SAND Number 2011-1524

Applications of Self-Correcting Causal-Learning Systems to Opinion Dynamics on Networks

Alexander Outkin
Robert J. Glass
Nancy Brodsky

Eighth International Conference on Complex Systems
Quincy, MA
June 26 – July 1, 2011

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Motivation

- Social systems can be represented as:
  - Engineered component - interaction structure and rules
    - Enforces constraints and processes
  - Decisional component - agent's strategies
    - Drives the system

Networks and Markets

- Agent-based models can represent both engineered and decisional components (the causal structure) of socio-economic-technical systems in which we live. Examples include:
  - Social networks
    - Dynamics are driven by agents' decisions within social or technological constraints
  - Markets
    - Agents have heterogeneous strategies
    - Operate within rules, regulations, and existing infrastructure

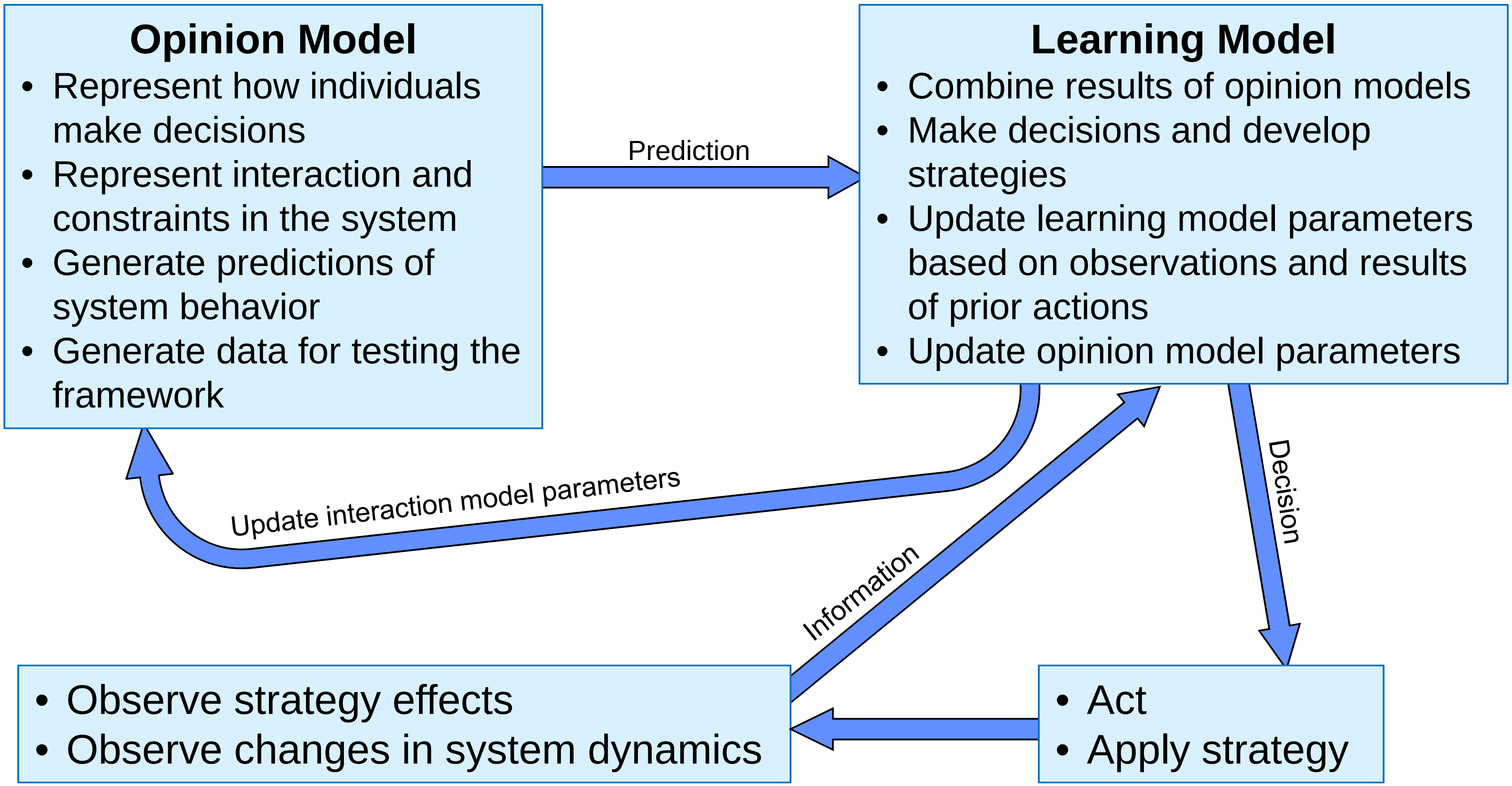
Effective Decision-making Requires an Understanding of Causal Mechanisms

- Challenges:
  - Social systems are only partially observable
    - Motivations for behaviors are typically deduced, not measured.
    - Behavioral data are often aggregated, and specificity is lost.
    - Therefore: Matching output of causal models to observed data is difficult.
  - Both the causal structure of a system and decision-making strategies of agents can change abruptly.
- Questions:
  - Is there a way of automating the competition among causal models to better our predictive capabilities?
    - Can it account for partially observable data (incomplete input)?
    - Can it be useful if system changes substantially?

Approach

- Create causal models of the system
- Incorporate results of multiple models into a learning framework.
- Learning framework maps predictions into actual observations
- Evaluate individual model contributions
- Re-calibrate models as new data arrives

Conceptual Framework: Act / Observe / Select / Recalibrate



Opinion Dynamics

- Real-world applications: identify a (set of) causal models that predict future system performance and response to interventions
- Test case:
  - Create a randomly generated opinion model and its underlying network
  - Apply the learning framework to predict its dynamics

Test Case: Existing Opinion Dynamics Model

- Complex Adaptive Systems of Systems Initiative codified and expanded opinion dynamics modeling framework*
- Network of agents share and affect each others' opinions
  - Weisbuch updating rule, for example
- Network topology may vary: scale-free, random, ring...

** Brodsky, N. et al. "Application of Complex Adaptive Systems of Systems Engineering to Tobacco Products", ICCS 2011.*

Test Case Parameters

- What we can observe:
  - Average behaviors
  - Number of nodes (more or less)
  - Partial topological or connectivity parameters
- What we cannot observe:
  - Individual strategies
  - Exact topology
  - All/complete opinions and behaviors for individual agents

Learning Framework

Individual causal models

- Represent causal structures and interactions that give rise to their dynamics
- Are calibrated to available data
- Attempt to predict system output

- A set of inputs $X = x_1, \ldots, x_K$ represent past system behaviors.
- A set of outputs $Y = y_1, \ldots, y_L$ represent system responses to inputs.
- A set of causal models: $M = \ m_1, \ldots, m_N$, each represent a (sub) set of the systems' behavior.
  - Each model m_i has a set of inputs x_{ij}, where $i \in K_i \subseteq \{1, \ldots, K\}$.
  - Each model m_i has a set of outputs: y_{il}, where $l \in L_i \subseteq \{1, \ldots, L\}$.

Learning Framework

- In our learning framework, output is a function of the set of all causal models

$$Y(t+1) = F\big(m_1(t), \dots, m_N(t)\big) + \epsilon(t)$$

In comparison, a standard linear model is of the form

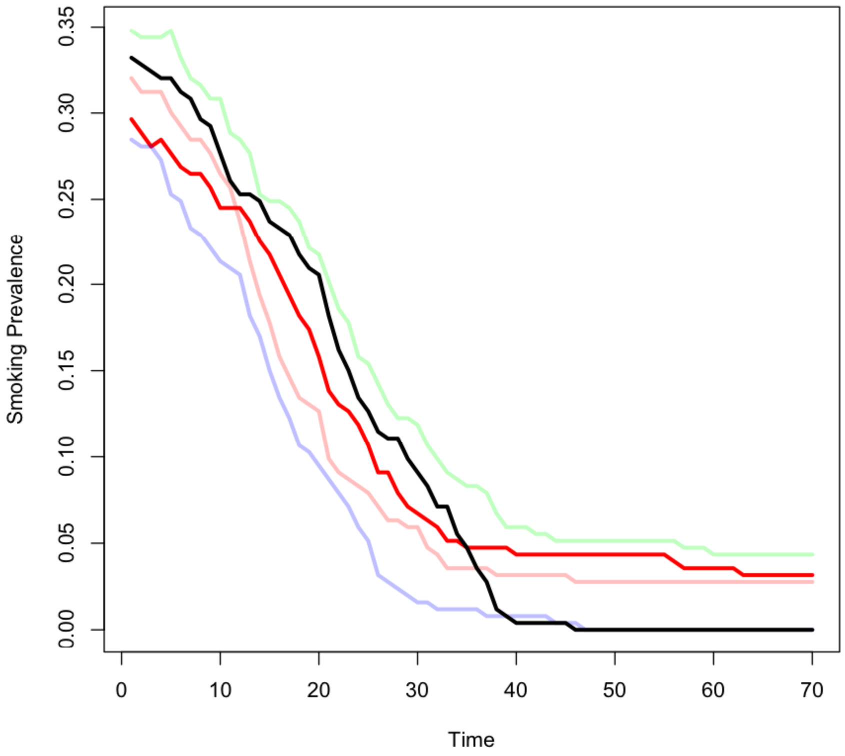
$$Y(t+1) = \sum_i w_i x_i(t) + \varepsilon(t)$$

- In a real time setting a key question is also how to re-calibrate the causal models with newly arriving data. This work lays the foundation for real time analysis.

Experimental Setup

- Create a single instance of a network
- Create a set of families of models using known parameters:
  - Average opinion
  - Number of nodes
  - Information on network structure
- Two randomly generated models are sufficient to represent the test model output when the "learning" function is a simple linear regression

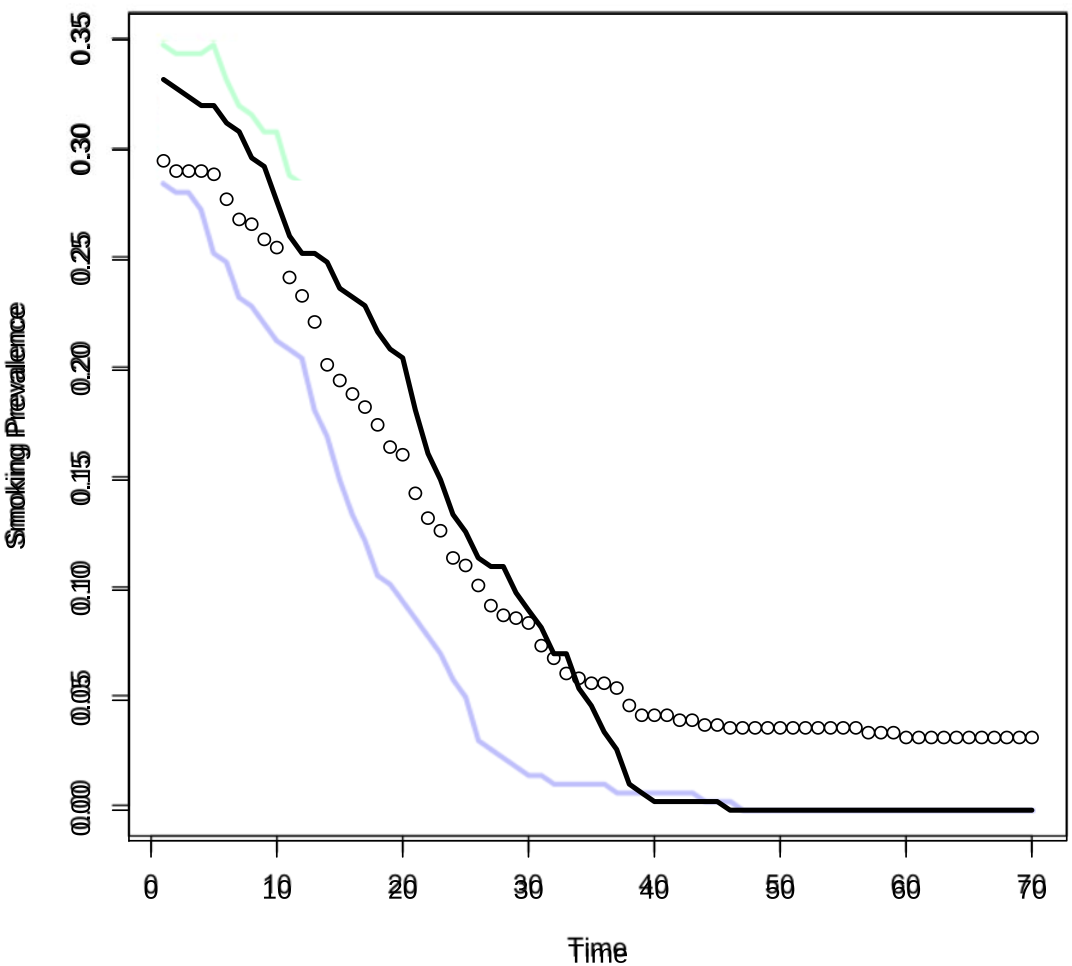
$$Y(t+1) = \sum_i w_i m_i(t) + \varepsilon(t)$$



- Example objective function – red
- Basis models are drawn from "same" distribution

Test Case Results

- Experimental results



"Red" – the "true" function

Points – estimated value

Adjusted R-squared: 99.66

Online Notes

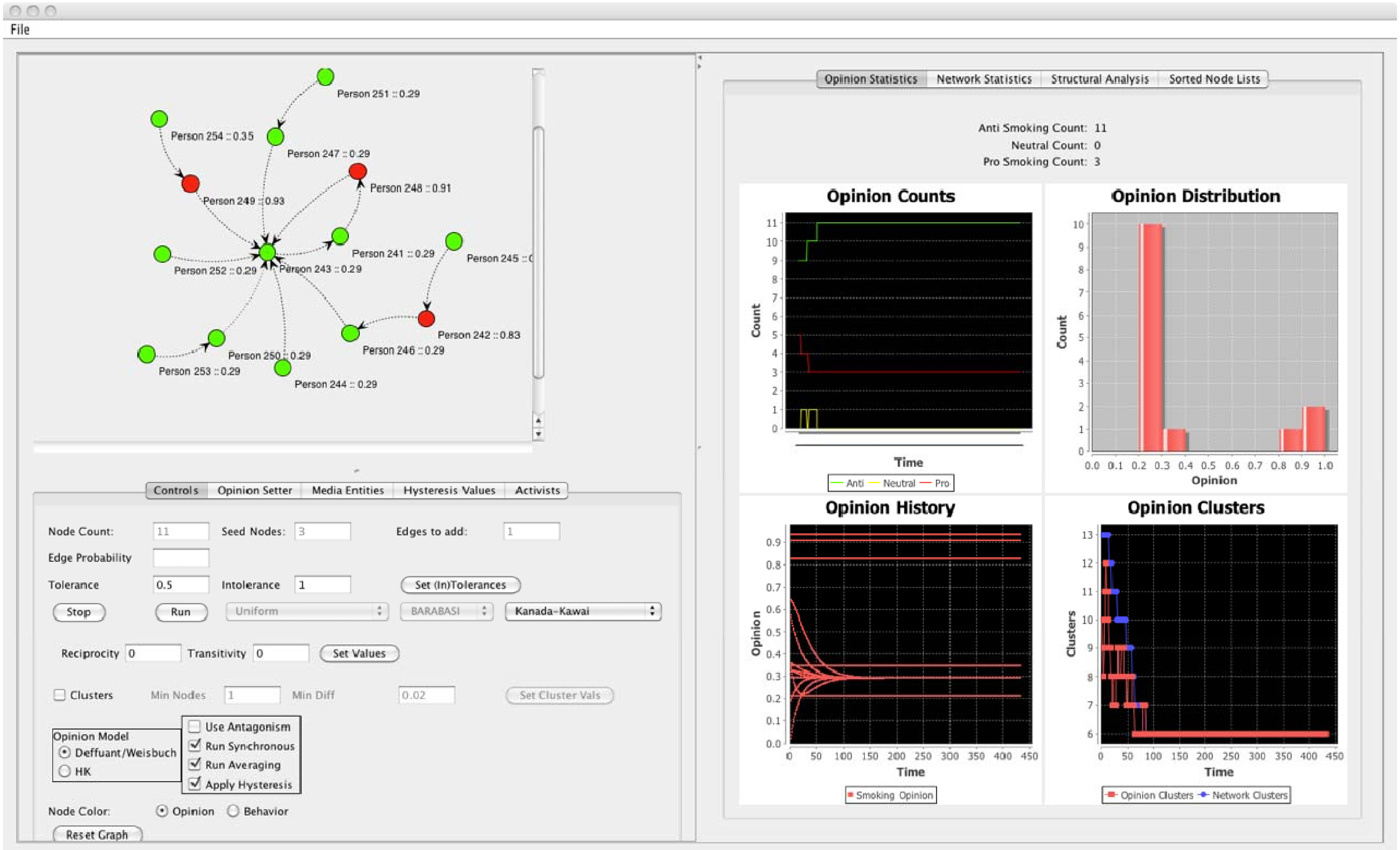
- In an on-line or real-time setting, the key usefulness of causal model learning is in the ability to select from among the “basis” models, and to recalibrate based on newly arrived information
- Future tasks include:
  - Identify the “basis” models that best represent the observed data
  - Identify when the basis model ensemble is inadequate; e.g., regime change.

Conclusions

- Connecting causal models to a learning framework provides an ability to use incomplete causal information for prediction, and the ability to select across many possible causal models for best applicability.
- Preliminary results from a simple opinion dynamics application demonstrate the utility of using randomly generated causal models and the learning framework for prediction.
- Future directions include online estimation and model selection, and regime change identification

Extra Slides

Opinion Dynamics Model Graphical User Interface



Test Case Results Statistics

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
Experiment.1  0.31113    0.03716   8.373 1.46e-11 ***
Experiment.2  0.56198    0.02927  19.197  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.008718 on 58 degrees of freedom
Multiple R-squared: 0.9968, Adjusted R-squared: 0.9966
F-statistic:  8914 on 2 and 58 DF,  p-value: < 2.2e-16
```

Diagnostics