NSLAB TSINGHUA UNIVERSITY

RBRA: A Simple and Efficient *R*ating-*B*ased *R*ecommendation *A*lgorithm to Cope with Sparsity in Recommender Systems

Feng Xie, Ming Xu, and Zhen Chen

Present by *Feng Xie*

December 13, 2011

Content

- Background
- RBRA
  - Similarity Measure
  - Rating Prediction
- Experimental Results
- Conclusion

Background

The world is an over-crowded place

Background

They all want to get our attention

Who can help us?

- Can google help?
  - Yes, but only when we really know what we are looking for
- Can experts help?
  - Yes, but it won't scale well
    - Everyone receives exactly the same advice!
- Recommender System (RS) can help us
  - With information overload
    - Based on our history selections
    - Based on other people with similar interests

Approaches of RS

- Collaborative filtering
  - User-based (1994, GroupLens)
  - Item-based (2001, Amazon)
- Content-based filtering
- Hybrid
  - Linear
  - Switching combination
  - Sequential

Collaborative Filtering

- Collaborative filtering
  - Widely used in practice
- Classification
  - User-based
  - Item-based
- Rating based
  - user-item rating matrix

Existed Problems

- Data sparse with less correlation
- Similarity computation in different dimensions
- Relevance among data
- Cold start
- ……

RBRA: *R*ating-*B*ased *R*ecommender *A*lgorithm

- Contributions
  - New similarity computation model
  - Adaptively integrate user-based and item-based
    - Entropy
- Advantages
  - Not care about the correlation
  - Similarity computation in the same dimension
  - Eliminate the relevance
  - Performance improved

RBRA Construction

- Similarity Measure
- Rating Prediction

Similarity Measure

- Mean, Variance and Range

$$\bar{r}_u = \frac{\sum_{i \in I(u)} r_{u,i}}{n} \qquad S_u = \frac{1}{size(I(u))} \sum_{i \in I(u)} (r_{u,i} - \bar{r}_u)^2 \qquad R_u = \max(r_{u,i}) - \min(r_{u,i})$$

- Cosine-based

$$sim(u, v) = \cos(\vec{u}, \vec{v}) = \frac{\vec{u} \cdot \vec{v}}{\| \vec{u} \| * \| \vec{v} \|} = \frac{\bar{r}_u \bar{r}_v + S_u S_v + R_u R_v}{\sqrt{\bar{r}_u^{\,2} + S_u^{\,2} + R_u^{\,2}} \cdot \sqrt{\bar{r}_v^{\,2} + S_v^{\,2} + R_v^{\,2}}}$$

Rating Prediction

- User-based

$$Ur_{u,i} = \bar{r}_u + \frac{1}{\sum_{v \in S(u)} sim(u,v)} \sum_{v \in S(u)} sim(u,v)(r_{v,i} - \bar{r}_v)$$

- Item-based

$$Ir_{u,i} = \frac{1}{\sum_{j \in S(i)} sim(i,j)} \sum_{j \in S(i)} sim(i,j) r_{u,j}$$

- Final prediction

$$\hat{r}_{u,i} = \alpha \cdot Ur_{u,i} + \beta \cdot Ir_{u,i} \quad , \quad \beta = 1 - \alpha \quad \text{initialization,} \quad \alpha = \beta = 0.5$$

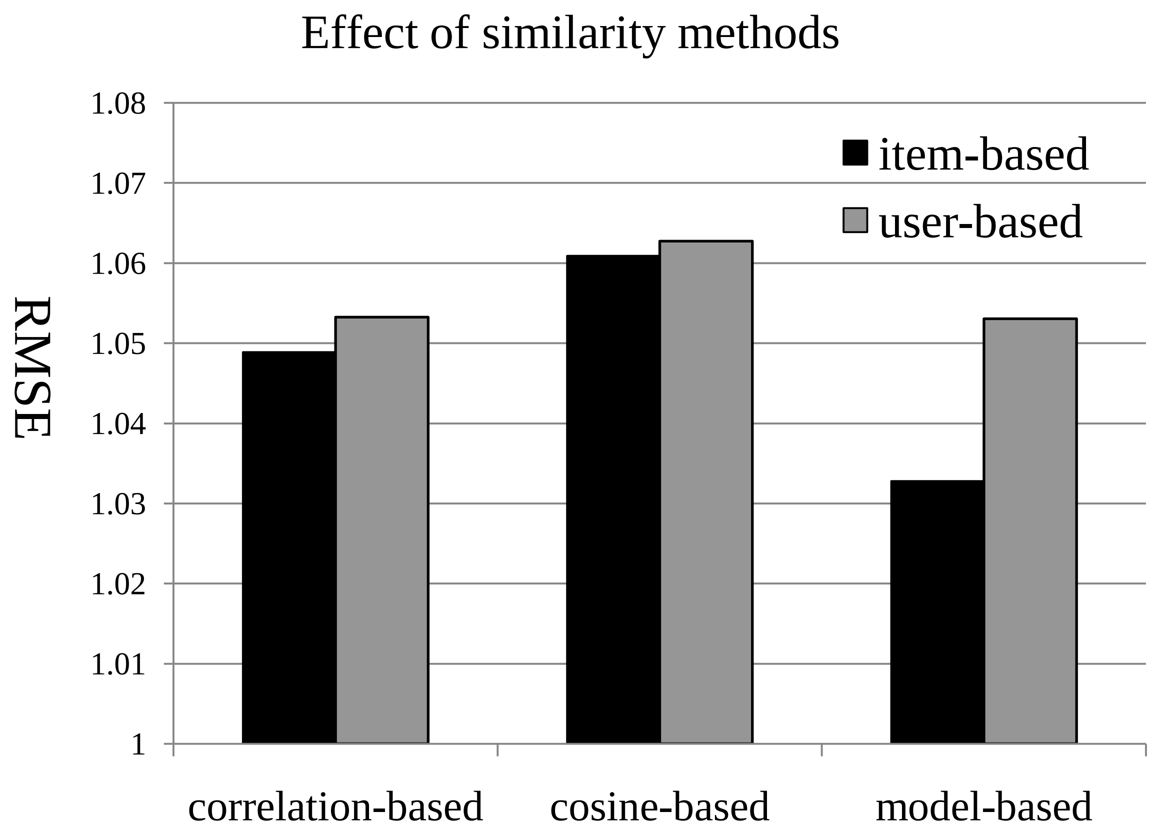
$$I = \frac{\sum_k num_i}{n * k} \qquad \alpha = 0.5 + \mathrm{sgn}(I_{user-based} - I_{item-based}) \cdot step$$

Experiments Design

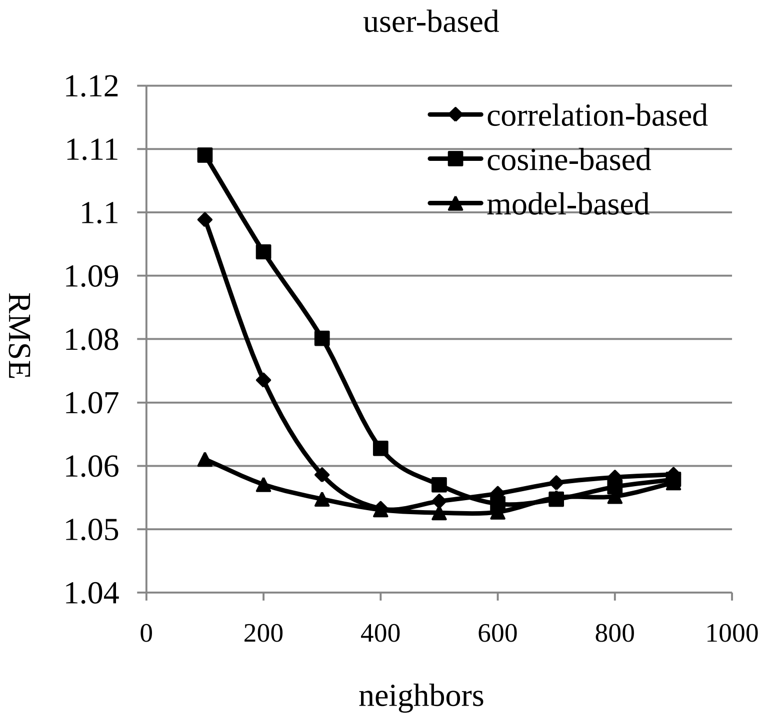
- Dataset
  - MovieLens 100K (sparsity level 93.7%)
  - 10,000 ratings of 943 users on 1682 movies
  - 80% for training, 20% for test
- Evaluation Metric
  - RMSE

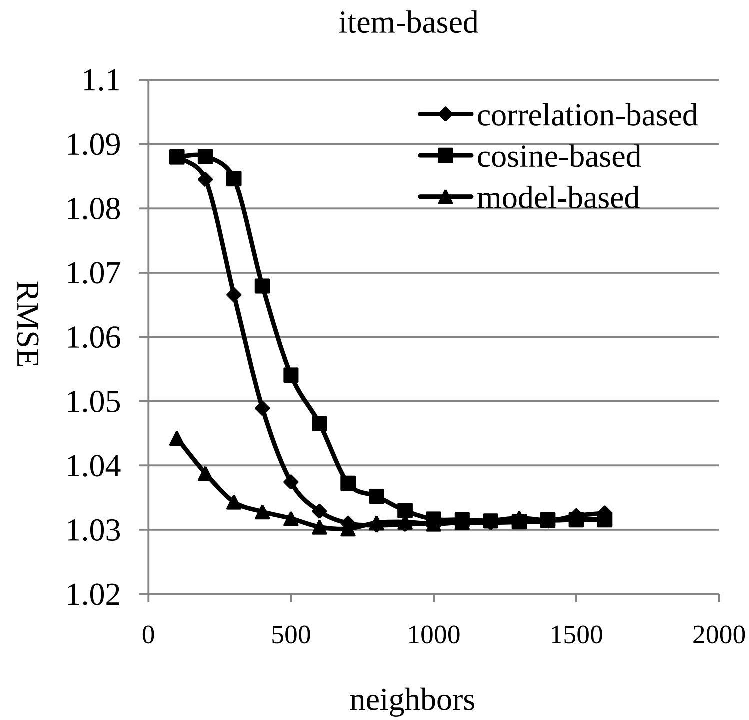
$$RMSE = \sqrt{\frac{\sum_{i=1}^{size(TEST)} (prate_i - rate_i)^2}{size(TEST)}}$$

Experimental Results 1

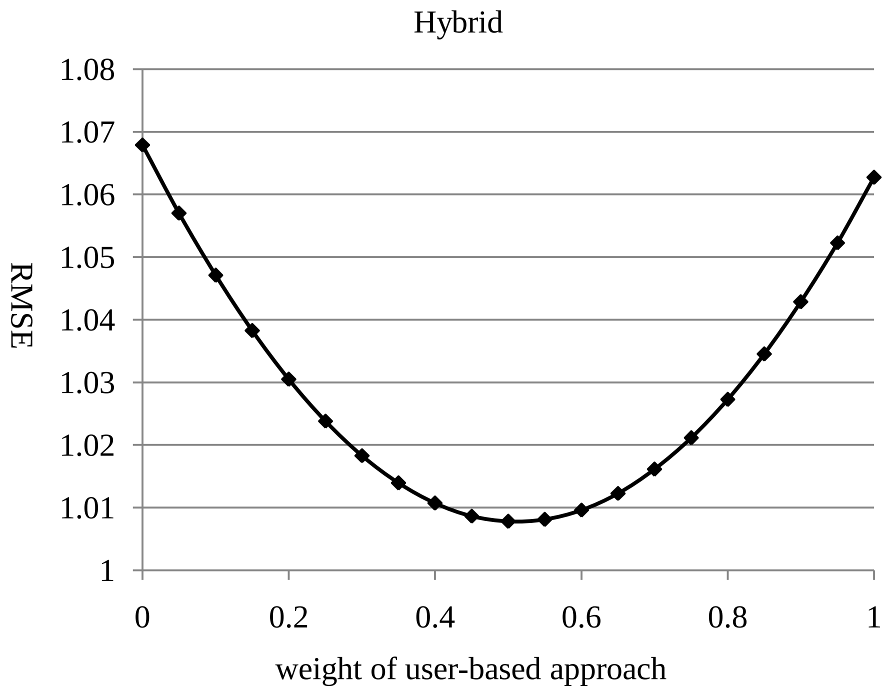


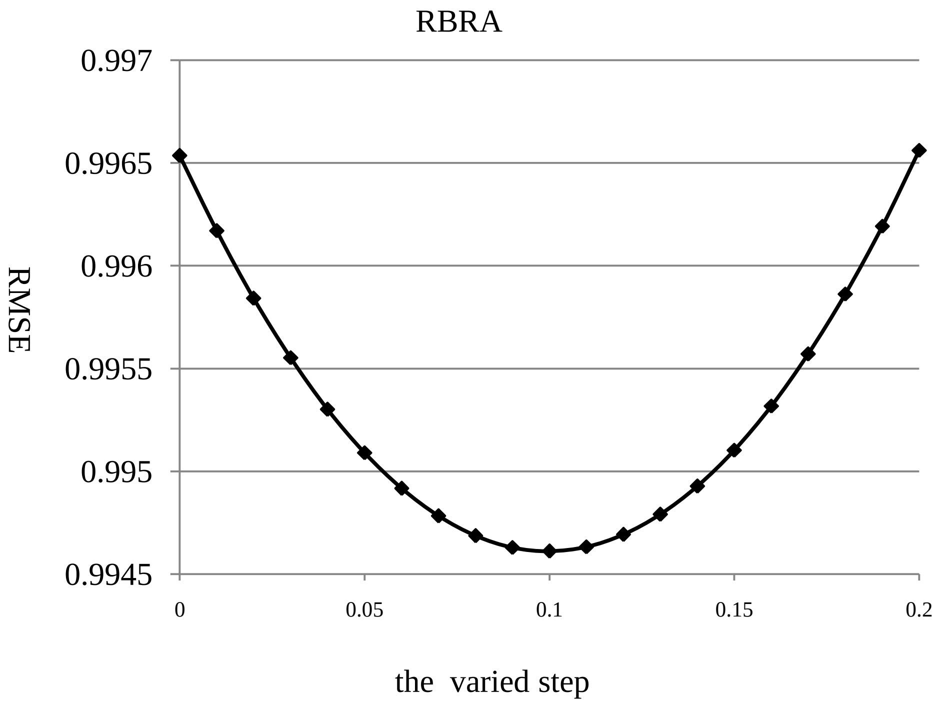
Experimental Results 2





Experimental Results 3





Conclusion

- New similarity computation model
  - Not care about the correlation
  - Similarity computation in the same dimension
  - Eliminate the relevance
- Adaptively weighted user-based and item-based
  - More information used
  - Performance improved
- RBRA achieves 400% faster speed in our experiments

Any Questions?

Thanks