

# 基于 BERT 的盗窃罪构成要件识别方法<sup>①</sup>



费志伟<sup>1</sup>, 艾中良<sup>2</sup>, 张可<sup>1</sup>, 曹禹<sup>1</sup>

<sup>1</sup>(华北计算技术研究所, 北京 100083)

<sup>2</sup>(中国司法大数据研究院, 北京 100043)

通信作者: 费志伟, E-mail: jewelpei@126.com

**摘要:** 随着人工智能技术的发展, 人工智能技术在生活中被广泛使用, 并逐步深入到司法审理中. 但在实际应用中存在着可解释性不足, 不能有效的辅助审理这一问题. 针对这一问题, 本文结合刑事案件审理过程中依据犯罪构成采用的四要件理论, 从犯罪构成的四要件角度, 设计了构成要件识别任务. 筛选了盗窃罪中一些构成要件, 构建盗窃罪构成要件数据集. 并基于预训练语言模型 BERT (bidirectional encoder representations from transformers), 设计了构成要件识别模型, 对该模型在本文构建的数据集上进行测试, 模型识别准确率达到 93.54%. 在构成要件基础上构建量刑辅助算法能提高现有算法的解释性, 更有效的辅助法官审理案件.

**关键词:** 四要件; 盗窃罪; BERT; 文本分类; 深度学习; 自然语言处理

引用格式: 费志伟, 艾中良, 张可, 曹禹. 基于 BERT 的盗窃罪构成要件识别方法. 计算机系统应用, 2022, 31(4): 229-237. <http://www.c-s-a.org.cn/1003-3254/8428.html>

## Constitutive Elements Identification Method of Theft Crime Based on BERT

FEI Zhi-Wei<sup>1</sup>, AI Zhong-Liang<sup>2</sup>, ZHANG Ke<sup>1</sup>, CAO Yu<sup>1</sup>

<sup>1</sup>(North China Institute of Computing Technology, Beijing 100083, China)

<sup>2</sup>(China Justice Big Data Institute, Beijing 100043, China)

**Abstract:** With the development of artificial intelligence technology, it has been widely used in life and gradually penetrated judicial proceedings. However, there is insufficient interpretability in practical applications and thus it cannot effectively assist trials. In light of the four-element theory used in criminal case trials according to the constitution of a crime, this study addresses the above problem by designing an identification task of the four elements constituting a crime. Some constituent elements of crimes of theft are screened, and a data set of the constituent elements is constructed. Moreover, a constitutive elements identification model is developed on the basis of the pre-trained language model BERT and then tested on the data set constructed in this study, with the identification accuracy reaching 93.54%. Constructing an auxiliary sentencing algorithm based on the constituent elements can improve the interpretability of the existing algorithm and more effectively assist judges in hearing cases.

**Key words:** four elements; crime of theft; bidirectional encoder representations from transformers (BERT); text classification; deep learning; natural language processing (NLP)

近年来, 人工智能技术取得了长足的进步, 并深入到生活的方方面面, 在司法领域也是如此. 2018 年, 司法部印发《“十三五”全国司法行政信息化发展规划》, 明确提出我国到 2020 年全面建成智能高效的司法行

政信息化体系 3.0 版. 在顶层政策大力推动下, 全国各级法院参与研制了多种司法智能裁判辅助系统和装备, 例如, 北京市高级人民法院的“睿法官”智能研判系统、上海市法院的“上海刑事案件智能辅助办案系

<sup>①</sup> 基金项目: 国家重点研发计划 (2018YFC0832306, 2018YFC0831203, 2018YFC0831206)

收稿时间: 2021-07-05; 修改时间: 2021-07-30; 采用时间: 2021-08-12; csa 在线出版时间: 2022-03-22

统”、杭州互联网法院的“智能立案”系统、浙江省高级人民法院的小 AI 做庭审笔录等。

本文认为现有的司法领域中的人工智能系统可分为两种:

(1) 利用人工智能技术建立的信息化系统. 此类系统主要通过目前成熟的人工智能技术将案卷文书, 庭审信息等材料信息化, 录入到系统中. 如通过 OCR 技术, 识别审问讯问笔录录入系统. 通过语音识别技术将庭审过程中庭审语音转换成文字记录下来, 辅助记录员记录信息<sup>[1]</sup>等. 这种系统不涉及审理过程, 不需要结合法律相关知识. 直接利用现有成熟的人工智能技术, 在这些任务上效果也很好.

(2) 结合人工智能技术辅助法官审理的智能审判系统. 此类系统面向司法领域中的任务, 设计相关算法, 来辅助法官审理案件, 如量刑系统, 通过人工智能算法根据案件事实得出量刑结果. 这类系统需要面向司法领域的需求, 结合法律知识和计算机知识设计相应算法, 并设计相应的交互流程, 以法官判案逻辑为主导, 结合审理流程, 来实现智能化判案的效果.

目前利用人工智能技术建立司法信息化系统已经取得了很好的效果, 上述的庭审语音辅助系统切实的减少了记录员的工作量, 让法院庭审过程更加高效. 但在智能化辅助审理上效果不理想, 如类案推荐系统一些法官表示无法提供精准类案, 类案没有起到真正的参考价值. 在江苏智慧审判系统的应用情况来看, 部分法官甚至表示未使用该系统<sup>[2]</sup>. 本文认为目前智能化辅助审理系统主要面临着以下挑战:

计算机系统设计与司法领域知识结合不足. 现有的辅助审判算法在设计时未考虑司法判案过程, 未结合司法审判知识. 如 2018 年“法研杯”量刑预测任务中, 在算法设计时使用自然语言处理相关技术基于案件事实直接得到罪名以及判刑刑期结果, 与法院通常的审理流程不符, 及不具备法理上的解释性, 在刑期任务中结果仍有很大的提升空间<sup>[3]</sup>.

面对上述问题, 本文认为在实际量刑人工智能系统的构建中需要引入量刑理论, 依据司法判案中审理流程来构建相应算法. 在我国, 刑事案件判案过程中通常根据犯罪构成理论, 目前主流的犯罪构成理论为四要件理论与三阶层理论, 在实际司法审判中四要件理论使用更为广泛. 犯罪构成是指依照我国刑法规定, 决定某一行为的社会危害性及其程度而为该行为构成犯

罪所必须的一切客观和主观要件的有机统一<sup>[4]</sup>, 依据四要件理论, 主要包括犯罪客体、犯罪客观方面、犯罪主体和犯罪主观这 4 方面. 我国犯罪四要件理论起源于苏联, 司法实践至今仍在沿用这一理论. 在构建智能审判系统时应延续司法中四要件理论, 将四要件识别引入系统设计中, 在构成要件的基础上进一步来做刑期预测、类案推送等任务, 为法官提供司法上的解释, 来进一步辅助法官量刑.

本文主要贡献如下:

(1) 梳理了盗窃罪构成要件标签体系, 分析了构成要件识别的难点和挑战, 并详细分析了识别构成要件所需的前置条件和内容.

(2) 设计了构成要件识别模型, 利用机器学习技术来识别构成要件, 包括数据集的构建和构成要件模型的设计.

(3) 对比了常见的方法与本文设计的方法, 在本文构建的数据集上, 对现有方法做了充分的测试.

本文设计的构成要件识别任务符合当下司法审理流程, 在构成要件的基础上进一步来做刑期预测、类案推送等任务能提高现有方法的可解释性, 能更加有效的辅助法官审理案件.

## 1 相关工作

案件构成要件识别根据构成要件理论, 从案件事实中识别出构成要件. 目前在构成要件识别中的工作较少, 与之有一定关联的是案情关键要素识别, 在计算机任务上可以看做文本分类任务. 在案件要素识别方面主要有 CAIL2019 法研杯提出的案情要素识别任务, 该任务在案情描述中重要事实描述基础上, 识别案情要素. 基于速裁案件要素式审判的理论, 法律专家梳理了婚姻家庭、劳动争议和借款合同 3 个领域的案情要素体系. 该数据集包含民事案件内容, 不包括刑事案件内容. 王得贤<sup>[5]</sup>提出基于层次注意力的模型在该数据集上进行测试. 刘海顺等人<sup>[6]</sup>提出编码器解码器结构, 利用 BERT 对文本内容进行编码, 提出 BERT 模型后三层参数融合策略, 然后使用 LSTM 作为解码器得到最终的预测结果. 其他工作在可解释性上, 也使用了案情要素这一概念, Devlin 等人<sup>[7]</sup>构建了一个可解释的智能量刑模型, 该模型一定程度上能解释模型如何确定其最终输出. 但机器关注的关键部分与司法审理中关注的焦点并不相同, 现有的通过注意力机制等方法

提取出的案件要素并不具备司法上的解释性。钟皓曦等<sup>[8]</sup>基于强化学习和深度学习,提出通过问答的方式,获取案件事实中的影响定罪的元素,并通过这些元素来确定最终量刑并提供一定的解释性。但是该文并未详细阐述如何筛选案件要素,在单个罪名下的案件要素量少,该文选出的案件要素在司法上的支撑薄弱,在实际审理中的参考价值模糊。在实际司法中需要考虑更多的信息,按照本文所述的构成要件来进行审理。

案件构成要件识别的难点在于不同的案件构成要素不同,在梳理时需要法律人士参与,需要针对每个罪名梳理对应的构成要件体系。其次是利用大数据技术或深度学习技术依赖大规模标注数据集,如何利用现有的公开信息,构建一个构成要件数据集,减少人工标注工作量也是案件构成要件识别面临的挑战。

## 2 盗窃罪构成要件分析

### 2.1 盗窃罪构成要件

根据刑法第二百六十四条,盗窃罪是指以非法占有为目的,盗窃公私财物数额较大或者多次盗窃、入户盗窃、携带凶器盗窃、扒窃公私财物的行为。审理一起盗窃案件时,根据犯罪构成中四要件理论,要判断一起案件的主体要件、主观要件、客体要件和客观要件这4个构成要件是否存在。主体要件刻画了行为人是否具有刑事责任能力,主观要件分析了行为人对自身实施的危害社会的行为及其结果所持的心理态度,客体要件描述犯罪侵犯的法益,客观要件描述了犯罪行为事实。

本文详细梳理了盗窃罪的构成要件部分,并阐述了构成要件与法院认定事实之间的关系。在盗窃罪主体要件中主要从主体行为人是否具有刑事责任能力,是否为未成年人,是否有犯罪前科和是否为中国居民这些方面来判断。

主观要件可从5点进行考察,故意、非法占有为目的、牟利为目的、对特殊情节的明知和转化为其他罪名。故意主要从犯罪嫌疑人有主动参与作案的动机、犯罪过程中有无策划、是否事先通谋、事后销赃的行为人,通谋的内容包括盗窃行为和有无共犯这4个方面进行识别。非法占有为目的的对他人财物的明知和对盗窃后果的明知两点进行识别。牟利为目的,主要从盗窃他人通信线路、复制他人电信码或者明知是盗窃、复制的电信设备、设施使用来判断。对特殊情

节的明知,可依常识推论犯罪嫌疑人是否知道被害人是残疾人、孤寡老人或者丧失劳动能力的人、被盗财物是否为珍贵文物,或者救灾、抢险、防汛、优抚、扶贫、移民、救济款物,盗窃地点是自然灾害、事故灾害、社会安全事件等突发事件期间的事件发生地。转化为其他罪名从犯罪嫌疑人对行为对象的性质、功能等特征是否存在明确的认知来判断是否构成特别罪名。

客体要件从他人占有的财物和价值来考察,他人占有的财物主要从实践中的表现和学理解释来认定,如被害人对所盗财物拥有合法权利。学理解释根据法学理论判断客体的一些属性如是否为遗忘物,基于委托关系的占有等。价值在确定刑事处罚时有重要意义,根据被盗物品价值确定案件的基准刑以及在相应的量刑格中确定增减刑。

客观要件从客观行为事实来对案件进行考量,不同的客观行为事实会影响盗窃罪的认定。在盗窃罪的客观行为事实中需要考虑时间、地点、犯罪参与人、动机、手段、方法、被害人、行为对象、情节和后果这些方面。

法官依据证据以构成要件为指导归纳、认定案件事实,在司法证据中识别案件构成要件,并最终归纳总结出案件事实,最终认定的案件事实包含这起案件所涉及的全部构成要件。盗窃罪构成要件标签体系如表1中所示。

表1 盗窃罪构成要件标签体系

构成要件	标签
主体要件	16周岁以上
	具备刑事责任能力
	国籍的认定 前科
主观要件	故意
	非法占有为目的
	牟利为目的 对特殊情节的明知 转化为其他罪名
客体要件	他人占有的财物 价值
客观要件	客观行为事实 转化为其他犯罪

本文选取了客观要件作为识别对象,进一步分析客观行为事实,筛选识别的要件内容。在案件审理中,法官以客观要件中的客观行为事实来判定罪名以及量

刑.在上述客观行为事实中主要考察时间、地点,犯罪参与人、动机、手段方法、被害人、行为对象、情节和后果这些内容,其中手段方法在审理中对案件的罪名有关键的作用,是判断一起案件是否构成盗窃罪,区别此罪与比罪的核心.

审理过程中犯罪情节和后果对最终处罚的基准刑有着关键作用.参考《最高人民法院量刑规范化的指导意见》第三节盗窃罪相关量刑基准中涉及到的情节,本文认为在情节和后果中可将盗窃情节分为一般盗窃行为、量刑从轻、减轻的情节和从重处罚情节这3个方面.一般盗窃行为中根据盗窃数额来判断对该案件处以多重的刑罚.如设立盗窃财物金额标准,划分数额较大,数额巨大和数额特别巨大这几档来确定基准刑.对被盗物品和时间以及行为人和后果进行划分,可设立量刑从轻、减轻情节,如盗窃近亲属财物,初犯、偶犯,未成年人犯罪等.对盗窃行为进行划分可确定从重处罚情节,如多次盗窃,入户盗窃,教唆未成年人盗窃等.

## 2.2 盗窃罪构成要件识别任务

本文提出的构成要件识别任务根据法院认定的犯罪事实来识别包含的构成要件.在形式上通过给系统输入案件的事实描述部分,得到该事实中包含的构成要件标签.给定法官认定的事实句子序列 $X = \{x_1, x_2, x_3, \dots, x_m\}$ ,预测与 $X$ 对应的构成要件标签集合 $\hat{Y} \subseteq Y, \dots$ 其中 $m$ 是序列 $X$ 的长度, $x_i$ 表示序列中的第 $i$ 个词. $Y = \{y_1, y_2, y_3, \dots, y_n\}$ 为构成要件标签集合. $n$ 为构成要件类别总数,一个案件事实至少对应一个构成要件标

签,可能对应多个构成要件标签,所以 $\hat{Y}$ 是 $Y$ 的子集.

## 2.3 盗窃罪构成要件数据集构建

在构建数据集时,本文充分利用公开信息,在上述分析基础上,首先确定构成要件标签,之后在裁判文书网上筛选相应的数据作为本文的训练和测试数据.主要方法如下:

(1) 本文在盗窃罪客观构成要件客观行为事实中情节和后果的基础上,依据无需金额标准的情形构建数据集,无需金额标准的情形主要包含四类,依据《最高人民法院、最高人民检察院关于办理盗窃刑事案件适用法律若干问题的解释》第三条,分别是多次盗窃,入户盗窃、携带凶器盗窃和扒窃.选取入户盗窃、携带凶器盗窃、多次盗窃和扒窃作为本文分析的盗窃罪构成要件中客观要件需要识别的部分.并加入其他标签来区分其他构成要件和其他情形.

(2) 从中国裁判文书网<sup>[9]</sup>,通过下载裁判文书网上刑事案件中的盗窃罪一审判决书,分析判决书中本院认为部分,查找上述分析出的关键词.图1所示,一则盗窃罪案例中,本院认为部分中包含“以非法占有为目的,在公共场所扒窃他人财物,其行为已构成盗窃罪”.从中获取关键词“扒窃”.将该文书的认定事实部分、被告人基本信息部分抽取出来,作为数据集中文本的内容.

上述构成要件识别主要依据认定的犯罪事实.无需个人信息、过往犯罪经过等信息,本文在构建数据集时,将法院裁判文书中认定的事实部分作为数据输入,将构成要件作为标签.盗窃罪构成要件数据实例如表2所示.



图1 裁判文书中构成要件示例

## 3 基于BERT的构成要件识别模型

### 3.1 模型整体流程

本文先构建盗窃罪构成要件数据集,然后基于预训练模型设计构成要件识别模型,在本文构建的盗窃

罪构成要件数据集上进行训练,并取验证集上最优模型作为最终的模型,整体流程如图2所示.

### 3.2 BERT-BiLSTM-Att模型

本文提出基于BERT的盗窃罪构成要件识别模型

首先通过 BERT 获取语句的向量表示,之后运用 BiLSTM 模型提取句特征,并结合注意力机制进一步

获取对分类结果有重要影响的特征.具体流程如下:

(1) 通过预训练语言模型获取句向量.预训练模型的大小会影响下游任务的效果.本文选用了 BERT-base 模型,能接受的最长文本长度  $m = 512$  字符.隐藏层维度  $d = 768$ ,编码层层数为  $l = 12$ .在预训练模型上

使用了目前最新的公开中文预训练数据集.

(2) 构建盗窃罪客观要件识别模型.使用 BERT 最后一层输出向量作为文本表示,并拼接 BiLSTM-Att 模型,下游任务设置为多分类任务,通过下游任务对预训练神经网络进行微调.具体结构如图 3 所示,将得到的文本向量送入 BiLSTM-Att 模型中,通过该模型识别盗窃罪构成要件并在最后使用交叉熵损失函数进行训练.

表 2 盗窃罪构成要件数据集示例

案件事实描述	构成要件标签
2018年6月10日凌晨1时许,被告人吉某某来到安顺市西秀区东屯乡高关村胡某某家小卖铺,徒手将被害人胡某某家窗户护栏掰开后翻进其家中...	入户盗窃
2018年11月7日晚,被告人王某某、陈某某采用镰刀割取的方式,窃得浙江...	携带凶器盗窃
2019年12月19日10时许,被告人张某甲在荣成市**镇**村集市上,趁连某某不备,扒窃其放在上衣口袋内的...	扒窃
2012年1月至2019年8月,被告人徐宁至四川省会理县绿水镇、...、杭州市xx,实施多起盗窃,具体犯罪事实如下...	多次盗窃
2018年8月12日凌晨,到泉州市洛江区公共文化活动中心大楼盗窃品牌均为太阳牌的五芯铜芯电缆线64.8米...	其他

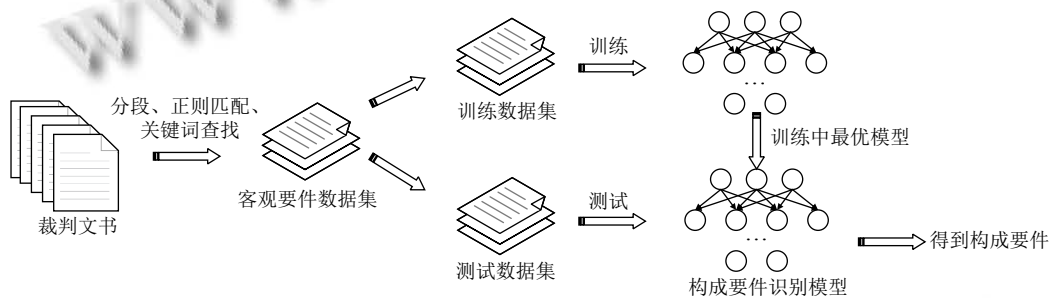


图 2 构成要件识别流程图

在训练时设置学习率为  $3e^{-5}$ ,参数优化使用 BERT-Adam 优化器.训练 epoch 为 60, batch-size 大小为 16.

### 3.3 BERT 模型获取句向量

BERT 模型<sup>[10]</sup>在双向 Transformer<sup>[11]</sup>编码器的基础上实现,其中每个 Transformer 编码单元由 6 个 Encoder 堆叠在一起,Transformer 编码器单个 Encoder 架构图如图 4 所示.

图 4 中  $N$  代表编码器层数,一个 Transformer 编码器包含两层,一个是多头的自注意力层,另一个是前馈神经网络层.多头自注意力层中自注意力机制能弥补循环神经网络面临的长依赖问题,不仅关注当前几个词,能够获取更长的全文信息.并通过多头的方式获取不同的交互关系.

自注意力机制的可以看做在一个线性投影空间中建立模型输入中不同向量之间的交互关系.自注意力

机制的运算过程中,首先会计算出 3 个新的向量:  $Q$  (query),  $K$  (key),  $V$  (value), 这 3 个向量是词嵌入向量与一个矩阵相乘得到的结果,该矩阵是随机初始化的维度为  $(64, 512)$  的矩阵.当输入一个句子时,该句子中的每个词都与其他词进行 Attention 计算, Attention 的计算公式如下:

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

其中,  $d_k$  表示每个字的 query 和 key 向量的维度,  $\text{Softmax}()$  是归一化指数函数.最终得到的 Attention 值是一个矩阵值,矩阵值的每一行代表输入句子中相应字的 Attention 向量,其中包含了句子中该词和其他位置的词的相互关系信息,是一个新的向量表示.由此,我们可以看到, BERT 模型使用带有自注意力机制的双向 Transformer 模型获得了句子的前后语义关系,从而更好地获得了一个句子的语义表达.

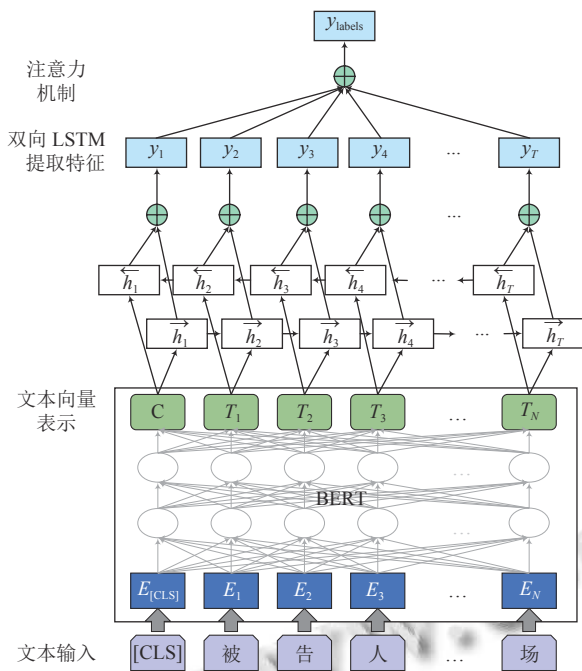


图3 BERT-BiLSTM-Att 模型结构图

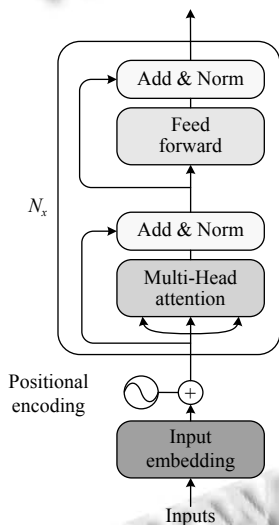


图4 Transformer encoder 模块结构图

在预训练过程中, BERT 使用 MLM (masked language model) 任务和 NSP (next sentence prediction) 任务进行预训练. 一般使用 BERT 做文本分类任务时使用 BERT 最后一层池化后的输出, 在其基础上使用全连接层和交叉熵损失函数进行训练, 如图 5 所示. 文本选取 BERT 输出作为文本向量表示.

### 3.4 BiLSTM-Att 分类模型

LSTM (long short-term memory network)<sup>[12,13]</sup> 是循环神经网络的一种变体, 通过引入门控机制来控制信

息的累计速度, 包括有选择地加入新的信息, 并有选择地遗忘之前累积的信息来改善循环网络的长依赖问题. BiLSTM-Attention<sup>[14]</sup> 模型在 LSTM 基础上, 以其作为基础的网络层, 通过增加一个逆句子顺序的网络层, 来获取一个词的上下文关系, 增强网络的表示能力. 并在获取的表示后加入注意力机制来更好的获取关键的信息.

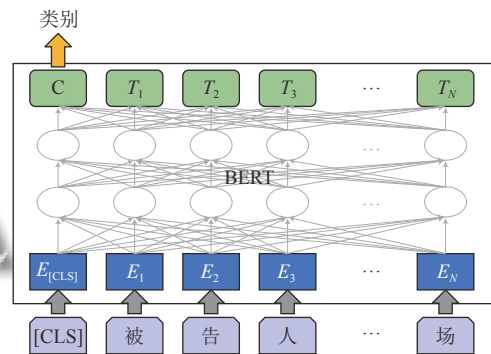


图5 BERT 文本分类模型

BiLSTM-Att 模型如图 6 所示, 其中  $e_t$  为词向量,  $\vec{h}_t$  为某一顺序上 LSTM 在该时刻的向量,  $y_t = [\vec{h}_t \oplus \overleftarrow{h}_t]$ , 为将两个向量拼接后的向量, 最后将  $y_t$  通过注意力层得到该句子的表示. 在得到的句子表示上加入 Attention 层之后为 BiLSTM-Attention 模型.

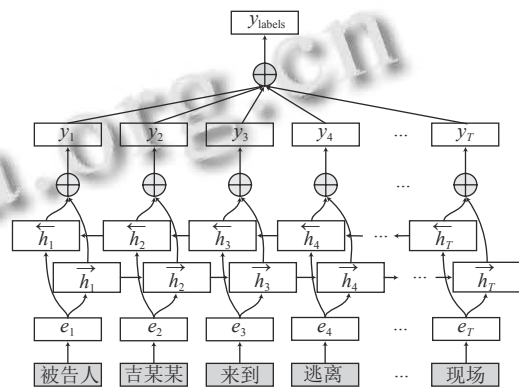


图6 BiLSTM-Att 文本分类模型

## 4 实验结果分析

### 4.1 测试数据

本文从裁判文书网下载盗窃罪一审判决书 1 万份, 通过正则的方式获取文书中案件事实部分, 之后对判决书提取判决结果. 通过本院认为部分, 提取判决结果中包含上述标签的案件构建 2400 条数据, 筛选标签示例如图 7 所示, 样例数据如表 2 所示.



图7 判决书中标签示例

通过对案件事实和标签进行统计分析, 本文构建的数据集在文本长度上按字符统计, 平均文本长度为235字, 最大文本长度653字, 最小文本长度为124字. 统计相应的占比, 本文构建的数据集中其他文书600份, 入户盗窃587份, 携带凶器盗窃122份, 扒窃836份, 多次盗窃483份. 占比如图8所示.

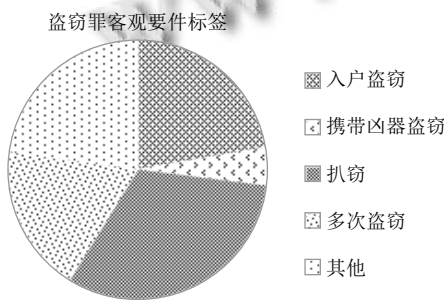


图8 各标签占数据集比例

### 4.2 测试环境

实验硬件设备如表3所示.

实验环境	具体配置
操作系统	CentOS 7.7
CPU	Intel(R) Xeon(R) CPU E5-2640 v4 @ 2.40 GHz × 8
GPU	NVIDIA TITAN RTX × 4
PyTorch	1.5.1
内存容量 (GB)	256

### 4.3 评价指标

将数据集划分为2000条作为训练数据, 200条作为评测在200份测试数据中进行, 评测上述所有的标签分类结果, 从精度 (*precision*), 召回率 (*recall*),  $F_1$ 值3方面评测算法的结果.

精度 (*precision*) 是指标记为正类的元组实际为正类的百分比, 计算方法为:

$$precision = \frac{T_p}{T_p + F_p}$$

其中,  $T_p$ 为被分类器正确分类的正元组个数,  $F_p$ 为错误标记为正元组的负元组个数.

召回率计算方法为:

$$recall = \frac{T_p}{T_p + F_n}$$

其中,  $T_p$ 为模型预测出的标签正确的标签个数,  $F_n$ 为被错误标记为负元组的正元组个数, 即假负例个数.

$F_1$ 值计算方式为:

$$F_1 = \frac{2 \times precision \times recall}{precision + recall}$$

其中, *precision*为上述精度, *recall*为上述召回率.

对上述3个指标从micro指标进行考察, 计算方式如下:

$$B_{micro} = B \left( \sum_{i=1}^q tp_i, \sum_{i=1}^q fp_i, \sum_{i=1}^q tn_i, \sum_{i=1}^q fn_i \right)$$

其中,  $B(\cdot)$ 为计算*precision*, *recall*和 $F_1$ 算符.

### 4.4 结果分析

数据集划分为训练数据2000条, 验证数据200条, 测试数据200条. 实验对比了目前文本分类中常用的模型和最新的模型. 对比了多个预训练模型, 主要有科大讯飞开源的中文预训练语言模型<sup>[15]</sup>, 和清华大学开源的司法数据上的预训练语言模型.

常用的模型比较了LSTM和基于LSTM的一些改进模型, 主要有LSTM、BiLSTM和BiLSTM-Att. 模型输入的词向量本文使用Google团队发布的Word2Vec工具<sup>[16]</sup>, 通过在CAIL2018罪名预测数据集<sup>[17]</sup>的事实部分作为训练语料库, 使用结巴分词将语料分词后进行训练. Word2Vec中选择的方法为CBOW, 该方法通过中心词周围的词来预测中心词. 预训练词向量维度为300维, 训练设置的相关参数如表4所示.

表4 Word2Vec参数设置

Window size	Dim size	Low-frequency word	Iteration	Negative sampling
5	300	10	10	5

主要比较的预训练模型如下:

**BERT-xs:** 该预训练模型在663万篇刑事文书上进行预训练,未采用全词覆盖训练策略,训练时以字为粒度进行切分。

**BERT-wwm:** 该预训练模型在中文维基百科上进行训练,采用全词覆盖训练策略,训练时一个完整的词的部分子词被覆盖,则同属该词的其他部分也会被覆盖。

**BERT-wwm-ext:** 该预训练模型在上述预训练方法的基础上增加了数据,其中EXT数据包括:中文维基百科,其他百科、新闻、问答等数据,总词数达5.4B。

**RoBERTa-wwm-ext:** 使用RoBERTa并使用上述的训练策略和训练数据,将模型换为RoBERTa, RoBERTa相比于原始的BERT做了如下改进:训练时间更久,并增大了batch size;移除了BERT预训练任务中的NSP任务;训练了更长的序列和动态调整mask策略。

经计算得到本文使用的模型 $precision_{micro}$ 为93.54%,  $recall_{micro}$ 为95.75%,  $F_{1micro}$ 为94.63%。

表5中BiA表示BiLSTM-Att部分, BERT-xs-BiA代表使用BERT-xs预训练模型得到文本句向量,再送入BiLSTM-Att中做分类,识别构成要件。BERT-xs表示只使用BERT做分类,识别构成要件,其他模型标识同理可得。

表5 测试结果(%)

模型	精度	召回率	$F_1$
LSTM	87.08	85.84	86.45
BiLSTM	87.62	86.79	87.20
BiLSTM-Att	89.52	88.67	89.09
BERT-xs	90.47	89.62	90.04
BERT-wwm	89.72	90.56	90.13
BERT-wwm-ext	91.58	92.45	92.01
RoBERTa-wwm-ext	92.62	94.81	93.70
BERT-xs-BiA	91.43	90.56	90.99
BERT-wwm-BiA	90.65	91.51	91.07
BERT-wwm-ext-BiA	92.99	93.86	93.42
RoBERTa-wwm-ext-BiA	<b>93.54</b>	<b>95.75</b>	<b>94.63</b>

从实验结果可以看到,相比于BiLSTM和带注意层层的BiLSTM, BERT预训练模型优于传统的模型,这表明BERT模型能够在训练数据较少的情形下,通过面向下游任务进行微调,取得较好的结果。

通过BERT-xs和BERT-xs-BiA对比和其他预训练模型间比较可知使用了BERT提取句向量做为输入比使用Word2Vec训练得到词向量得到的结果好。这表明通过BERT模型获取的文本向量能够更好的表示文本,通过与BiLSTM-Att结合能进一步提高识别效果。

在基于预训练语言模型的对比中,预训练语言模型的选择也相当重要,选择合适的预训练模型能提高一定的准确率,如BERT-wwm、BERT-wwm-ext和RoBERTa-wwm-ext模型三者模型大小相差不大,使用RoBERTa-wwm-ext能提高一定的准确率。在训练时BERT-xs能够更快的学习到司法任务相关的内容,在前几个epoch结果优于其他模型。但在最终结果上并非最优,本文认为这与下游任务有关,在司法文本分类任务上通用语料库上预训练的语言模型能达到司法文本上预训练的语言模型的结果。

## 5 总结

本文设计了盗窃罪构成要件识别任务,结合量刑理论,提出结合构成要件的司法智能系统构建思路。从案件审理的四要件角度,详细梳理了盗窃罪的构成要件和识别该要件所需的前置条件。之后构建了首个盗窃罪的构成要件数据集,从公开数据上利用搜索和正则匹配等方式筛选数据,构建了数据集。最后设计了基于BERT模型的构成要件识别模型,对该数据集进行分类,并测试了相关结果。在本文构建的数据集上,该模型达到了93.54%的准确率,优于传统模型。本文提出的构成要件识别任务有很强的司法理论支撑,能够指导规范案情要素识别的内容,并且在本文工作基础上构建智能审判相关算法,能够更好的为法官提供指引,有很强的实际意义。

## 参考文献

- 陆淳. “小庭审”展现“大智慧”. 人民法院报, 2021-03-20(002). [doi: 10.28650/n.cnki.nrmfy.2021.001177]
- 左卫民. 热与冷: 中国法律人工智能的再思考. 环球法律评论, 2019, 41(2): 53-64.
- Zhong HX, Xiao CJ, Guo ZP, et al. Overview of CAIL2018: Legal judgment prediction competition. arXiv: 1810.05851, 2018.
- 高铭喧, 马克昌. 刑法学. 北京: 北京大学出版社, 2018: 52.
- 王得贤. 法律文书中的要素识别方法研究 [硕士学位论文]. 太原: 山西大学, 2020.



- 6 Liu HS, Wang L, Chen YG, *et al.* A method for case factor recognition based on pre-trained language models. Proceedings of the 19th Chinese National Conference on Computational Linguistics. Haikou: Chinese Information Processing Society of China, 2020. 743–753.
- 7 Jiang X, Ye H, Luo ZC, *et al.* Interpretable rationale augmented charge prediction system. Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations. Santa Fe: Association for Computational Linguistics, 2018. 146–151.
- 8 Zhong HX, Wang YX, Tu CC, *et al.* Iteratively questioning and answering for interpretable legal judgment prediction. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(1): 1250–1257. [doi: [10.1609/aaai.v34i01.5479](https://doi.org/10.1609/aaai.v34i01.5479)]
- 9 中国裁判文书网. <https://wenshu.court.gov.cn/>.
- 10 Devlin J, Chang MW, Lee K, *et al.* BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv: 1810.04805, 2018.
- 11 Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need. arXiv: 1706.03762, 2017.
- 12 Graves A. Supervised Sequence Labelling with Recurrent Neural Networks. Berlin, Heidelberg: Springer, 2012. 1735–1780.
- 13 Hochreiter S, Schmidhuber J. Long short-term memory. Neural Computation, 1997, 9(8): 1735–1780. [doi: [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735)]
- 14 Zhou P, Shi W, Tian J, *et al.* Attention-based bidirectional long short-term memory networks for relation classification. Proceedings of the 54th annual meeting of the association for computational linguistics (Volume 2: Short Papers). Berlin: Association for Computational Linguistics, 2016. 207–212.
- 15 Cui YM, Che WX, Liu T, *et al.* Revisiting pre-trained models for Chinese natural language processing. arXiv: 2004.13922, 2020.
- 16 Mikolov T, Chen K, Corrado G, *et al.* Efficient estimation of word representations in vector space. arXiv: 1301.3781, 2013.
- 17 Xiao CJ, Zhong HX, Guo ZP, *et al.* CAIL2018: A large-scale legal dataset for judgment prediction. arXiv: 1807.02478, 2018.