FIGHTING SPAM ON SOCIAL WEBSITES AND SOCIAL I.P. IN NEWS

Mohammadmohammad Mohammadmohammadmohammad

March 28th, 2012

Andrea
LERMAN: SOCIAL INFORMATION PROCESSING IN SOCIAL NEWS AGGREGATION

**Premise:** The web has now become a participatory medium. Users create, evaluate, and distribute info.

**Findings:** Social networks play an important role in document recommendation. We can mathematically determine what a group’s evaluation of a document is based on an aggregate of individual choices.

**DIGG:** [http://digg.com/](http://digg.com/)
COMMON CHARACTERISTICS OF SOCIAL MEDIA

Users can:

- Create or Contribute
- Annotate
- Evaluate
- Create Social Network
Digg

A social news aggregator
- Users submit stories and moderate them
  - ~ 1 to 2 news per minute

When a story is submitted:
- Goes to the upcoming stories queue
- Users can:
  - Digg the story (+)
  - Bury the story
    - With enough bury votes → the story is removed
**Technology News**  Most Recent  All Stories

### Carpathia wants to delete orphaned Megaupload data, pay the bills

*engadget.com* — The Federal shut down of Megaupload did more than jail its founders, scare its competitors and worry its users -- it also left Carpathia Hosting for...

12 min ago via Tool  submitted in Technology and Gadgets

45  256  9 Comments

### Study: Forget sex, women want you to cook them dinner, clean the bathroom and mop the floor

*myfoxdc.com* — Married women prefer their husbands to cook them dinner than have sex, and more than half would choose food over time in bed, an Australian survey...

1 hr 11 min ago via cool  submitted in Technology, Science, Entertainment and 3 more

308  0  7 Comments

### College men walk a mile in high heels

*cnn.com*—Western Kentucky University students walk a mile "in her shoes" to raise awareness about violence against women.

Sponsored by CNN - U.S. News

### If Google were invented in the 80s

*youtube.com*  

3 hr 22 min ago via Ant Dude  submitted in Technology

604  192  8 Comments
**DIGG PARTS**

**Emergent front page**
- Story promoted to front page when it gets enough votes
- Consensus between many independent users

**Social networks**
- Users can designate others as their friends
  - Digg tracks user’s friends’ activities
- Users help each other find interesting news
  - Social filtering

**Top users**
- Ranking based on how many stories on the front-pg
  - Marketers paid top users for promotion
- Discontinued after Feb. 2007
SOCIAL FILTERING

• Out of 2858 stories
  ◦ Submitted by 1570 distinct users
  ◦ Only 98 made it to front page
    • By 60 users

Dynamics of votes on select stories
  ▪ Over a period of 4 days
  ▪ Small rectangle $\rightarrow$ In the upcoming Q
  ▪ Dashed line $\rightarrow$ transition to the front page

Interestingness
  ▪ When the votes saturate
USER RANK VS. INTERESTINGNESS

Top-ranked users are not submitting stories that get the most votes.
SOCIAL NETWORKS AND RECOMMENDATION

Top-ranked users
- not submitting the most interesting stories
- why successful?
  - social filtering, promotes their stories

Friend relationship → asymmetric

User’s success
- defined as: fraction of front-paged stories submitted by user
- correlated with his social network size
FRIENDS VS. REVERSE FRIENDS

Top users have larger social networks

Top 1020 users
- Black circles → top 33 users
CORRELATION BTWN SUCCESS & NETWORK

The larger the social network the more successful.
DIGGING BEHAVIOR OF USERS

Users digg stories that:
- Their friends submit
- Their friends have dug

- # of voters that are among the reverse friends of submitter
CHANGE IN PROMOTION ALGORITHM

Digg’s goal: feature only the most interesting on front page
- Takes into account opinions of thousands of users
  - Not just a few dedicated users

Problem: top users had a cabal
- Game system by automatically vote for each other

New algorithm
- Looks at unique digging diversity of the individuals
- Reduces dominance of top users
  - 3072 stories submitted by 1865 users
  - 71 promoted by 63 users
RESULTS OF NEW ALGORITHM

Rank distr. Less skewed towards top-ranked users

- 50% of stories from users with rank < 300
- Rather than rank < 25
MATHEMATICAL MODEL OF RATING

Interestingness coefficient $r$
- Probability of receiving a positive vote once seen

# of votes depends on visibility
- # of people who can see and follow the story

Types of visibility
- Visibility on the front page ($V_f$)
- Visibility in the upcoming stories queue ($V_u$)
- Visibility through the Friends interface
  - Friends of the submitter ($V_s$)
  - Friends of voters ($V_m$)
VISIBILITY CALCULATION

t: time since the story’s submission

If \( m(t) > h \) → visible on front page

Story stays:
- In the upcoming Q for 24h
- In Friends interface for 48h

\[
\begin{align*}
v_f &= c_f^{p(t)-1} N \Theta(m(t) - h) \\
v_u &= c_c^{q(t)-1} N \Theta(h - m(t)) \Theta(24hrs - t) \\
v_s &= a \Theta(S - at) \Theta(48hrs - t) \\
v_m &= b S_m \Theta(h - m(t)) \Theta(48hrs - t)
\end{align*}
\]
CONCLUSIONS

• Social information processing allows users to solve hard information processing problems like document recommendation and filtering, and quality evaluation.
• Social networks as a good basis for recommendation.
• Modeled collective voting using ratings on Digg.
• As we have seen with crowdsourcing, it is possible to exploit user activity to solve hard problems. There are more potential applications in personalization, search, and discovery.
QUESTIONS

• What are your overall impressions of the paper? Which elements of this paper did you most like or dislike?

• Can you think of any improvements or future directions?

• How might this be relevant to our projects?
Problem: With the rise of social websites, we also see a growing amount of spam. Spam undermines the trustworthiness and overall experience of these sites.

Findings: Spam on social websites must be addressed differently from spam in other contexts. Three types of countermeasures were investigated:

• Detection
• Demotion
• Prevention

Using network information improves spam detection models. Synthetic spam detection models have been underutilized.
Spam Characteristics:
- Goals fall under advertising, self-promotion, disruption/destruction, information gathering, and/or disparaging a competitor.
- Manifestations include audio, documents, annotations/comments, profiles, images, links. Depends on the attacker's goal.

Social Website Characteristics:
- One controlling entity, well-defined interactions, identity, multiple interfaces.

In social media sites, spammers have more avenues for attack, but service providers have more defense strategies available. They have better enforceability and better knowledge of user identity and history than in other contexts.

Choice of social bookmarking as an example is interesting, since these sites seem to have died out, or at least have been modified considerably.
imagenes, internet opportunities, ip, iphone cingular, ipod auction, ipod downloads, ipod generation, ipod link, ipod nano, ipod news update, ipod photos, ipod software, ipod video, inn mortage, japan, juegos, learning, lindsay lohan, local, london, low, make money online, manufactured homes, marketing research, medicine news, mobile marketing, model, monitor, mortgage calculator, mortgage calculator, mortgage foreclosure, mortgage jumbo, mortgage payments, mortgage rate, mortgage bond, mortgage calculators, mortgage rates, motor air, motorcycle helmet for sal., motor media, motor terra,
motor water, movie, movie2006, movie 2006, mycames, nse, online business, palm harbor homes, photos, pic, poker, pre foreclosure, quintet organ, reality, recorder, references, rental unit, scooter, second life, second mortages, sex, social networking, sofa, texas, transfer, unity, wallpapers, web cam, west nile, west nile virus, wrestling

(b)
IDENTIFICATION-BASED STRATEGIES (DETECTION)

1. Identify material that is likely spam (many possible object types)
   1. Manually by users of trusted monitors
   2. Automatically through source, text, behavioral, or link analysis.
      1. Text analysis of terms, stylistic elements, or with a corpus.
      2. Behavioral analysis can be very useful because typical interactions on social website are often quite well-defined.

2. Delete the information or mark that it is probably spam.
   Methods are commonly used on email as well.
RANK-BASED STRATEGIES (DEMOPTION)

• Reduce the prominence of context likely to be spam.
• Generally done through an automated system, but can also be done manually (Ex: Reddit).
• Creating personalized ranks by incorporating elements such as geography can make this more robust.
• Difficult to implement on time-ranked interfaces.

This method is also commonly applied on the web.
INTERFACE- OR LIMIT-BASED (PREVENTION)

• Make the contribution of spam difficult through secret details or limiting automated interaction.

• **Interface based or limit-based.**
  - Create restrictions on access to actions (ex: CAPTCHA, graylisting)
    • Common in other types of sites and email as well. Fairly universal.
  - Place restrictions on number of actions allowed (hard-number limits; charges)
    • Not often implemented

Which tactics do you think most effectively minimize social spam?
Are there other possibilities?
EVALUATION MODELS AND METRICS

- Evaluation is important because many of these strategies are costly to implement.
- The subjectivity of spam and malicious activity can make modeling difficult.
- Synthetic spam models versus trace-driven models. Both require annotation.
  - Trace-driven is generally preferred and considered more realistic.
  - Authors argue synthetic models should be utilized to avoid sample bias and time consumption, and to account for site growth/change. However, these are more difficult to interpret.

- Metrics:
  - Classification (precision and recall)
  - Ranking (SpamFactor, mean average precision), broader problems.

What did you think about the synthetic and trace-driven models? What about the metrics used by the author?
EXPERIMENT

• Created a synthetic model on a tagging system. Implemented a tag budget. The system returns ranked lists of objects from single-tag queries.
  • Boolean model
  • Occurrence-based model (counting tuples)
  • Coincidence-based model (user reliability)

• Leveraging social knowledge such as tag coincidences improved spam detection by a factor of 2 (when compared to SpamFactor). This is only if no collusion takes place. Also, what about users with mixed “good” and “bad” behavior?

Coincidence-based > Occurrence-based > Boolean
CONCLUSIONS

• The solutions detailed in the study will likely work better for social media sites than in other contexts due to their highly structured nature.
  • Importance of not annoying users with too much spam prevention.

• In the future, spammers will probably start coming up with more specialized spam methods, and these techniques may not be specialized enough.
  • Cycle of improved sophistication on both sides.

• Which combinations of countermeasures will be most effective in which environments? What is the appropriate combination of manual and automated work?
MORE QUESTIONS

• Which elements of this paper did you most like or dislike?

• Do you think spammers have evolved since 2007? Can you think of any examples? Is it still appropriate or useful to group all social media sites together?

• How are these concepts relevant to our projects? Are there any in particular that you think we should consider implementing?