

# 基于案件属性感知的案件倾向性抽取式摘要

闫婉莹<sup>1,2</sup>, 郭军军<sup>1,2</sup>, 余正涛<sup>1,2</sup>, 田应飞<sup>1,2</sup>

1昆明理工大学信息工程与自动化学院 昆明 650504

2昆明理工大学云南省人工智能重点实验室 昆明 650504

## 引言

涉案新闻文本摘要是从法院舆情数据文本中提取与案件相关的文本信息。面向涉案新闻的文本摘要方法受限于法律领域知识及开放数据集规模，其技术难点在于涉及法院新闻的数据包含大量的案件细节及案件属性词。

针对涉案新闻文本的差异性，本文提出一种基于案件属性感知的案件倾向性抽取式文本摘要网络架构，该网络结构主要是案件倾向性预测和文本摘要任务联合学习的模型，同时学习和预测每个句子的案件相关性，从而抽取出文档的候选摘要句。

## 模型

模型由三个模块组成：案件信息库编码器，案件属性感知文档编码器和句子提取器。

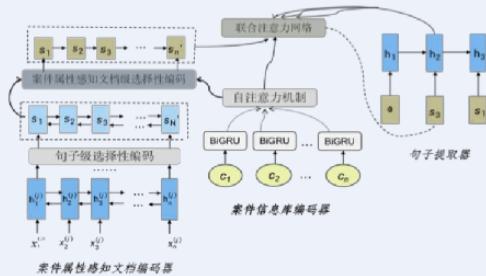


图1 案件属性感知的案件倾向性抽取式摘要模型

## 方法

④ 案件信息库编码器模块：通过BiGRU网络获得案件信息库每个单词的隐层向量并利用self-attention layer获得案件信息库的多语义信息的表征；

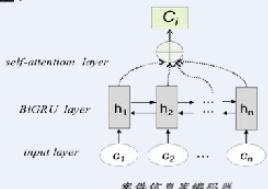


图2 案件信息库编码模型

④ 案件属性感知文档编码器模块：主要由句子级选择性编码、案件属性感知文档级选择性编码、案件倾向性联合注意力网络三个网络模块组成，如下图3所示；

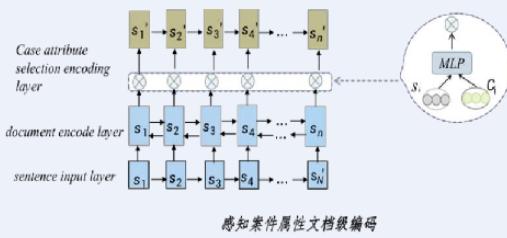


图3 案件属性感知文档编码模型

➤ 句子级选择性编码是对原文句子依次进行编码并选择重要信息获得句子级向量的语义表征；

➤ 案件属性感知文档级选择性编码模块首先

通过BiGRU网络获得文档级语义向量，然后通过感知案件属性选择编码策略将特定案件信息库已获得的语义向量与文档级语义向量共同学习得到案件属性感知的文档级语义向量；

➤ 案件倾向性联合注意力网络将案件属性感知文档级选择性编码向量表示与案件信息库编码语义表征进行联合学习，得到更倾向于案件信息的文档表示。

④ 句子提取器模块：我们使用联合打分与选择框架进行摘要的提取并使用KL损失函数优化模型。

## 实验

本文采用Rouge作为评价指标，本文模型Ours与五个baseline在涉案新闻文本数据集的对比实验如下表所示：

表3.涉案新闻文本数据集实验.

Model	Rouge-1	Rouge-2	Rouge-L
Lead3	26.71	27.10	24.69
TextRank	16.04	26.39	14.56
Nmf	27.64	19.43	22.08
SummaRunner	27.23	19.34	21.29
NeuSum	27.83	18.15	18.78
Ours	30.31	21.06	21.06

实验结果表明，在涉案新闻文本数据集中，文章的主要信息都在开头，并且对于新闻文本数据来说，主题信息可以提升文本摘要的性能。我们提出将案件信息库作为外部知识融入的感知案件属性的案件倾向性文本摘要模型可以较好的提高模型的摘要性能，能使抽取出的摘要更贴近涉案新闻文本主题，相关性更高。

➤ 消融实验. (1) 融合案件信息库的有效性分析.

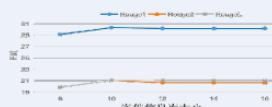
表4. 网络结构实验.

Model	Rouge-1	Rouge-2	Rouge-L
Ours_case	28.95	19.75	19.43
Ours_(select_case)	29.80	20.22	20.32
Ours	30.31	21.06	21.06

Ours\_case代表本文模型只加入案件信息库的语义表征,Ours\_(select\_case)意指模型对案件信息库进行语义的表征与对涉案新闻文本进行案件属性感知的文档级选择性编码并表征文档的语义向量.

(2) 案件信息库大小性能分析.

表5.不同案件信息库大小实验.



案件信息库大小主要是从涉案新闻文本与标题分别抽取不同关键词并进行数据的清洗组成。

## 结论

针对涉案新闻文本摘要任务，本文模型考虑案件信息属性和全局上下文信息，将案件信息库融入到文本摘要中，利用与涉案新闻文本的相关性，设计了句子级选择性编码与感知案件属性的文档级选择性编码网络结构完成特定领域数据的语义向量表示。在涉案新闻文本数据集上，加入案件信息库可以有效的提升新闻摘要句与领域的相关性。