

# 基于多模板模糊竞争的涉案财物关系抽取方法

李攀锋, 林 锋, 蒋宗神

(四川大学计算机学院, 成都 610065)

**摘 要:** 从法律法规中自动抽取涉案财物处置关系并构建知识库, 对司法实践中涉案财物的智能管理有着重要意义. 但该场景下训练语料少且单一以及抽取准确率要求高, 机器学习的关系抽取方法不适用. 针对这个问题, 基于模板匹配的关系抽取思想, 结合模糊逻辑机制, 研究了一种适用于涉案财物关系抽取的方法. 该方法考虑了涉案财物知识库的实际需求, 在传统三元组关系的基础上增加了二元属性, 扩充为五元组关系模式, 从不同维度设计了三个抽取模板, 并借助模糊逻辑对抽取的结果进行评判, 竞争出效果较优的五元组关系, 为涉案财物知识库的构建提供了支持.

**关键词:** 关系抽取; 模板匹配; 模糊逻辑; 涉案财物

**中图分类号:** TP391 **文献标识码:** A **DOI:** 10.19907/j.0490-6756.2021.042002

## Relation extraction method for property involved based on multiple templates and fuzzy competition

LI Pan-Feng, LIN Feng, JIANG Zong-Shen

(College of Computer Science, Sichuan University, Chengdu 610065, China)

**Abstract:** Extracting the relation of the property involved from laws and regulations to build a knowledge base is of great significance to the intelligent management of the property involved in the judicial practice. However, in this scenario, the training corpus is small and single, while high extraction accuracy required, so the relation extraction method of machine learning is not applicable. Aiming at addressing this problem, we propose a new relation extraction approach that extracts relations by template matching with fuzzy logic mechanism. In this approach, in order to cater to the actual needs of the knowledge base for property involved, two attributes are added to the traditional three-tuple relation to obtain the five-tuple relation model. Three extraction templates were designed from different dimensions, and the extraction results were judged with the help of fuzzy logic to compete for a better five-tuple relation, which provided support for the construction of the knowledge base for property involved.

**Keywords:** Relation extraction; Template matching; Fuzzy logic; Property involved

## 1 引 言

关系抽取是知识工程领域的重要任务之一,也是知识图谱构建的核心步骤. 它的目的在于从无结

构的自然语言文本中抽取出结构化的知识, 得到文本内含的语义关系, 进而用于知识库的构建、智能问答、推荐系统等.

目前常见的关系抽取方法主要分为三种: (1)

收稿日期: 2020-10-29

基金项目: 国家重点研发计划 (2018YFC0832300, 2018YFC0832303)

作者简介: 李攀锋(1996—), 男, 四川宜宾人, 硕士研究生, 研究方向为自然语言处理. E-mail: lipanfeng@stu.scu.edu.cn

通讯作者: 林锋. E-mail: linfeng@scu.edu.cn

基于模板匹配的方法<sup>[1]</sup>; (2) 基于监督学习的方法<sup>[2]</sup>; (3) 基于半监督或无监督的方法。

随着机器学习和深度学习的发展, 基于监督学习和半监督学习的关系抽取方法应用十分广泛<sup>[3]</sup>。Zeng 等<sup>[4]</sup>在 2014 年首次提出采用卷积神经网络来提取句子的语义特征, 进而实现关系抽取; Zhang 等<sup>[5]</sup>在 2015 年采用了双向长短期记忆网络并借助依存分析和命名实体识别来解决关系抽取问题; 谌予恒等<sup>[6]</sup>在 2020 年采用了结合注意力机制与残差网络进行远程监督关系抽取。这些方法大多用于处理如人物关系之类的关系抽取问题, 因为在这种场景下, 关系类别明确、训练语料丰富<sup>[7]</sup>。但本文提出的关系抽取方法是为了解决涉案财物知识库构建的特定问题, 即是立足于抽取“业务单位实体”与“财物实体”之间存在的“处置方式”关系。受限于涉案财物知识库场景训练语料少且单一以及识别准确率要求高的特点, 机器学习的关系抽取方法在涉案财物领域并不适用, 而采用基于模板匹配的方法更为合理。

本文提出的基于多模板模糊竞争的涉案财物关系抽取方法, 根据涉案财物处置的实际场景需求, 在常规三元组关系的基础上扩充了“财物状态”和“处置条件”两个属性元素, 定义为五元组关系, 以不同维度设计了三个关系抽取模板, 从法律法规文本自动抽取五元组关系, 并运用模糊逻辑计算各个模板抽取结果的置信度, 使得三个模板间相互竞争, 进一步提高抽取结果的准确度。

## 2 基本概念与任务描述

### 2.1 涉案财物知识库

涉案财物知识库指在根据现有法律法规自动完成刑事案件中涉案财物处置的相关知识融合, 为司法实践中公检法等执法司法单位的办案人员提供支持<sup>[8]</sup>。核心的工作是从法律法规出发, 抽取“业务单位实体”与“财物实体”之间的处置关系。由于该领域的特殊性, 在创建知识库过程中完成关系抽取面临着挑战如下:

(1) 训练语料少且单一。涉案财物知识库的构建目标是基于法律法条自动完成知识抽取, 语料主要来源于正式实施的法律法规中与涉案财物处置相关的法条, 这造成进行关系抽取时训练语料不仅远远少于通用知识库, 也少于一般的法律知识库。

(2) 识别准确率要求高。涉案财物知识库的应用目标是为司法实践中的一线办案人员提供支持,

这对知识库中知识的准确性提出了极高的要求。为保证知识库的正确性和减少后续工作, 关系抽取算法的抽取准确率应尽可能提高。

上述问题使得涉案财物知识库构建过程中的关系抽取区别于一般关系抽取, 成为了一个独特的、挑战性的问题。

### 2.2 涉案财物关系概述

实体关系三元组是由头实体、尾实体、实体间关系组成<sup>[9]</sup>。在人物关系领域, 即是形如[“姚明”, “叶莉”, “夫妻”]这样的人物关系三元组, 其中“姚明”为头实体, “叶莉”为尾实体, “夫妻”为实体间关系<sup>[10]</sup>。

在涉案财物领域, 关系三元组稍有不同。在法律法规文本中, 财物实体和业务单位实体直接存在着某种处置关系, 如[“公安机关”, “涉案财物”, “扣押”], 其中“公安机关”为头实体, “涉案财物”为尾实体, “扣押”关系同时作为一种处置方式, 本质上是“公安机关”作为主语, 对宾语“涉案财物”实施“扣押”动作。由此形成了包括“业务单位实体”、“处置方式”、“财物实体”的三元组关系模式, 对应于司法实践的具体任务, 即是公安机关在办理各类刑事案件以及检察机关在职务犯罪案件侦查过程中, 对与案件有关的物品、款项等依法进行扣押、查封、冻结等操作。

基于法律法规的涉案财物关系抽取即是从法律法规文本中抽取涉案财物处置的规则信息。这种规则信息对于司法实践中涉案财物的智能管理有着重要的意义。考虑到司法实践的严谨性, 单纯的三元组信息并不能较好地反映业务单位执行涉案财物处置的具体情形。因此, 本文在传统三元组关系的基础上, 针对涉案财物处置实际场景, 为“财物实体”增加了“财物状态”属性, 为“处置方式”增加了“处置条件”属性, 形成了形如[“业务单位实体”, “处置方式”, “处置条件”, “财物实体”, “财物状态”]的五元组关系模式。

### 2.3 涉案财物关系抽取任务

在与涉案财物处置相关的法律文件中, 法条准确地描述了在何种情况下某个机构可以对某种财物实施特定的处置。涉案财物关系抽取任务即是从法条文本中自动抽取这种处置规则。即是抽取形如[“业务单位实体”, “处置方式”, “处置条件”, “财物实体”, “财物状态”]的五元组关系。

在目前常规的关系抽取任务中, 往往采用基于神经网络的方法, 本质上是将抽取问题转化为了分

类问题<sup>[11]</sup>,在诸如人物关系这种关系类型确定的场景下效果较好,由于本文研究的涉案财物关系抽取任务特殊,难以转化为分类问题,采用模板匹配的方法进行抽取,抽取的五元组示例如下。

对法条文本:“人民法院在必要的时候,可以采取保全措施,查封、扣押或者冻结被告人的财产。”内含如表 1 所示的五元组关系。

| 表 1 五元组关系                  |                 |
|----------------------------|-----------------|
| Tab. 1 Five-tuple relation |                 |
| 元素名称                       | 元素内容            |
| 业务单位实体                     | 人民法院            |
| 财物实体                       | 财产              |
| 处置方式                       | 查封 扣押 冻结        |
| 财物状态                       | 缺省              |
| 处置条件                       | 在必要的时候,可以采取保全措施 |

在上述例子中,实体与关系之间的相对位置较为常规,且各元素成分相对独立,抽取过程中干扰信息较少。

但对另一个法条文本:“对查封、扣押的财物、文件、邮件、电报或者冻结的存款、汇款、债券、股票、基金份额等财产,经查明确实与案件无关的,应当在三日以内解除查封、扣押、冻结,予以退还。”

该法条中语法结构有所不同,财物实体进行了前置,且财物状态搭配复杂,抽取难度较大。

可见,在涉案财物关系抽取任务中,由于语言习惯的差异,单个模板难以较好地应对不同结构的法条,因此有必要从不同维度设计多个模板,以应对不同的语言现象。但由此产生了另一个问题,即多个模板抽取的结果如何整合的问题。本文为此提出了一种基于模糊逻辑的方法,以评判多个模板抽取结果的置信度,竞争得出质量较优的结果。

### 3 基于多模板模糊竞争的关系抽取

#### 3.1 数据预处理

涉案财物关系抽取的初始数据是法律文件,但模板匹配的处理对象是单句法条,因此,需要对初始法律文件进行预处理,以适配模板匹配,处理流程如图 1 所示。

首先,对输入的法律文件进行段落解析,得到段落集 A,对于每一个段落,判断其是否是居中的标题,若是则忽略该段落,否则利用正则表达式提取法条序号及法条内容。通过以上步骤得到了带有序号标记的法条集 A'。后续根据涉案财物处置相

关的特征词过滤掉与涉案财物处置无关的法条,得到最终有效的法条集 A''。

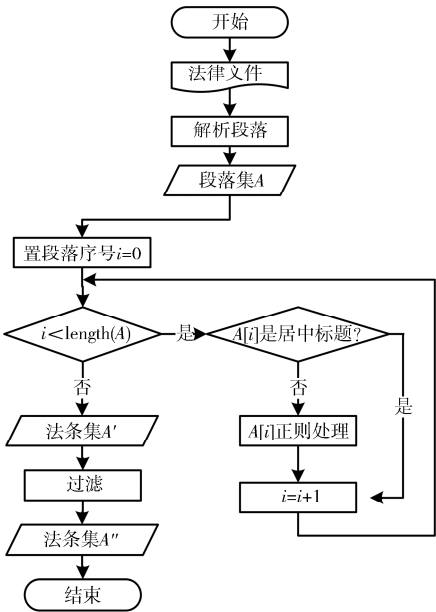


图 1 预处理流程图  
Fig. 1 Flow chart of preprocessing

#### 3.2 模板设计

3.2.1 模板设计基础说明 本文以不同的维度设计了三个抽取五元组关系的模板。下面就三个模板的公共部分做说明。

模板输入输出: 1) 输入:  $T, W_1, W_2, W_3, W_4, W_5$ ; 2) 输出:  $R$ 。其中,  $T$  为输入的法条文本;  $W_1, W_2, W_3, W_4, W_5$  为人工整理的词典<sup>[12]</sup> ( $W_1$  为财物实体词典;  $W_2$  为业务单位实体词典;  $W_3$  为触发词典;  $W_4$  为处置方式词典;  $W_5$  为财物状态词典)。  $R$  为输出的 5 元组关系集合。

模板伪代码中将用到的重要函数如表 2。

表 2 函数说明

Tab. 2 Functions description

| 函数名   | 说明                                                                               |
|-------|----------------------------------------------------------------------------------|
| DTW   | 输入起始位置 $s$ 、结束位置 $e$ , 词列表 $W$ , 词典 $A$ , 输出在 $W$ 的 $(s, e)$ 区间内 $A$ 中元素首次出现的位置。 |
| PTW   | 输入起始位置 $s$ 、结束位置 $e$ , 词列表 $W$ , 词典 $A$ , 输出在 $W$ 的 $(s, e)$ 区间内出现的 $A$ 中元素的集合。  |
| GCT   | 输入起始位置 $s$ 、结束位置 $e$ , 词列表 $W$ , 输出在 $W$ 的 $(s, e)$ 区间内出现的处置条件短语。                |
| cut_w | 分词函数, 导入了涉案财物词集的结巴分词。                                                            |
| cut_c | 分句函数, 以“ $,$ ; $;$ ”划分子句。                                                        |

3.2.2 模板 1 设计 模板 1 是以词为单元进行匹配,首先确定财物实体的位置,以此为中心检索其

余 4 个元素. 具体匹配规则如下.

|                                                              |
|--------------------------------------------------------------|
| <b>输入:</b> $T, W_1, W_2, W_3, W_4, W_5$                      |
| <b>输出:</b> $R$                                               |
| (1) $W \leftarrow \text{cut\_w}(T)$                          |
| (2) $p \leftarrow \text{DTW}(0, \text{len}(W), W, W_1)$      |
| (3) $t \leftarrow \text{DTW}(0, \text{len}(W), W, W_3)$      |
| (4) $P \leftarrow \text{PTW}(0, \text{len}(W), W, W_1)$      |
| (5) if $t \leq p$ then $f \leftarrow \text{True}$            |
| (6) end if                                                   |
| (7) $C \leftarrow \{\text{GCT}(\min(t, p), \max(t, p), W)\}$ |
| (8) $S \leftarrow \text{PTW}(p, 0, W, W_5)$                  |
| (9) if $f$ is not True then                                  |
| (10) $M \leftarrow \text{PTW}(t, \text{len}(W), W, W_4)$     |
| (11) else                                                    |
| (12) $M \leftarrow \text{PTW}(p, \text{len}(W), W, W_4)$     |
| (13) if $M == \emptyset$ then                                |
| (14) $M \leftarrow \text{PTW}(t, p, W, W_4)$                 |
| (15) end if                                                  |
| (16) end if                                                  |
| (17) $G \leftarrow \text{PTW}(t, 0, W, W_2)$                 |
| (18) $R \leftarrow P \times G \times M \times S \times C$    |
| (19) return $R$                                              |

3.2.3 模板 2 设计 模板 2 同样是以词为单元进行匹配,但首先确定的是业务单位实体的位置,以此为中心检索其余 4 个元素. 匹配规则如下.

|                                                                             |
|-----------------------------------------------------------------------------|
| <b>输入:</b> $T, W_1, W_2, W_3, W_4, W_5$                                     |
| <b>输出:</b> $R$                                                              |
| (1) $W \leftarrow \text{cut\_w}(T)$                                         |
| (2) $G \leftarrow \text{PTW}(0, \text{len}(W), W, W_2)$                     |
| (3) if $G \neq \emptyset$ then                                              |
| (4) $n_1 \leftarrow \text{DTW}(0, \text{len}(W), W, W_2), n_2 \leftarrow 0$ |
| (5) else $n_1 \leftarrow 0, n_2 \leftarrow \text{len}(W)$                   |
| (6) end if                                                                  |
| (7) $P \leftarrow \text{PTW}(n_1, n_2, W, W_1)$                             |
| (8) $t \leftarrow \text{DTW}(0, \text{len}(W), W, W_3)$                     |
| (9) $S \leftarrow \text{PTW}(t, 0, W, W_5)$                                 |
| (10) $M \leftarrow \text{PTW}(t, \text{len}(W), W, W_4)$                    |
| (11) $s \leftarrow \text{DTW}(t, 0, W, W_5)$                                |
| (12) $C \leftarrow \{\text{GCT}(s, \text{len}(W), W)\}$                     |
| (13) $R \leftarrow P \times G \times M \times S \times C$                   |
| (14) return $R$                                                             |

3.2.4 模板 3 设计 模板 3 是以子句为单元进行

匹配. 模板 3 中认为业务单位实体与处置方式应当同属一个子句,财物实体与财物状态属性应当同属一个子句,处置条件属性单独属于一个子句. 具体匹配规则如下.

|                                                           |
|-----------------------------------------------------------|
| <b>输入:</b> $T, W_1, W_2, W_3, W_4, W_5$                   |
| <b>输出:</b> $R$                                            |
| (1) $Q \leftarrow \text{cut\_c}(T)$                       |
| (2) select $q$ in $Q$ that 包含业务单位实体或处置方式                  |
| (3) $Q \leftarrow Q - \{q\}$                              |
| (4) $W \leftarrow \text{cut\_w}(q)$                       |
| (5) $G \leftarrow \text{PTW}(0, \text{len}(W), W, W_2)$   |
| (6) $M \leftarrow \text{PTW}(0, \text{len}(W), W, W_4)$   |
| (7) select $q$ in $Q$ that 包含财物实体或财物状态                    |
| (8) $Q \leftarrow Q - \{q\}$                              |
| (9) $W \leftarrow \text{cut\_w}(q)$                       |
| (10) $P \leftarrow \text{PTW}(0, \text{len}(W), W, W_1)$  |
| (11) $S \leftarrow \text{PTW}(0, \text{len}(W), W, W_5)$  |
| (12) select $q$ in $Q$ that 包含处置条件                        |
| (13) $Q \leftarrow Q - \{q\}$                             |
| (14) $W \leftarrow \text{cut\_w}(q)$                      |
| (15) $P \leftarrow \{\text{GCT}(0, \text{len}(W), W)\}$   |
| (16) $R \leftarrow P \times G \times M \times S \times C$ |
| (17) return $R$                                           |

### 3.3 模糊竞争

3.3.1 数值化 本文拟利用模糊逻辑对模板抽取出的五元组关系进行打分,进而实现多个模板间抽取结果的奖惩机制,综合胜出置信度较高的五元组关系. 鉴于模糊逻辑适用于数值计算<sup>[13]</sup>,而五元组关系为文本数据,加之初始抽取的五元组关系存在空值干扰,因此,首先定义五元组各元素补全方法及数值化方法.

- (1) 初始抽取数据如下所示.
- 模板 1

模板 2

模板 3

$[a_1, b_1, c_1, d_1, e_1]$

$[a_2, b_2, c_2, d_2, e_2]$

$[a_3, b_3, c_3, d_3, e_3]$
- (2) 补全过程:
- 1) 通过实验数据确定三个模板初始置信度:  
 $w_1, w_2, w_3$ ;
- 2) 对于  $\{a, b, c, d, e\}$  中的每一种元素  $x$ :
- (a) 筛选出  $x_1, x_2, x_3$  中的非空元素,记作集合  $R$ ;

- (b) 选出集合  $R$  中对应模板置信度最大的元素  $r$ ;
- (c) 用  $r$  补全  $x_1, x_2, x_3$  中的空值元素;
- (d) 若  $R$  为空集, 则  $x_1, x_2, x_3$  置为“空”。
- (3) 补全后进行数值化, 对于每一个元素  $x_i$ , 其数值化结果  $v$  计算公式如下。

$$v = \frac{\text{sim}(x_i, x_{[(i+1)\%3]+1}) + \text{sim}(x_i, x_{(i\%3)+1})}{2}$$
$$x \in \{a, b, c, d, e\}, i \in \{1, 2, 3\}$$

(1)

其中,  $\text{sim}$  为文本相似度计算函数, 经实验尝试, 摒弃了常规的基于词袋的余弦相似度方法, 实际采用的相似度方法如下。

输入:  $s_1, s_2$

输出:  $t$

- (1)  $Q \leftarrow \{\}, P \leftarrow \{\}, m \leftarrow 0$
- (2) for  $i$  in  $s_1$  do
- (3)  $Q[i] \leftarrow Q.\text{get}(i, 0) + 1$
- (4) end if
- (5) for  $j$  in  $s_2$  do
- (6) if  $P.\text{contains}(j)$  then
- (7)  $n \leftarrow P[j]$
- (8) else
- (9)  $n \leftarrow Q.\text{get}(j, 0)$
- (10) end if
- (11)  $P[j] \leftarrow n - 1$
- (12) if  $n > 0$  then  $m \leftarrow m + 1$
- (13) end if
- (14) end for
- (15)  $t \leftarrow 2 * \lceil m / (\text{len}(s_1) + \text{len}(s_2)) \rceil$
- (16) return  $t$

3.3.2 模糊化 通过数值化处理后, 每个模板抽取的五元组关系均如以下格式:  $[a\ b\ c\ d\ e]$ , 其中, 各元素均为 0 到 1 之间的浮点数。

在模糊化阶段, 定义每个元素  $x$  均隶属于  $P$ 、 $A$ 、 $G$  三个集合。其中,  $P$  集合和  $G$  集合采用梯形隶属函数,  $A$  集合采用三角形隶属函数<sup>[14]</sup>。如图 2 所示。

图 2 中, 横轴为输入的元素浮点数值, 纵轴为对应的各集合的隶属度。  $p_1, p_2, a, d, g_1, g_2$  为各隶属度函数的参数。

通过隶属函数模糊化后, 得到五元组各元素隶属于  $PAG$  三个集合的隶属度, 如图 3 所示。

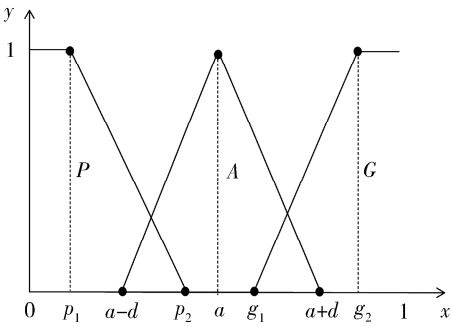


图 2 隶属函数图  
Fig. 2 Image of membership functions

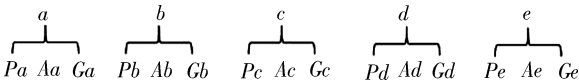


图 3 隶属示意图  
Fig. 3 Diagram of affiliation

3.3.3 规则化 通过模糊化处理后, 需要根据模糊规则和模糊逻辑的运算进行重新组合。五元组各元素均隶属于  $PAG$  三个集合, 五种元素不同隶属集合组合情况共有  $3^5$  种, 如下所示。

$$\begin{bmatrix} P_a \\ A_a \\ G_a \end{bmatrix} \times \begin{bmatrix} P_b \\ A_b \\ G_b \end{bmatrix} \times \begin{bmatrix} P_c \\ A_c \\ G_c \end{bmatrix} \times \begin{bmatrix} P_d \\ A_d \\ G_d \end{bmatrix} \times \begin{bmatrix} P_e \\ A_e \\ G_e \end{bmatrix}$$
$$\vdots \vdots \vdots \vdots \vdots$$

$3^5$  种组合

对于每一种组合, 通过规则指定最终的隶属集合以及相应的隶属度。为减少模糊规则数量, 本文简化规则如下。

- 1) 定义。
- $$\omega(P_x) = 0, \omega(A_x) = 1, \omega(G_x) = 2,$$
$$x \in \{a, b, c, d, e\}$$

- 2) 对于任意一种组合:
- $$S = (Y_a, Y_b, Y_c, Y_d, Y_e), Y \in \{P, A, G\},$$
$$\text{sum}(S) = \sum_{i \in S} i$$

- 3) 隶属集合:
- $$R(S) = \begin{cases} P, & \text{sum}(S) \leq k_1 \\ A, & k_1 < \text{sum}(S) < k_2 \\ G, & \text{sum}(S) \geq k_2 \end{cases}$$

- 4) 隶属度:
- $$V(S) = \min(Y_a, Y_b, Y_c, Y_d, Y_e)$$

3.3.4 去模糊 通过规则化处理后, 得到了各种组合情况下对应的隶属集合以及隶属度, 借助去模糊化将其转化为最终评判五元组关系质量的数值。本文采用加权平均判决法。

$$Output = \frac{\sum_i FS_i * OW_i}{\sum_i FS_i}$$

(6)

其中,  $FS_i$  为规则化阶段得到的隶属度;  $OW_i$  为对应隶属集合的权重系数. 在本文中, 取值如下.

$$OW(P) = \frac{p_2}{2},$$

$$OW(A) = a,$$

$$OW(G) = \frac{g_1 + 1}{2}.$$

(7)

其中,  $p_2, a, g_1$  为图 2 中隶属度函数参数.

3.3.5 模板竞争 三个模板抽取的五元组关系通过上述模糊计算, 得到三个 output 值, 最高值对应的模板胜出, 本次抽取结果以该模板为准. 同时更新三个模板置信度, 更新规则如下.

(1) 胜出的模板:

$$w_i = w_i + (1 - w_i) * 0.001$$

(2) 其余模板:

$$w_j = w_j - w_j * 0.0005$$

(8)

(9)

4 实验与评估

4.1 实验数据

本文的实验数据来自“法律法规数据库”, 共选取了 10 个与涉案财物处置相关的法律文件, 其中 1~4 号文件用于模板初值置信度确定, 5~10 号文件用于算法效果测试.

通过人工对上述法律文件进行整理, 共标定五元组数据 1 450 条. 数据格式: [文件名, 法条序号, 法条内容, 五元组].

4.2 评价指标

本文使用正确率作为评价指标对关系抽取的效果进行评估. 考虑到五元组关系的特殊性, 定义: 有测试结果  $S = [s_1, s_2, \dots, s_i]$ , 标定数据  $K = [k_1, k_2, \dots, k_i]$ . 若匹配度  $q > 0.6$ , 则认为抽取成功.

$$q = (\sum_{j=0}^i \text{sim}(s_j, k_j)) \div i$$

(10)

正确率的计算方法如下.

$$c = \frac{\text{抽取成功的关系数量}}{\text{标定数据的关系数量}}$$

(11)

本文同时进行了 3 元组和 5 元组抽取效果的评估, 在计算 3 元组正确率时,  $i=3$ , 计算 5 元组正确率时,  $i=5$ .

4.3 实验结果

本文利用选取的涉案财物处置相关的法律文

件, 对基于多模板模糊竞争(MTFC)的方法进行了实验, 同时将其与单模板抽取方法(模板 1、模板 2、模板 3)、基于非空元素数量的投票方法(NONV)进行对比.

首先, 我们运用三个模板分别对 1~4 号文件进行关系抽取, 三元组抽取正确数目的比值约为 33 : 37 : 30. 因此, 我们设定三个模板初始置信度:  $w_1=0.33$ ;  $w_2=0.37$ ;  $w_3=0.3$ .

在基于多模板模糊竞争方法的实验中, 共有如下 8 个参数:  $p_1, p_2, a, d, g_1, g_2$  (图 2 隶属度函数参数);  $k_1, k_2$  (式(4)模糊规则参数).

为降低参数选取难度, 设定整数参数步距为 1, 浮点数参数步距为 0.1, 并将单轮实验参数调整数量限制为 2 个. 实验表明, 当  $p_1=0.3$ ,  $p_2=0.5$ ,  $a=0.5$ ,  $d=0.2$ ,  $g_1=0.7$ ,  $g_2=0.9$ ,  $k_1=5$ ,  $k_2=8$  时, 多模板模糊竞争方法有较优的效果. 在 5~10 号法律文件上, 各方法实验结果如表 3 和表 4 所示. 可以看出, 在 8 号文件上, 多模板模糊竞争方法的五元组正确率没有明显的提升, 但在其他情况下, 多模板模糊竞争方法的正确率较其余方法, 均有显著优势. 总体来看, 如图 4 所示, 多模板模糊竞争方法对三元组及五元组关系抽取效果明显.

表 3 三元组正确率  
Tab. 3 Accuracy of three-tuple

| 文件序号 | 5    | 6    | 7    | 8    | 9    | 10   | 总正确率 |
|------|------|------|------|------|------|------|------|
| 模板 1 | 0.40 | 0.78 | 0.60 | 0.61 | 0.41 | 0.36 | 0.51 |
| 模板 2 | 0.26 | 0.66 | 0.58 | 0.45 | 0.49 | 0.39 | 0.48 |
| 模板 3 | 0.13 | 0.63 | 0.47 | 0.31 | 0.42 | 0.42 | 0.41 |
| NONV | 0.49 | 0.81 | 0.68 | 0.65 | 0.50 | 0.46 | 0.58 |
| MTFC | 0.72 | 0.82 | 0.70 | 0.68 | 0.55 | 0.48 | 0.62 |

表 4 五元组正确率  
Tab. 4 Accuracy of five-tuple

| 文件序号 | 5    | 6    | 7    | 8    | 9    | 10   | 总正确率 |
|------|------|------|------|------|------|------|------|
| 模板 1 | 0.49 | 0.51 | 0.28 | 0.50 | 0.27 | 0.29 | 0.35 |
| 模板 2 | 0.06 | 0.27 | 0.20 | 0.27 | 0.25 | 0.23 | 0.23 |
| 模板 3 | 0.06 | 0.33 | 0.11 | 0.22 | 0.26 | 0.23 | 0.21 |
| NONV | 0.49 | 0.49 | 0.30 | 0.52 | 0.35 | 0.30 | 0.38 |
| MTFC | 0.66 | 0.53 | 0.39 | 0.51 | 0.43 | 0.31 | 0.43 |

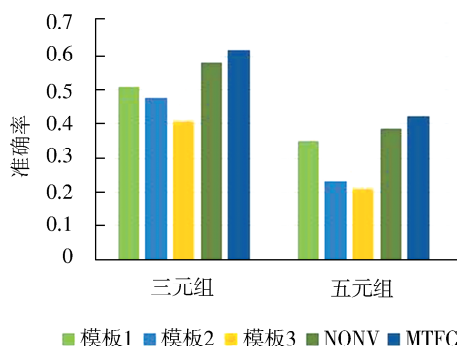


图 4 综合正确率

Fig. 4 Comprehensive accuracy

## 5 结 论

本文提出了一种基于多模板模糊竞争的涉案财物关系抽取方法,从不同维度设计了三个涉案财物五元组关系抽取的模板,并借助模糊逻辑算法,竞争出较优的结果.实验表明,在涉案财物关系抽取任务中,基于多模板模糊竞争的方法效果优于单模板,也优于基于非空元素数量的投票方法.因此,可以在涉案财物知识库构建过程中引入该方法,以较好地适应其训练语料少且单一以及识别准确率高要求高的特点,为后续知识推理奠定基础.

## 参考文献:

[1] 姜芳,李国和,岳翔,等.语义关系自动抽取方法[J].山西大学学报:自然科学版,2015,38:560.

[2] 鄂海红,张文静,肖思琪,等.深度学习实体关系抽取研究综述[J].软件学报,2019,30:1793.

[3] 李冬梅,张扬,李东远,等.实体关系抽取方法研究综述[J].计算机研究与发展,2020,57:1424.

[4] Zeng D J, Liu K, Lai S W, *et al.* Relation classi-

cation via convolutional deep neural network [C]//Proceedings of the 25th Int Conf on Computational Linguistics. Stroudsburg: ACL, 2014.

[5] Zhang S, Yang M, Zheng D Q, *et al.* Bidirectional long short-term memory networks for relation classification [C]//Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation. Shanghai: Shanghai Jiao Tong University, 2015.

[6] 湛予恒,王峥.结合注意力机制与残差网络的远程监督关系抽取[J].计算机与数字工程,2020,48:909.

[7] 司文豪,贾雷萍,戚银城.基于卷积神经网络的中文人物关系抽取方法[J].计算机与现代化,2018:17.

[8] 李奋军.治安案件办理中涉案财物处置应注意的问题[J].派出所工作,2020(10):74.

[9] 王磊,刘露,牛亮,胡封晔,彭涛.基于关系触发词与单层 GRU 模型的关系抽取方法[J].吉林大学学报:理学版,2020,58:95.

[10] 陈刚,徐星羽.基于维基百科信息框的本体信息提取[J].吉林大学学报:理学版,2020,58:355.

[11] 汪嘉伟,杨煦晨,琚生根,等.基于卷积神经网络和自注意力机制的文本分类模型[J].四川大学学报:自然科学版,2020,57:469.

[12] 胡思才,孙界平,琚生根,等.基于扩展的情感词典和卡方模型的中文情感特征选择方法[J].四川大学学报:自然科学版,2019,56:37.

[13] 裴道武.模糊逻辑中的一些问题与研究进展[J].四川师范大学学报:自然科学版,2017,40:411.

[14] 刘开旻,吴小俊.一种基于新隶属度函数的模糊支持向量机[J].计算机工程,2016,42:1559.

### 引用本文格式:

中 文:李攀峰,林锋,蒋宗神.基于多模板模糊竞争的涉案财物关系抽取方法[J].四川大学学报:自然科学版,2021,58:042002.

英 文:Li P F, Lin F, Jiang Z S. Relation extraction method for property involved based on multiple templates and fuzzy competition [J]. J Sichuan Univ: Nat Sci Ed, 2021, 58: 042002.