The Effects of Feedback on Human Behavior in Social Media: An Inverse Reinforcement Learning Model

Allen Lavoie
with Sanmay Das

Washington University in St. Louis
How do social interactions change us?
How do social interactions change us?

in social media
USA Archery reports membership doubles in last two years thanks to Hunger Games - talk about the power of books!  (npr.org)

TIL each year Canada Post receives a million letters addressed to "Santa Claus, The North Pole, H0H 0H0". They reply to everyone.  (canadapost.ca)

Bitcoin hits $1000  (businessinsider.com)
The Effects of Feedback on Human Behavior in Social Media: An Inverse Reinforcement Learning Model

Das, Lavoie (Wash. U.)
news cats compsci cats compsci
Replies to comment in subreddit

Behavior changes by reply count

Fraction after / fraction before

Behavior changes by reply count

Similar effects in e.g. Wu et al on YouTube and Digg.

Das, Lavoie (Wash. U.)

The Effects of Feedback on Human Behavior
Similar effects in e.g. Wu et al on Youtube and Digg
Introduction

Modeling behavior changes

Evaluation

Collective behavior
How do humans play games?

<table>
<thead>
<tr>
<th>Row player</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strategy 1</td>
</tr>
<tr>
<td>(3 3)</td>
</tr>
<tr>
<td>Strategy 2</td>
</tr>
<tr>
<td>(1 5)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Column player</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strategy 1</td>
</tr>
<tr>
<td>(2 4)</td>
</tr>
<tr>
<td>Strategy 2</td>
</tr>
<tr>
<td>(6 0)</td>
</tr>
</tbody>
</table>

Simple model which fits human learning (Erev and Roth)

Want behavioral model

Agnostic on concepts of rationality/equilibrium

Uniform initial preferences

\[ q_k(1) = Q \]

Update on reward

\[ q_k(t+1) = (1 - \phi)q_k(t) + (1 - \epsilon)R \]

Strategies not picked:

\[ q_k(t+1) = (1 - \phi)q_k(t) + R \epsilon / (M - 1) \]

Forgetting \( \phi \), exploration \( \epsilon \), number of strategies \( M \)

Linear probabilistic choice

\[ p_k(t) = \frac{q_k(t)}{\sum_k q_k(t)} \]

Das, Lavoie (Wash. U.)
How do humans play games?

Simple model which fits human learning (Erev and Roth)

- Want behavioral model
  - Agnostic on concepts of rationality/equilibrium
How do humans play games?

<table>
<thead>
<tr>
<th>Column player</th>
<th>Row player</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Strategy 1</strong></td>
<td><strong>Strategy 2</strong></td>
</tr>
<tr>
<td>(2 4)</td>
<td>(6 0)</td>
</tr>
<tr>
<td>(3 3)</td>
<td>(1 5)</td>
</tr>
</tbody>
</table>

Simple model which fits human learning (Erev and Roth)

- Want behavioral model
  - Agnostic on concepts of rationality/equilibrium
- Uniform initial preferences $q_k(1) = Q$
How do humans play games?

<table>
<thead>
<tr>
<th></th>
<th>Column player</th>
<th>Row player</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strategy 1</td>
<td>(2 4)</td>
<td>(6 0)</td>
</tr>
<tr>
<td>Strategy 2</td>
<td>(3 3)</td>
<td>(1 5)</td>
</tr>
</tbody>
</table>

Simple model which fits human learning (Erev and Roth)

- Want behavioral model
  - Agnostic on concepts of rationality/equilibrium
- Uniform initial preferences $q_k(1) = Q$
- Update on reward $q_k(t + 1) = (1 - \phi)q_k(t) + (1 - \epsilon)R$
  - Strategies not picked: $q_k(t + 1) = (1 - \phi)q_k(t) + R\epsilon/(M - 1)$
  - Forgetting $\phi$, exploration $\epsilon$, number of strategies $M$
How do humans play games?

Simple model which fits human learning (Erev and Roth)

- Want behavioral model
  - Agnostic on concepts of rationality/equilibrium
- Uniform initial preferences $q_k(1) = Q$
- Update on reward $q_k(t + 1) = (1 - \phi)q_k(t) + (1 - \epsilon)R$
  - Strategies not picked: $q_k(t + 1) = (1 - \phi)q_k(t) + R\epsilon/(M - 1)$
  - Forgetting $\phi$, exploration $\epsilon$, number of strategies $M$
- Linear probabilistic choice $p_k(t) = q_k(t) / \sum_k q_k(t)$
Social media complexities

- Initial preferences
  - Living in St. Louis, active on /r/stlouis
Social media complexities

- Initial preferences
  - Living in St. Louis, active on /r/stlouis

- Infinite strategy space
  - New communities every day
Social media complexities

- Initial preferences
  - Living in St. Louis, active on /r/stlouis
- Infinite strategy space
  - New communities every day
- Unknown rewards
  - Regression using future behavior
Proposed generative model

Concentration Global popularities

\( q^0 \sim \text{Dirichlet}(\alpha_0\beta) \)

\( q \leftarrow q^0 \)  ▶ Initial propensities
Proposed generative model

Concentration Global popularities

\[ q^0 \sim \text{Dirichlet}(\alpha_0, \beta) \]

\[ q \leftarrow q^0 \]

for \( i \in C_u \) do  \hspace{1cm} ▶ For each of this user’s actions (in order)

\[ s_i \sim \text{Categorical}(q / \sum_j q_j) \]

▶ Initial propensities

▶ For each of this user’s actions (in order)

▶ Strategy picking
Proposed generative model

Concentration Global popularities

\[ q^0 \sim \text{Dirichlet}(\alpha_0, \beta) \]
\[ q \leftarrow q^0 \]

for \( i \in C_u \) do

\[ s_i \sim \text{Categorical}(q/\sum_j q_j) \]
\[ q \leftarrow q(1 - \phi) \]
\[ q_{s_i} \leftarrow (1 - \epsilon)R(r_i) + q_{s_i} \]
\[ q \leftarrow q + \epsilon R(r_i)q^0 \]

▷ Initial propensities

▷ For each of this user’s actions (in order)

▷ Strategy picking

▷ Forgetting

▷ Direct reward

▷ Exploration

- \( r_i \): Feature vector
  - We use normalized karma, reply count

- \( R \): Reward function; assumed linear
Inverse reinforcement learning

- Given an agent's (optimal) behavior, what is their utility function?
Inverse reinforcement learning

- Given an agent’s (optimal) behavior, what is their utility function?
- Here, humans not necessarily acting optimally
  - Do know something about the learning process
Inference strategy

- Basic strategy: separate initial propensities from user learning
  - Equivalent to original model
Inference strategy

- Basic strategy: separate initial propensities from user learning
  - Equivalent to original model

- Gibbs sampling
  - Assume every variable is known, sample each sequentially
  - Want to estimate probability distribution
Simulating learning

$q_{\text{init}} \leftarrow 1$
$q \leftarrow 0$

\textbf{for} $i \in C_u$ \textbf{do}

\begin{align*}
q^i, q_{\text{init}}^i &\leftarrow q, q_{\text{init}} \\
q &\leftarrow q(1 - \phi) \\
q_{\text{init}} &\leftarrow q_{\text{init}}(1 - \phi) \\
q_{s_i} &\leftarrow (1 - \epsilon)R(r_i) + q_{s_i} \\
q_{\text{init}} &\leftarrow q_{\text{init}} + \epsilon R(r_i)
\end{align*}

\begin{itemize}
\item For each of this user’s actions (in order)
\item Record current weights
\item Forgetting
\item Direct reward
\item Exploration
\end{itemize}
Introduction

Modeling behavior changes

Evaluation

Collective behavior
Prediction task

- Given past behavior and social interaction, what will I do next?
Prediction task

- Given past behavior and social interaction, what will I do next?
- Inherently probabilistic (from our perspective)
  - Evaluate with proper scoring rule
Models, baselines, and results

- **Reinforcement**: Proposed model
- **UserAll**: User’s empirical distribution of past choices
- **UserKMax**: Past $K$ choices, with $K$ maximizing performance
- **Global**: Collective empirical distribution of choices
- **Initial**: Proposed model with no learning
- **InitKMax**: No learning, over past $K$ choices, using best $K$
- **True**: For synthetic data, the true choice distribution
Models, baselines, and results

- **Reinforcement**: Proposed model
- **UserAll**: User’s empirical distribution of past choices
- **UserKMax**: Past $K$ choices, with $K$ maximizing performance
- **Global**: Collective empirical distribution of choices
- **Initial**: Proposed model with no learning
- **InitKMax**: No learning, over past $K$ choices, using best $K$
- **True**: For synthetic data, the true choice distribution

---

![Real performance by data set size](image1)

![Synthetic performance by data set size](image2)
Generative model (review)

\[ q^0 \sim \text{Dirichlet}(\alpha_0 \beta) \]
\[ q \leftarrow q^0 \]
\[ \text{for } i \in C_u \text{ do} \]
  \[ s_i \sim \text{Categorical}(q/ \sum_j q_j) \]
  \[ q \leftarrow q(1 - \phi) \]
  \[ q_{si} \leftarrow (1 - \epsilon)R(r_i) + q_{si} \]
  \[ q \leftarrow q + \epsilon R(r_i)q^0 \]

▷ Initial propensities

▷ For each of this user’s actions (in order)
  ▷ Strategy picking
    ▷ Forgetting
      ▷ Direct reward
        ▷ Exploration

- \( r_i \): Feature vector
  - We use normalized karma, reply count
- \( R \): Reward function; assumed linear
Introduction
Modeling behavior changes
Evaluation
Collective behavior
How do new social media sites start?
Users vote and reply to comments in a chosen community
Users vote and reply to comments in a chosen community

- K seed users for S rounds, voting up every comment in D
  - 100 regular users
Users vote and reply to comments in a chosen community

K seed users for S rounds, voting up every comment in D
  ▶ 100 regular users

Can D become dominant and self-sustaining?
Seeding outcomes

Community seeding outcomes (averaged)

- Successful
- Late failure
- No traction

Forced community phase transitions
- 100 seed rounds
- 200 seed rounds
- 50 seed rounds
Conclusions

- Simple model predicts user behavior changes in social media
Conclusions

- Simple model predicts user behavior changes in social media
- Basis for studying complex collective dynamics