Participation Maximization Based on Social Influence in Online Discussion Forums

Tao Sun\textsuperscript{1,2}, Wei Chen\textsuperscript{2}, Zhenming Liu\textsuperscript{2,3}, Yajun Wang\textsuperscript{2}, Xiaorui Sun\textsuperscript{2}, Ming Zhang\textsuperscript{1}, Chin-Yew Lin\textsuperscript{2}

\textsuperscript{1}Peking University
\textsuperscript{2}Microsoft Research Asia
\textsuperscript{3}Harvard School of Engineering and Applied Sciences
Outline

- Motivation
- User Model based on social influence
- Participation Maximization
  - Problem formulation
  - Algorithms
- Summary
Motivation

- Increase participation in online forums
  - Users expect a wide audience
  - Forum owners care about the traffic

- Users influence each other
  - Users tend to post after others

- Goal
  - Maximize participants based on social influence
Influence Network in forums

- Given an Influence network \( G = (\mathcal{U}, E, w) \), directed, asymmetric and weighted
  - social ties, \( u \rightarrow v \)
  - the corresponding influence values, \( W_{u,v} \)

- Add a thread user \( \tau \) to incorporate the influence from the original thread
  - \( \tau \rightarrow v \): influence from the thread content
  - One \( \tau \) for a set of threads on a specific topic, e.g., one category
User posting model

- Enter unposted thread?
- Read unvisited post in order
- Triggered?
- Last post?
- Leave
- Write a new post
Allocate B threads to each user, e.g., in his sidebar

Users visit their suggested threads with a higher probability

Users post in these threads with a higher probability

Their posts further influence more users to post

Increase the overall number of participants

How to allocate?

To maximize the total #participants through influence propagation.
Participation Maximization

A special case of Social welfare maximization

Monotonicity & submodularity

Optimize the allocation to maximize social welfare (participants)
## Comparison With Related Problems

<table>
<thead>
<tr>
<th>Recommender Systems</th>
<th>Participation Maximization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Focus on users who are likely to post <strong>NOW</strong></td>
<td>Consider <strong>FUTURE</strong> users who will be influenced to post</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Influence Maximization (viral marketing)</th>
<th>Participation Maximization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximize participants in <strong>ONE</strong> specific threads</td>
<td>Maximize <strong>TOTAL</strong> participants in <strong>ALL</strong> threads</td>
</tr>
</tbody>
</table>
Thread Allocation Algorithms

- **Random allocation** - simple and fast
  - When all the consumers have *the same utility function*, achieve \( (1 - \frac{1}{e}) \) – approximation
  - Works for new threads
    - Not for threads with participants already, different utility functions

- **Approximation algorithm** (RPA)
  - \( (1 - \frac{1}{e}) \) – approximation, Dobzinski and Schapira 2006
  - Assume the computation of utility function is done by an oracle
  - Require > \((mn)^7\) iterations, infeasible for a real forum
Thread Allocation Algorithms

- Our heuristic algorithm:
  Thread Allocation based on Influence (TABI)
  - Symbols
    - $EP_j$ - Existing Participants in $T_j$
    - $I_v$ - the set of $v$'s in-neighbors
    - $O_v$ - the set of $v$'s out-neighbors
    - $\delta_k$ - the original visit probability in $kth$ time slot (discretize continuous time)
    - $\delta^*$ - the boosted probability that $v$ visits threads in $v$'s SideBar
  - $v$ is influenced/activated in a suggested thread $j$
    - $\delta^* \left( 1 - \prod_{u \in EP_j \cap I_v} (1 - w_{u,v}) \right)$
  - $v$ influences others in $(k+1)th$ time slot
    - $\delta_{k+1} \sum_{x \in O_v \setminus EP_j} w_{v,x} \prod_{u \in EP_j \cap I_x} (1 - w_{u,x})$
the additional inf \( \Delta inf_v^j \) that brought by displaying thread \( T_j \) to user \( v \)'s sidebar

(omitted visit pr which is same for all users)

\[
\Delta inf_v^j = \left( 1 - \prod_{u \in EP_j \cap I_v} (1 - w_{u,v}) \right) \left( 1 + \sum_{x \in O_v \setminus EP_j} w_{v,x} \prod_{u \in EP_j \cap I_x} (1 - w_{u,x}) \right)
\]

\( v \) is activated \( v \) itself \( v \)'s out-neighbors who are activated

---

**Algorithm 2 TABI**

1: for each \( v \in U \) do
2:     for each \( j \in T \) do
3:         calculate \( \Delta Inf_v^j \) as Equation 2
4:     Rank threads by \( \Delta Inf_v^j \) in descending order
5: Select top \( B \) threads to display in \( v \)'s sidebar
Thread Allocation Algorithms

- **Personalized recommendation**
  - Topic sensitive early adoption based information flow --- **TEABIF**
  - Song, X., Tseng, B. L., et al. Personalized recommendation driven by information flow. In SIGIR '06.

- **Target at different goals:**
  - individual recommendation vs. overall participation

- **Outcomes are comparable:** #participants
Simulation based on User Model

- Parameter Extraction (TripAdvisor)
  - Implicit Influence Network
    - $G_\tau = (U_\tau, E_\tau, w)$
    - Keep edge $u \rightarrow v$ iff $v$ follows $u$ to post in at least $N$ threads ($N=2$)
    - EM-algorithm to learn edge weight $w_{u,v}$ (Gruhl and Guha 2004)
  - Visit Probabilities
    - Note Random, TEABIF and TABI don’t rely on visit probabilities
    - $\delta_\tau$, drawn from empirical distribution

*Thread rank*: thread $T_j$’s rank in chronological order of all threads at a certain time
Different Allocation Algorithms

Intend to verify that TABI could perform consistently better under different numbers of threads (30, 40 and 50)

NYC
London
Orlando

TABI outperforms all

Parameter setting:
B=5, allocation time slot s=2
\( \delta_r \) to approximate \( \delta_t \)
Boosted visit probability \( \delta^* = 0.8 \).
The value of \( \delta^* \) won’t affect thread allocation of Random, TEABIF and TABI
Simulation shows total participants under RPA is still linear to \( \delta^* \)
Intend to verify that TABI could perform consistently better under different original visit probabilities

- **PowerLaw**
  - NYC: 867.85, 1002.91, 1036.26
  - London: 982.09, 1052.77, 1158.83
  - Orlando: 1499.98, 1783.89, 1819.04

- **Constant(=0.1)**
  - NYC: 867.85, 945.28, 1024.48
  - London: 982.09, 1132.05, 1258.76
  - Orlando: 1499.98, 1789.68, 2050.08

**TABI outperforms all**

**Parameter settings:**
1) **Power Law:** $\delta_t = kt^{-\alpha}$ ($k = 0.3, \alpha = 0.6$), to simulate the decreasing trend with a larger visit probability compared to $\delta_r$
2) **Constant Value:** $\delta_t = 0.1$ for all $t$
Different Allocation Time Slots

Intend to verify that TABI could perform consistently better when allocating in different time slots.

**TABI outperforms all**

Seems that participation increases as $s$.
A larger boost in visit probabilities may provide more participation, but the selection of $s$ should consider other factors such as user experience.

Parameter settings:
Vary allocation time slot $s$ from $s = 2$ to $s = 10$. 
Other Applications

- Advertisements in Facebook
- Posts in Google Buzz
- Comments in YouTube
- Etc.
Propose a personalized allocation mechanism to maximize total participation through influence propagation

- Formulate the problem of participation maximization
  - Monotonicity and submodularity
- Propose a heuristic algorithm--- TABI
  - Efficient, effective and robust
THANK YOU

Twitter: @GabriellaTaoSun