Integrating NLP and Structured Occurrence Nets for Crime Modelling: A Pattern-based Approach

Tuwailaa Alshammari[1]

[1]School of Computing, Newcastle University, Science Square, Newcastle upon Tyne, NE4 5TG, United Kingdom

Keywords
structured occurrence net, acyclic nets, communication structured acyclic net, natural language processing, event extraction

Structured Occurrence Nets (SONs) [1, 2] are a Petri net-based formalism for representing the behaviour of complex systems consisting of interdependent subsystems that proceed concurrently and interact with each other. In this work we aim at constructing SONs capturing crime reports expressed in natural language. This requires the identification of the subsystems (acyclic nets ANs) and their component nodes (in particular, transitions). Identifying relevant words that identify these components from the unstructured text can be challenging. Here we propose a method based on the automatic identification of subsystem components derived from manual modelling experiments.

The focus of the experiments are crime stories obtained from [3]. The participants were asked to extract and represent crime events using the SONs framework. In addition, our interest was stimulated by the prospect of observing the manual identification of ANs and corresponding events/transitions to evaluate the uniformity and consistency of the resulting models. The proposed methodology employs Natural Language Processing (NLP) to automatically detect and extract crime events, using two fundamental principles. Initially, it involves the extraction of potential entities, as represented by acyclic nets. Afterwards, it identifies transitions within these ANs in order to capture the system's behaviour for further analysis and evaluation. To accomplish this, a sentence pattern-based approach is proposed to process sentences and identify events, resulting in the generation of a comprehensive representation of the system.

Approach. In our previous work [4], we discussed three approaches to extract useful information from written crime stories, recognising root verbs, root verbs alongside common crime verbs, and the inclusion of all verbs to represent events. We concluded that, of the three approaches, using root verbs alongside common verbs was the most effective approach. However, it was limited by a list of common verbs that are usually found in crime stories. The current work involves two distinct phases. Initially, we analysed manual models created by SON experts. The second phase involved attempting to identify similar outputs automatically, i.e., by discovering methods to automatically identify and capture system components and behaviour - which, in the context of this experiment, refers to the behaviour of crime participants.

Manual modelling. Modellers generally selected entities such as people's names and locations to

International Workshop on Petri Nets and Software Engineering, June 26-27, 2023, Lisbon, Portugal
t.t.t.alshammari2@ncl.ac.uk (T. Alshammari)

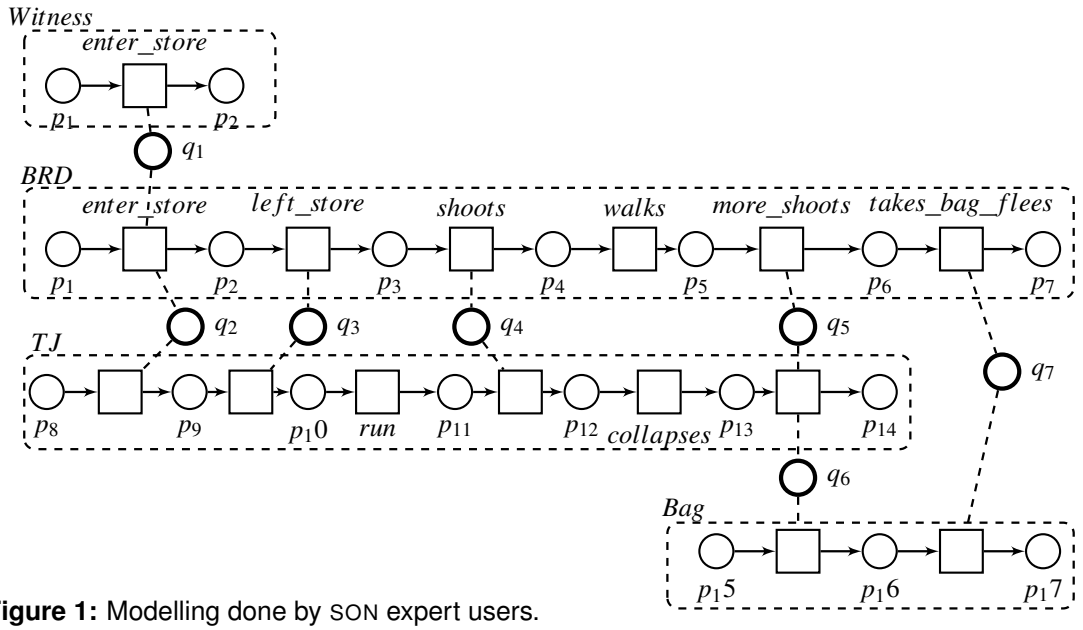

Figure 1: Modelling done by SON expert users.

represent different ANs. For this experiment, modellers selected: people's names, weapons, law enforcement, relative names, co-workers and professions, transportation, locations, and witnesses. The inconsistency in entity identification and, consequently, modelling was observed among the modellers. Events are critical in crime reports as they provide information about the actions that occurred during a crime. We observed that modellers generally represented events/transitions with verbs. We also found that modellers varied in the number of events selected.

Proposed modelling approach (automatic extraction). In order to create an automatic modelling approach that mimics human modelling, we analysed and used manually created models. As a result, and for more accurate and consistent extraction, rules were established to base our assumptions for linking natural language with SONs.

- ENTITIES in crime text are the represention of SON acyclic nets.
- VERBS represents SON EVENTS within acyclic nets.
- verbs shared by several entities will result in the formation of communication buffer places.

After analysing the manual models, we observed that specific entities are of more concern with crime investigation than others. Therefore, new NER labels were introduced [WEAPON, WITNESS, PROFESSION, TRANSPORTATION, RELATIVES].

We then implemented a pattern-based approach to identify verbs representing events from manually-created models. To be more precise, we examined the verbs indicated by most of the modellers and, using dependency parsing, we then determined where these verbs are located in the sentence in terms of the relationship between verbs and surrounding words.

Our objective was to find a pattern that would allow us to identify events by imitating human events recognition from verbs based on the manually-created SON models without the use of predetermined words in [4]. In this approach, we identified seven patterns, one example is when a

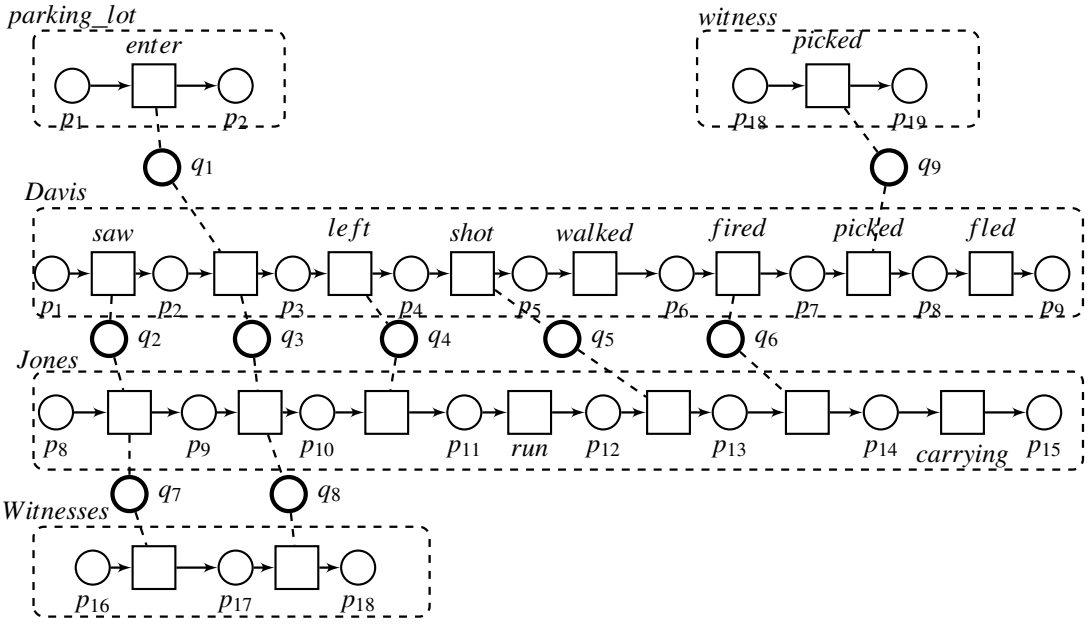

Figure 2: Potential SON model based on the results of the extraction

subject noun, proper noun, or pronoun precedes a verb that is followed by an object noun, proper noun, or pronoun.

Results. We conducted an experiment on one of the stories to test and compare the extractor output with human models. Figure 1 shows the human modelling and Figure 2 is the comparable model using the output from the extractor. The identified entities were [*parking lot, Davis, Jones, Witnesses, witness*]. Moreover, ten events were identified [*saw, enter, left, ran, shot, walked, fired, picked, carrying, fled*] compared to eight events identified manually.

Conclusion. This work proposes a sentence pattern-based verb detection (events) based on models generated manually. Initially, entities and events are identified to represent and describe acyclic nets and their components in order to construct formal SON models which can then be analysed Petri net-based methods.

References

[1] M. Koutny, B. Randell, Structured occurrence nets: A formalism for aiding system failure prevention and analysis techniques, Fundam. Informaticae 97 (2009) 41–91.

[2] B. Randell, Occurrence nets then and now: The path to structured occurrence nets, in: PETRI NETS 2011, volume 6709 of *Lecture Notes in Computer Science*, Springer, 2011, pp. 1–16.

[3] Violence Policy Center, [online] available at: https://vpc.org/., 2022.

[4] T. Alshammari, Towards Automatic Extraction of Events for SON Modelling, CEUR Workshop Proceedings 3170 (2022) 188–201.