Activity prediction: A Twitter-based exploration
Wouter Weerkamp, Maarten de Rijke
Life mining
timeu.se: What Do People Do All Day?

What Do People Do All Day? Let's find out! By collecting millions of messages from Twitter, we can explore in great detail exactly how people report spending their time. learn more

(1) Enter query words: breakfast

(2) Choose a plot type:
Mouse-over each plot type for details.
more info

Submit Query

breakfast

Sun Mon Tue Wed Thu Fri Sat

1.5e-02 1.0e-02 5.0e-02

Permalink: http://timeu.se/a840741d

timeu.se was created by Scott Golder, a graduate student in the Social Dynamics Lab in the Department of Sociology at Cornell University. On Twitter, he is @redlog.

→ Like the website? Read the paper! S. Golder, M. Macy. Diurnal and Seasonal Mood Vary with Work, Sleep and Daylength Across Diverse Cultures. Science. 9/30/2011.
I did not sleep at all last night. I must be that excited for Peter and Chris tonight.

Working on my day off.... I wouldn’t have it any other way!!!

Just finished watching Wrath Of The Titans! #GoodMovie :)

I am gonna wrestle a midget tonight... uberexcited

Come on Germany tonight.. like to see u in the final.... watching tv tonight at home... have the dogs.. kids out public viewing

Future activities

Exited to dance with the girls tonight:) #wewillrockyou
Activity prediction

- Given a future timeframe and a stream of messages, determine (popular) future activities

- Instance of “life mining”
  - Extracting useful knowledge from digital trails of online users
Why predict activities?

- Users
  - Activity recommendation
- Social media monitors
  - Crowd control
- Marketeers
  - Advertisement strategy
- Media
  - Ad pricing
Preliminary experiment

- **Method**
  - Tweet selection
  - Term extraction
  - Summarization

- **Data**
  - 7.0m tweets referring to “tonight”
  - 8.3m tweets as background corpus

- **Evaluation**
  - Predict activities for 3 days at 6pm
  - Ask 2 judges to check proposed activities
Results

- Inter-annotator agreement
  - Good (>0.65) for summary yes/no and popularity
  - Fair (0.37) for summary quality

- After discarding bad summaries
  - Agreement rises substantially
  - 66% of summaries describe an activity

- Using “to watch” in tweet selection
  - Very high agreement
  - More activities (75%)
  - Higher percentage of popular (>60% vs 30%)
Examples

**Sunday June 3**
I’m going to bed early tonight
the final episode of “peter r de vries”
tonight to “wtt”
I tonight pizza
watching a movie

**Monday June 4**
first evening of the “avondvierdaagse”
the finals of “in love with sterretje”
the first episode of “vioranje”
tonight squad training
Tonight to “guns n roses”
I have to watch “gtst” tonight
final episode of

**Tuesday June 5**
tonight to the fun fair
to the “tros” music festival
tonight working out with
tonight at 8.30pm it’ll be raffled off
tonight “gtst”
tonight once again watching at practice

“A sentence like ’tonight gtst’ is not really a sentence, let alone something you can do, even though ’watching gtst tonight’ would be.”
Discussion

- **Activity classification**
  - Does a tweet mention an activity?

- **Summarization**
  - Very important. Small number of tweets, include verbs

- **Term extraction**
  - Not only unexpected activities, but also recurring ones

- **Time indication**
  - How do people refer to timeframes ("tonight," "2nite," etc)

- **Tweet segmentation**
  - Now working on X, tonight I'll do Y

- **Evaluation**

- **Combining multiple sources**
  - Not just Twitter, also Google Calendars, FB event pages, etc.
Conclusions

- We introduced **activity prediction**
  - What will be popular activities in the (near) future?

- Initial experiments using a naive method
  - Activity prediction is feasible
  - Summarization is very important
  - Verbs are key to activities

- We identified 7 future research directions