MODELING DOMAIN SPECIFIC AUTHOR REPUTATION AND TRUST IN WIKIPEDIA

Matthew Molek

CMSC601 Spring 2011

Introduction

- Why is trust important in Wikipedia?
  - Vandalism Prediction
  - Predicting information quality
    - One implementation: text highlighting
  - Highly controversial or frequently vandalized pages can be locked to only allow trusted users to makes edits

How is trust currently modeled?

- Natural language processing analysis of text
- Meta-data analysis
  - Including information quality metrics
- Analysis of content-persistence
- Citation analysis
  - Similar in function to page rank

Previous work

□ Content persistence based trust.

- B.T. Adler and L. De Alfaro. A content driven reputation system for the Wikipedia. WWW '07: *Proceedings of the 16ᵗʰ international conference on World Wide Web. 2007.*

□ *Wikipedia stores all versions of every page enabling revision history to be used to calculate reputation.*

□ *Reputation is based on:*

□ *Text survival*

□ *Edit survival*

Strengths and weaknesses of Adler and De Alfaro's system

- Strengths
  - Has both prescriptive and descriptive value to Wikipedia
  - Trust values are predictive
    - The edits of low reputation users were 4.2 times more likely to be reverted on the next edit, and 32.2% of short lived edits came from low reputation users
- Weaknesses
  - Trust is modeled globally, rather than by topic.
  - The trust update function is a little unnatural.

Improving the trust update function

- Currently uses 10 successive edits to adjust the users reputation
  - Improvement: Use an exponential decay function to weigh the influence of each successive edit.

Domain-specific Trust

- The current system has no way to model trust on a per-domain basis

- Knowing that an author is trusted on one subject, but has never before contributed to some other subject should provide the trust function increased predictive abilities

Implementing Domain-specific Trust

- Categories will be taken from Wikipedia's "primary categories" list.
  - Every Wikipedia article falls into one of these categories
  - E.g., Business, Technology, Arts, and History

Implementing Domain-specific Trust

- Every author will start will a low reputation score in all categories

  - Anonymous users will be stuck at this score

- Update domain trust as the author makes contributions/edits.

  - The trust update will give more weight to revisions made by users trusted within the domain, and less to users from other domains

Test Methodology

- Data: Early full site dumps of the Italian (2005) and French (2006)

- Starting at the beginning, run through the edit logs in chronological order, updating trust values at each time step.

  - Want to get user trust vs. edit longevity statistics

Test Methodology

- Analyze the results of the test to determine the predictive accuracy of the trust system.
  - What is the likelihood that an untrusted user in a domain will have their text/edit reverted
  - What % of short lived edits come from users not trusted in the domain of the edit?
  - What is the likelihood that an trusted user in a domain will make contributions that are long lived?
  - What % of long lived edits come from users trusted in the domain of the edit?

Comparing results

- Results will be compared to
  - Adler and De Alfaro's system
  - A simple "number of edits" reputation system

QUESTIONS?