The ACM Conference Series on Recommender Systems

NANYANG TECHNOLOGICAL UNIVERSITY

Integrating Trust and Similarity to Ameliorate the Data Sparsity and Cold Start for Recommender Systems

Guibing Guo

School of Computer Engineering
Nanyang Technological University, Singapore

0 About me

- Ph.D candidate in NTU, Singapore
- Research Interest
  - Trust-aware RecSys
    - Data sparsity
    - Cold start
  - Trust & trust management
- More …
  - http://trust.sce.ntu.edu.sg/~gguo1/
  - gguo1@e.ntu.edu.sg

NANYANG TECHNOLOGICAL UNIVERSITY

1 Target Research Problems

- **Data Sparsity:** many empty entries (? marked)
- **Cold Start:** new users rate only few items

Table: User-item rating matrix

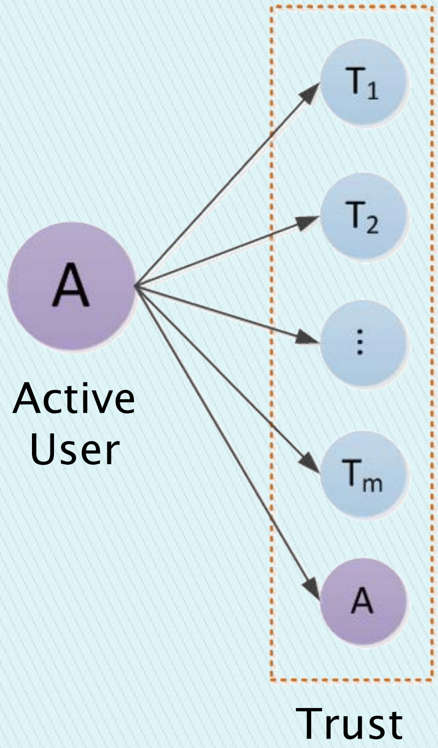
	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}
A	3	4	5	?	?	?	?	3	?	?
B	?	4	4	5	3	?	?	?	?	1
C	5	5	3	4	?	?	2	?	?	?

NANYANG TECHNOLOGICAL UNIVERSITY

1 Current Research Progress

- Incorporate additional information:
  - Trust-aware profile merging (UMAP 2012)
- Utilizing existent ratings
  - Bayesian similarity measure (IJCAI 2013)
- Eliciting more kinds of ratings
  - Prior ratings (RecSys 2013)

2 The Merge Method (Guo et al., UMAP 2012)



	P_1	P_2	P_3	…	P_n
T_1	$r_{1,1}$	$r_{1,2}$	$r_{1,3}$	…	$r_{1,n}$
T_2	$r_{2,1}$	$r_{2,2}$	$r_{2,3}$	…	$r_{2,n}$
⋮	⋮	⋮	⋮	⋮	⋮
T_m	$r_{m,1}$	$r_{m,2}$	$r_{m,3}$	…	$r_{m,n}$
A			$r_{a,3}$	…	

A	$(\tilde{r}_{a,1}, c_{a,1})$	$(\tilde{r}_{a,2}, c_{a,2})$	$(r_{a,3}, 1.0)$	…	$(\tilde{r}_{a,n}, c_{a,n})$

New Rating Profile

NANYANG TECHNOLOGICAL UNIVERSITY

2 Merging process

- Merging ratings:

$$\tilde{r}_{u,i} = \frac{\sum_{v \in TN_u} t_{u,v} r_{u,i}}{\sum_{v \in TN_u} t_{u,v}}$$

- Confidence:

$$c_{u,i} = \frac{1}{2} \int_0^1 \left| \frac{x^r (1-x)^s}{\int_0^1 x^r (1-x)^s dx} - 1 \right| dx$$

  - r: # positive ratings
  - s: # negative ratings

2 Integrating with Collaborative Filtering

- Similarity measure:

$$s_{u,v} = \frac{\sum_{i \in I_{u,v}} c_{u,i} (\tilde{r}_{u,i} - \bar{r}_u)(r_{v,i} - \bar{r}_v)}{\sqrt{\sum_{i \in I_{u,v}} c_{u,i}^2 (\tilde{r}_{u,i} - \bar{r}_u)^2} \sqrt{\sum_{i \in I_{u,v}} (r_{v,i} - \bar{r}_v)^2}}$$

- Prediction generation:

$$\hat{r}_{u,j} = \frac{\sum_{v \in NN_u} s_{u,v} r_{v,j}}{\sum_{v \in NN_u} |s_{u,v}|}$$

2 Evaluation

- Results:

Table: The predictive performance on the Flixster data set

Views	CF	MT1	MT2	MT3	RN	TCF1	TCF2	Merge1	Merge2	Merge3
All Users	0.928 0.686 0.736	1.060 0.124 0.213	0.932 0.714 0.751	0.862 0.907 0.855	0.858 0.004 0.008	0.870 0.809 0.808	0.850 0.852 0.831	0.890 0.896 0.847	0.877 0.949 0.869	0.875 0.950 0.872
Cold Users	1.153 0.033 0.063	1.127 0.081 0.146	1.005 0.527 0.628	0.934 0.796 0.794	NaN 0.0 NaN	1.047 0.130 0.222	0.923 0.214 0.337	1.008 0.630 0.696	0.960 0.831 0.808	0.949 0.852 0.819

MT: MoleTrust
RN: Reconstruct Network
TCF: Trust-enhanced CF

Row 1: mean absolute error
Row 2: rating coverage
Row 3: F-measure

3 Bayesian Similarity Measure (Guo et al., IJCAI 2013)

- Make better use of ratings
  - Traditional measures fails in cold condition
  - Direction only, length ignored

- Bayesian similarity measure
  - Both direction and length
  - Evidence based
  - Chance correlation removed
  - User bias considered

NANYANG TECHNOLOGICAL UNIVERSITY

3 Dirichlet distribution

- The probability density

$$p(x|\alpha) = \frac{\Gamma(\alpha_0)}{\prod_{i=1}\Gamma(\alpha_i)}\prod_{i=1} x_i^{\alpha_i - 1}$$

where

$$\alpha_i = \begin{cases} \sum_{j=1}^{n} n^2 p_j^2, & if\ i = 1 \\ 2\sum_{j=1}^{n-i+1} n^2 p_j p_{j+i-1}, & if\ 1 < i \le n \end{cases}$$

- Evidence weight:

$$e_i = \begin{cases} 1 & if\ c\sigma_k = 0 \\ 1 - \dfrac{d_i}{c\sigma_k} & if\ 0 \le d_i < 2c\sigma_k \\ -1 & otherwise \end{cases}$$

3 User distance

- New data probability

$$E\left(x_i \middle| \alpha_i + \gamma_i^0\right) = \frac{\alpha_i + \gamma_i^0}{\alpha_0 + \gamma^0}$$

Where

$$\gamma_i^0 = \sum_{j=1}^{N} \gamma_i^j e_i^j \text{ and } \gamma^0 = \sum_{i=0}^{N} \gamma_i^0$$

- User distance

$$d_{u,v} = \frac{\sum_{i=1}^{n} w_i d_i}{\sum_{i=1}^{n} |w_i|}$$

where

$$w_i = E\left(x_i \middle| \alpha_i + \gamma_i^0\right) - E(x_i|\alpha_i) > 0$$

NANYANG TECHNOLOGICAL UNIVERSITY

3 Similarity Measure

- Raw similarity

$$s'_{u,v} = 1 - \frac{d_{u,v}}{d_n}$$

- Chance correlation

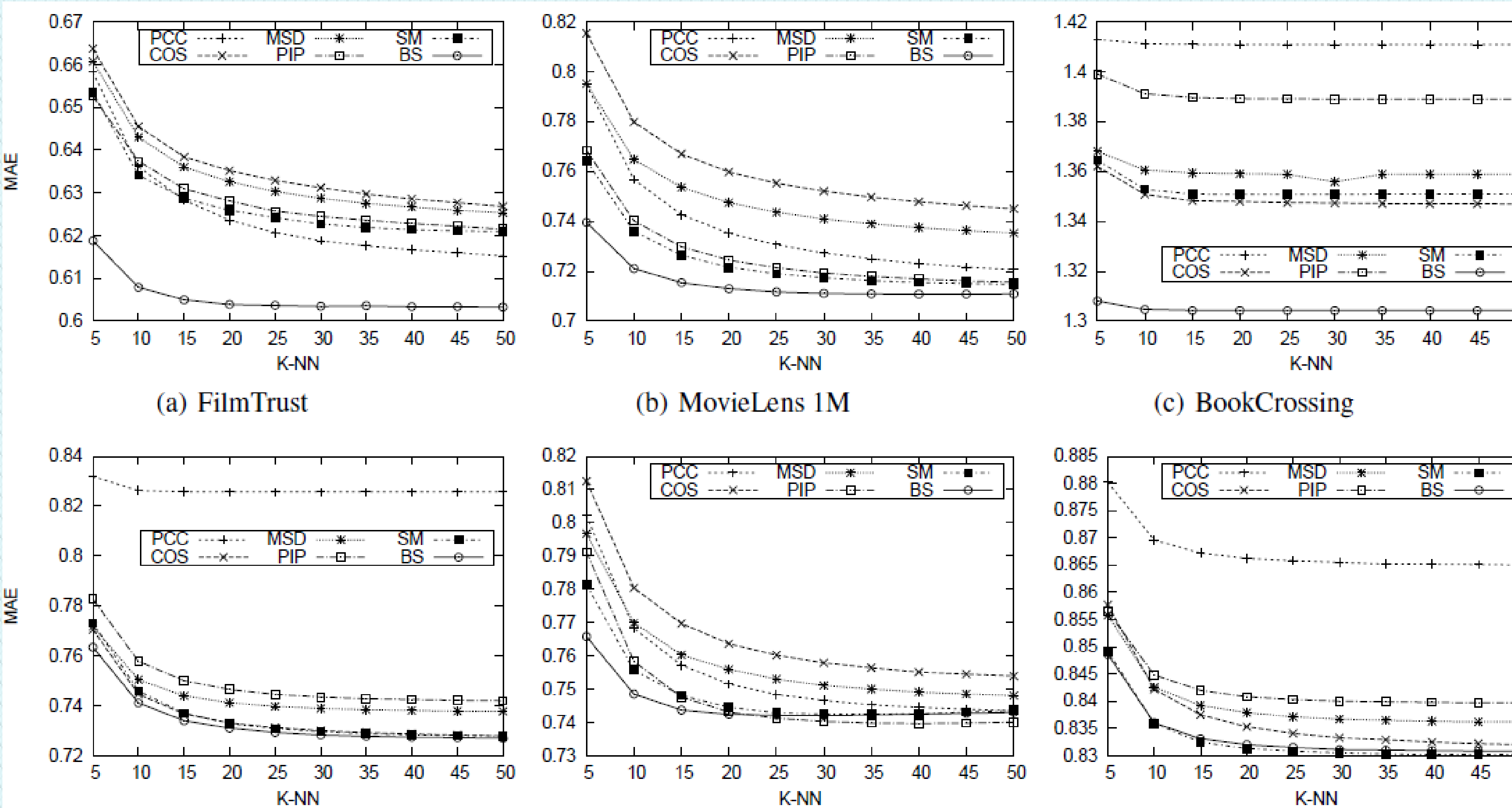
$$s''_{u,v} = \prod_{i=1}^{n} \left(\frac{\alpha_i}{\alpha_0}\right)^{\gamma_i^0}$$

- User similarity

$$s_{u,v} = \max(s'_{u,v} - s''_{u,v} - \delta, 0)$$

δ: user bias

3 Evaluation



(a) FilmTrust (b) MovieLens 1M (c) BookCrossing

(d) Flixster (e) MovieLens 100K (f) Epinions

4 Prior Ratings (Guo et al., RecSys 2013)

- Motivation
  - Users lack of incentives to rate
  - Rate often after purchase

- Prior ratings
  - Rate before purchase
  - Interactive virtual product experience
  - More joyful experience raises incentives

4 Conceptual Model

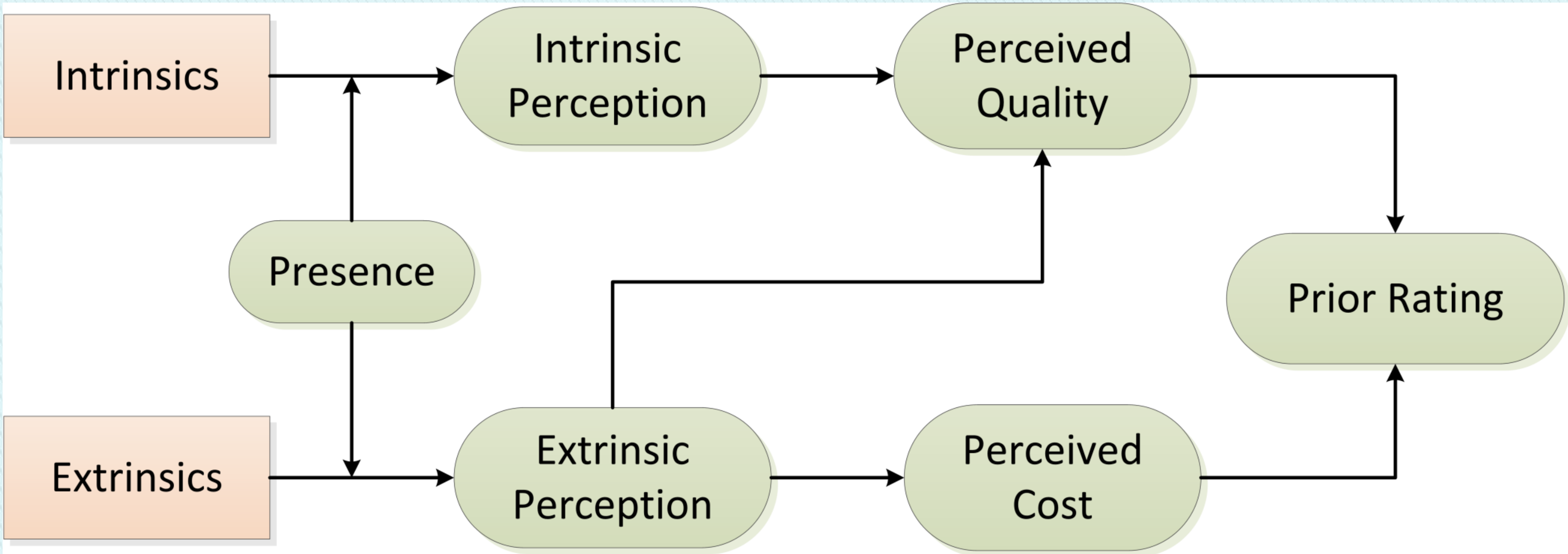
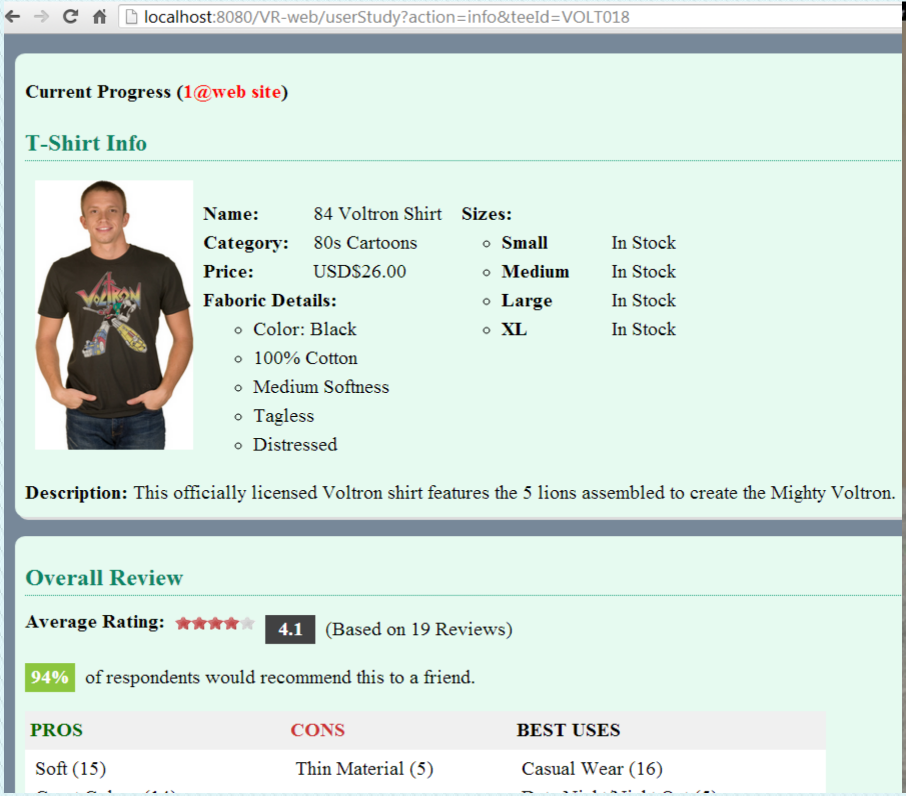


Figure: the conceptual model of prior ratings

4 User Study



1. Website (WS)

2. Virtual Reality (VR)

4 Conclusions

1. Users are willing to provide prior ratings in VR.
2. High confidence and closer to the posterior ratings.
3. Presence: positive influence on the perceptions of both intrinsic and extrinsic attributes.
4. Users depend more on extrinsic attributes in WS but more on intrinsic attributes in VR.
5. Both perceived quality and cost have positive influence on the prior ratings.

5 Questions w.r.t Future Work

- Infer implicit trust
  - Learn trust factors from explicit trust
  - Incorporate distrust information
  - Generalize to applns wo trust information

- Model-based approaches
  - Incorporate contextual information
  - Balance accuracy and diversity

Thank You! & Questions?