Gender and Interest Targeting for Sponsored Post Advertising at Tumblr

Mihajlo Grbovic, Vladan Radosavljevic, Nemanja Djuric, Narayan Bhamidipati, Ananth Nagarajan

Yahoo Labs
Ad Targeting Team
Overview

- **Tumblr basics**
  - Overview of the network
  - Advertising on Tumblr

- **Tumblr data**
  - Data sources
  - Tumblr user profiles

- **Tumblr gender prediction**
  - Approach
  - Results

- **Tumblr interest prediction**
  - Approach
  - Results
Tumblr basics

- Tumblr statistics
  - 249 million blogs
  - 117 billion posts
  - 90 million daily posts
  - 13 languages

- Source
  - http://www.tumblr.com/about
Tumblr basics – Blog

1 user has 1 primary blog (user=blog)

- And potentially more than 1 secondary blog
Blogs have informative descriptions:

- Tristan, i’m 15, Canada... Snowboarding – Travel - Football – FTB
- Hello, I’m Tess, I post a lot of stuff and Spot Conlon is my bae. Musicals are rad and Shawn Hunter is forever golden
- Alyssa | 18 | California. I like bands, books, shows, and random things. And geese. One bit me in the crotch once. Good times, good times
- I’m Carla // 19yrs old // Texas y’all listen, I just like to blog about anime and cute animals and and video games.
- My name’s Kierstin. I love basketball
As blog owner you can create posts
- You can also reblog/follow other people’s posts

Post types:
- text: 14.13%
- photo: 78.11%
- quote: 2.27%
- link: 0.46%
- chat: 0.85%
- audio: 2.01%
- video: 1.35%
As a blog owner you view posts of blogs you follow
Tumblr basics – Post examples

regular post

Scribble pen copies and draws in any color.
The Scribble features a scanner at one end, which can be pressed against any surface to capture its color, which is stored as a standard six digit hex code. The pens five refillable cartridges allow it to print in the scanned color, or any of its possible 16 million colors.
A Kickstarter campaign is planned to launch the pen ‘in the near future’.

#gadgets #tech #technology #kickstarter #pen #scribble

108 notes

like

reblog

sponsored post

indicates advertisement

indicates advertisement

#StreetArtbyPerrier, net art by @strancilevitori
See his other designs or learn how to submit your own net art here
See more artists who’ve created GIFs for Perrier’s Street Art Limited Edition here

@strancilevitori #Perrier #flashbreeden #socialaperrier #streetart #gif #uk
5,380 notes

like
Advertising on Tumblr

8 of 10 most valuable global brands are on Tumblr†

3 Billion + paid ad impressions

18 minutes per visit* Visitors spend more time on Tumblr per visit than any of the Top 100 sites

400M hours per month Every month, more time is spent on Tumblr than watching the top 10 primetime TV shows combined

*Tumblr internal stat and comScore ranking †Interbrand Best Global Brands 2012
Advertising on Tumblr

How to enhance it?

Targeting – Reach only users that are highly interested in the product or product category

1. Gender targeting
   - Most common form of ad targeting
   - Shown to work better than targeting random users

2. Interest targeting
   - More involved – find users with interest in specific category, e.g. travel, fashion, sports, etc.
   - Shown to work better than pure gender targeting
Tumblr data sources

- Firehose (user actions + post details)
  - User actions – post, reblog, like, unlike
  - Post details:
    - photo posts: caption, tags
    - text posts: title, tags
    - audio posts: artist, tags
  - Blog details – title, description
- Available online: gnip.com/sources/tumblr
Follower graph
- 96.9M nodes (users)
- 5.1B edges (follows)
- 18.2M blogs follow each other
- Average user follows 58.9 blogs
Creating user profiles

User profile created from counts of:

1. Declared features
   - Text from Blog Title
   - Text from Blog Description

2. Content features
   - Tags from Blog Posts
   - Text from Blog Post content
   - Artist names from audio posts

3. User actions
   - Like
   - Follow
   - Reblog

User vector:

```
vector = [0, 1, 7, 0, 3]
```
Main goal
- Assign gender to Tumblr users
- Based on the results serve targeted ads

Steps
1. Use **golden set** (of known gender) + **user profiles** to train a logistic regression model
2. Score all users
3. Apply threshold to keep only most certain predictions
Tumblr gender prediction

Golden set:

- Based on user’s declared first names
  - Extract first names from blog descriptions

<table>
<thead>
<tr>
<th>regex</th>
<th>count</th>
<th>regex</th>
<th>count</th>
</tr>
</thead>
<tbody>
<tr>
<td>my name is*</td>
<td>783,564</td>
<td>mi chiamo*</td>
<td>9,181</td>
</tr>
<tr>
<td>my name’s*</td>
<td>291,811</td>
<td>mein name ist*</td>
<td>1,025</td>
</tr>
<tr>
<td>me llamo*</td>
<td>47,663</td>
<td>meu nome e*</td>
<td>512</td>
</tr>
<tr>
<td>the name’s*</td>
<td>38,065</td>
<td>mon nom est*</td>
<td>215</td>
</tr>
<tr>
<td>mi nombre es*</td>
<td>9,751</td>
<td>mio nome e*</td>
<td>185</td>
</tr>
</tbody>
</table>

- Use US census data (years 1880 to 2013) to get probability of gender given a name

<table>
<thead>
<tr>
<th></th>
<th>male</th>
<th>female</th>
</tr>
</thead>
<tbody>
<tr>
<td>golden set size</td>
<td>395K</td>
<td>564K</td>
</tr>
</tbody>
</table>
Tumblr gender prediction

Results:

- On hold-out set:
<table>
<thead>
<tr>
<th>Gender</th>
<th>Precision</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>female</td>
<td>0.806</td>
<td>0.838</td>
</tr>
<tr>
<td>male</td>
<td>0.794</td>
<td>0.689</td>
</tr>
</tbody>
</table>

- Editorial evaluation of 1,000 random blogs:
<table>
<thead>
<tr>
<th>Gender</th>
<th>Correct</th>
<th>Wrong</th>
<th>Don’t know</th>
</tr>
</thead>
<tbody>
<tr>
<td>female</td>
<td>429</td>
<td>4</td>
<td>298</td>
</tr>
<tr>
<td>male</td>
<td>144</td>
<td>5</td>
<td>127</td>
</tr>
</tbody>
</table>

- Coverage: The classified users cover >95% actions (posts, reblogs, likes, etc.)
Main goal
- Assign interest categories to Tumblr users
  - For example: user x is interested in fashion
  - Based on the results serve targeted ads
  - Interests come from IAB Taxonomy

Main challenge
- Labeled examples are not available
## Tumblr interest categories

**Level 1**

- Arts and Entertainment
- Style and Fashion
- Pets
- Shopping
- Food and Drink
- Home and Garden
- Health and Fitness
- Beauty and Personal Care
- Education
- Society
- Sports
- Technology and Computing
- Travel
- Automotive

**Level 2**

- Arts and Entertainment/Movies
- Arts and Entertainment/Television
- Style and Fashion/Clothing
- Hobbies and Interests/Photography
- Food and Drink/Dining Out
- Family and Parenting
- Food and Drink/Dining Out - Fast Food
- Education/K to 12 Education
- Beauty and Personal Care/Face and Body Care
- Arts and Entertainment/Music
- Arts and Entertainment/Books and Literature
- Beauty and Personal Care/Hair Care
- Style and Fashion/Footware
- Arts and Entertainment/Movies
Tumblr interest prediction

Approach (more details in paper):

1. Categorize keywords from post content (post tags, post text) and blog titles and descriptions

2. Predict user interest categories based on the categorized tags and text in posts, blog titles and descriptions they use

\[ u_{i, \text{cat}=k}^t = \sum_{\text{feat} \in \mathcal{A}_i} \alpha^{(t-t_{\text{feat}})} w_{\text{feat}} I(\text{feat is of class } k). \]

3. Leverage follower graph and like actions to categorize users who do not create much content
Tumblr interest prediction

Approach:
- Categorize keywords from post content (post tags, post text)
- Use 8,400 manually categorized tags as seeds

Tumblr interest prediction

Approach:

1. Categorize keywords from post content (semi-supervised word2vec)
We inferred interests for majority of Tumblr users

Examples of resulting user interest assignments:

<table>
<thead>
<tr>
<th>User</th>
<th>Category</th>
<th>Categorized Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>user 3</td>
<td>Food and Drinks</td>
<td>txt food:16 photo food:11 txt meals:6 tag dessert:4 follows user31:1 follows user43:1 tag soup:1 tag brunch:1 tag fruit:2 tag chicken:3 tag smoothie:1 tag cake:2 tag breakfast:2 tag ginger:2 tag salad:5 tag avocado:1</td>
</tr>
</tbody>
</table>
A/B tests

- With 8 advertisers we ran consecutive untargeted and targeted campaigns
- On average **20% lift in engagement** (i.e., likes, reblogs, follows)

<table>
<thead>
<tr>
<th>Campaign</th>
<th>Control</th>
<th>Targeted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Home &amp; Garden</td>
<td>-</td>
<td>+9.71%</td>
</tr>
<tr>
<td>Style &amp; Fashion</td>
<td>-</td>
<td>+42.53%</td>
</tr>
<tr>
<td>Sports/Outdoor</td>
<td>-</td>
<td>+19.86%</td>
</tr>
<tr>
<td>Arts &amp; Enter./Television</td>
<td>-</td>
<td>+24.37%</td>
</tr>
<tr>
<td>Arts &amp; Enter./Video Games</td>
<td>-</td>
<td>+19.02%</td>
</tr>
<tr>
<td>Pets/Dogs</td>
<td>-</td>
<td>+27.21%</td>
</tr>
<tr>
<td>Arts &amp; Enter. (1)</td>
<td>-</td>
<td>+9.08%</td>
</tr>
<tr>
<td>Arts &amp; Enter. (2)</td>
<td>-</td>
<td>+6.54%</td>
</tr>
</tbody>
</table>
Evaluation

- Sanity check – Tested on Mihajlo’s blog

  - Gender Prediction

    | Score     | #features | inferred gender |
    |-----------|-----------|-----------------|
    | 1.330301  | 236       | male            |

  - Interest Prediction

    | Category            | #features | why                                           |
    |---------------------|-----------|-----------------------------------------------|
    | Sports              | 111       | He follows a lot of soccer related blogs      |
    | Arts and Entert./TV | 107       | He follows and reblogs Game of Thrones blogs |
    | Photography         | 95        | In his description he mentions to like photography |
    | Science             | 29        | He reblogs Yahoo Labs blogs and has it in description |
    | Advertising/Marketing | 7        | He follows advertising related blogs         |
Product Impact

- Gender- and interest-targeted Tumblr campaigns can now be booked on Yahoo Gemini ad platform
- Since the launch, majority of Tumblr campaigns now use our targeting
Thank you!

Question and/or suggestions?