

Real-World Causal Relationship Discovery from Text

Constantine Lignos, Chester Palen-Michel, Oskar Singer, Pedro Szekely, and Elizabeth Boschee

Information Sciences Institute
University of Southern California
Marina del Rey CA 90292, USA
{lignos, cpm, osinger, pszekely, boschee}@isi.edu

Abstract. Automatic extraction of causal relations from text has the potential to aid in the understanding of complex scenarios, but to date there has been limited work exploring extraction from natural data at scale. We describe a system that implements a rich language processing pipeline for the purpose of extracting causal relations between events described in text. The system uses a syntactic pattern-based approach to causality, using mutual bootstrapping to expand a set of seed patterns by discovering additional high-reliability patterns through a human-in-the-loop approach. We evaluate the performance of the system on newswire data and explore the properties of the causal relations it identifies.¹

Keywords: Causal relationship extraction · Information extraction · Natural language processing

1 Introduction

Existing manual methods to holistically represent the causal relationships contained in complex geopolitical, sociopolitical, and economic environments are both labor and time intensive. Our goal is to empower the understanding of such environments by automatically organizing the information available from vast and diverse data sources, ultimately allowing decision makers to explore the impact of their possible actions via quantitative analysis and what-if simulations. For this poster, we focus specifically on the extraction and organization of causal information from text, which is critical to downstream planning.

Causality extraction from text has been previously explored in both unsupervised and supervised settings, but rarely in the context of a real-world application. For instance, the natural language processing (NLP) community developed a 2010 shared task [6] that provided data annotated with semantic relations, including

¹ Copyright (c) 2019 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0). This material is based upon work supported by United States Air Force under Contract No. FA8650-17-C-7715.

causality, between nominal phrases in sentences, e.g. *the outbreak is the biggest ever caused by the vaccine*. However, a model can be very successful on these curated sentences but perform extremely poorly “in the wild”. A simple model we trained achieved a precision of 0.88 and recall of 0.91 on this data set, but precision plummeted to < 0.10 when run directly on a broad sample of newswire, where sentences were significantly more complex and true causality more rare.

For this poster, we describe two critical elements of our approach to real-world extraction of causal relationships from text: a bootstrapping approach to causal pattern discovery and a flexible end-to-end document processing pipeline that integrates a complex variety of customized and off-the-shelf NLP components.

2 Causal Relationship Discovery

The core concepts of mutual bootstrapping have been well-studied; Riloff and Jones provide an excellent retrospective summary [11] in honor of their seminal work from 1999 [10]. To be effective, the techniques we use here require two different “views” of the data: the lexical and ontological content of two possibly causally-related events and the syntactic connections between them.

We begin by manually generating a small set of simple causal sentences, e.g. *the protests caused violence*, which we use to generate an initial set of syntactic patterns. We use these patterns as the seed for our automatic, iterative pattern discovery process, applying them to a large unannotated corpus. At each iteration a small amount of human feedback can be included to reduce semantic drift.

To begin an iteration, the system applies all already-discovered patterns (including the seed set) to generate pairs of events likely to be causally related, e.g. *Attack(‘bombing’)* and *Death(‘fatalities’)* from the sentence *The bombing caused three fatalities*. Our hypothesis is that these pairs of events may contain characteristics that can predict these causal relationships; perhaps *Attacks* are likely to be causes, and particularly likely to be causes of *Deaths*. The use of ontological classes is critical in moving beyond the simple patterns that capture the low-hanging fruit of this task (e.g. *X causes Y*). For this effort, we use the ontology developed for DARPA’s Causal Exploration program, which contains ~ 500 event types, mostly in the domain of sociopolitical and military events, e.g. *CounterTerrorismOperation* or *TariffOrTradeSanctions*. At the same time, our system also finds causal relationships between unontologized events, e.g. *expressing fluorescent protein in some tissue allows us to see individual cells*.

For example, in our first iteration, the system estimates that *Attack/Death* pairs are twice as likely as a random pair of events to be causally related; if the *Attack* is also ontologically classified as a *MilitaryAction*, a causal relationship is forty-one times more likely. We assign each event pair a score that reflects that likelihood, combining a measure of the “causality” of each individual word (e.g. *bombing* as a cause and *fatalities* as an effect), the word pair together, and if available the same for their ontological classes (e.g. *cause=event:Attack*, *effect=event:Death*, as discussed above). Event pairs with scores above an experimentally-tuned threshold are deemed *causally-likely* for this iteration.

The second stage of each iteration then generates new patterns using all instances of these causally-likely event pairs in the corpus. For instance, the *bombing/fatalities* pair might suggest a new pattern (cause) -[nsubj]-> result <-[prep]- in <-[pobj]- (effect) from the sentence *The bombing resulted in five deaths*. Each pattern represents a dependency path between two nodes, where each node consists of the original text, stemmed text, and optional *input* and *output* labels. The input labels categorize the node as a member of a particular ontological class (e.g. *Person* or *Attack*). The output labels indicate the role that a node plays in a causal relation (e.g. cause, precondition, mitigating factor, effect). The labels on the paths (e.g. [nsubj]) represent universal dependency relations. Each node attribute and dependency edge label can become a wildcard, allowing one path to become many patterns with varying levels of granularity.

Each pattern is evaluated on the basis of the event pairs it generates and how causally-likely each one is deemed by the system. The highest scoring patterns are retained, and their matches can then be used to inform the scoring decisions in the next iteration. In practice, we use the scoring mechanism to generate a ranked list of candidate patterns, and use a human-in-the-loop approach to select new patterns to avoid semantic drift by only adding high-precision patterns. Here is a selection of patterns suggested by our system at an early iteration, along with examples of causal relationships they extract.

(effect) -[ccomp]-> mean <-[nsubj]- (cause)

Safina's ranking means the Williams sisters will be in the same half of the draw.

(effect / event:Death) <-[advcl]- (cause / event:Attack)

Two police officers died after a car rigged with explosives was detonated.

(cause / event:Decrease) -[advcl]-> (effect / event:Decrease)

Automakers are expected to reduce vehicle production by 25 percent from last year, when auto sales fell 18 percent from 2007 levels.

The first and second are good, high-precision additions to our collection, the former applying to all possible pairs and the latter restricted by ontological classes (*Attack/Death*). The third, however, is too general, since it frequently produces false positives where the two events are merely coincident, for example *Domestic traffic fell 4.6 percent while international traffic fell 11.2 percent*.

We evaluated our system on 1,000 newswire documents using a baseline of 660 seed patterns developed by pattern-writing experts over several months. Even on top of this strong baseline, our iteratively-discovered patterns improve recall by 6.6% while maintaining the precision of the seed patterns (~70%).

3 End-to-End System

To provide the rich structure required for applying causal patterns, we developed an NLP pipeline which combines software and components developed specifically for this system. Documents are processed in the following sequence: DetectorMorse [5] identifies sentence boundaries, spaCy [7] provides tokenization, part of speech tagging, and dependency parses, a conditional random field-based system trained on ACE [2] data identifies named entities, a sieve-based [9] system provides

entity coreference, Joint IE [8] and BBN ACCENT [1] provide ACE and CAMEO [4] ontology events, and JAMR [3] provides abstract meaning representation (AMR) parses. We use a new, flexible Python NLP framework (ISI VistaNLP) to integrate all components, implement the named entity extraction, coreference, and causal pattern discovery and matching systems, and perform experiments. This framework enables the simultaneous representation of multiple analyses of a document (e.g. events extracted by multiple systems using different ontologies), allowing the merging of information extracted by all components to a single representation using a single ontology.

4 Conclusion

The system described here provides the structure needed to identify causal relations between linguistically-rich events and a framework for matching and discovering causal patterns. This precision-centric pattern-based approach can be bootstrapped from a relatively small set of seed patterns and human-in-the-loop filtering, enabling the identification of relations without annotated data.

References

1. Boschee, E., Lautenschlager, J., O'Brien, S., Shellman, S., Starz, J., Ward, M.: BBN ACCENT event coding evaluation. In: ICEWS Coded Event Data. Harvard Dataverse (2015), <https://doi.org/10.7910/DVN/28075/GBAGXI>
2. Doddington, G.R., Mitchell, A., Przybocki, M.A., Ramshaw, L.A., Strassel, S.M., Weischedel, R.M.: The automatic content extraction (ACE) program-tasks, data, and evaluation. In: Proc. LREC. vol. 2, p. 1 (2004)
3. Flanagan, J., Thomson, S., Carbonell, J., Dyer, C., Smith, N.A.: A discriminative graph-based parser for the abstract meaning representation. In: Proc. ACL. vol. 1, pp. 1426–1436 (2014)
4. Gerner, D.J., Schrodt, P.A., Yilmaz, O., Abu-Jabr, R.: Conflict and mediation event observations (CAMEO): A new event data framework for the analysis of foreign policy interactions (2002), presented at the Annual Meeting of the ISA
5. Gorman, K.: DetectorMorse (2014), <https://pypi.org/project/DetectorMorse/>
6. Hendrickx, I., Kim, S.N., Kozareva, Z., Nakov, P., Séaghdha, D.O., Padó, S., Pennacchiotti, M., Romano, L., Szpakowicz, S.: SemEval-2010 task 8: Multi-way classification of semantic relations between pairs of nominals. In: Proc. International Workshop on Semantic Evaluation. pp. 33–38 (2010)
7. Honnibal, M., Montani, I.: spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing (2017)
8. Li, Q., Ji, H., Huang, L.: Joint event extraction via structured prediction with global features. In: Proc. ACL. vol. 1, pp. 73–82 (2013)
9. Raghunathan, K., Lee, H., Rangarajan, S., Chambers, N., Surdeanu, M., Jurafsky, D., Manning, C.: A multi-pass sieve for coreference resolution. In: Proc. EMNLP. pp. 492–501 (2010)
10. Riloff, E., Jones, R.: Learning dictionaries for information extraction by multi-level bootstrapping. In: Proc. AAAI. pp. 474–479 (1999)
11. Riloff, E., Jones, R.: A retrospective on mutual bootstrapping. *AI Magazine* **39**(1), 51–61 (2018)