Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18), New Orleans, Louisiana, February 1–7, 2018

University of Science and Technology of China (USTC)

Confidence-aware Matrix Factorization for Recommender Systems

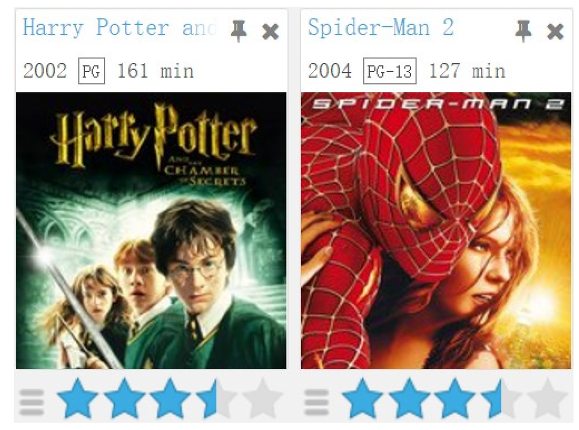
Chao Wang[1], Qi Liu[1], Runze Wu[1], Enhong Chen[1]*, Chuanren Liu[2], Xunpeng Huang[1], Zhenya Huang[1]

[1]Anhui Province Key Lab. of Big Data Analysis and Application, University of Science and Technology of China,

[2]Decision Science and Management Information Systems Department, Drexel University

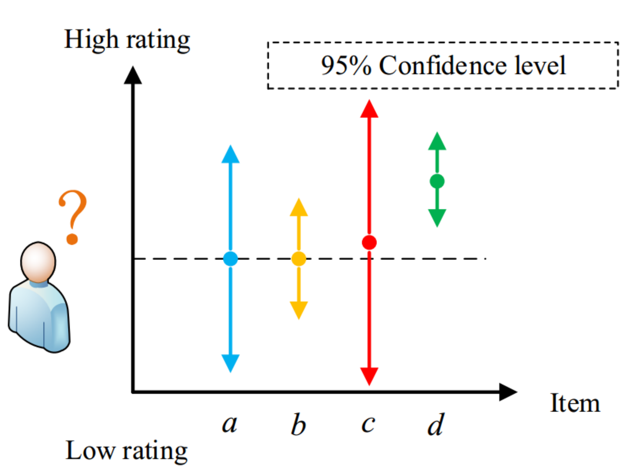
Introduction

Motivation: Existing accuracy-oriented MF methods cannot always meet the expectation of end-users. Users do not necessarily prefer the items with higher predicted ratings.



Fact : In practice, a user often has several alternatives in the recommendation. Confidence weighs in the user's decision significantly when the rating itself is not sufficient to make conclusive decisions.

Definition: The **confidence of rating prediction** is defined as the recommender system's trust in its prediction.



How to build confidence-aware recommender systems?

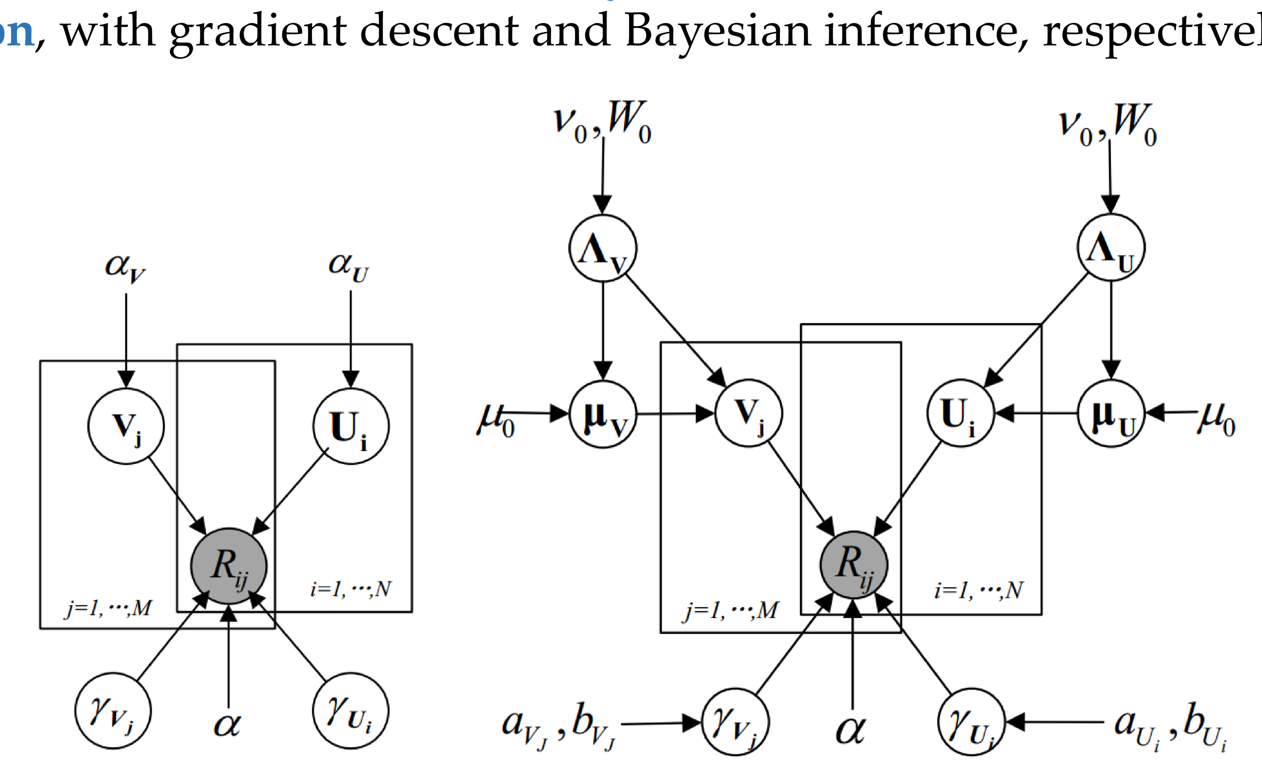
Goal: Optimize accuracy of rating prediction and measure the prediction confidence simultaneously in a general framework.

Challenges:

- Requires **an unified way** to understand and measure the ratings and confidence from an overall perspective.
- **Influence of users and items** on rating variances should be both considered.
- The confidence of ratings should be employed to **improve the accuracy** of rating prediction.

Implementation

We propose two implementations, i.e., **Confidence-aware Probabilistic Matrix Factorization** and **Confidence-aware Bayesian Probabilistic Matrix Factorization**, with gradient descent and Bayesian inference, respectively.



(a) CPMF (b) CBPMF

Algorithm 1 Gibbs sampling for CBPMF

1: Initialize factor parameters U^0, V^0.
2: **for** $t = 1$ to T **do**
3: Draw factor hyperparameters Θ_U^t and Θ_V^t from Gaussian-Wishart distributions: $P(\Theta_U^t|U^{t-1})$ and $P(\Theta_V^t|V^{t-1})$ like Equation (11).
4: Draw variance parameters γ_U^t and γ_V^t from Gamma distributions: $\Gamma(\gamma_{U_i}^t|a_{U_i}^{*t}, b_{U_i}^{*t})$ and $\Gamma(\gamma_{V_j}^t|a_{V_j}^{*t}, b_{V_j}^{*t})$.
5: Draw factor parameters U^t and V^t from Gaussian distributions: $\mathcal{N}(U^t|\mu_U^{*t}, [\Lambda_U^{-1}]^{*t})$ and $\mathcal{N}(V^t|\mu_V^{*t}, [\Lambda_V^{-1}]^{*t})$.
6: Draw each rating R_{ij} from $\mathcal{N}(R_{ij}|(U_i^t)^T V_j^t, (\gamma_{U_i}^t \gamma_{V_j}^t \alpha)^{-1})$.
7: **end for**

Speaker: Chao Wang
Email: wdyx2012@mail.ustc.edu.cn
Lab Homepage: http://bigdata.ustc.edu.cn/

Confidence-aware Matrix Factorization Framework

- Input space : $\{R_{ij}\}$: The rating of user i for item j.
- Latent space : $\{U_i$ and $V_j\}$: User-specific and item-specific latent feature vectors for user i and item j.
  $\{\gamma_{U_i}$ and $\gamma_{V_j}\}$: Variance parameters of user i and item j.
- Output space: Rating prediction and prediction interval.

$$P(R|U,V,\alpha,\gamma_U,\gamma_V) = \prod_i^N \prod_j^M \left[\mathcal{P}(R_{ij}|U_i^T V_j, (F(\gamma_{U_i},\gamma_{V_j})\cdot\alpha)^{-1})\right]^{I_{ij}},$$

$$U_i \sim \widetilde{\mathcal{P}}(U_i|0,\alpha_U^{-1}I), \quad \gamma_{U_i} \sim \Gamma(\gamma_{U_i}|a_{U_i}, b_{U_i}),$$

$$V_j \sim \widetilde{\mathcal{P}}(V_j|0,\alpha_V^{-1}I). \quad \gamma_{V_j} \sim \Gamma(\gamma_{V_j}|a_{V_j}, b_{V_j}).$$

Prediction interval :

$$\left[U_i^T V_j - z(\tfrac{p}{2}) * (\gamma_{U_i}\gamma_{V_j}\alpha)^{-\frac{1}{2}}, U_i^T V_j + z(\tfrac{p}{2}) * (\gamma_{U_i}\gamma_{V_j}\alpha)^{-\frac{1}{2}}\right],$$

Confidence-aware Ranking:

1) A candidate list of more than K elements is generated by ranking mean of the predicted rating;

2) The candidate list is rearranged by ranking in the order of the defined value of **Sharpe Ratio**.

$$SharpeRatio = \frac{\hat{R}_{ij} - R_0}{\sigma_{ij}},$$

- R_0 is a constant and represents the benchmark rating which means items below this boundary may be unacceptable.

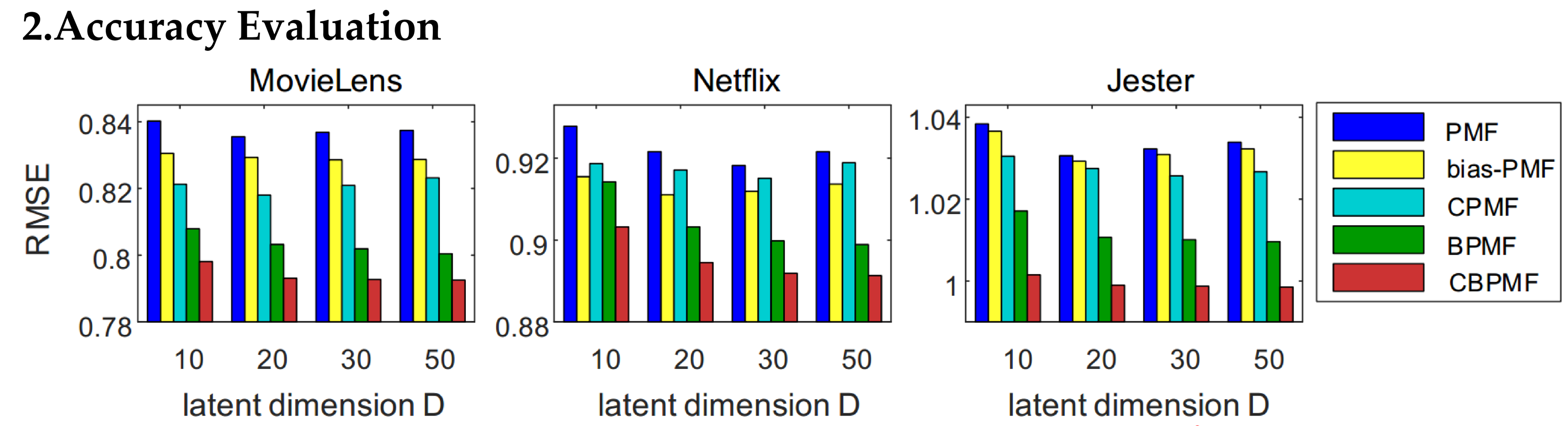
Experiments

1.Dataset: We conduct experiments on three real-world datasets.

Table 1: The statistics of the datasets.

Dataset	Users	Items	Ratings	sparsity
MovieLens	6,040	3,952	1M	4.19%
Netflix	480,198	17,770	100M	1.18%
Jester	59,132	140	1.8M	21.28%

2.Accuracy Evaluation

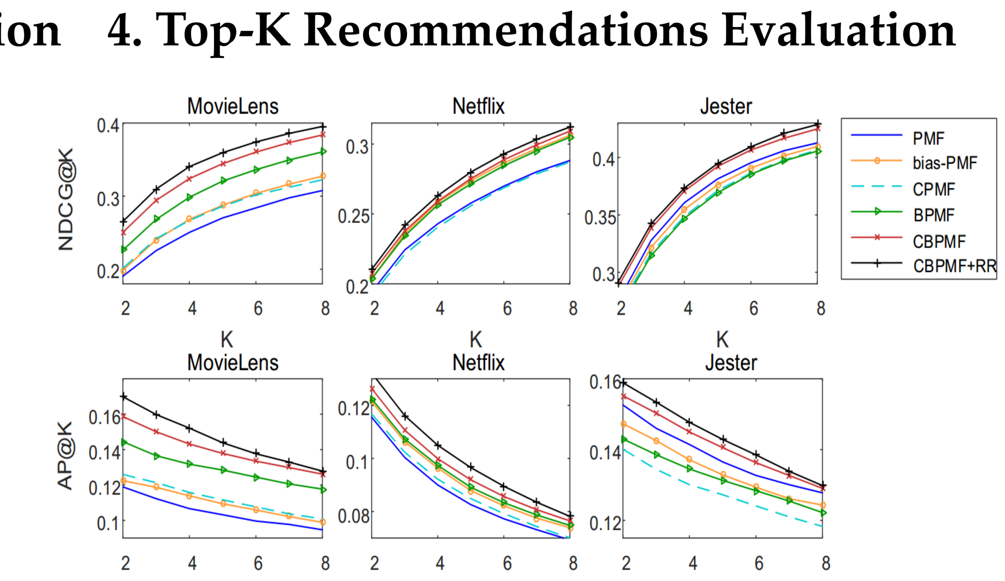


3. Confidence Measurement Evaluation

Table 2: The comparison results of CP.

Dataset	confidence level 90%	
	BPMF	CBPMF
MovieLens	86.26% (3.74%)	**89.24% (0.76%)**
Netflix	82.16% (7.84%)	**86.48% (3.52%)**
Jester	78.06% (11.94%)	**86.57% (3.43%)**
Dataset	**confidence level 95%**	
	BPMF	CBPMF
MovieLens	91.14% (3.86%)	**94.26% (0.74%)**
Netflix	87.88% (7.12%)	**91.73% (3.27%)**
Jester	84.15% (10.85%)	**92.56% (2.44%)**

4. Top-K Recommendations Evaluation



- By **combining accuracy and confidence**, our model outperformed alternative methods on prediction accuracy, confidence measurement and top-K recommendations.
- Extensive experiments on three real-world datasets demonstrate the **effectiveness** of our framework from **multiple perspectives**.

5. Analyzing the Variance Parameters on MovieLens

- Elder users tend to have bigger variance parameters, which indicates less uncertainty on ratings.
- Males tend to have bigger variance parameters than females.
- Movies with more non-realistic elements seem to show more uncertainty on ratings.