Analyzing Community Preference of Digg Comments

Elham Khabiri, Chiao-Fang Hsu, and James Caverlee
Department of Computer Science and Engineering
khabiri,drakihsu,caverlee@cse.tamu.edu

Digging into Digg

- A social media website
- A social news aggregator
- Collaborative filtering
- Creation, Evaluation
- Users can "Post Stories"
- "Comment" on stories
- "Digg" (rank) the comments and stories

Digg Analysis

- Features based on the visibility of the comment in the community
  - The Digg score of the article the comment is attached to
  - The order of the posted comment

- Features based on the influence of the commenter
  - Number of Articles Submitted
  - Number of Diggs
  - Number of Articles Appeared on Digg Front Page
  - Number of Profile Views
  - History of Received Comment Ratings
  - History of Received Comment Replies

- Linear regression Classifier

\[
F = W - S = \sum_{i=1}^{12} f_i \times w_i = S
\]

\[
F = W - S = \sum_{i=1}^{12} f_i \times w_i = S
\]

- Quadratic Classifier

\[
g(x) = -\frac{1}{2}(x - \mu)\Sigma^{-1}(x - \mu) - \frac{1}{2}\log(\Sigma) + \log(P(w_i))
\]

Summary

- Quadratic classifier has a higher classification rate than regression method.
- Precision for the Fair and Good categories is high (0.82 and 0.695) relative to the precision for the Bad and Excellent categories (0.25 and 0.019). These two categories are relatively small and difficult to predict.
- Digg users prefer short, simple, and readable comments.
- The so-called power users in the community do not wield considerable influence over the scores of comments in the community.

References