Predictability of User Behavior in Social Media: 
*Bottom-Up v. Top-Down Modeling*

David Darmon and Jared Sylvester

in collaboration with
Bill Rand and Michelle Girvan
Predictability of User Behavior in Social Media

Social Media

Twitter, Facebook, Google+, Instagram, Path

Unprecedented access to

millions of people’s behavior

at second-level resolution.
Predictability of User Behavior in Social Media

Individual as a Computational Unit

Treat individuals as information processing units.

Not a new idea:
Claude Shannon (1948)
Information Theory and Channels

Simon DeDeo (2012)
Markov Models of Wikipedia Activity
Predictability of User Behavior in Social Media
Individual as a Computational Unit
Predictability of User Behavior in Social Media

Individual as a Computational Unit

Input  

Output  

Processing
The Setup
Predictability of User Behavior in Social Media

The Setup

<table>
<thead>
<tr>
<th>Timestamp</th>
<th>Tweet Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013-08-22 12:54:06</td>
<td>Is Your Gmail Social? How to Use Gmail Daily to Build an Engaged Community</td>
</tr>
<tr>
<td>2013-08-22 13:11:22</td>
<td>Facebook's Embedded Posts Now Available to Everyone</td>
</tr>
<tr>
<td>2013-08-22 13:14:06</td>
<td>The Credible Hulk [link]</td>
</tr>
<tr>
<td>2013-08-22 13:29:02</td>
<td>25 Things You Didn’t Know About Ninjas</td>
</tr>
<tr>
<td>2013-08-22 13:48:46</td>
<td>10 Brilliant Facebook Marketing Tactics</td>
</tr>
<tr>
<td>2013-08-22 14:17:11</td>
<td>Google Now Adds Cards for NCAA Football, Concert Tickets, Car Rentals, and More</td>
</tr>
<tr>
<td>2013-08-22 15:18:03</td>
<td>What is the NSA Really Up To? [Comic]</td>
</tr>
<tr>
<td>2013-08-22 15:39:04</td>
<td>6 Things Every Good Business Blog MUST Have</td>
</tr>
</tbody>
</table>

User: DanielZeevi
Predictability of User Behavior in Social Media

The Setup

Bin (in time) Twitter data, giving a discrete time series for each user $v$ at time $t$:

$$X(v, t) = 0 \quad \text{— user } v \text{ doesn’t tweet}$$

$$X(v, t) = 1 \quad \text{— user } v \text{ tweets}$$
Models
Predictability of User Behavior in Social Media

Modeling Framework — A Predictive View

A model that predicts well captures something about the computational capabilities of a user. Necessary but not sufficient.
Predictability of User Behavior in Social Media

Modeling Framework — A Predictive View

Simplifying assumption: a process with self-feedback.

**Observe:** \( X_{i-L}^{i-1} = (X_{i-L}, \ldots, X_{i-2}, X_{i-1}) \).

**Predict:** \( \hat{X}_i = \arg \max_{x \in \{0,1\}} r(x; X_{i-L}^{i-1}) \).
Predictability of User Behavior in Social Media

Modeling Framework — A Predictive View

Our goal: Learn $r$.
Learn the function mapping us from the past to the future.

In essence, a problem in autoregression.

**Observe:** $X_{i-L}^{i-1} = (X_{i-L}, \ldots, X_{i-2}, X_{i-1})$.

**Predict:** $\hat{X}_i = \arg \max_{x \in \{0,1\}} r(x; X_{i-L}^{i-1})$. 
Predictability of User Behavior in Social Media
Modeling Framework — A Predictive View

Two approaches to learning $r$:

- Computational Mechanics
  “Bottom Up”

- Reservoir Computing
  “Top Down”
Models

Computational Mechanics
Assume $\{X_i\}_{i=1}^N$ was generated by a conditionally stationary stochastic process.

Explicitly learn the predictive distribution

$$P(X_i|X_{i-L}^{i-1} = x)$$

by grouping together pasts $x$ that give equivalent predictions.
Predictability of User Behavior in Social Media

Computational Mechanics

The sets of equivalent pasts induce an auxiliary (hidden) process \( \{ S_i \}_{i=1}^{N} \) that is:

Markov

Prescient for prediction
We only need to know that hidden state to perform prediction.
Predictability of User Behavior in Social Media

Computational Mechanics

A —— 1 —— B

B —— 1 —— C

C —— 1 —— A

0 —— 1 —— 0

state
Predictability of User Behavior in Social Media

Computational Mechanics

Call the model learned the

*causal state model (CSM)*

for each user.

Learn this state-space representation of the process using

Causal State Splitting Reconstruction (CSSR).
Models

Echo State Networks
Predictability of User Behavior in Social Media

Feedforward Nets

*Easy learning rules*

Recurrent Nets

*Good with sequences*
Predictability of User Behavior in Social Media

Echo State Networks

Q: How do we learn these weights?
Predictability of User Behavior in Social Media

Echo State Networks

Q: How do we learn these weights?

A: We don’t.
Predictability of User Behavior in Social Media

Echo State Networks

Random, Fixed (sparse)

Trained

Q: How do we learn these weights?
A: We don’t.
Echo State Networks

$$y_t = \sigma(W_{in}x_t + W_y y_{t-1} + W_{fb} z_{t-1})$$

$$z_t = \sigma \left( W_{out} \begin{bmatrix} x_t \\ y_t \end{bmatrix} \right)$$
Data Collection and Processing
Predictability of User Behavior in Social Media

The Dataset

Twitter users embedded in a 15k user follower network.

Statuses of all users collected over 7 weeks.

Select 3k subset of most frequently tweeting users.
Predictability of User Behavior in Social Media

The Dataset — Coarsening

Need to looking $L$ steps back in time.

Dimensionality of predictive space grows like $2^L$.

To deal with this limitation:
coarsen users’ time series.
Predictability of User Behavior in Social Media

Example

```
0 0 0 1 1 1 1 0 0 1 1 0
...
0 1 1...
```

“Does the user tweet during each binning?”
bin size = 600 s = 10 min
Results
Predictability of User Behavior in Social Media

Testing Procedure

• Build model for each user separately
• Training: 45 days
• Testing: 4 days
• Look back 10 steps
• Predict ahead 1 step
• 0-1 Loss
• Compare to “majority vote” baseline
ESN vs. Baseline

Accuracy Rate vs. Tweet Rate

- Red dots: Baseline
- Blue dots: ESN
CSM vs. ESN

The scatter plot shows a strong correlation between CSM Improvement and ESN Improvement, with a regression line indicating a nearly perfect linear relationship. The correlation coefficient is 0.6.
Case Studies
User: DanielZeevi

Base Rate: 0.4506
CSM Rate: 0.9477
ESN Rate: 0.9419
User: LiveFreeRadio

Base Rate: 0.2122
CSM Rate: 0.7035
ESN Rate: 0.7936
Conclusions and Future Directions
Predictability of User Behavior in Social Media

Conclusions

Many users on Twitter are well-modeled as processes with self-feedback.
  Didn’t need social information.

Computational Mechanics and Echo State Networks performed similarly on a large proportion of users.
  Despite very dissimilar modeling paradigms.
Predictability of User Behavior in Social Media

Future Work

Consider:

- Network effects
- Explicitly consider social dynamics

Content from Tweets
- Sentiment, etc.

Longitudinal studies
- Do users change over time?
Thanks!

Questions?