

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.

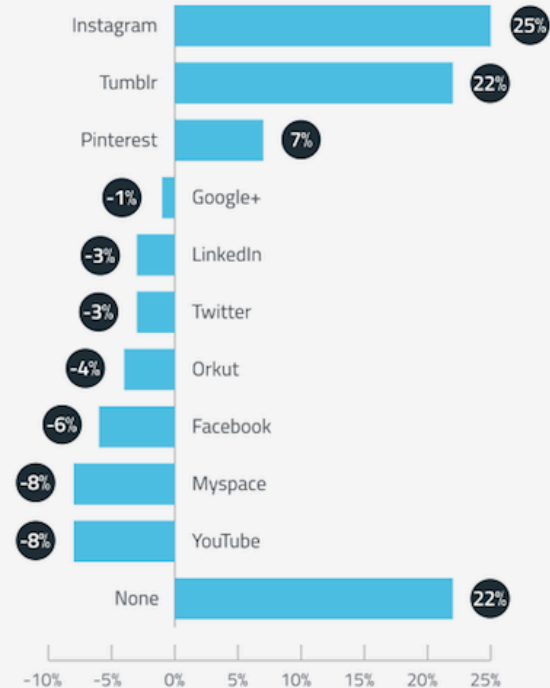
■ Tumblr statistics

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

■ Source

- <http://www.tumblr.com/about>

TOP 10 SOCIAL PLATFORMS: CHANGES IN ACTIVE USAGE IN LAST SIX MONTHS

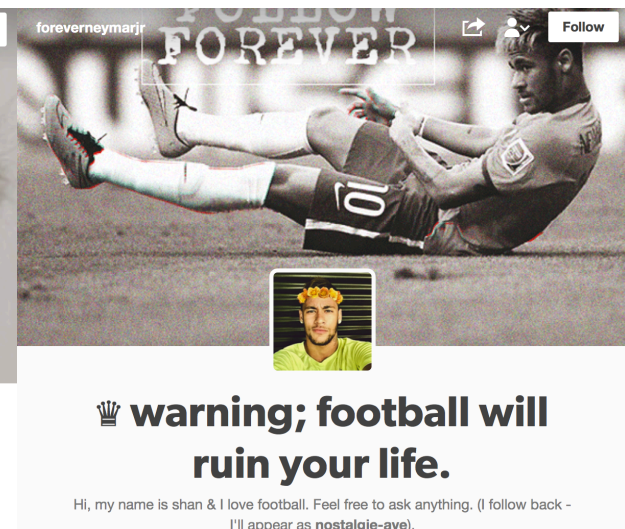
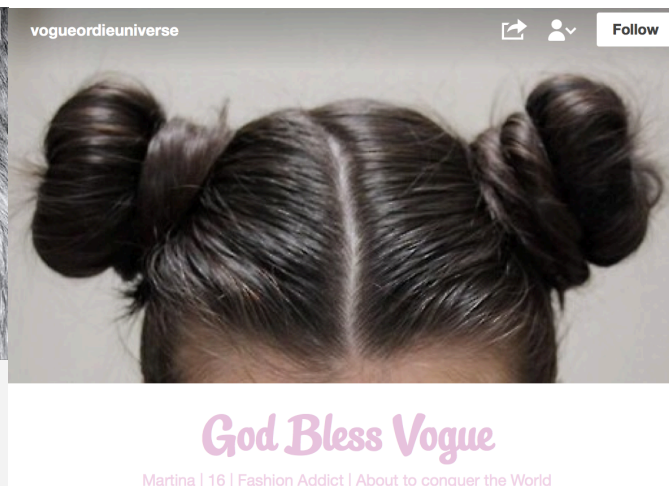
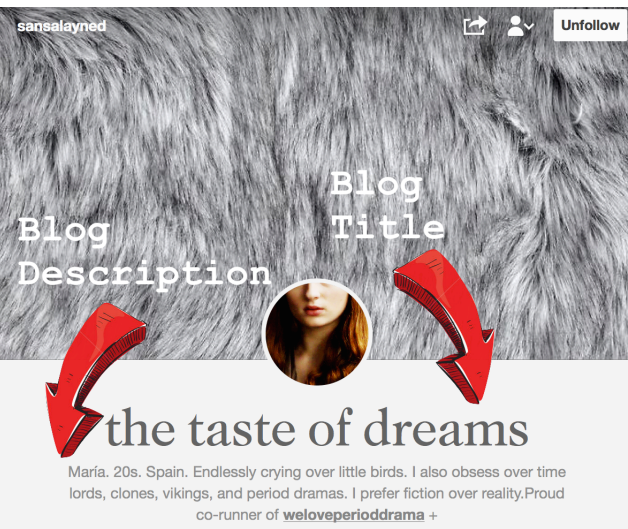


Question: Which of the following services have you used or contributed to in the past month using any type of device? e.g. PC/Laptop, Mobile phone, Tablet, etc. // Source: GlobalWebIndex Q3 2013 - Q1 2014 // Base: Internet users aged 16-64, exc. China

Tumblr basics – Blog

- 1 user has 1 primary blog (user=blog)

- And potentially more than 1 secondary blog



Tumblr basics – Blog description

▣ 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

Tumblr basics – User actions

■ As blog owner you can create posts

■ You can also reblog/follow other people's posts

■ Post types:

Aa

text

14.13%



photo

78.11%



quote

2.27%



link

0.46%



chat

0.85%



audio

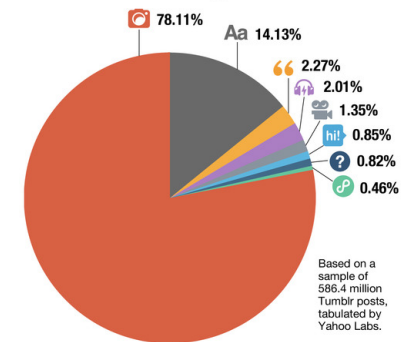
2.01%



video

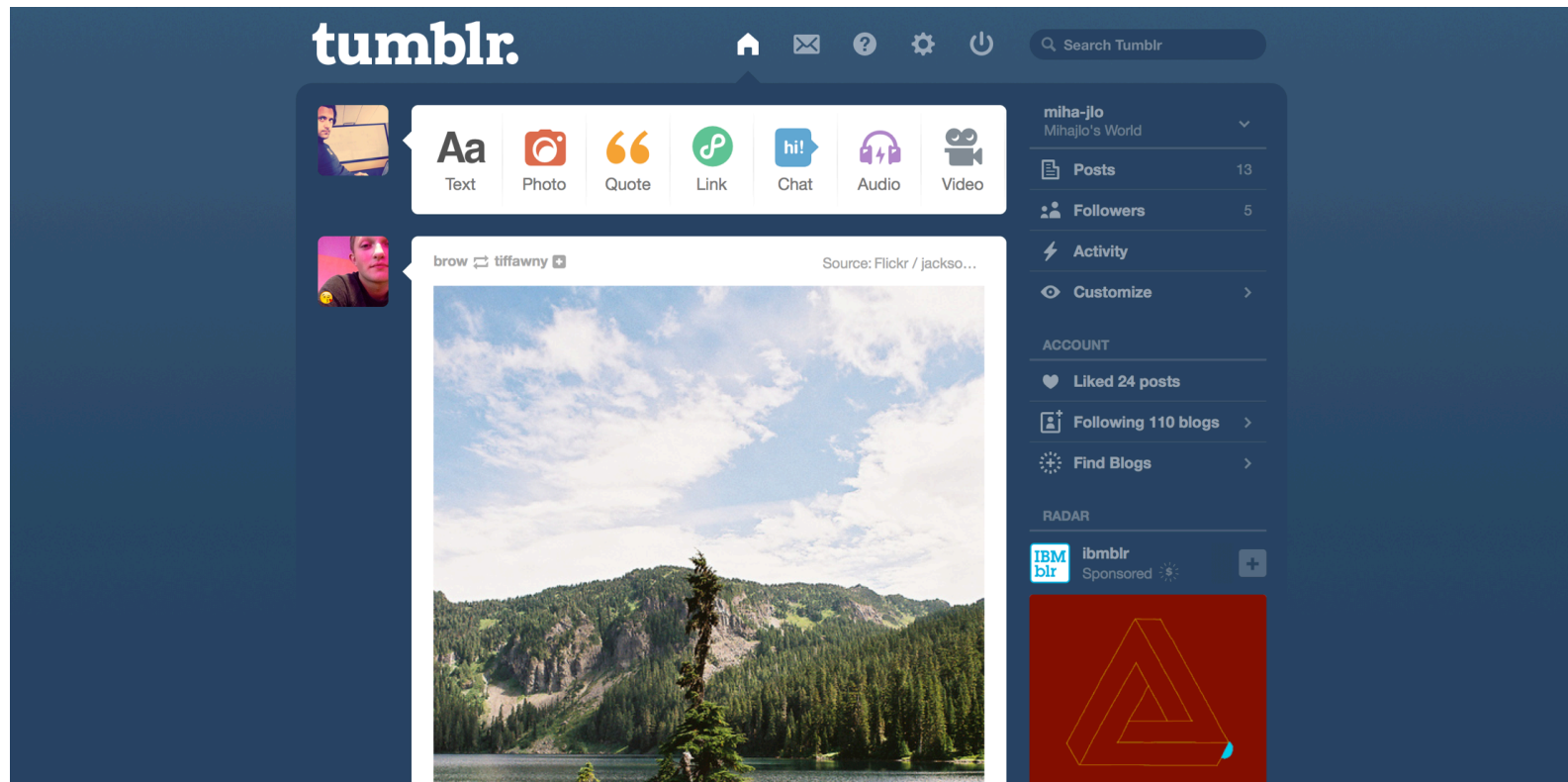
1.35%

tumblr. Post Types Breakdown



Tumblr basics - Dashboard

- As a blog owner you view posts of blogs you follow



Tumblr basics – Post examples

regular post



8bitfuture Source: springwise.com



title → **Scribble pen copies and draws in any color.**

body → 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.

tags → 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



societeperrier Sponsored  Follow

indicates advertisement



indicates advertisement

StreetArtByPerrier, net art by stfranciselevatorride

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](#)

#stfranciselevatorride #perrier #josh breeden #societeperrier #streetart #gif #su

5,390 notes

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

Tumblr data sources

- 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

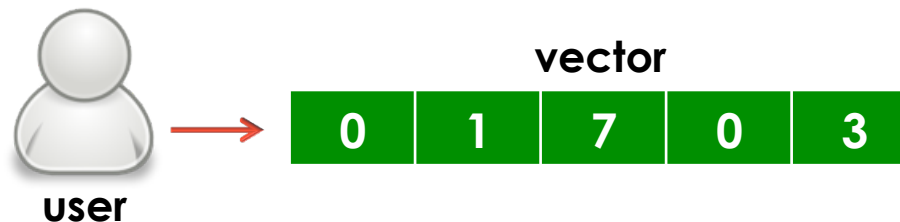
- ▣ Text from Blog Title
- ▣ Text from Blog Description

2. Content features

- ▣ Tags from Blog Posts
- ▣ Text from Blog Post content
- ▣ Artist names from audio posts

3. User actions

- ▣ Like
- ▣ Follow
- ▣ Reblog



Tumblr gender prediction

▣ 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

regex	count	regex	count
my name is*	783,564	mi chiamo*	9,181
my name's*	291,811	mein name ist*	1,025
me llamo*	47,663	meu nome e*	512
the name's*	38,065	mon nom est*	215
mi nombre es*	9,751	mio nome e*	185

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

golden set size

male	female
------	--------

395K	564K
------	------

Tumblr gender prediction

Results:

- On hold-out set:

Gender	Precision	Recall
female	0.806	0.838
male	0.794	0.689

- Editorial evaluation of 1,000 random blogs:

Gender	Correct	Wrong	Don't know
female	429	4	298
male	144	5	127

- Coverage: The classified users cover >95% actions (posts, reblogs, likes, etc.)

Tumblr interest prediction

- ▣ 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/Footwear
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

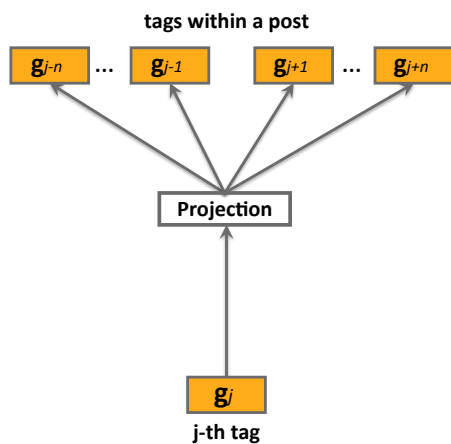
$$u_{i,cat=k}^t = \sum_{feat \in \mathcal{A}_i} \alpha^{(t-t_{feat})} w_{feat} I(feats \text{ is of class } k).$$

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

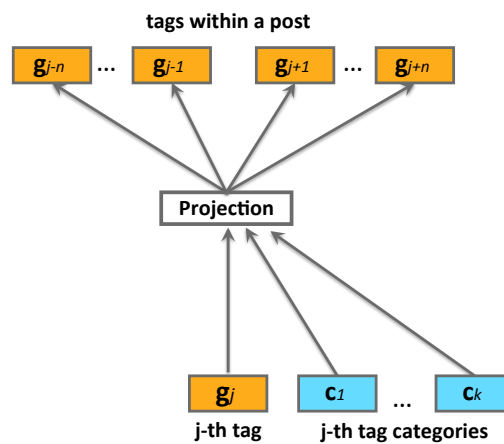
Tumblr interest prediction

Approach:

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



word2vec [1]



semi-supervised word2vec [2]

[1] Mikolov et. al. Distributed Representations of Words and Phrases and their Compositionality, NIPS 2013

[2] Grbovic et. al. queryCategorizr: A Large-Scale Semi-Supervised System for Categorization of Web Search Queries, WWW 2015

Tumblr interest prediction

Approach:

1. Categorize keywords from post content (semi-supervised word2vec)



Nearest tags to “Health & Fitness/Weight Loss” (left) and “Food & Drink/Desserts” categories (right)

Tumblr interest prediction

- We inferred interests for majority of Tumblr users
- Examples of resulting user interest assignments:

User	Category	Categorized Features
user 1	Arts and Entertainment/ Movies	tag spoilers:30 tag shrek:18 tag hercules:12 desc dvd:1 tag pokemon:7 tag thor:58 tag cinderella:3 tag hobbit:123 desc comedy:1 txt movies:100 desc movie:1 tag hulk:21 photo aladdin:28 tag disney:500 photo batman:10 txt bambi:12 desc animation:12 tag pixar:87 tag tarzan:8 tag marvel:385 tag wolverine:21 desc oscar:1 tag twilight:2 tag godzilla:1 tag x-men:53 tag pocahontas:4 tag avengers:134
user 2	Style and Fashion	txt fashion:108 tag curls:6 tag womensfashion:110 tag fashiondiaries:133 tag redhair:2 tag menswear:125 tag springfashion:50 tag style:132 tag streetstyle:132 tag hairstylist:134 tag dapper:3 tag mensfashion:124 tag chanel:4
user 3	Food and Drinks	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
user 4	Home and Garden	txt garden:32 tag daisies:2 txt interior:17 tag kitchen:20 tag chair:3 tag art:81 tag outdoor:20 tag chandelier:12 tag lamp:8 tag window:2 txt home:41 tag floral:17 tag bath:1 tag home:3 tag wildflowers:1 tag flowers:102 tag interior:201 tag tree:1 tag flower:49 tag table:1 tag stairs:2 tag bedroom:56 tag wood:2 tag bathroom:26

Tumblr interest prediction

■ A/B tests

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

Campaign	Control	Targeted
Home & Garden	-	+9.71%
Style & Fashion	-	+42.53%
Sports/Outdoor	-	+19.86%
Arts & Enter./Television	-	+24.37%
Arts & Enter./Video Games	-	+19.02%
Pets/Dogs	-	+27.21%
Arts & Enter. (1)	-	+9.08%
Arts & Enter. (2)	-	+6.54%

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?