LA-CTR: A Limited Attention Collaborative Topic Regression for Social Media

Jeon-Hyung Kang and Kristina Lerman
USC Information Sciences Institute
Information Overload in Social Media

We are bombarded with so much data that we’re on **information overload**

2,500,000,000,000,000,000 (2.5 quintillion) bytes of data is created everyday! - IBM estimates 2012

27% of total U.S. internet time is spent on social networking sites - Experian 2012

In some sense, **Social Media** provides a solution by allowing users to **subscribe/follow updates from specific users**.

But **Information Overload** will comeback when users have so many friends.

Filtering & Recommendation!

Ruff, J. (2002). Information Overload: Causes, Symptoms and Solutions. Harvard Graduate School of Education’s Learning Innovations Laboratory (LILA)
Solution?: Social Recommendation

User Actions
- **Share** (submit, tweet, status update…) an item
- **Adopt** (re-share, retweet, digg, …) an item shared by a friend

Actions are broadcast to all user’s followers (Information diffusion)
Solution?: Social Recommendation

1. Apply statistical model to observations
   - LDA: Blei, Ng & Jordan (2003)
   - PMF: Salakhutdinov & Mnith (2008)

2. Learn interests (users) & topics (items)

3. Filter and personalize streams of new information

Social Media Observations

Recommendations
Selective Attention

What’s missing?

1. Users have finite attention, which limits their ability to process shared items.
   - Some close or influential friends’ shared items may receive more attention.

2. Users divide their attention non-uniformly over their friends and interests.
   - Users may preferentially pay more attention to each friend depending on topic.

Generic models of information diffusion ignore user’s selective attention

Example: Selective Attention

Bob

- **Foundation work**
  - Amazing leadership from Norway to increase foreign aid budget for 2013. Great to see immunization is a key priority. b-gat.es/QEBloa
- **Microsoft**
  - Microsoft To Host Free App Labs For Windows 8 Developers In 30 Cities Around The World tcm.ch/U2phqX by @federici
- **Sports**
  - Updated game recap from LAC’s 107-100 win over Miami with quotes and notes. READ on nba.com/TkSdGf
  - “Practical Bayesian Optimization of Machine Learning Algorithms” prsm.tc/frXiWQ - recommended via @Prismatic
- **Apple & Google**
  - Blogger’s Android and iOS apps get landscape composing, Google integration, and iPad support tw.hk/IDUMn by @emilproteinski
  - Twitter Adds “Share Tweet Via Email” So You Can Loop In People Not On Twitter tcm.ch/U2rUJo by @joahconstine

John

- **Game Business**
  - Lockout Gaming Takes $1.25M Seed Round To Accelerate Indie Game Revenues tcm.ch/QJSHLd by @steve2007
  - The .Co Domain Launches Membership Program With Free Events, Classes, And Consultation tcm.ch/TPHNFC by @anthonyha
- **Window Dev**
  - Microsoft To Host Free App Labs For Windows 8 Developers In 30 Cities Around The World tcm.ch/U2phqX by @federici
- **Foundation work**
  - Amazing leadership from Norway to increase foreign aid budget for 2013. Great to see immunization is a key priority. b-gat.es/QEBloa
- **Policy**
  - Watch President Obama discuss his plans for the next four years—and his mandate to help the middle class: OFA.BO/VShtw
  - “My favorite moment: Travelling to Florida from California to get out the vote. We did it” — Catherine, CA pic.twitter.com/smWykN9g
- **Daily Life**
  - Policy
Learning what users are interested in!

\{b, c, g\} \{a, c, f\} \{d, e, g\}

user’s interests

attention to friends

\{a, b, d, e, g\}

user’s attention
Limited Attention CTR

LA-CTR: incorporates users’ limited attention into collaborative topic regression model to learn their attentions from observations

For each user $i$
- Generate $u_i \sim N(0, \lambda_u^{-1} I_K)$
- Generate $s_i \sim N(0, \lambda_s^{-1} I_N)$
- For each user $l$
  - Generate $\phi_{il} \sim N(g_{il}^T u_i, \epsilon_{il}^T \lambda_{il}^{-1} I_K)$

For each item $j$
- Generate $\theta_j \sim Dirichlet(\alpha)$
- Generate $\epsilon_j \sim N(0, \lambda_v^{-1} I_K)$ and set $v_j = \epsilon_j + \theta_j$
  - For each word $w_{jm}$
    - Generate topic assignment $z_{jm} \sim Mult(\theta)$
    - Generate word $w_{jm} \sim Mult(\beta_{z_{jm}})$

For each user $i$
- For each attention friend $l$
  - For each adopted item $j$
    - Choose the rating $r_{ijl} \sim N(\phi_{il}^T v_j, \epsilon_{ijl}^{-1})$
Baseline

LDA & PMF

CTR

[Salakhutdinov et al. 08]

[Ma et al. 08]

CTR-smf

[Ma et al. 08]

LDA & PMF & SMF

LDA

PMF

SMF

LDA

PMF

SMF
## Experiment Results: Dataset

<table>
<thead>
<tr>
<th></th>
<th>2009 Dataset</th>
<th>2010 Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Time</strong></td>
<td>Jun 2009 (1 month)</td>
<td>Jul – Dec 2010 (6 months period)</td>
</tr>
<tr>
<td><strong># of users</strong></td>
<td>70K active users</td>
<td>12K users</td>
</tr>
<tr>
<td><strong># of links</strong></td>
<td>1.7M</td>
<td>1.3M</td>
</tr>
<tr>
<td><strong># of stories</strong></td>
<td>3.5K (promoted to Front page)</td>
<td>48K</td>
</tr>
<tr>
<td><strong># of votes</strong></td>
<td>2.1M</td>
<td>1.9M</td>
</tr>
<tr>
<td><strong>Assigned topics</strong></td>
<td>8 topics</td>
<td>10 topics</td>
</tr>
</tbody>
</table>

We examine only the votes that the story accrued before promotion to the front page.
Limited Attention CTR

Evaluation: Predict which items will be shared on a social media site

Diagram showing the topic model and relationships between user’s interests, attention, and number of predicted items vs. fraction correct items.
Example: Learned user’s attention,…

User’s attention (top 3):
{b, c, g}  {a, c, f}  {d, e, g}

Attention to friends:

User’s interests (top 5):
{a, b, d, e, g}

<table>
<thead>
<tr>
<th>IDX</th>
<th>Bag of words</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>pic, christmas, guy, holiday, girl, hot, cool, tree, season, mark, geek, santa, mario, gallery, hottest, super, gadget</td>
</tr>
<tr>
<td>b</td>
<td>found, city, dead, body, fall, lost, walk, teen, zombie, beach, pound, fish, feet, shark, swim, bone, pool</td>
</tr>
<tr>
<td>d</td>
<td>team, player, point, run, season, pick, center, football, score, play, field, start, nfl, nba, fan, fantasy, league, injury, talent, coach, yard</td>
</tr>
<tr>
<td>e</td>
<td>guide, culture, lady, collection, style, tradition, beer, suit, fashion, survive,</td>
</tr>
<tr>
<td>g</td>
<td>women, kid, sex, men, children, young, parent, baby, child, toy, mother, adult, relationship, mom</td>
</tr>
</tbody>
</table>
Related Work

• Traditional topic models have been extended to networks
  – hyperlinks between documents [Nallapati and Cohen 08]
  – varying vocabularies and style of different authors [Rosen-Zvi et al 04]
  – Relational topic models for document networks [Chang and Blei 09]

• Collaborative filtering methods for item recommendations
  – Probabilistic Matrix Factorization [Salakhutdinov & Mnith 08]
  – recommend new items that were liked by similar users [Ma et al. 08, Chua et al 12]
  – improve explanatory power of recommendations by extending LDA & SN matrix factorization [Wang and Blei 11, Prushotham et al 12]

• Limited attention in social media
  – [Weng et al. 12, Hodas & Lerman 12]
Conclusions and Future Work

• Introduced LA-CTR, a novel collaborative topic regression model that takes into account social media users’ limited attention.
• Showed empirically that learning users’ limited attention help to understand user’s behavior better on social media.
• Demonstrates the importance of modeling psychological factors in social media analysis.

[QR code for LA-CTR paper]
[QR code linking to LinkedIn profile]