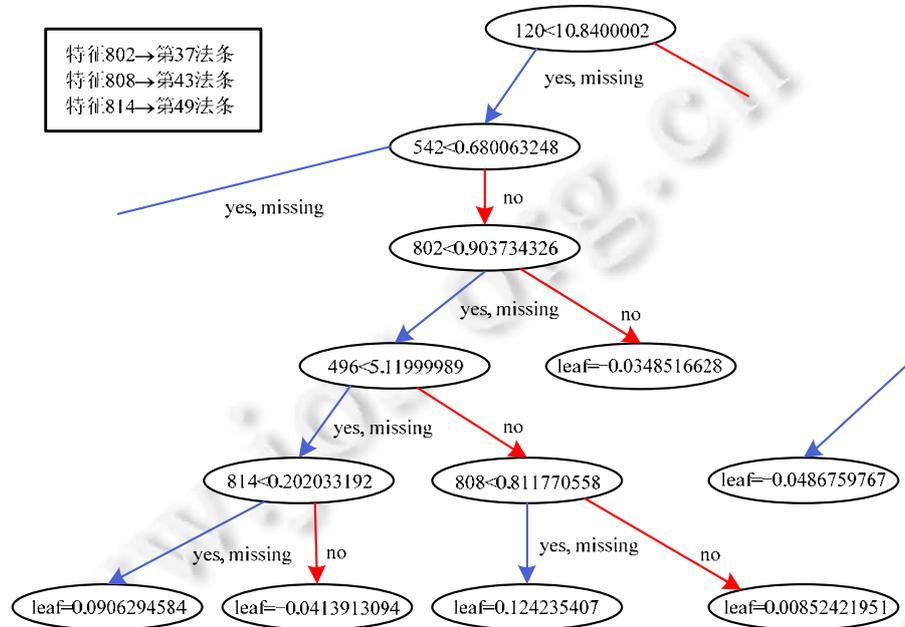


图 6 基于回归树的法条推荐过程

4 总结和展望

本文提出的基于深度语义匹配的法条推荐方法将法条推荐问题看成案例描述与法条的匹配度计算,与传统的文本分类思路相比,能够更好地利用法条自身的语义信息。通过搭建句子级编码和卷积计算结合的网络模型,将案例描述与法律条文信息进行语义匹配学习,从而完成对候选法条的匹配度计算;再进而提出全局语义推荐模块,从匹配的候选法条中再次进行选择,提升准确率和可解释性。在真实数据集上测试结果表明,本文提出的法条推荐模型与文本分类模型以及语义匹配算法相比,显著提升了法条推荐质量。

在今后的研究中,还可以进一步探索如何引入领域知识辅助法律条文推荐任务。同时,现有的研究工作表明,罪名预测、法条推荐和刑期预测这 3 个任务之间存在关联,可在多任务框架中通过联合学习相互提升任务质量^[38]。此外,关于词向量模型的研究,在基于预训练模型 BERT 上已经有了很多改进的工作,如 ALBERT, XLNET 和 ERNIE 等,可进一步应用到本文提出的方法中。

References:

- [1] Jiang X, Ye H, Luo Z, *et al.* Interpretable rationale augmented charge prediction system. In: Proc. of the Int'l Conf. on Computational Linguistics: System Demonstrations. 2018. 146–151.

- [2] Ye H, Jiang X, Luo Z, *et al.* Interpretable charge predictions for criminal cases: Learning to generate court views from fact descriptions. In: Proc. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2018. 1854–1864.
- [3] Liu CL, Liao TM. Classifying criminal charges in Chinese for Web-based legal services. In: Proc. of the Asia-Pacific Web Conf. 2005. 64–75.
- [4] Zhang H, Wang X, Wang C, *et al.* Law article prediction method for legal judgment documents. *Computer Science*, 2019, 46(9): 211–215 (in Chinese with English abstract).
- [5] Wang S, Jiang J. A compare-aggregate model for matching text sequences. In: Proc. of the Int'l Conf. on Learning Representations. 2017. <https://openreview.net/forum?id=HJTzHtqee>
- [6] Feng M, Xiang B, Glass MR, *et al.* Applying deep learning to answer selection: A study and an open task. In: Proc. of the 2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU). IEEE, 2015. 813–820.
- [7] Liu YH, Chen YL, Ho WL. Predicting associated statutes for legal problems. *Information Processing & Management*, 2015, 51(1): 194–211.
- [8] Kort F. Predicting supreme court decisions mathematically: A quantitative analysis of the “right to counsel” cases. *American Political Science Review*, 1957, 51(1): 1–12.
- [9] Ulmer SS. Quantitative analysis of judicial processes: Some practical and theoretical applications. *Law and Contemporary Problems*, 1963, 28(1): 164–184.
- [10] Nagel SS. Applying correlation analysis to case prediction. *Texas Law Review*, 1963, 42(7): 1006–1017.
- [11] Keown R. Mathematical models for legal prediction. *Computer/Law Journal*, 1980, 2: 829–862.
- [12] Segal JA. Predicting supreme court cases probabilistically: The search and seizure cases, 1962–1981. *American Political Science Review*, 1984, 78(4): 891–900.
- [13] Lauderdale BE, Clark TS. The supreme court’s many median justices. *American Political Science Review*, 2012, 106(4): 847–866.
- [14] Liu CL, Hsieh CD. Exploring phrase-based classification of judicial documents for criminal charges in Chinese. In: Proc. of the Int'l Symp. on Methodologies for Intelligent Systems. 2006. 681–690.
- [15] Lin WC, Kuo TT, Chang TJ, *et al.* Exploiting machine learning models for Chinese legal documents labeling, case classification, and sentencing prediction. *Int'l Journal of Computational Linguistics & Chinese Language Processing*, 2012, 17(4): 40–68.
- [16] Aletras N, Tsarapatsanis D, PreoŃiuc-Pietro D, Lampos V. Predicting judicial decisions of the European court of Human rights: A natural language processing perspective. *PeerJ Computer Science*, 2016, 2: No. e93. <https://doi.org/10.7717/peerj-cs.93>
- [17] Sulea OM, Zampieri M, Malmasi S, *et al.* Exploring the use of text classification in the legal domain. arXiv: 1710. 09306, 2017.
- [18] Zhang Y, Wallace B. A sensitivity analysis of (and practitioners’ guide to) convolutional neural networks for sentence classification. In: Proc. of the Int'l Joint Conf. on Natural Language Processing. 2017. 253–263.
- [19] Khan A, Baharudin B, Lee LH, *et al.* A review of machine learning algorithms for text-documents classification. *Journal of Advances in Information Technology*, 2010, 1(1): 4–20.
- [20] Tang D, Qin B, Liu T. Document modeling with gated recurrent neural network for sentiment classification. In: Proc. of the Empirical Methods in Natural Language Processing. 2015. 1422–1432.
- [21] Luo B, Feng Y, Xu J, *et al.* Learning to predict charges for criminal cases with legal basis. In: Proc. of the Empirical Methods in Natural Language Processing. 2017. 2727–2736.
- [22] Long SB, Tu CC, Liu ZY, Sun MS. Automatic judgment prediction via legal reading comprehension. In: Proc. of the CCL 2019. 2019. 558–572.
- [23] Chen S, Qin J, Ji X, *et al.* Automatic scoring of multiple semantic attributes with multi-task feature leverage: A study on pulmonary nodules in CT images. *IEEE Trans. on Medical Imaging*, 2016, 36(3): 802–814.
- [24] Li S, Zhao Z, Hu RF, *et al.* Analogical reasoning on Chinese morphological and semantic relations. In: Proc. of the 56th Annual Meeting of the Association for Computational Linguistics. 2018. 138–143.
- [25] Hu Z, Li X, Tu C, *et al.* Few-shot charge prediction with discriminative legal attributes. In: Proc. of the Int'l Conf. on Computational Linguistics. 2018. 487–498.
- [26] Zhong H, Xiao C, Guo Z, *et al.* Overview of CAIL2018: Legal judgment prediction competition. arXiv: 1810.05851, 2018.

- [27] Xiao C, Zhong H, Guo Z, *et al.* CAIL2018: A large-scale legal dataset for judgment prediction. arXiv: 1807.02478, 2018.
- [28] Kim Y. Convolutional neural networks for sentence classification. In: Proc. of the 2014 Conf. on Empirical Methods in Natural Language Processing (EMNLP). 2014. 1746–1751.
- [29] Sepp Hochreiter and Jurgen Schmidhuber. Long short-term memory. *Neural Computation*, 1997, 9(8): 1735–1780.
- [30] Cho K, Van Merriënboer B, Gulcehre C, *et al.* Learning phrase representations using RNN encoder-decoder for statistical machine translation. In: Proc. of the Empirical Methods in Natural Language Processing. 2014. 1724–1734.
- [31] Luo B, Feng Y, Xu J, *et al.* Learning to predict charges for criminal cases with legal basis. In: Proc. of the Empirical Methods in Natural Language Processing. 2017. 2727–2736.
- [32] Yang WM, Jia WJ, Zhou XJ, Luo YT. Legal judgment prediction via multi-perspective bi-feedback network. In: Proc. of the IJCAI 2019. 2019. 4085–4091.
- [33] Xu N, Wang PH, Chen L, Pan L, Wang XY, Zhao ZJ. Distinguish confusing law articles for legal judgment prediction. In: Proc. of the ACL 2020. 2020. 3086–3095.
- [34] Radford A, Narasimhan K, Salimans T, *et al.* Improving language understanding by generative pre-training. 2018. <https://www.gwern.net/docs/www/s3-us-west-2.amazonaws.com/d73fdc5ffa8627bce44dca2fc012da638ffb158.pdf>
- [35] Devlin J, Chang MW, Lee K, *et al.* Bert: Pre-training of deep bidirectional transformers for language understanding. In: Proc. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2018. 4171–4186.
- [36] Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need. In: Advances in Neural Information Processing Systems. 2017. 5998–6008.
- [37] Liu Y, Ott M, Goyal N, *et al.* RoBERTa: A robustly optimized BERT pretraining approach. arXiv: 1907.11692, 2019.
- [38] Duan W. Intelligent judgment prediction for legal text [MS. Thesis]. Wuhan: Wuhan University of Technology, 2020 (in Chinese with English abstract).

附中文参考文献:

- [4] 张虎, 王鑫, 王冲, 等. 面向法律裁判文书的法条推荐方法. *计算机科学*, 2019, 46(9): 211–215.
- [38] 段围. 面向法律文本的智能判决预测方法研究 [硕士学位论文]. 武汉: 武汉理工大学, 2020.



李琳(1977—), 女, 博士, 教授, 博士生导师, CCF 高级会员, 主要研究领域为数据挖掘, 推荐系统, 信息检索.



周栋(1979—), 男, 博士, 教授, 博士生导师, CCF 高级会员, 主要研究领域为自然语言处理, 信息检索.



段围(1994—), 男, 博士生, 主要研究领域为人工智能, 自然语言处理.



袁景凌(1975—), 女, 博士, 教授, 博士生导师, CCF 高级会员, 主要研究领域为机器学习, 分布式并行处理, 智能分析.