Who Are Experts Specializing in Landscape Photography?
Analyzing Topic-specific Authority on Content Sharing Services

Bin Bi
University of California, Los Angeles

Ben Kao
The University of Hong Kong

Chang Wan
The University of Hong Kong

Junghoo Cho
University of California, Los Angeles
Content Sharing Service

Resource
• Video
• Photo
• Travel blog
• ......
Explosion of User-generated Content

sheer amount of UGC

Blessing

- Investigating topics of interest
- Checking facts
- Getting advice about problems

Curse

- Confusion
- Sub-optimum decisions
- Dissatisfaction

Solution: Discover a set of authorities who create high-quality resources
Prior work on authority identification

- Primarily in the context of social network and network structure analysis, e.g., PageRank
- Only global authorities are identified

**Topic-specific** authority analysis

Users have different topical needs

- Who’s a master of sunset photography?
- Who’s expert in portrait photography?
Topic-specific Authority Analysis

- Prior work on authority identification
  - Primarily in the context of social network and network structure analysis, e.g., PageRank
  - Only global authorities are identified

- **Topic-specific** authority analysis

No one is authoritative on every topic
Roadmap

- Motivation
- Topic-specific Authority Analysis
- Experimental Results
- Conclusion
LDA-based Naïve Solution

Adapt *Latent Dirichlet Allocation* to data in sharing log

- User $\rightarrow$ Document
- Tag $\rightarrow$ Word

Two implications:

- A user is interested in topic $T$ $\iff$ She frequently posts photos with tags specific to $T$

- More frequently a user uses tags covering $T$ $\iff$ More authoritative she should be on $T$
Favorite Click

- A supplementary source about content quality is needed
- Favorite log provides valuable signal

**Challenge**: Users don’t specify topical causes behind favorite clicks

<table>
<thead>
<tr>
<th>User ID</th>
<th>Favorited Photo</th>
</tr>
</thead>
<tbody>
<tr>
<td>82310</td>
<td>![Image]</td>
</tr>
<tr>
<td>185963</td>
<td>![Image]</td>
</tr>
<tr>
<td>28737</td>
<td>![Image]</td>
</tr>
<tr>
<td>49856</td>
<td>![Image]</td>
</tr>
<tr>
<td>93274</td>
<td>![Image]</td>
</tr>
<tr>
<td>...</td>
<td>![Image]</td>
</tr>
<tr>
<td>...</td>
<td>![Image]</td>
</tr>
</tbody>
</table>
Topic-specific Authority Analysis (TAA) Model

- Jointly model topical interest and topical authority:

  **Sharing Log**

<table>
<thead>
<tr>
<th>User ID</th>
<th>Tag</th>
<th>Photo</th>
</tr>
</thead>
<tbody>
<tr>
<td>14529</td>
<td>king's gate castle</td>
<td>![image]</td>
</tr>
<tr>
<td>14839</td>
<td>board stairs</td>
<td>![image]</td>
</tr>
<tr>
<td>14694</td>
<td>beach</td>
<td>![image]</td>
</tr>
<tr>
<td>319526</td>
<td>vanrhum forest</td>
<td>![image]</td>
</tr>
<tr>
<td>319526</td>
<td>sunrise</td>
<td>![image]</td>
</tr>
</tbody>
</table>

  **Favorite Log**

<table>
<thead>
<tr>
<th>User ID</th>
<th>Favorited Photo</th>
</tr>
</thead>
<tbody>
<tr>
<td>82210</td>
<td>![image]</td>
</tr>
<tr>
<td>189543</td>
<td>![image]</td>
</tr>
<tr>
<td>26737</td>
<td>![image]</td>
</tr>
<tr>
<td>49956</td>
<td>![image]</td>
</tr>
<tr>
<td>93274</td>
<td>![image]</td>
</tr>
</tbody>
</table>

- Diagram illustrating the relationship between sharing and favoriting.
Intuition for Characterizing Authoritativeness

- **Users’ authority**
  - Different from each other

- **Each user’s authority**
  - Specific to individual topics

- **Introduce** $\eta_u$ **to characterize topical authority**
  - K-dimensional random vector over topics
  - Specific to individual user $u$
  - $\eta_u \sim \text{MVN}(\mu, \Sigma)$

User $u$’s authority: $\eta_u$

<table>
<thead>
<tr>
<th>$\eta_{u1}$</th>
<th>$\eta_{u2}$</th>
<th>$\eta_{u3}$</th>
<th>$\eta_{u4}$</th>
<th>......</th>
<th>$\eta_{uK}$</th>
</tr>
</thead>
</table>

K-dimensional vector
Intuition for Characterizing Favorite Clicks

- Introduce $f_{ur}$ to represent favorite feedback
  - Binary random variable
  - Specific to user $u$ and resource $r$

$$f_{ur} = \begin{cases} 1 & \text{if } u \text{ favorited } r \\ 0 & \text{otherwise} \end{cases}$$

- $u$ favorites $r$, if topical authority of $r$’s owner exhibited by $r$ matches $u$’s topical interest
Identify hidden topical causes behind favorite clicks

- Designed a model for topic discovery

With the topics, we specify the likelihood:

\[
p(f_{ur} = 1|\theta_u', \hat{z}_u, \hat{z}_{ur}) = \frac{1}{1 + e^{-\eta_{u'}^T(\hat{z}_u \circ \hat{z}_{ur})}}
\]

\[
p(f_{ur} = 0|\theta_u', \hat{z}_u, \hat{z}_{ur}) = 1 - \frac{1}{1 + e^{-\eta_{u'}^T(\hat{z}_u \circ \hat{z}_{ur})}}
\]
Intuition behind Topic Discovery for Favorite Clicks

Logistic likelihood function:

\[
p(f_{ur} = 1 | \eta_u', \hat{z}_u, \hat{z}_{u'r}) = \frac{1}{1 + e^{-\eta_u^T (\hat{z}_u \odot \hat{z}_{u'r})}}
\]

- \( f_{ur} = 1 \) indicates that poster \( u' \) should be expert in the topics prominent in both \( u' \)’s interest and resource \( r \), so the weights of these topics have to be boosted.

- \( \eta_u' \)'s topical authoritativeness captures similarity between topic distributions for resource \( r \) and \( u' \)'s interest.
For user $u$:

- Pick a topic distribution $\theta_u$
- Pick an authority vector $\eta_u$ from $\text{MVN}(\mu, \Sigma)$
- To generate the $n^{th}$ tag:
  - Pick a topic $z$ from topic distribution $\theta_u$
  - Pick a tag $t$ from word distribution $\phi_z$

To generate favorites:

- Pick a favorite response $f_{ur}$ from $\text{Bernoulli}\left(\frac{1}{1 + e^{-\eta_u^T(s_u \circ \tilde{z}_u, t)}}\right)$
Quantities of Interest

- $\eta_u$ quantifies user $u$’s unique authoritativeness over topics
- $\theta_u$ characterizes user $u$’s topical interest
- $\varphi_t$ indicates probabilities of tag $t$ belonging to individual topics
Preference Modeling

- Modeling of the preferences of favorites

\[
p(D|\Theta) = \prod_{(u,r_i,r_j) \in D} p(r_i > u r_j | \eta_{u'}, \hat{z}_u, \hat{z}_{u' r_i}, \hat{z}_{u' r_j})
\]

\[
= \prod_{(u,r_i,r_j) \in D} \frac{1}{1 + e^{-\eta_{u'} (\hat{z}_u \circ \hat{z}_{u' r_i} - \hat{z}_u \circ \hat{z}_{u' r_j})}}
\]
Inference for TAA (cont’d)

- Bayesian inference for a model with logistic likelihood has long been recognized as a hard problem

- We extend recent work [Polson et al. 2013] for inference of TAA
  - Introduce Polya-Gamma variables to posterior distribution
Gibbs Sampler for TAA

Conditionals for Gibbs sampling:

\[ p(\eta_x | \bullet) \propto p(\eta_x) \prod_{r_i \in R(x) \land r_j \in R(x)} e^{\frac{\eta_i^T z_{ur_{ij}} - \delta_{ur_{ij}} (\eta_i^T z_{ur_{ij}})^2}{2}} \]

\[ p(z_{un} = k | \bullet) \propto \frac{(c_{ku}^{-(un)} + \alpha_k)(g_{kt}^{-(un)} + \beta_{tun})}{\sum_{t=1}^{V} g_{kt}^{-(un)} + \sum_{t=1}^{V} \beta_t} \times \prod_{(u, r_i, r_j) \in D} p(r_i \succ u, r_j | \eta_{u'}, z_{-(un)}, z_{un} = k) \]

\[ p(\delta_{ur_{ij}} | \bullet) \propto e^{-\frac{\delta_{ur_{ij}} (\eta_i^T z_{ur_{ij}})^2}{2}} p(\delta_{ur_{ij}} | 1, 0) = PG(1, \eta_i^T z_{ur_{ij}}) \]
Roadmap

- Motivation
- Topic-specific Authority Analysis
- Experimental Results
- Conclusion
Experiments

Data collections

<table>
<thead>
<tr>
<th>Data</th>
<th>#users</th>
<th>#photos</th>
<th>#tag asgmts</th>
<th># fav. clicks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flickr</td>
<td>21,054</td>
<td>204,335</td>
<td>3,014,813</td>
<td>1,562,805</td>
</tr>
<tr>
<td>500px</td>
<td>33,581</td>
<td>318,906</td>
<td>3,520,179</td>
<td>1,837,049</td>
</tr>
</tbody>
</table>

Metrics for effectiveness

- (Mean Reciprocal Rank) MRR

\[
MRR = \frac{1}{|Q|} \sum_{q \in Q} \frac{1}{\text{rank}_q}
\]

- Spearman’s correlation coefficient

\[
\rho = \frac{1}{|Q|} \sum_{q \in Q} \rho_q
\]
Predictive Power

**Perplexity metric**

\[
\text{perplexity}(F_{\text{test}}) = \exp \left\{ -\frac{\sum_{f \in F_{\text{test}}} \log p(f)}{|F_{\text{test}}|} \right\}
\]
Case visualization

**Query topic**: waterscape

<table>
<thead>
<tr>
<th>User ID</th>
<th>Rank</th>
<th>Photos</th>
</tr>
</thead>
<tbody>
<tr>
<td>87620688</td>
<td>Rank 1</td>
<td><img src="image1" alt="Photos" /></td>
</tr>
<tr>
<td>25355186</td>
<td>Rank 100</td>
<td><img src="image2" alt="Photos" /></td>
</tr>
<tr>
<td>50701553</td>
<td>Rank 1000</td>
<td><img src="image3" alt="Photos" /></td>
</tr>
</tbody>
</table>

**Query topic**: winter snow landscape

<table>
<thead>
<tr>
<th>User ID</th>
<th>Rank</th>
<th>Photos</th>
</tr>
</thead>
<tbody>
<tr>
<td>29762217</td>
<td>Rank 1</td>
<td><img src="image4" alt="Photos" /></td>
</tr>
<tr>
<td>25355186</td>
<td>Rank 100</td>
<td><img src="image5" alt="Photos" /></td>
</tr>
<tr>
<td>11052010</td>
<td>Rank 1000</td>
<td><img src="image6" alt="Photos" /></td>
</tr>
</tbody>
</table>
Roadmap

- Motivation
- Topic-specific Authority Analysis
- Experimental Results
- Conclusion
Conclusion

- Propose a novel TAA model for topic-specific authority analysis on content sharing services
  - Leverage both *sharing log* and *favorite log*

- Propose a method to learn from preferences of favorites
  - Embed a new logistic likelihood

- Extend Gibbs sampling by data augmentation for inference
Thank you!