A Hybrid Recommender System based on Weighted Tags

Huizhi (Elly) Liang

Queensland University of Technology, Brisbane, Australia

Agenda

- Introduction
- Related Work
- The Proposed Approach
- Experiments
- Conclusions
- Questions & Answers

Introduction

- Recommender systems
  - User information
    - Explicit rating
    - Implicit rating
- Web 2.0: Read and Write web (O'Reilly Media, 2004)
  - A platform for users to conduct online participation, collaboration and interaction.
  - Expressing opinions, sharing information, building networks
  - Plenty of new user information
    - Tags, reviews, networks, blogs etc.
- Opportunities
  - Providing possible new solutions to profile users

Introduction 2

- Social tags
  - Keywords given by users to describe/organize items

The World Is Flat 3.0: A Brief History of the Twenty-first Century by Thomas L. Friedman
(Paperback - July 24, 2007)

Buy new: ~~$16.00~~ **$8.40**

149 new from $4.44 340 used from $2.50

Get it by **Tuesday, Apr. 20** if you order in the next **22 hours** and choose one-day shipping.

Category

Books > History > Americas > United States > 21st Century
Books > History > World > 21st Century
Books > Nonfiction > Politics > Globalization
Books > Nonfiction > Politics > International > Relations
Books > Nonfiction > Social Sciences
Books > Science > Technology > History of Technology
Books > Science > Technology > Social Aspects

Tags

- globalization (77)
- economics (42)
- business (32)
- technology (32)
- history (27)
- capitalism (22)
- sociology (20)
- flat world (10)
- freakonomics (9)
- book (5)
- outsourcing (4)
- careers (3)
- politics (3)
- internet (2)
- it (2)
- navy reading list (2)
- nonfiction (2)
- rtfm (2)
- thomas l friedman (2)
- 0312425074 (1)

Introduction 3

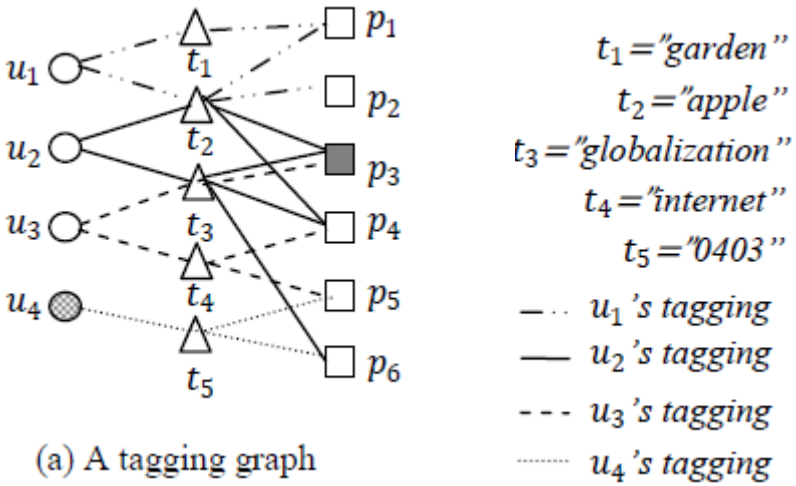
- Features
  - Given by users
  - lightweight textural information
  - Independent with the content types of items
  - Implying users' interests and preferences
  - Three dimensional relationships
- Challenges
  - Free chosen arbitrary words
    - 60% are personal tags
    - Semantic ambiguity
    - Tag synonyms
  - Inaccurate user profiling & content description

Related work

- Tag recommendation
  - Tensor based approach *(KDD, 2009)*
  - Graph based approach *(SIGIR, 2009)*
- User recommendation
  - Siersdorfer S. *(HT, 2009)*
- Item recommendation using tags
  - Tso-Sutter etc. *(SAC, 2008)*
  - Previous Work *(WI, 2009)*
  - Shilad Sen etc. *(WWW,2009)*
  - Siersdorfer S. *(HT, 2009)*

The Multiple relationships

- Two dimensional relationships
  - User-Item relationship
    - User-Item mapping → Standard user based CF approach
    - Item-User mapping → Standard item based CF approach
  - User-Tag relationship
  - Item-Tag relationship
- Three dimensional relationships
  - User-Tag-Item relationship: records each user's personal tagging relationships
    - *(User×Tag)-Item mapping*
    - *Item-(User×Tag) mapping*



(a) A tagging graph

The Proposed Approach

- Tag representation
- User profiling
- Neighborhood forming
- Recommendation Generation

Tag representation

- Eliminate the noise of tags

  Find the meaning of each tag for each user individually

  - The relevance of a tag to an item
    - The number of users used the tag for the item/ the total number of users tagged the item

$$\mathcal{P}(t_x \mid p_k) = \frac{\mathcal{P}(p_k,\ t_x)}{\mathcal{P}(p_k)} = \frac{|U_{p_k,t_x}|}{|U_{p_k}|} \qquad (1)$$

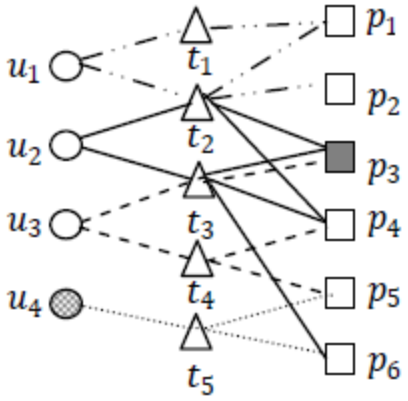
  - The relevance of two tags in terms of each individual user
    - Each item is equally important
    - The average relevance of a tag to the collected items

$$r_{u_i,t_x}(t_y) = \sum_{p_k \in P_{u_i,t_x}} \frac{\mathcal{P}(t_y \mid p_k)}{|P_{u_i,t_x}|} \qquad (3)$$

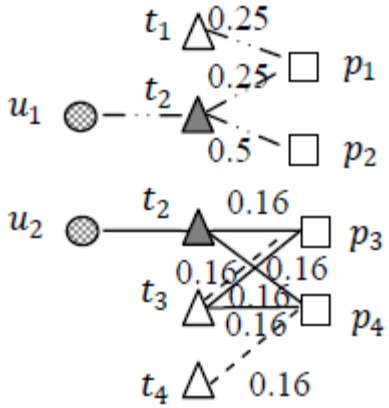
  - Tag representation
    - Represents each tag $t_x \in T$'s relevance to each tag $t_y \in T$ with respect to a individual user
    - A set of tags with weights $RT(u_i, t_x) = \{(t_y, r_{u_i,t_x}(t_y)) \mid t_y \in T\}$

Tag representation 2

- Example



(a) A tagging graph



(b) Tag representation

t_1 = "garden"
t_2 = "apple"
t_3 = "globalization"
t_4 = "internet"
t_5 = "0403"

$$\mathcal{P}(t_2 \mid p_3) = 1/3, \mathcal{P}(t_3 \mid p_3) = 2/3 \qquad r_{u_2,t_2}(t_3) = \frac{1}{2}\cdot\frac{1}{3} + \frac{1}{2}\cdot\frac{2}{3} = 0.5$$

$$RT(u_1, t_2) = \{(t_1, 0.25), (t_2, 0.75)\}$$

$$RT(u_2, t_2) = \{(t_1, 0.0), (t_2, 0.16), (t_3, 0.5), (t_4, 0.34), (t_5, 0.0)\}$$

$$RT(u_4, t_5) = \{(t_1, 0.0), (t_2, 0.0), (t_3, 0.25), (t_4, 0.25), (t_5, 0.5)\}$$

Item representation

- Expand the tags of each item
  - Get the tag representation of the user-tag pair
  - Calculate the weights of the tags
    - M: the number of user-tag pairs
    - Each user-tag pair is equally important

$$wp(t_y) = \sum_{u_i \in U_{p_k}, t_x \in T_{p_k}} \frac{1}{M} \cdot r_{u_i,t_x}(t_y) \qquad (4)$$

  - The inverse item frequency of each tag

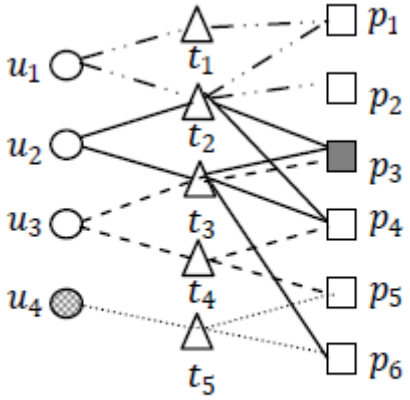
$$w_y^p = wp(t_y) \cdot iif(t_y) \quad (5)$$

  - Item representation
    - Represents each item's relevance to each tag $t_y \in T$
    - A set of tags with weights

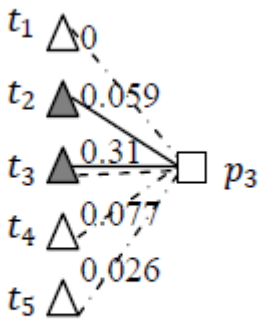
$$RP(p_k) = \{(t_y, w_y^p) | \, t_y \in T\}$$

Item representation 2

- Example



(a) A tagging graph



(c) Item representation

t_1 = "garden"
t_2 = "apple"
t_3 = "globalization"
t_4 = "internet"
t_5 = "0403"

$$wp(t_5) = \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{1}{2} = 0.05 \qquad iif(t_5) = \frac{1}{log\,(e+4)} = 0.52$$

$$RP(p_3) = \{(t_1, 0.0), (t_2, 0.059), (t_3, 0.31), (t_4, 0.077), (t_5, 0.026)\}$$

User Profiling

- User profile
  - Item preferences
  - Tag preferences:
    - Get the weight of each tag for a user
      - Number of items in the tag/Total tagging number $\mathcal{P}(t_x \mid u_i) = \frac{|P_{u_i,t_x}|}{|P_{u_i}|}$ (2)
    - Get the weight of the related tag
      - Get the representation of each user-tag pair
      - The total relevance weight in all user-tag pairs

$$wu(t_y) = \sum_{t_x \in T} \mathcal{P}(t_x \mid u_i) \cdot r_{u_i,t_x}(t_y) \quad (6)$$

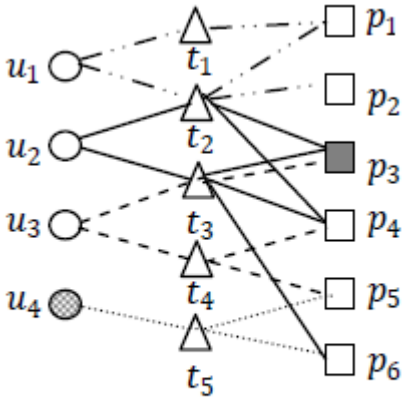
    - The inverse user frequency of each tag

$$w_y^u = wu(t_y) \cdot iuf(t_y) \quad (7)$$

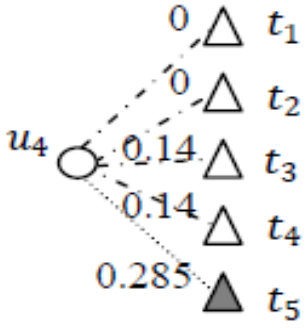
  - User representation
    - Represents each user's preferences to each tag $t_y \in T$
    - A set of tags with weights $RU(u_i) = \{(t_y, w_y^u) \mid t_y \in T\}$

User Profiling 2

- Example



(a) A tagging graph



(d) User representation

t_1 = "garden"
t_2 = "apple"
t_3 = "globalization"
t_4 = "internet"
t_5 = "0403"

$$wu(t_5) = 0.5 \quad iuf(t_5) = \frac{1}{log\,(e+3)} = 0.57$$

$$RU(u_4) = \{(t_1,\ 0.0),\ (t_2,\ 0.0),\ (t_3,\ 0.14),\ (t_4,\ 0.14),\ (t_5\ 0.285)\}$$

Neighbourhood Formation

- K-Nearest Neighbourhood
  - Similar users
    - Tag preference
      - User representation
      - Cosine similarity

$$cosine(v_i, v_j) = \frac{\sum_{y=1}^{|T|} v_{i,y} \cdot v_{j,y}}{\sqrt{(\sum_{y=1}^{|T|} {v_{i,y}}^2) \cdot (\sum_{y=1}^{|T|} {v_{j,y}}^2)}}$$

    - Item preference
      - Users' implicit ratings

$$sim_u^P(u_i, u_j) = \frac{\sum_{p_k \in P_{u_i} \cap P_{u_j}} iuf(p_k)}{\sqrt{|P_{u_i}| \cdot |P_{u_j}|}}$$

    - Linear combination

$$sim_u(u_i, u_j) = (1-\eta) \cdot sim_u^T(u_i, u_j) + \eta \cdot sim_u^P(u_i, u_j) =$$
$$(1-\eta) \cdot cosine(u_i, u_j) + \eta \cdot \frac{\sum_{p_k \in P_{u_i} \cap P_{u_j}} iuf(p_k)}{\sqrt{|P_{u_i}| \cdot |P_{u_i}|}} \qquad (16)$$

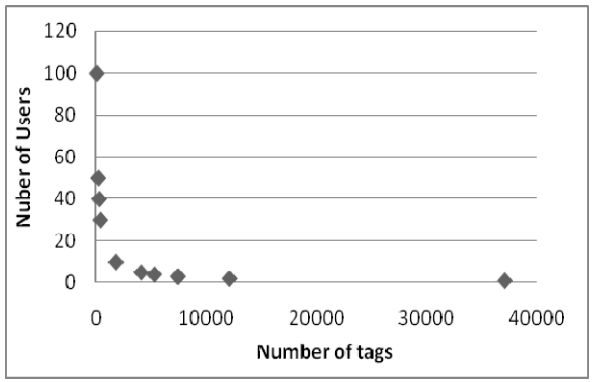
Recommendation Generation

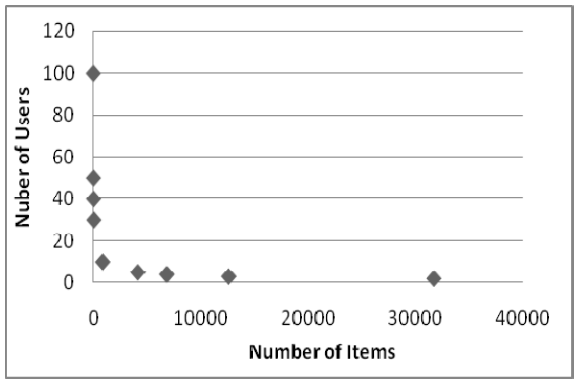
- Candidate items
  - Neighbour items
  - Not collected by the target user
- Recommend top *N* items
  - Collaborative Filtering
    - Neighbour users
  - Content filtering
    - Item representation
    - User representation
  - Hybrid

$$A_u(u_i, p_k) = sim_{u,p}(u_i, p_k) \cdot \sum_{u_j \in \check{\mathrm{N}}(u_i) \cap U_{p_k}} sim_u(u_i, u_j) =$$
$$cosine(u_i, p_k) \cdot \sum_{u_j \in \check{\mathrm{N}}(u_i) \cap U_{p_k}} sim_u(u_i, u_j)$$

Experiments

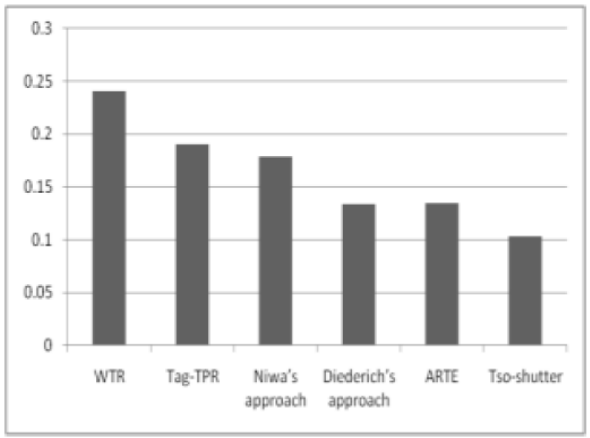
- Datasets
  - Amazon.com
    - 4112 users, 34201 tags, 30467 items, 9919 taxonomic topics
- Evaluation methods
  - Precision
  - Recall
- The distributions of tags and items

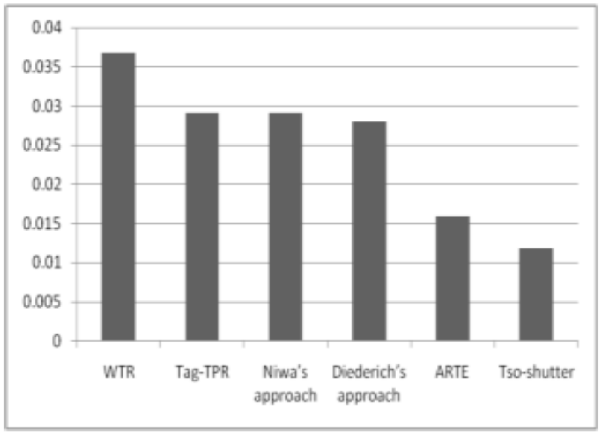




Experiments 2

- Comparisons *(Top 3 Precision and Recall)*
  - *WTR: the proposed approach*
  - *ARTE: association rule*
  - *Tag-TPR: taxonomic topics*
  - *Tso-Sutter's approach: user-item-tag matrix*
  - *Diederich's approach: tag tf-iuf*
  - *Niwa's approach: clustering*





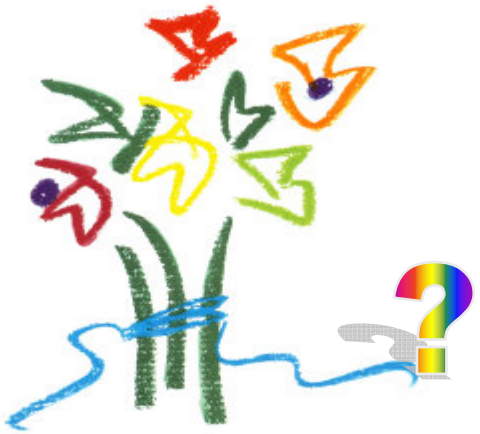
Conclusions

- New user information in Web 2.0
- Tag quality
- The multiple relationships
- Hybrid the Collaborative Filtering and Content filtering approach based on tags

Acknowledgements

- Travel support of QCIF (Queensland Cyber Infrastructure Foundation) and HPC Group of QUT
- The feedbacks of reviewers.
  - *Connecting Users and Items with Weighted Tags for Personalized Item Recommendations*
  - Accepted as a full paper by Hypertext 2010

Questions & Answers



huizhi.liang@student.qut.edu.au