Learning to Extract International Relations from News Text
Brendan O’Connor,† Brandon M. Stewart,‡ Noah A. Smith†
†School of Computer Science, Carnegie Mellon University ‡Government Department, Harvard University
More information: http://brenocon.com/irevents

Event data in international relations
What are the causes of war and peace? Do democracies engage in fewer wars? Why do some crises spiral into conflict, but others are resolved peacefully? Can we forecast future conflicts?

To help answer these questions, political scientists use event data: historical datasets of friendly and hostile interactions between countries, as reported in news articles. How can we extract this structured information, from millions of news articles?

Previous work: knowledge engineering
Besides manual coding (which is too labor-intensive at scale), previous work in political science uses a knowledge engineering approach: a manually defined ontology of event types and 15,000 textual patterns to identify events. This took decades of knowledge engineering to construct, is very difficult to maintain and must be completely rebuilt for new domains (e.g. domestic politics, commercial news, literature...)

We seek to automate some of this process: from the textual data, is it possible to automatically learn the semantic event types, and extract meaningful real-world political dynamics?

Our approach: learning both event types and political dynamics

Event Tuples

Model Inferences

Model

Quantitative evaluations

Conclusions

Our method simultaneously
(1) extracts a database of political events
(2) infers latent sociopolitical context
(3) organizes insightful summaries of this large and high-dimensional textual data.

Next steps include semi-supervised methods to exploit previously built knowledge bases, which will greatly help political science researchers, the incorporation of temporal and location textual analysis, and discovery of new actors and their properties.

More generally, event data analysis from political science is an interesting and exciting application area of NLP. It combines traditional concerns in text mining with information extraction and semantics. Numerous techniques and approaches are possible.