In recent years the content of the web is changing. User-generated content includes web pages and content created and published by users of websites, rather than website owners. Everyone is a contributor. People use the web to publish and share photos, videos, audio, and various forms of written material. A particular type of user-generated content is the blog.

Blogs are frequently updated web pages, with entries shown in reverse chronological order. Many blogs function as online diaries. Personal-oriented content written in an informal manner invites interaction with readers. It offers a unique window into people’s areas of interest, thoughts, feelings.

Content includes personal nature, thoughts, emotions, commentary. Structure involves the social network of people. Timeline has every blog post, comment, and has a timestamp attached. Growth involves the growth of the blogspace as a corpus.

Our interest in weblogs addresses the information overload problem. We are no longer able to digest all material we have access to. A need for mechanisms for searching, categorizing, summarizing, and... simply understanding this data.

Blog research at ILPS involves mood/views - blog mood analysis, advertisement placement, product recommendation, opinion search in blogs, people search in blogs, recognizing social roles, blog community discovery, blog genre classification, faceted search in blogs, impact of cultural events (e.g., a concert in Paradiso), tracking the Dutch elections in news and blogs (verkiezingsviewer.nl), and correlations between financial news, financial blogs, and the stockmarket.
MoodViews.com
Gilad Mishne, Krisztian Balog, Maarten de Rijke
Collection of tools for tracking the stream of mood annotated text made available by Livejournal

Livejournal
- Popular blogging platform
- Users can annotate their posts with their mood at the time of writing
- Choose from a list of 132 moods or enter free-text
- ~1 million posts/month

Aggregating moods
- The totality of mood reports gives an "internet global mood"
- Over time: "internet mood swings"

Moodgrapher
- Simply tracks mood levels as they develop
- Some moods display a cyclic behavior
- Awake, sleepy, hungry, drunk, stressed, …
- Observe strong responses to global world events

Questions
- Can we predict the mood levels?
- Can we identify and explain peaks?

Moodteller
- Can we predict the mood from the text of a blog post?
- A text classification problem
- Individual post classification is hard
- Even with many, diverse features
- Main setback: short text → meaningless statistics
- Aggregation solves sparseness
- Can mood levels be derived from the language used by bloggers?
**Mood level prediction**

- Estimate intensity level (for each mood)
- Mood "recipe"
  - temporal metadata (hour, day of week)
  - most indicative terms
- Build linear regression model
  - Models must be constantly updated

**Evaluation**

- **Corpus**
  - 39 days of blog posts from LiveJournal
  - 8.1M posts, of these 3.5M indicate moods
  - 2.2GB of text
- **Experiments**
  - 10-fold cross validation
  - Measure: Pearson’s correlation, Relative Error
  - Baseline: use only temporal data

**Results**

- Baseline alone gets 0.71 correlation and 64% relative error
- Strong prior bias
- Regression improves both by ~20%, achieving 0.83 correlation and 52% relative error
- Some moods have correlations >0.95
  - “bored”, “happy”

**moodviews.com/Moodteller**

**Moods signals**

- Moods often reflect people’s responses to global world events
- Detect unusual behavior (swings in mood levels)
- Explain peaks

**Moods signals**

- Detect peaks
  - need to deal with cyclic events
  - calculate expected mood level based on historical data
  - If divergence exceeds a threshold a spike has occurred
    - divergence = actual / expected mood level
Explaining peaks

- Overused words
  - Identify changes in language usage
  - Compare word frequencies of peak period vs all blog posts
- Finding explanations
  - Use over-used terms and start/end dates of peak period to generate a query against a news archive
  - Return headline(s) found

Moodsignals in action

Exploring relationship between mood levels and the content of the mood tagged blog posts

Moodspotter: returns the moods associated with the topic

Use language models to relate topics and moods

Moodspotter in action

Looking forward ...

- At present over 30M blogposts indexed
- Not just tracking but also searching
  - rank along objective dimensions: time, relevance
  - rank along subjective dimensions: rank by mood
  - view it as an experiment (we don’t know how to rank here)
- Language models ...
  - observe how the language is being used
  - discover patterns
  - profiles of language usage around a topic or around a mood

Wrap-up

- Content of the web is changing
- Blogs offer a unique look into people’s reactions and feelings
- Some blogging environments (e.g. Livejournal) allow users to tag posts with their mood
- Moodviews
  - Collection of tools for tracking the stream of mood-annotated text
  - Tracking, Predicting, Explaining, Searching
Questions?

www.moodviews.com