

A Knowledge Environment for Social Science Exploration

Kurt Rohloff
BBN Technologies
10 Moulton St.
Cambridge, MA, USA
krohloff@bbn.com

Wayne Thornton
BBN Technologies
10 Moulton St.
Cambridge, MA, USA
wthornton@bbn.com

ABSTRACT

The ability to understand social processes has improved dramatically with the application of emerging information technologies to social science analysis. Key technologies include automated data collection, knowledge representation, model integration and data visualization. We discuss an end-to-end distributed knowledge system that supports:

- Automated collection and classification of unstructured data (such as raw news feeds and communications data) and collection of structured data
- Automated fusing of extracted information with historical structural and geospatial datasets using Semantic Web technologies to support distributed modeling and analysis.
- Context-dependent data visualizations, including faceted browsing and spatio-temporal displays, to reveal underlying structures, patterns, and correlations.

Our system utilizes technologies for automated collection and classification of unstructured data including a set of named entity, relationship, and event extraction capabilities that operate over the entire content of articles and can “learn” or evolve over time. We applied these technologies to analyze a voluminous corpus of news feed data that covered a wide geographic region over a period of ten years. The extracted information formed the basis of theory-based independent variables (such as general tension metrics, non-state actor attributes, and leadership characteristics) as well as augmenting the more stale historical factors extracted from existing social science and econometric datasets.

Using Semantic Web technologies, our knowledge system fuses information extracted via the natural language tools described with data from numerous social science datasets to develop a knowledge environment. This capability also provides a basis for automated reasoning and inference for

model and analysis results integration. By directly encoding the semantics typically stored in dataset codebooks, the system fuses multiple datasets that goes beyond superficial dataset alignment (such as merely sorting data from various datasets by county and year). Our system’s knowledge infrastructure supports model-agnostic access to the stored data for further manipulation and analysis. Modeling and analysis results can be fed back into the knowledge system for access by other models.

The capstone interface of our knowledge system employs data visualization techniques to display data analysis results and provide interactive “drill-down” capabilities to better study results. Faceted browsing of factors and patterns based on these data values allows a user to select different events and the associated variables associated with the events in various countries.

Categories and Subject Descriptors

H.4 [Information Systems Applications]: Miscellaneous;
D.2.8 [Software Engineering]: Metrics—*complexity measures, performance measures*

General Terms

Delphi theory

Keywords

ACM proceedings, L^AT_EX, text tagging