



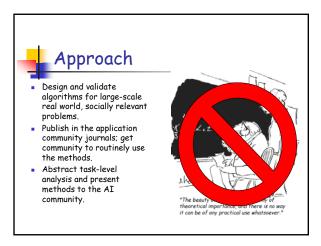
Why embed learning?

- We cannot calculate and implement an action-choice/decision-making strategy for the system at design time.
 - System dynamics are unknown/partially known.
 - System dynamics change with time.
 - A one-size-fits-all solution is not appropriate - customization is needed.



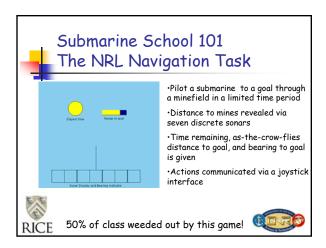
Research questions in adaptive embedded system design

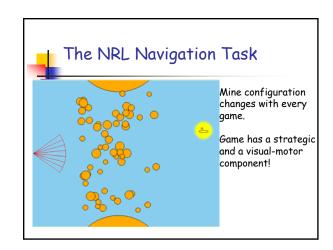
- Representation: What aspects of the task, environment and system dynamics do we need to observe and model for decision making?
- Learning: How can we build and maintain embedded models in changing environments?
- Decision making/acting: How can we use these models effectively to make decisions with scarce resources in changing environments?

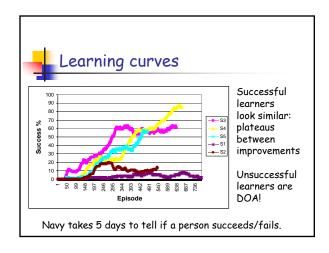


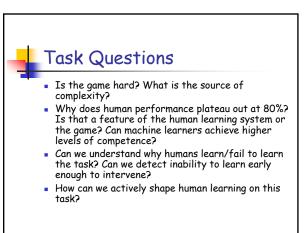


- - Unknown system, changing dynamics
 - Tracking human learning on a complex visual-
 - Predicting the evolution of international conflict.





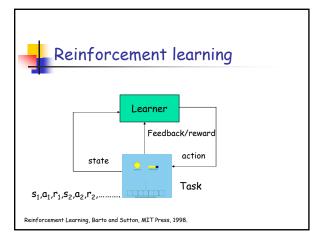






Mathematical characteristics of the NRL task

- A partially observable Markov decision process which can be made fully observable by augmentation of state with previous action.
- State space of size 10¹⁴, at each step a choice of 153 actions (17 turns and 9
- Feedback at the end of up to 200 steps.
- Challenging for both humans and machines.





Reinforcement learning/NRL task



- Representational hurdles
 - State and action spaces have to be manageably
 - Good intermediate feedback in the form of a non-deceptive progress function needed.
- Algorithmic hurdles
 - Appropriate credit assignment policy needed to handle the two types of failures (timeouts and explosions are different).
 - Q-learning is too slow to converge (because there are up to 200 steps in a single training episode).



State space design

- Binary distinction on sonar: is it > 50?
- Six equivalence classes on bearing: 12, {1,2}, {3,4}, {5,6,7},{8,9}, {10,11}
- State space size = 2⁷ * 6 = 768.
- Discretization of actions
 - speed: 0, 20 and 40.
 - turn: -32, -16, -8, 0, 8, 16, 32.

Automated discovery of abstract state spaces for reinforcement learning, Griffin and Subramanian, 2001.



The dense reward function

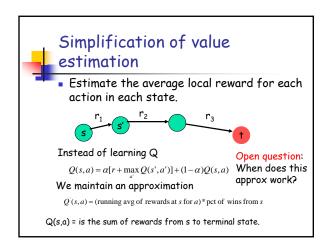
r(s,a,s') = 0 if s' is a state where player hits mine.

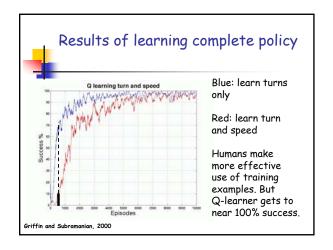
- = 1 if s' is a goal state = 0.5 if s' is a timeout state
- = 0.75 if s is an all-blocked state and s' is a not-all-blocked state
- = 0.5 + Diff in sum of sonars/1000 if s' is an all-blocked state = 0.5 + Diff in range/1000 + abs(bearing 6)/40 otherwise

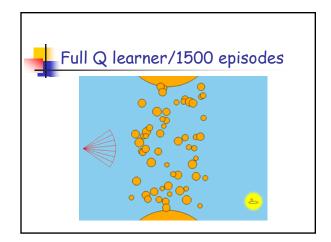


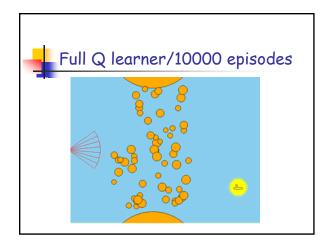
Credit assignment policy

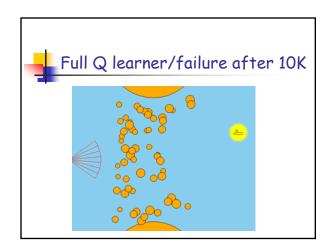
- Penalize the last action alone in a sequence which ends in an explosion.
- Penalize all actions in sequence which ends in a timeout.

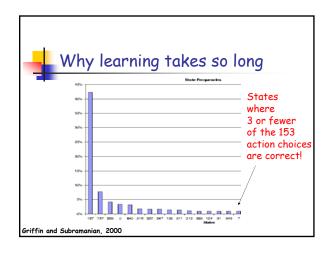














Lessons from machine learning

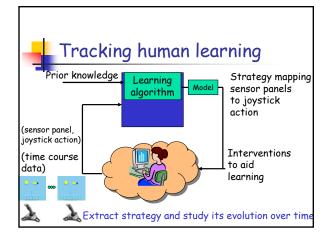
- Task level
 - Task is hard because states in which action choice is critical occur less than 5% of the time.
 - Staged learning makes task significantly easier
 - A locally non-deceptive reward function speeds up learning.
- Reinforcement learning
 - Long sequence of moves makes credit assignment hard; a new cheap approximation to global value function makes learning possible for such problems.
 - Algorithm for automatic discretization of large, irregular state spaces.

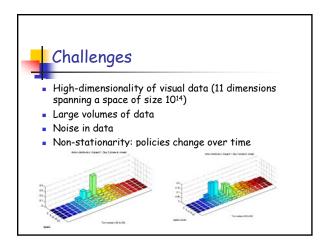
Griffin and Subramanian, 2000, 2001



Task Questions

- Is the game hard? Is it hard for machines? What is the source of complexity?
- Why does human performance plateau out at 80%? Is that a feature of the human learning system or the game? Can machine learners achieve higher levels of competence?
- Can we understand why humans learn/fail to learn the task? Can we detect inability to learn early enough to intervene?
- How can we actively shape human learning on this task?

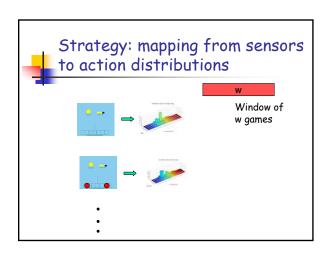






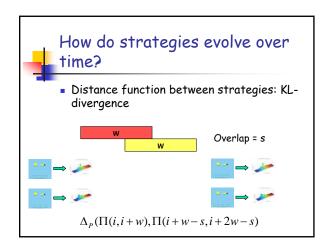
Embedded learner design

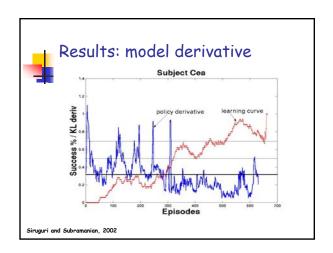
- Representation
- Use raw visual-motor data stream to induce policies/strategies.
- Learning
 - Direct models: lookup table mapping sensors at time t and action at t-1 to distribution of actions at time t. (1st order Markov model)
- Decision-making
 - Compute "derivative" of the policies over time, and use it (1) to classify learner and select interventions, (2) to build behaviorally equivalent models of subjects

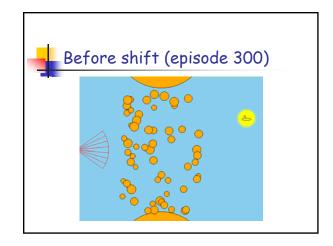


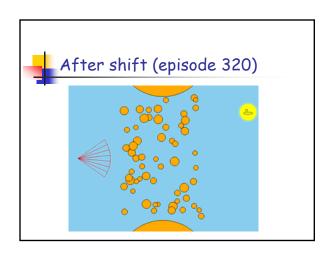


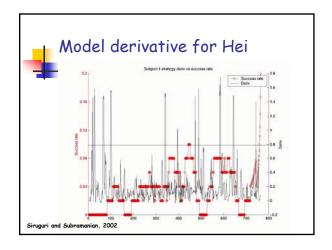
- There are 10¹⁴ sensor configurations possible in the NRL Navigation task.
- However, there are between 10³ to 10⁴ of those configurations actually observed by humans in a training run of 600 episodes.
- Exploit sparsity in sensor configuration space to build a direct model of the subject.







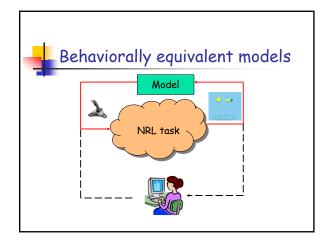






How humans learn

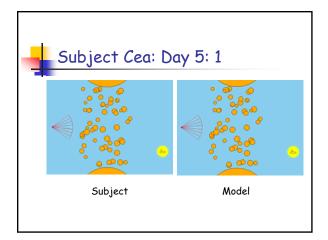
- Subjects have relatively static periods of action policy choice punctuated by radical shifts.
- Successful learners have conceptual shifts during the first part of training; unsuccessful ones keep trying till the end of the protocol!

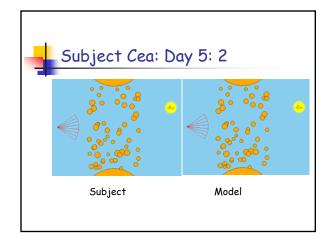


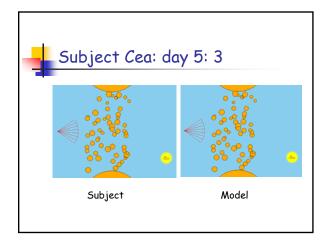


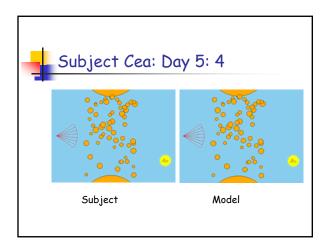
Generating behaviorally equivalent models

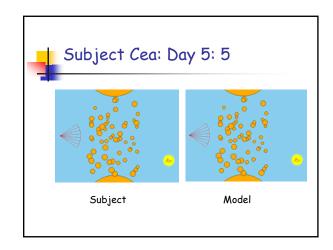
- To compute action a associated with current sensor configuration s in a given segment,
 - take 100 neighbors of s in lookup table.
 - perform locally weighted regression (LWR) on these 100 (s,a) pairs.

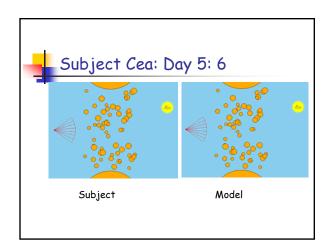


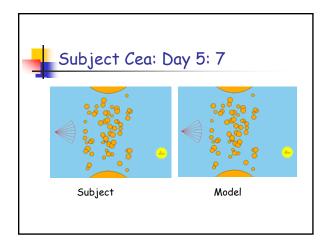


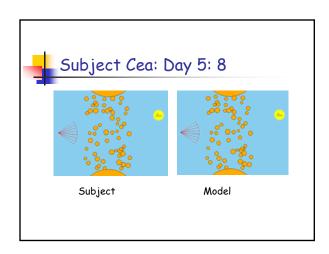


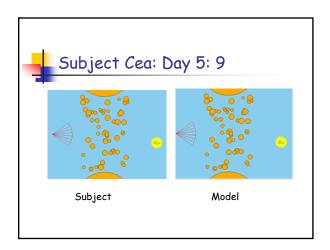


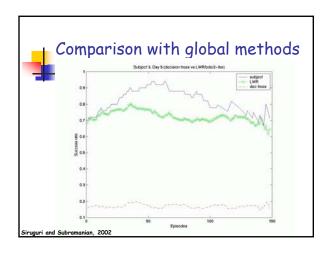














Summary

- We can model subjects on the NRL task in real-time, achieving excellent fits to their learning curves, using the available visualmotor data stream.
- One of the first in cognitive science to directly use objective visual-motor performance data to derive evolution of strategy on a complex task.





Lessons

- Learn simple models from objective, lowlevel data!
- Non-stationarity is commonplace, need to design algorithms robust with respect to it.
- Fast new algorithms for detecting changepoints and building predictive stochastic models for massive, noisy, non-stationary, vector time series data.



Neural correlates

 Are there neural correlates to strategy shifts observed in the visual-motor data?



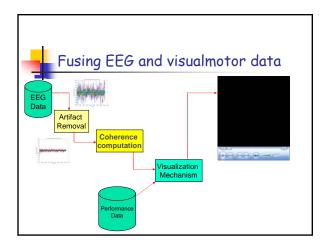


Task Questions

- Can we adapt training protocols in the NRL task by identifying whether subjects are struggling with strategy formulation or visual-motor control or both?
- Can we use analysis of EEG data gathered during learning as well as visual-motor performance data to correlate 'brain events' with 'visual-motor performance events'? Can this correlation separate subjects with different learning difficulties?









The coherence function

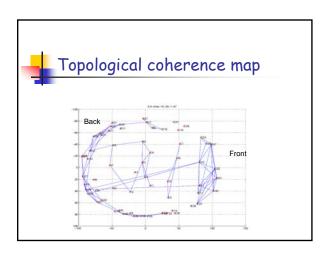
- Coherence provides the means to measure synchronous activity between two brain areas
- A function that calculates the normalized cross-power spectrum, a measure of similarity of signal in the frequency domain

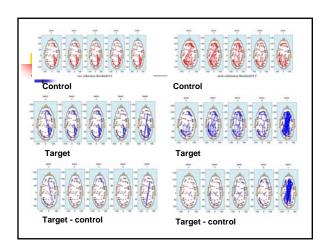
$$C_{xy}(f) = \frac{|S_{xy}(f)|^2}{[S_{xx}(f)S_{yy}(f)]}$$

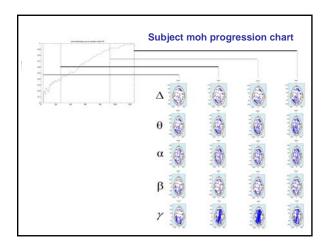


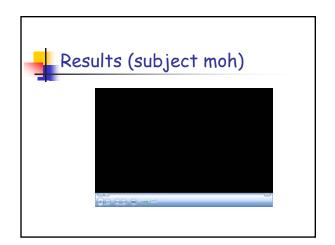
Frequency bands

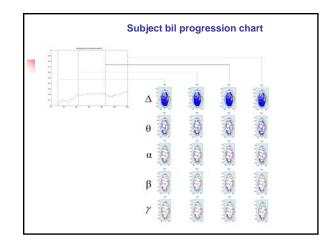
- Coherence map of connections in each band
 - ∆ (0-5 Hz)
 - θ (5-9 Hz)
 - a (9-14 Hz)
 - β (14-30 Hz)
 - γ (40-52 Hz)

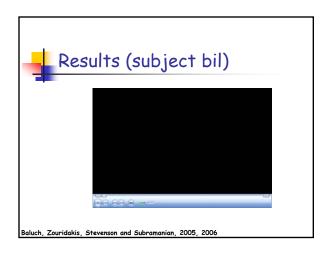


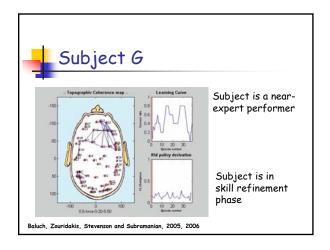


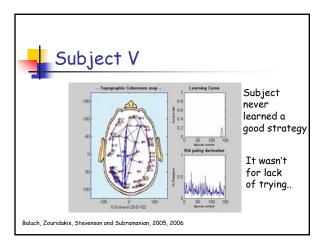














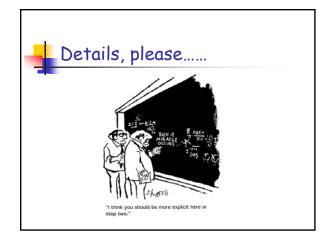
Results

- There are distinct EEG coherence map signatures associated with different learning difficulties
- Lack of strategy
- Shifting between too many strategies
- Subjects in our study who showed a move from a low level of performance to a high level of performance show front to back synchrony in the gamma range or long range gamma synchrony (LRGS).
- We are conducting experiments on more subjects to confirm these findings. (14 subjects so far, and more are being collected right now.)



What else is this good for?

- Using EEG readouts to analyze the effectiveness of video games for relieving pre-operative stress in children (A. Patel,
- Using EEG to read emotional state of players in immersive video games (M. Zyda,
- Analyzing human performance on any visualmotor task with significant strategic component.





Publications

- <u>Human Learning and the Neural Correlates of Strategy Formulation</u>, F. Baluch, D. Subramanian and G. Zouridakis, 23rd Annual Conference on Biomedical Engineering Research, 2006.
- Understanding Human Learning on Geompta Tasis by Functional Real Transing, D. Subramanian, R. Bandyapdhyay and G. Zourladkis, 20th Annual Conference on Biomedical Engineering Research, 2003. Tookins the evolution of Learning on a visualmost roak Devike Subramanian and Sameer Singury, Technical report TR02-401, Department of Computer Science, Rice University, August 2002. Tracking the evolution of Learning on a visualmost roak Sameer Siruguri, Master's thesis under the supervision of Devike Subramanian, May 2001.

- supervision or Devina outomanian, May Catalantia, S. Griffin and D. Subramanian, Technical report, Department of Computer Science, Rice University, June 2000.

 <u>Inducing hybrid models of learning from visualmator data</u>, Proceedings of the 22nd Annual Conference of the Cognitive Science Society, Philodelphia, PA, 2000.

- Completine Virtual Conference of the NRL Navigation 16st, Proceedings of the 20th Annual Conference of the Cognitive Science Society, Madison, WI, 1998 (with D. Gordon).

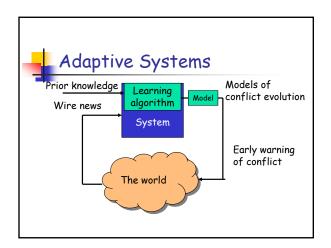
 A cognitive model of learning to reviousle, Proceedings of the 19th Annual Conference of the Cognitive Science Society, Stanford, CA, 1997 (with D. Gordon).

 Science Society, Stanford, CA, 1997 (with D. Gordon).
- Cognitive modeling of action selection learning, Proceedings of the 18th Annual Conference of the Cognitive Science Society, San Diego, 1996 (with D. Gordon)



Roadmap of talk

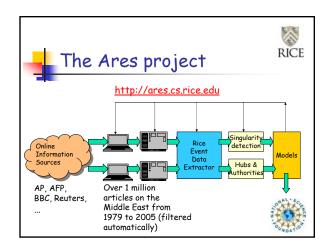
- Four case-studies
 - Unknown system, changing dynamics
 - Tracking human learning on a complex visualmotor task.
 - Predicting the evolution of international conflict.

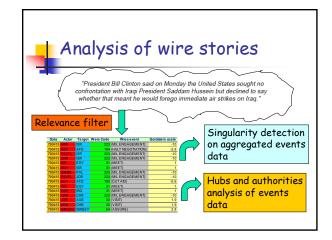




Task Question

Is it possible to monitor news media from regions all over the world over extended periods of time, extracting low-level events from them, and piece them together to automatically track and predict conflict in all the regions of the world?







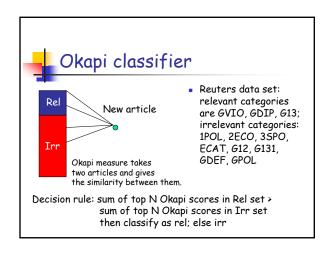
Embedded learner design

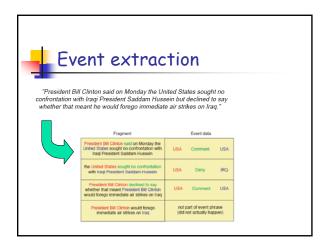
- Representation
 - Identify relevant stories, extract event data from them, build time series models and graphtheoretic models.
- Learning
 - Identifying regime shifts in events data, tracking evolution of militarized interstate disputes (MIDs) by hubs/authorities analysis of events data
- Decision-making
 - Issuing early warnings of outbreak of MIDs

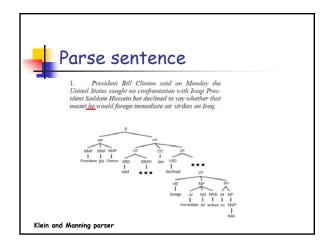


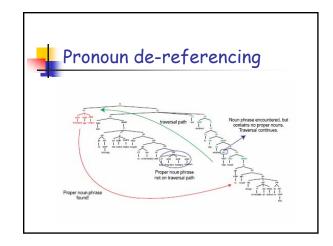
Identifying relevant stories

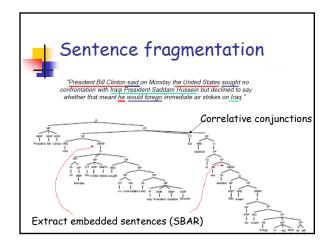
- Only about 20% of stories contain events that are to be extracted.
 - The rest are interpretations, (e.g., op-eds), or are events not about conflict (e.g., sports)
- We have trained Naïve Bayes (precision 86% and recall 81%), SVM classifiers (precision 92% and recall 89%) & Okapi classifiers (precision 93% and recall 87%) using a labeled set of 180,000 stories from Reuters.
- Surprisingly difficult problem!
 - Lack of large labeled data sets;
 - Poor transfer to other sources (AP/BBC)
 - The category of "event containing stories" is not well-separated from others, and changes with time

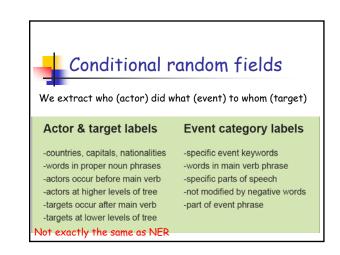


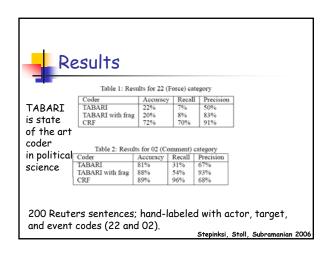


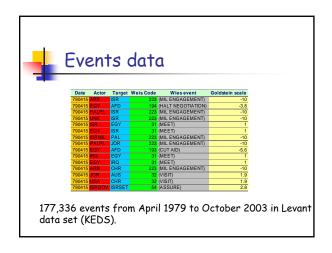




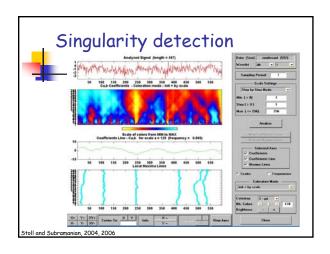


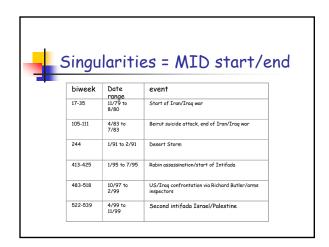


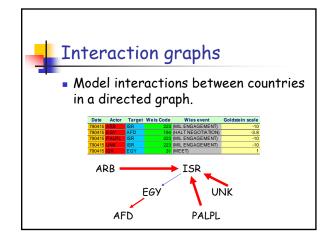








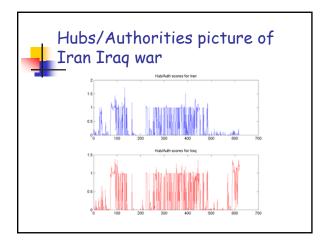


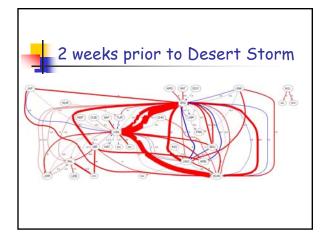




Hubs and authorities for events data

- A hub node is an important initiator of events
- An authority node is an important target of events.
- Hypothesis:
 - Identifying hubs and authorities over a particular temporal chunk of events data tells us who the key actors and targets are.
 - Changes in the number and size of connected components in the interaction graph signal potential outbreak of conflict.



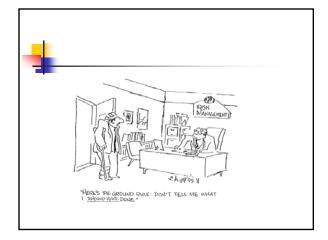




Validation using MID data

- Number of bi-weeks with MIDS in Levant data: 41 out of 589.
- Result 1: Hubs and Authorities correctly identify actors and targets in impending conflict.
- Result 2: Simple regression model on change in hubs and authorities scores, change in number of connected components, change in size of largest component 4 weeks before MID, predicts MID onset.
- Problem: false alarm rate of 16% can be reduced by adding political knowledge of conflict.

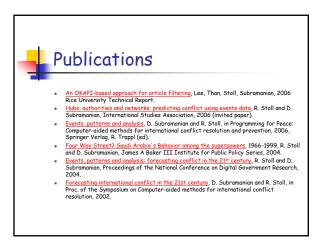
Stoll and Subramanian, 2006



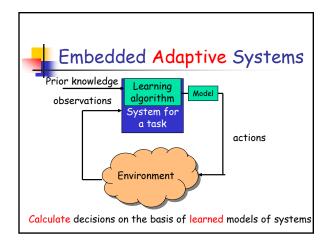


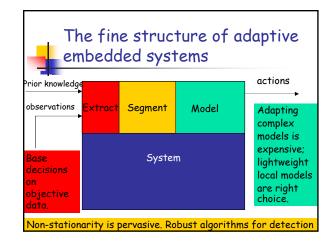
Current work

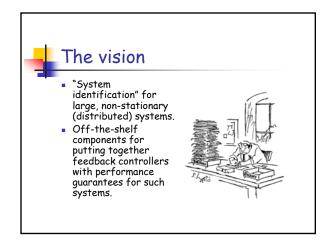
 Extracting economic events along with political events to improve accuracy of prediction of both economic and political events.



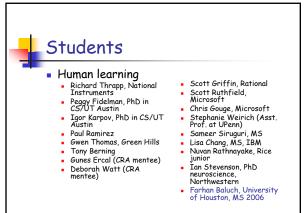


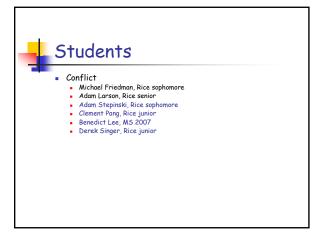














Sponsors

- Conflict analysis: NSF ITR 0219673
- Human learning: ONR N00014-96-1-0538



