Learning proximity of events from context in news

Sen Xu
GeoVISTA Center, Department of Geography, Pennsylvania State University
senxu@psu.edu
Motivation
Web: *information and knowledge exchange platform*

- Also: a *data repository* for analysts
- Semi-structured, constantly updating, open, accessible,
- A variety of data sources
Challenging topic: Proximity

- How close something is to something else (Encycl. of GIScience)
- Relates to distance, but hard to be defined quantitatively
  - 1 mile is near, how about 1.01 miles?
- Context-dependent
Why proximity?

Pullar and Egenhofer’s geographical scale spatial relations (1988)

- Direction: north, northwest
- Topological: disjoint, touches
- Ordinal: in, at
- Distance: far, near
- Fuzzy: next to, close
Existing research on proximity

- Lundberg and Ekman, 1973; Guttman, 1968
  - Proximity and distance have linear relationship
  - Relative distance
  - Perceived scale (larger area has larger distance for proximity)
- Gahegan, 1995 proposed a combined proximity measure to integrate absolute and relative proximity
  - Scale
  - Attractiveness
  - Reachability
- Sadalla et al. 1980: asymmetrical
- Worboys 2001: a fuzzy membership function for nearness neighborhood using experiment data
Inspiration

- Use rich content available from the web
  - Corpus about proximity

- Study different attributes that influence proximity
  - Type: point, line, polygon
  - Popularity/attractiveness (Gahegan, 1995): status of liberty, local grocery store
  - Semantics: event type
Iranian forces kill Kurdish rebels near Iraq border
Fires near Houston force evacuations, strain resources
Two students wounded at party near USC campus
Research Question

- Is there difference in proximity given different context?
- how to define context?
- Is it possible to summarize the proximity difference (from web corpus) into a context ontology?
- How to prove the summarized proximity ontology matches how human perceive proximity?
Methods

- Corpus-based linguistic analysis
- Spatial entity extraction and geocoding
- Behavioral experiments
Building and Analyzing Proximity Corpus

**Data Collection**
- RSS crawler
- Document classification

**Data Processing**
- HTML/XML parsing
- Spatial entity extraction

**Proximity Instances (Distance)**
- Place Type
- Event Type
- Popularity/attractiveness
Building Proximity Corpus

- RSS web-crawler for keyword news search
  - “near”, “close to”, “in the neighborhood of”
- However, not all search result are regarding spatial proximity
  - “near death experience”, “close to becoming history”, “in the neighborhood of 10 million dollars”
- Spatial preposition has extended meaning that are widely used in everyday life
- Binary Document Classifier (Zhang, 2010)
Interpretation

- Meta-data
  - pubDate, news source
- Geocoding
  - From spatial entity in text to <lat, lon>
- Spatial type recognition
  - Popularity/Attractiveness
- Event type categorization
  - Crime, nature disaster
- Scale
  - Local, region, national, international
Validation

- Paragraphs that contextualize proximity
- Geo-referenced location presented in maps
- Ask participants degrees of proximity
- Correlates with result from web corpus
Discussion
Outcome

- Proximity criteria for different attributes
  - Assist analysis of events
  - Provide guidelines for fuzzy spatial reasoning
  - Interpretation of spatial language
- Data collection/processing workflow can be extended to other data sources
  - Difference in proximity criteria from news vs blogosphere
Questions / Comments?

Special Thanks to

Advisor:
Alexander Klippel

Committee members:
Alan MacEachren
Donna Peuquet
Prasenjit Mitra

Colleagues at the
Human Factors in GIScience Lab