

# 犯罪行为演化图谱的自动构建方法研究

段锡辉, 黄瑞章

(贵州大学计算机科学与技术学院, 贵州 贵阳 550025)

✉1416831033@qq.com; rzhuang@gzu.edu.cn



**摘要:** 随着国家司法信息化建设的推进, 促进司法公平正义、提高司法审判工作效率成为必然的趋势。本文将事件图谱的构建方法引入司法领域, 将裁判文书案情描述中的犯罪行为视为节点, 犯罪行为间的逻辑关系视为边, 构建了一种犯罪行为演化图谱, 以此来将案情中孤立的犯罪行为利用逻辑关系连接起来, 从而描述案情的发展过程而不是独立的要素, 最大限度地保留了案情的语义信息。本文提出的自动构建方法在五个案由中的抽取结果比传统方法平均提高了7.8%, 此方法构建的犯罪行为演化图谱可为法律从业人员研究案情提供参考, 同时为推进智慧法院建设如罪名预测、类案推荐等提供了新的思路与方法。

**关键词:** 犯罪行为; 关系识别; 犯罪行为链; 犯罪行为演化图谱

**中图分类号:** TP399 **文献标识码:** A

## Research on Automatic Construction Method of Criminal Act Evolution Graph

DUAN Xihui, HUANG Ruizhang

(School of Computer Science and Technology, Guizhou University, Guiyang 550025, China)

✉1416831033@qq.com; rzhuang@gzu.edu.cn

**Abstract:** With the advancement of national judicial information construction, it has become an inevitable trend to promote judicial fairness and justice, and improve the efficiency of judicial trials. This paper proposes to introduce construction method of event graph into the judicial field and construct an evolution graph of criminal acts, where criminal act in case fact description of the judgment document is regarded as a node, and the logical relationship between criminal acts as an edge. Isolated criminal acts are thus connected by logical relationship, so as to describe the development process of the case rather than independent elements, and preserve the semantic information of the case to the greatest extent. The extraction results of the five reasons of case by the proposed automatic construction method are improved by an average of 7.8% compared with the traditional method. The criminal act evolution graph constructed by this method can provide a reference for legal practitioners to study cases, and at the same time, it provides new ideas and methods for advancing the construction of smart courts, such as crime prediction and case recommendation.

**Keywords:** criminal act; relationship identification; criminal act chain; criminal act evolution graph

## 1 引言(Introduction)

裁判文书是我国司法案件的一种载体, 根据内容大致分为案件信息、案情描述、判决要素及审判结果四个部分, 如图1所示。法律从业人员在研究过往案件裁判文书的过程中往往需要获取案情的发展脉络, 以此来为当前案件提供论据。在过去的知识图谱分析方法中, 人们往往只关注裁判文书中独立的案情要素, 忽略了犯罪行为之间存在的逻辑与规

律<sup>[1-2]</sup>。在能够描述事件发展和演化的事件演化图谱<sup>[3]</sup>(Event Evolutionary Graph, EEG)的启发下, 本文提出了一种全新的案情结构化表示方法——犯罪行为演化图谱(Criminal Act Evolutionary Graph), 将犯罪行为视为事件(Event), 而犯罪行为之间的演化关系则可以用事件间的时序、因果等关系(Relation)来描述, 故犯罪行为演化图谱的构建方法可以表述为: 构建一个能够利用事件之间的时序及因果等关系来描

述事件演化规律的有向无环图。在以图结构描述案情的过程中，如何在清晰明确地描述案情发展及其演化过程的同时尽可能地保留上下文语义信息是工作的一大难点。

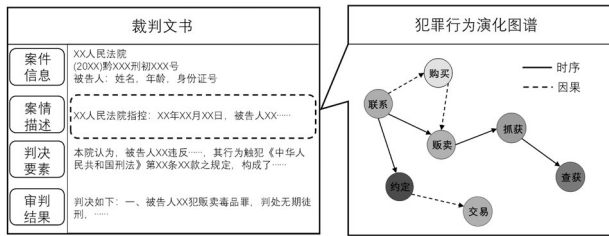


图1 犯罪行为演化图谱示例

Fig.1 Illustration of criminal act evolution graph

## 2 背景及相关工作(Background and related work)

近年来,国内外学者普遍认为计算机技术对法律行业产生了重大的影响<sup>[4]</sup>,我国在司法信息化建设过程中取得了不错的进展,并积累了相当多的经验。目前智能司法领域的相关工作主要集中在通过裁判文书、证词、庭审记录等司法大数据,利用大数据与人工智能技术达成辅助审判的应用<sup>[5]</sup>,同时在罪名预测、法条推荐、刑期预测、类案推荐及司法知识图谱等应用<sup>[6]</sup>上取得了令人瞩目的成果,但这些工作还只停留在对案情要素的应用上,没有对案情的发展和演化进行深入研究。

在过去对案情的分析方法中,常采用知识图谱的方法来表述裁判文书中包含的半结构化信息<sup>[7]</sup>,但这种方法通常只关注实体及关系等静态要素,难以获得案情部分具有序列特征的犯罪行为的演化与发展等动态信息。受知识图谱构建方法的启发,本文引入了一种特殊的知识图谱——事件图谱。在事件图谱的构建工作中,HOLME等人<sup>[8]</sup>将这种结构统称为时间网络(Temporal Networks),这是具有跨学科性质的研究内容;CHAMBERS等人<sup>[9]</sup>通过引入局部有序的时序关系,构建了一种新的结构化数据表示:叙事事件链(Narrative Event Chains);LI等人<sup>[10]</sup>在其基础上又引入了因果关系,称为事件演化图谱(Event Evolutionary Graph),使其能描述现实世界中事件的演化和发展逻辑。

本文将犯罪行为(事件)对应知识图谱中的实体,犯罪行为的演化关系对应知识图谱中的关系,将“实体-关系-实体”的三元组变为“犯罪行为-演化关系-犯罪行为”的三元组,犯罪行为之间用演化关系彼此连接,构建犯罪行为演化图谱,用来描述案情的演化与发展规律。

## 3 定义(Definition)

犯罪行为(Criminal Act)是指案情描述中的事件提及,即表示案情中事件发生的谓语动词及其所包含的参数,其中谓语动词称为“犯罪行为词”。在本文中,我们利用基于NLP(Natural Language Process)的语法分析工具提取犯罪行为,识别犯罪行为之间的时序、因果两种演化关系,构造出犯罪行为链(Criminal Act Chain),最后将多个犯罪行为链拼接为犯罪行为演化图谱。在本部分中,我们会对犯罪行为、犯罪行为链及犯罪行为演化图谱等概念做出系统性的定义与解释。

### 3.1 犯罪行为元组及演化关系

任何犯罪活动都是由人实施的,犯罪行为人是实施犯罪行为的主体,称为施事者(Agent)。犯罪行为词是指犯罪行为人在实施犯罪的过程中所做出的行为,是犯罪行为的触发词。被施加犯罪行为的对象称为受事者(Patient),间接对象称为间接受事者(Indirect-patient)。同时,犯罪行为中还存在时间约束与空间约束,用来明确犯罪行为发生的时间与地点。因此,犯罪行为元组(Criminal Act Tuple)可表示为犯罪行为词及其参数的集合: $CrATuple = CrA_{trigger} < P_{agt}, P_{pat}, C_{ipa}, C_{time}, C_{loc} >$ 。例如,“贩卖<张三,海洛因,李四,××年××月××日,××地>”为犯罪行为词,“贩卖”及其参数组成了犯罪行为元组。

由于裁判文书中对案情的描述多使用叙述语法,且注重对案情的时序及因果等演化过程的描述,故本文用识别出的犯罪行为及演化关系来表示犯罪行为之间的演化规律。时序关系是裁判文书文本中的主要演化关系,同时引入因果关系用以对犯罪行为的演化关系进行补充。因果关系虽然在所有演化关系中占比较小,但往往是案情发展描述的关键点与转折点。犯罪行为间的演化关系定义为:

时序(Time-sequence):表示两个犯罪行为之间顺序演化关系。在本文中,将时序演化关系分为两种情况,一种是同一句子中行为之间的“顺承关系”,另一种是不同句子的根触发词之间的“序列关系”。

因果(Cause-effect):表示两个犯罪行为“原因”和“结果”之间的演化关系,后一行为为前一行为的结果。本文中因果演化关系同样划分为两种情况,一种是基于文本语法规则的演化关系:(1)状语连词,如“因此”“所以”等;(2)因果动词,如“导致”;(3)介词、连词,如“因为”“由于”;(4)从句关系,如“……联系……购买……”。另一种为逻辑因果,在本文中只引入了一种逻辑因果,表示将要发生的某一行为与后续已经发生的此行为之间互为因果关系,例如,不同句子中的两个行为“约定交易”与“交易完成”之间为逻辑因果演化关系。因果演化的优先级高于时序演化。

犯罪行为演化关系(Criminal Act Evolutionary Relation)的结构为: $< CrA_{trigger,1}, CrA_{trigger,2} >_{cause}$ ,例如<联系,购买><sub>cause</sub>,表示犯罪行为“联系”与“购买”之间存在因果关系,即“联系”的原因是“购买”。

### 3.2 犯罪行为链

CHAMBERS等人首次提出了叙事事件链的概念,当一个实体作为一个动词的参数被提及时,视为其参与了这些事件,其定义的叙事事件链为共享有一个主角的谓语动词序列。GRANROTH等人<sup>[11]</sup>在CHAMBERS等人工作的基础上加入了谓语动词及其语法依赖关系作为一个事件,是一种由谓语动词及其语义依赖关系组成的事件的序列。本文在叙事事件链的基础上引入了犯罪行为演化关系构成犯罪行为链,作为犯罪行为演化图谱构建的前提,将犯罪行为链定义为演化关系的集合: $CrA_{chain,i} = \{ < CrA_{trigger,1}, CrA_{trigger,2} >_{time}, \dots,$

$\langle CrA_{trigger,i}, CrA_{trigger,j} \rangle_{cause}$ }, 其中 $\langle CrA_{trigger,i}, CrA_{trigger,j} \rangle_{cause}$ 表示 $CrA_{trigger,i}$ 到 $CrA_{trigger,j}$ 存在因果关系的有向边。

### 3.3 犯罪行为演化图谱

本文提出了一种犯罪行为演化图谱的自动构建框架,在抽取犯罪行为元组及演化关系生成犯罪行为链的基础上构建犯罪行为演化图谱。犯罪行为演化图谱是一个以犯罪行为元组为节点、演化关系为边的有向无环图: $CrA_{graph} = \langle V, E \rangle$ , 其中 $V$ 为犯罪行为元组的集合, $E$ 为演化关系的集合,能够最大限度地保留案情的上下文语境,从而描述案情的走向并为如罪名预测、犯罪行为预测等应用打下基础。

## 4 系统描述(System specification)

犯罪行为演化图谱的自动构建方法的总体结构如图2所示。裁判文书是一种半结构化文本,首先抽取裁判文书中的案情描述部分,通过基于文本语义的连续管道预处理;在获取语义角色标签并对其进行依存句法解析的基础上,对裁判文书进行犯罪行为提取和演化关系识别,生成犯罪行为链;最后通过泛化及降噪处理后,将犯罪行为链拼接成犯罪行为演化图谱。

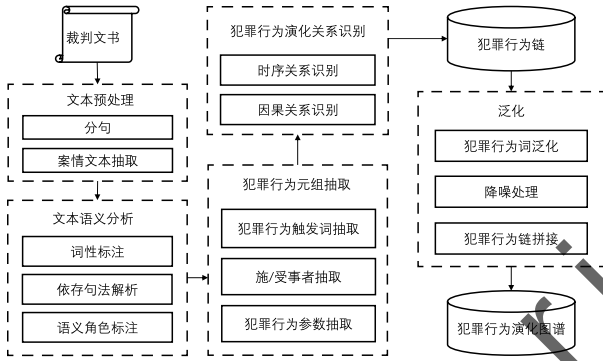


图2 犯罪行为演化图谱的自动构建方法

Fig.2 The automatic construction method of criminal act evolution graph

### 4.1 犯罪行为抽取

将裁判文书输入文本处理的管道过程中进行犯罪行为抽取。图3为使用LTP<sup>[12]</sup>(Language Technology Platform)工具对裁判文书中犯罪行为进行分析的示例。基于裁判文书的犯罪行为抽取方法用来提取犯罪行为词及其参数,构造犯罪行为元组。

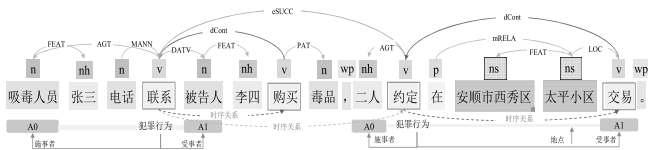


图3 利用LTP工具实现的犯罪行为分析实例

Fig.3 An example of criminal act analysis implemented by using LTP tools

以图3中的句子为例,在经过文本预处理后,获得句子的词性标注 $p_{os}$ 、语法依存分析标签 $S_{dp}$ 、语义角色标签 $S_{ri}$ 以及命名实体识别 $N_{er}$ 。将上述变量连同文本作为算法的输入,其输出为犯罪行为词及其对应的参数。首先通过 $S_{dp}$ 获得根触发词“联系”,由其对应的语义依赖树标签“AGT”可以得知犯

罪行为词“联系”对应的施事者(Agent)为“吸毒人员”,若其对应的标签 $S_{ri}$ 为“AO”,则识别正确。同时,由标签“DATV”抽取出受事者(Patient)为“被告人”,并根据 $S_{ri}$ 标签“A1”确定。将“吸毒人员”与“被告人”都作为广义的施事者与受事者,如图3中例子所示,我们可以利用 $p_{os}$ 和 $S_{dp}$ 识别到“吸毒人员”的具体姓名“张三”,其标签为“FEAT”;同理,可以识别到“被告人”的具体姓名“李四”,据此可以构建结构为“联系<吸毒人员张三,被告人李四>”的犯罪行为元组。对于间接受事者,可以用 $S_{dp}$ 标签“CONT”和 $S_{ri}$ 标签“A1”来抽取。而时间、地点的抽取方法需要先利用命名实体方法识别正确的实体后,再将句子重新进行词性标注,以此获得与犯罪行为词相关的参数。类似地,后续事件也可以被抽取出来,获得对应的犯罪行为元组。

### 4.2 演化关系识别

犯罪行为演化关系识别的目标是从犯罪行为抽取过程得到的行为集合中生成相应的子集。犯罪行为链的构建过程为:(1)从单独句子中识别行为触发词对;(2)通过两个连续句子的根行为识别演化关系;(3)将上述行为触发词对添加到行为链。

在案情描述的每一句话中,将提取的行为触发词根据定义的演化关系组合成一组此行为触发词对 $Tg_{pair}$ ,并根据对应的犯罪行为元组生成 $CrA_{pair} = \langle CrA_{trigger,i}, CrA_{trigger,j} \rangle_{cause}$ 。首先,每句话中都存在一个根行为词,作为语义依存分析的起点,根据根行为和后续行为之间的演化关系,从裁判文书的第一句案情描述中构造根行为和后续行为的 $Tg_{pair}$ 。如图3所示,根触发词为“联系”,由其语义标签“dCont”识别其后续行为“购买”为其从句(子行为),为因果演化,同时生成对应的 $CrA_{pair}$ 。对于同一裁判文书中连续的句子,可以根据每一句的顺承关系来连接两个独立句子的根触发词,视为时序演化。当获得第 $k$ 条句子的根触发词时 $Tg_{r<k>}$ ,将第 $k+1$ 条句子的根触发词 $Tg_{r<k+1>}$ 连接到 $Tg_{r<k>}$ 上,并添加相应的参数后得到根行为对 $R_{pair}$ 。最后,将 $CrA_{pair}$ 与 $R_{pair}$ 组合为行为链 $CrA_{chain}$ ,其为多条有序链组成的集合。

### 4.3 演化图谱的构建

在生成行为链 $CrA_{chain}$ 的基础上,本文引入了一种降噪策略Ordered-PMI<sup>[13]</sup>来去除行为链中的噪音,它是一种基于点互信息的犯罪行为排序函数 $f$ 。为了定义用于修正行为链的函数 $f$ ,我们将 $CrA_{chain}$ 中的除根触发词链外的行为链 $CrA'_{chain} = [Tg_1, Tg_2, Tg_3, \dots, Tg_n]$ 作为输入,其输出为行为的排序列表,排序越高的行为属于此链的概率越大。对于链中任一行为 $m$ , $f_m$ 是其自身与链中每个行为之间的点互信息(Pointwise Mutual Infomation)之和:

$$f(A_m) = \sum_{i=1}^m \ln \frac{p(A_i, A_m)}{p(A_i)p(A_m)} + \sum_{i=m+1}^n \ln \frac{p(A_m, A_i)}{p(A_m)p(A_i)}$$

其中, $p(x, y)$ 为裁判文书数据集中有序行为对 $(x, y)$ 出现的频数, $p(x)$ 是行为 $x$ 出现的频数。最后,为了识别行为链,使用行为链的计算得分 $f_m$ ,过滤掉小于阈值 $\theta$ 的行为,得到行为链 $CrA_{chain}$ 。利用中文近义词工具Synonyms计算犯罪行为词间的



在表4与表5中分别展示了犯罪行为抽取的结果与犯罪行为演化关系的示例。可以看出在表4中,犯罪行为元组包含犯罪行为词及其参数,犯罪行为元组可以简单明了地描述犯罪行为;在表5中展示了“犯罪行为-演化关系-犯罪行为”的犯罪行为三元组结构。

表4 犯罪行为元组示例

Tab.4 An example of a criminal act tuple

| 行为触发词 | 施事者  | 受事者   | 间接受事者    | 时间        | 地点 |
|-------|------|-------|----------|-----------|----|
| 付款    | 谭某   | 樊某    | 14,000 元 | 2017 年6 月 | —  |
| 指使    | 叶某   | 樊某    | —        | 2017 年6 月 | —  |
| 购买    | 樊某   | 甲基苯丙胺 | —        | 2017 年6 月 | 广东 |
| 返回    | 樊某   | —     | —        | 2017 年7 月 | 毕节 |
| 抓获    | 公安民警 | 樊某    | —        | 2017 年7 月 | 毕节 |

表5 犯罪行为演化关系示例

Tab.5 An example of a criminal act evolutionary relation

| 犯罪行为1 | 演化关系 | 犯罪行为2 |
|-------|------|-------|
| 付款    | 时序关系 | 指使    |
| 指使    | 时序关系 | 购买    |
| 购买    | 时序关系 | 返回    |
| 付款    | 因果关系 | 购买    |
| 返回    | 时序关系 | 抓获    |

## 6 结论(Conclusion)

本文通过犯罪行为抽取方法生成犯罪行为链,构建犯罪行为演化图谱,对裁判文书的案情描述进行层层抽象。在此过程中提出了一种新的案情描述框架——犯罪行为演化图谱,以此来对案情进行包含上下文语境的结构化表示,为警务工作和司法辅助审判的信息化建设提供了一种新的案情分析方法。在未来的工作中,我们将继续完善和推广犯罪行为演化图谱理论,并深入挖掘犯罪行为演化图谱在犯罪行为预测、类案推荐、罪名预测、法条推荐等方面的应用。

## 参考文献(References)

- [1] LIAN H, QIN Z, HE T, et al. Knowledge graph construction based on judicial data with social media[C]// WANG G, LIN X, HENDLER J, et al. 2017 14th Web Information Systems and Applications Conference (WISA). Piscataway, USA: IEEE, 2017:225-227.
- [2] CHEN B, LI Z, SHEN S, et al. Judicial knowledge reasoning based on representation learning[C]// MENG S, DAI Y, LUO L, et al. 2019 IEEE 19th International Conference on Software Quality, Reliability and Security Companion (QRS-C). Piscataway, USA: IEEE, 2019:84-88.
- [3] LI Z, DING X, LIU T. Constructing narrative event evolutionary graph for script event prediction[C]// LANG J. Proceedings of the 27th International Joint Conference on Artificial Intelligence. Menlo Park, USA: AAAI, 2018: 4201-4207.

- [4] XU A. Chinese judicial justice on the cloud: A future call or a pandora's box? An analysis of the 'Intelligent Court System' of China[J]. Information & Communications Technology Law, 2017, 26(1):59-71.
- [5] 杨凯.公共法律服务智能应用新视野——以人工智能技术与审判辅助办案机制建构为中心[J].湖北警官学院学报,2020,33(05):28-40.
- [6] 秦永彬,冯丽,陈艳平,等. “智慧法院”数据融合分析与集成应用[J].大数据,2019,5(03):35-46.
- [7] 陈彦光,刘海顺,李春楠,等.基于刑事案例的知识图谱构建技术[J].郑州大学学报(理学版),2019,51(03):85-90.
- [8] HOLME P, SARAMÄKI J. Temporal networks[J]. Physics Reports, 2012, 519(3):97-125.
- [9] CHAMBERS N, JURAFSKY D. Unsupervised learning of narrative event chains[C]// MOORE J, TEUFEL S, ALLAN J, et al. Proceedings of ACL-08: HLT. Stroudsburg, USA: ACL, 2008:789-797.
- [10] LI Z, ZHAO S, DING X, et al. EEG: Knowledge base for event evolutionary principles and patterns[C]// CHENG X, MA W, LIU H, et al. Chinese National Conference on Social Media Processing. Singapore: Springer, 2017:40-52.
- [11] GRANROTH-WILDING M, CLARK S. What happens next? Event prediction using a compositional neural network model[C]// ALTO P. Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence. Menlo Park, USA: AAAI, 2016:2727-2733.
- [12] CHE W, LI Z, LIU T. Ltp: A Chinese language technology platform[C]// LIU Y, LIU T. Coling 2010: Demonstrations. Stroudsburg, USA: ACL, 2010:13-16.
- [13] JANS B, BETHARD S, VULIĆ I, et al. Skip N-grams and ranking functions for predicting script events[C]// DAELEMANS W. Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics. Stroudsburg, USA: ACL, 2012:336-344.
- [14] RAJPURKAR P, JIA R, LIANG P. Know what you don't know: Unanswerable questions for SQuAD[C]// GUREVYCH I, MIYAO Y. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, USA: ACL, 2018:784-789.

## 作者简介:

段锡辉(1996—),男,硕士生.研究领域:数据挖掘,知识图谱,机器学习.

黄瑞章(1979—),女,博士,教授.研究领域:数据挖掘,机器学习,聚类.