Social Tag-Based Recommendation Services

Jordan Bentley
Advisor: Elke Rundensteiner
Collaborator: Dr. K. Claypool, MIT Lincoln Lab
Outline

- Introduction to Collaborative Filtering
- Tag based solutions
- Problems
  - Tag Unification
  - Recommendation Algorithm
Collaborative Filtering

- Predicts user interest in a given resource
  - i.e., Netflix, Amazon

Collaborative Filtering

- User-User model
  - The “Gold Standard” of collaborative filtering models
  - Finds users with similar interests

---

User-User Model

USER A

USER B

Reading

Writing

Arithmetic
Collaborative Filtering

- Ratings Based
  - This is the most common form
  - Takes into consideration how much a user likes an item but not why
  - Needs CONTEXT\(^1\)

---

Solution – TAGS

Web 2.0 tags include:
- Aggregators
- Blogs
- Folksonomy
- Participation
- Recommendation
- Social Software
- Wikis
- User Centered
- Joy of Use
- Usability
- Six Degrees
- Perpetual Beta
- Simplicity
- Widgets
- Browser
- AJAX
- Design
- CSS
- Pay Per Click
- UMTS
- Convergence
- Audio
- Video
- Mobility
- Atom
- RSS
- OpenAPIs
- Remixability
- Data Driven
- Accessibility
- Modularity
- SOAP
- SVG
- Ruby on Rails
- REST
- Web Standards
- Standardization
- Microformats
- Syndication
- XML
- UDDI
- Trust
- Affiliation
- Economy
- The Long Tail

Author: Luca Cremonini
Impact

- Tag based social bookmarking sites already exist without recommendation engines
  - Del.icio.us
  - Citeulike
  - Flickr

- Only tag searching

Objective

- Create a collaborative filtering algorithm that uses tags to take context into account
  - Targeting academic papers for researchers
The Process

Data → Recommender → User
Tag Unification

- Problems in the data:
  - Misspellings
  - Form
    - Walk -> Walking, Walked, Walker, Walks, etc
- Easy solution:
  - Use a dictionary and an existing spell checking and stemming algorithm
- But what if the words do not appear in any dictionary?
Tag Unification

- Tag changes are bi-directional
- Focus is on consistency
  - Not correct spelling

spelling ↔ spelin

spelin ↔ spelling
Tag Unification

- Proposed heuristic method
  - Starts with no predefined dictionary
  - Uses multiple measures to find good changes
    - Edit Distance
    - Co-Occurrence
    - Length
Edit Distance

- Traditional method of difference in spelling

- This example has an edit distance of 3
Tag Unification

distance_{t,p} = EditDistance_{t,p}
* \frac{\text{Resources} \in (t \cap p)}{\text{Resources} \in (t \cup p)}
* \sqrt{LEN(t) - 5}
Tag Unification

- Evaluation had to be done by hand
  - A ‘good’ match was any change that preserved meaning
    - Even if the change was to an *incorrect* spelling, it was a good change
Tag Unification

7.4%
Similarity Layers

User

Tag Vector

Tag
Tag Similarity

- Uses the co-occurrence from tag unification

\[ \text{sim}_{a,b} = \frac{\text{Resources} \in (a \cap b)}{\text{Resources} \in (a \cup b)} \]
Tag-Vector Similarity

Data \times Suggestions = Low Match
Data \times Information = High Match
Recommendations \times Suggestions = High Match
Recommendations \times Information = Low Match
User Similarity

$$\text{Sim}(A_{\text{writing}}, B_{\text{writing}}) + \text{Sim}(A_{\text{arithmetic}}, B_{\text{arithmetic}})$$
Evaluation

1. Remove random resource from a random user
2. Generate recommendations for that user
3. Evaluate how often the user is recommended the removed resource
Evaluation

- Dataset
  - Obtained from citeulike
  - 115,548 unique resources
  - 23,133 unique tags
  - 3,567 users

Evaluation

- 24% of removed resources were in the top 50 recommendations
- 50% were in the top 270
- 95% were in the top 2000
Future work

- Compare Tag based methods to ratings based methods
  - Survey in progress
  - Will be able to run both methods on the same data
  - Data will be reusable for future studies
Thank you

Questions?