Events, patterns and analysis: Forecasting conflict in the 21st century

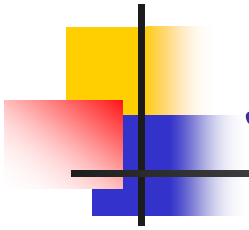
Devika Subramanian and Richard J. Stoll

Rice University

NSF- ITR- 0219673

Collaborations

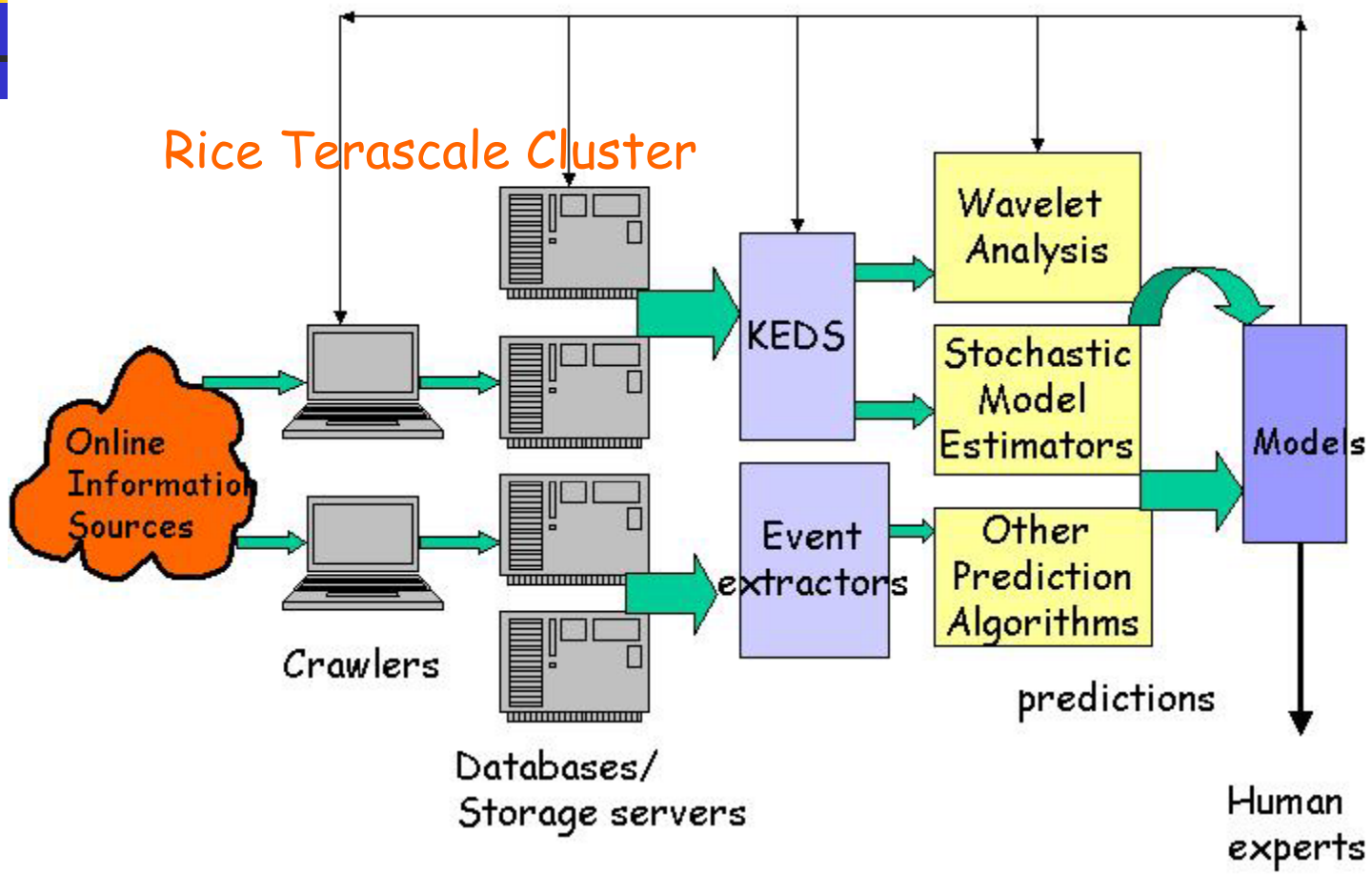
- No formal collaborations with government agencies yet.
- Informal interactions with government intelligence services and with the Center for the Scientific Study of International Conflict and Cooperation at Rice.
  - These interactions have helped us assess the relative merits of various online media sources.
- Support from the Computer and Information Technology Institute (CITI) at Rice in the form of cycles on the Rice Terascale cluster.

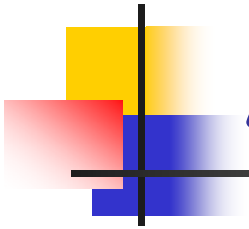


Scientific Research Objective

- With the explosion of online news over the past few years and with recent advances in extracting information from text, as well as the ready availability of compute power:
  - Is it possible to monitor news media from regions all over the world over extended periods of time, extracting low-level event information from them, and piece them together to automatically to track and predict conflict in all the regions of the world?
- This research is motivated by this question and we have completed some studies that suggest the power of this approach.

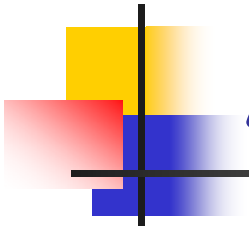
The Rice Conflict Prediction System





Accomplishments (Part 1)

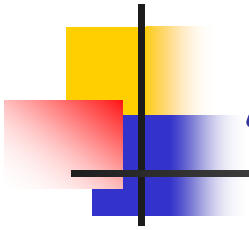
- Gathering and archiving news
  - We have designed and implemented a system that extracts stories from online news sources.
    - We have ALL Mid-east stories from the AP (1998 – present), BBC (1998-present), AFP (1998-present), Reuters (1996-1997), Washington Post (1977-2003), Boston Globe (1979-2003), Scripps Howard newswire (1990-2003), Houston Chronicle (1985-2003).
  - We have constructed a custom relational database system for archiving news stories. The system supports complex queries over stored stories.



Accomplishments (Part 2)

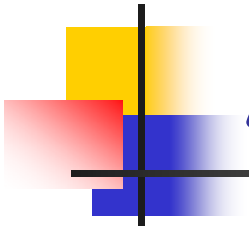
- Filtering stories not pertaining to conflict.
  - Only about 10% of the stories contain events that are to be extracted.
  - The rest are interpretations of conflictual events (e.g., op-eds), or are events not about conflict (e.g., sports, human-interest stories, etc.).
- We have trained a Naïve Bayes classifier (with specificity and sensitivity of about 80%) for this phase. Further improvements using semantic information about words (Wordnet) are being implemented.

Example of event extraction



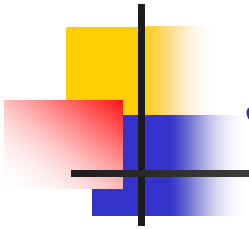
Accomplishments (Part 3)

- Coding relevant stories
  - Who did what to whom and when?
    - Identifying actors and targets.
    - Identifying events types from verbs.
  - Simple declarative sentences can be coded easily.



Accomplishments (Part 4)

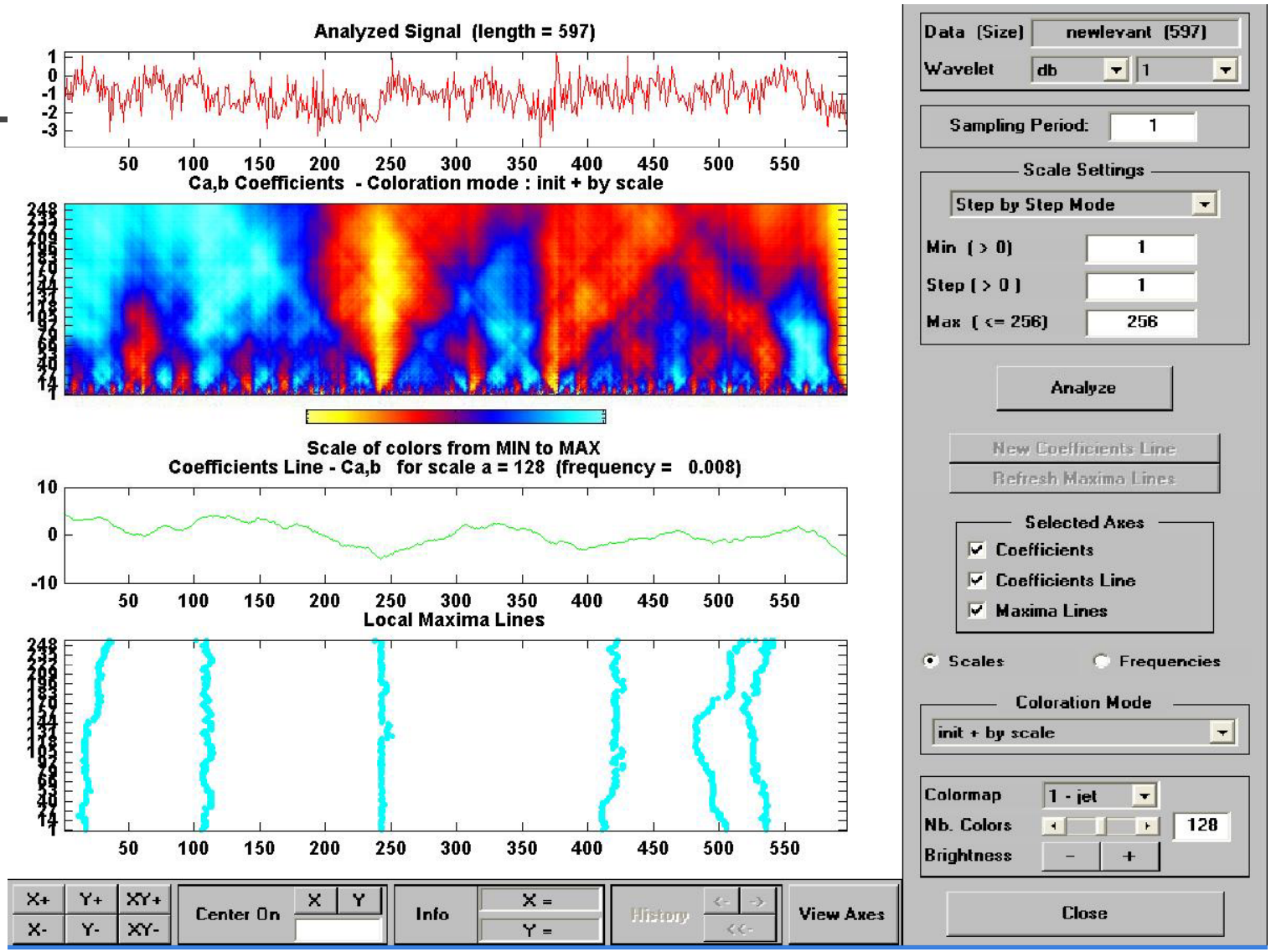
- Analysis of event data
  - Mid east event data from 1979-1999 (gathered from Reuters by KEDS).
  - Cold war event data from WEIS dataset.
- Questions
  - Is the time series of event data stationary? If not, does it have singularities/trends?
  - How predictable are conflict levels in the quasi-stationary segments of the data?



Success stories

- Event data time series are highly non-stationary. Surprisingly, singularities in the time series (detected by multi-scale wavelet analysis) correspond to major conflicts.
- Event streams before singularities have distinct temporal signatures and can be used for predicting when singularities could occur.

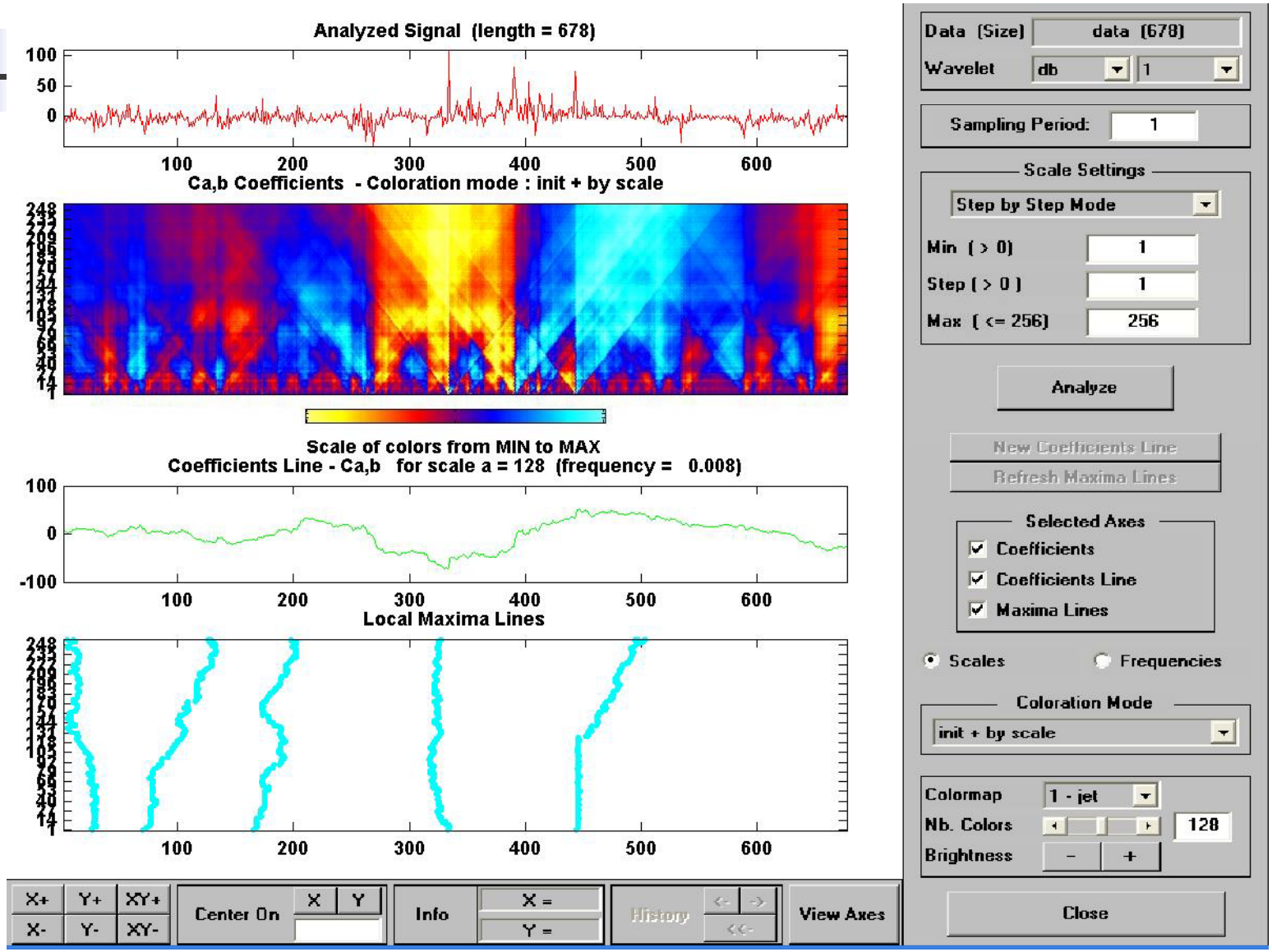
Analyzing conflict (Mid-east)



Mid-east analysis

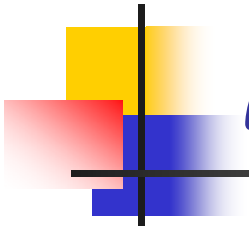
biweek	Date range	event
17-35	11/79 to 8/80	Start of Iran/Iraq war
105-111	4/83 to 7/83	Beirut suicide attack
244	1/91 to 2/91	Desert Storm
413-425	1/95 to 7/95	Rabin assassination/start of Intifada
483-518	10/97 to 2/99	US/Iraq confrontation via Richard Butler/arms inspectors
522-539	4/99 to 11/99	Camp David summit hosted by Clinton

Analyzing conflict (Cold war)



Cold war analysis

Week(s)	Date(s)	Comments
56 – 83	67/1 – 67/8	[Red] Six Day War
132	68/7	[Red] Czech Crisis (138; 68/8), but NPT signed
187-265	69/8 – 71/1	[Red] Cienfuegos (Cuban sub base)
271-289	71/3 – 71/7	[Blue] Soviet ships leave Cienfuegos.
295-312	71/8 – 71/12	[Red] Soviets criticize Nixon trip to China, Moscow summit
316-334	72/1 – 72/5	[Blue] Moscow summit, SALT, ABM Treaties
341-363	72/7 – 72/12	[Red] US "Christmas bombing" of North Vietnam
364-390	72/12 – 73/6	[Blue] US-Soviet Summit (Brezhnev to US)
391-585	73/6 – 77/3	[Red] October War Crisis; Angola; Carter criticizes Soviets on human rights
620-649	77/11 – 78/6	[Red] Soviet UN employees arrested. Carter criticizes dissident trials. Soviet stage large maneuvers in Central Europe. Ginzburg & Shchransky sentenced.

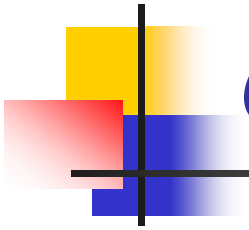


Management structure

- PIs: Devika Subramanian (Computer Science) and Richard J. Stoll (Political Science) at Rice University.
- Students: A. Larson (undergraduate), T. Maxwell, L. Chang (CS), M. Stepanyan, R. Sandoval-Bustos, D. Torres (Poli Sci).

Broad Impact

- Media coverage
  - Future Tense (MPR): radio interview w/ Jon Gordon (January 26, 2004).
    - http://www.publicradio.org/columns/futuretense/2004/01/index.shtml
  - The Digital Government Newsletter (January 2004)
    - http://www.diggov.org/news/dgonline/dgonline_0104.jsp
  - TV interviews with KPRC (Channel 2) and Fox TV (Channel 26) in Houston in 2003.
- Papers being prepared for PNAS, AIJ, and the Baker Institute.
- Paper to be presented at national Political Science forum (International Studies Association 2004).



Challenges

- Extracting events from stories is a very challenging problem. News wires no longer written in the 'rat-a-tat' subject-verb-object style anymore.
- New methods needed for segmenting non-stationary event time series data into quasi-stationary segments.
- New forecasting algorithms needed for making conflict predictions.

Research value in working in the DG domain

- Computer science
  - Will advance information extraction technology.
  - Will push state-of-the-art in statistical machine learning algorithms for prediction over long range time series data.
- Political science
  - Will fundamentally transform the production of event data.
  - Will allow formulation and testing of process theories of conflict based on comprehensive event data sets.

Recommendations for the DG program

- Consolidate information extraction projects in the DG program, and create a resource for researchers to share ideas and code for text mining.