210430 Special Lectures on Database (RecSys)

Collaborative Filtering for Implicit Feedback Dataset

Heeseung Yun

heeseung.yun@vision.snu.ac.kr

Contents

- Motivation
- Approach
- Experiment
- Conclusion

Motivation

Recommender Systems

- Content-based Approach
  - Requires external information
- Collaborative Filtering
  - Neighborhood models vs. Latent factor models
  - domain-free
  - cold-start problem

Recommender Systems

- Explicit Feedback
  - Content-based Approach
    - Requires external information
  - Collaborative Filtering
    - Neighborhood models vs. Latent factor models
    - domain-free
    - cold-start problem
- Implicit Feedback
  - Purchase history, browsing history, search patterns, etc.
  - We don’t always provide feedback ☺

Implicit Feedback

- Properties
  - No negative feedback
    - Dislike or inaccessible? “Asymmetry”
    - Implications of missing data
  - Inherently noisy
    - Implicit feedback may not represent preference
    - (ex) disappointing product, purchased as a gift

me: *turns off tv cause it was too loud*

my dad: hey i was watching that

him 2 seconds ago:

Implicit Feedback

- Properties
  - No negative feedback
  - Inherently noisy
  - Feedback indicates confidence
    - cf) Explicit feedback indicates preference
  - Tricky evaluation
    - Appropriate measure
    - Conflicts (ex. two concurrent channels)

Implicit Feedback

- Properties
  - No negative feedback
  - Inherently noisy
  - Feedback indicates confidence
  - Tricky evaluation

- Objective
  - Design framework for implicit feedback-oriented datasets
    - Factorization-based
    - TV-show dataset

Approach

Preliminaries

- Quantifying implicit feedbacks
  - Observation: r_{ui}
    - (ex) 2.5 if one watched a drama two times till the end and 50% of the show
    - Exact value is known for **every** entry
  - Preference: p_{ui}
    - $p_{ui} = 1$ if $r_{ui} > \tau$ $(\tau = 0)$
  - Confidence: $c_{ui} = 1 + \alpha r_{ui}$
    - Alternatively, $c_{ui} = 1 + \alpha \log\left(1 + \frac{r_{ui}}{\epsilon}\right)$
    - Empirically $\alpha = 40$
    - ϵ for normalizing repetition ($\epsilon = 10^{-8}$)

Model

- Objective

$$\min_{x_\star, y_\star} \sum_{u,i} c_{ui}(p_{ui} - x_u^T y_i)^2 + \lambda \left(\sum_u \|x_u\|^2 + \sum_i \|y_i\|^2 \right)$$

  - Predicting preference weighted by confidence
  - Accounting for all possible pairs
    - SGD not applicable
  - Solution: Alternating-Least-Squares (ALS) optimization

Model

- ALS Optimization
  - Update user-factors with fixed item-factors and vice versa
  - Repeat until convergence (i.e., minimized loss)

Model

- ALS Optimization
  - User-factor

$$x_u = (Y^T \underbrace{C^u}_{\substack{\text{User-specific confidence}\\ n \times n \text{ diagonal matrix}}} \overbrace{Y}^{\substack{\text{Item-factor}\\ n \times f \text{ matrix}}} + \underbrace{\lambda I}_{\text{Regularizer}})^{-1} Y^T C^u \overbrace{p(u)}^{\substack{\text{User preference}\\ n\text{-D vector}}}$$

Model

- ALS Optimization
    - User-factor

Confidence-weighted
user preference
(n_u non-zero entries)

$$x_u = (Y^T C^u Y + \lambda I)^{-1} Y^T C^u p(u)$$

Computational bottleneck: $O(f^2 n)$

Model

- ALS Optimization
  - User-factor
    - Complexity: $O(f^2 n_u + f^3)$

Confidence-weighted
user preference
(n_u non-zero entries)

$$x_u = (Y^T C^u Y + \lambda I)^{-1} Y^T C^u p(u)$$

Computational bottleneck: $O(f^2 n)$

$$Y^T C^u Y = Y^T Y + Y^T (C^u - I) Y$$

Precomputed

n_u non-zero entries

Model

- ALS Optimization
  - User-factor
    - Complexity: $O(f^2 n_u + f^3)$
    - For all users: $O(f^2 N + f^3 m)$ $(N = \sum_u n_u, 20 \leq f \leq 200)$
  - Linear complexity

Model

- ALS Optimization
  - Item factor

User-factor
$m \times f$ matrix

Item preference
m-D vector

$$y_i = (X^T C^i X + \lambda I)^{-1} X^T C^i p(i)$$

Regularizer

Item-specific confidence
$m \times m$ diagonal matrix

Model

- ALS Optimization
  - Item factor
    - Complexity: $O(f^2 n_i + f^3)$
    - For all users: $O(f^2 N + f^3 n)$ $(N = \sum_i n_i, 20 \leq f \leq 200)$
  - Linear complexity

  - Empirically 10 sweeps until convergence

Model

- Explanation
  - For user’s trust, debugging, etc.

$$
\begin{aligned}
\hat{p}_{ui} &= y_i^T x_u \\
&= y_i^T \underbrace{(Y^T C^u Y + \lambda I)^{-1}}_{\downarrow} Y^T C^u p(u)
\end{aligned}
$$

User-specific weight
$f \times f$ matrix (W^u)

$$
= \sum_{j:r_{uj}>0} \underbrace{s_{ij}^u}_{y_i^T W^u y_j} c_{uj}
$$

Model

- Explanation
  - For user’s trust, debugging, etc.

$$\begin{aligned} \hat{p}_{ui} &= y_i^T x_u \\ &= y_i^T (Y^T C^u Y + \lambda I)^{-1} Y^T C^u p(u) \\ &= \sum_{\underbrace{j:r_{uj}>0}_{p(u)}} \overbrace{s_{ij}^u}^{\text{Weighted item-item similarity } y_i^T W^u y_j} c_{uj} \end{aligned}$$

Model

- Explanation
  - For user’s trust, debugging, etc.
  - Item-level decomposition
  - Higher scalar value indicates more contribution

$$\hat{p}_{ui} = \sum_{j:r_{uj}>0} \underset{\underset{\text{User Contrib.}}{\swarrow}}{s_{ij}^{u}} \underset{\underset{\text{Item Contrib.}}{\searrow}}{c_{uj}}$$

  - Complexity: $O(f^2 n_u + f^3)$
  - Explainable item-oriented neighborhood CF + Learnt user-specific item similarity

Experiments

Dataset

- No dataset specifically designed for implicit feedback
- TV show dataset
    - 300k set top boxes (r_{ui})
        - aggregated or fully anonymized
    - 17k unique programs
    - 4 weeks (train) + 1 week (test)
        - Shorter period deteriorates prediction
        - Longer period does not add value (+seasonality)
        - 32 million train + 2 million test (non-zero)

Dataset

- Test split
  - Prediction target
    - Remove repetitive watching
    - Remove non-strong indicators $(r_{ui} < 0.5)$
  - Momentum effect
    - Decaying interest when watching single channel for long
    - For t-th show, weight $\frac{e^{-(at-b)}}{1+e^{-(at-b)}}$
    - Empirically $a = 2, b = 6$
      - Third consecutive show halves r_{ui}
      - Fifth consecutive show reduces by 99%
  - Zero explicit feedback

Evaluation

- Percentile Ranking task
  - Focusing on loved item
    - Due to asymmetry in implicit feedback
    - Recall instead of precision
  - $rank_{ui} = 0\%$ being the most desirable
  - $\overline{rank} = \frac{\sum_{u,i} r^t_{ui} rank_{ui}}{\sum_{u,i} r^t_{ui}}$

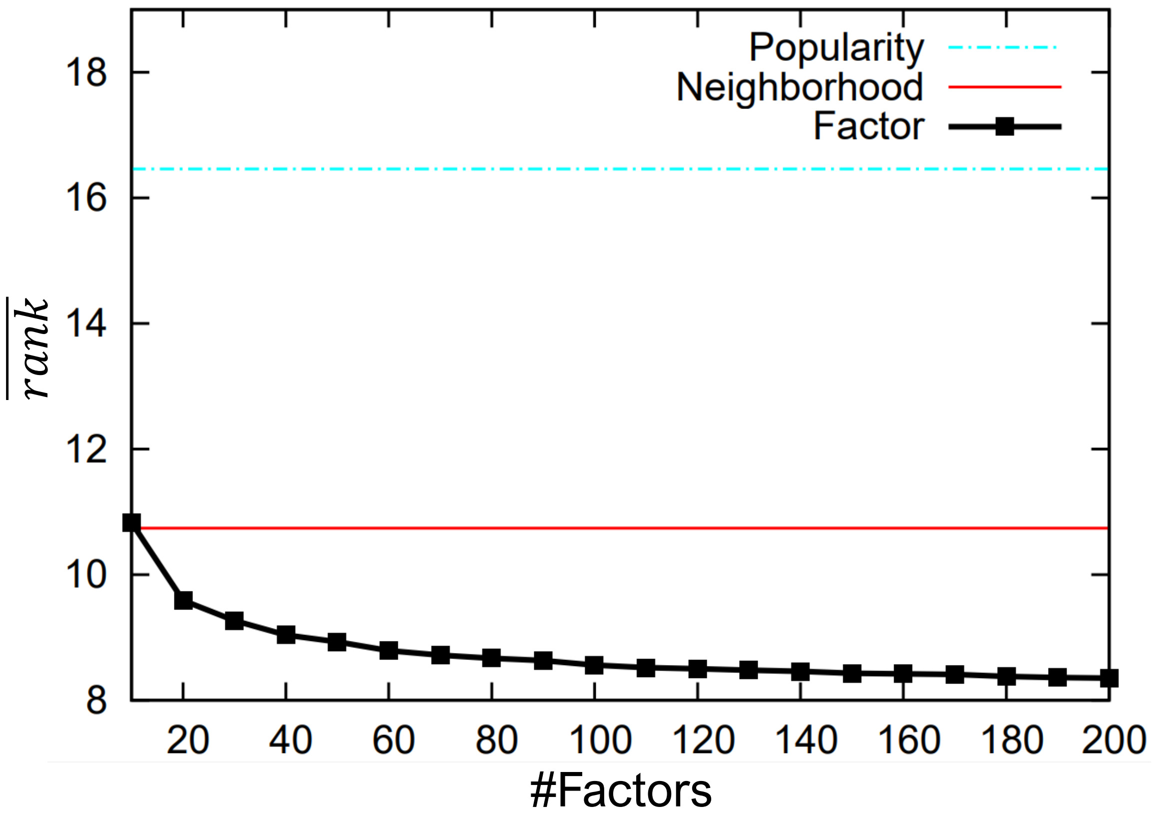
Experiments

- Baselines
  - Random prediction: $\overline{rank} = 50\%$
  - Popularity-based
  - Item-based neighborhood model
    - All items as neighbors + cosine similarity
    - cf) Explicit feedback neighborhood CF: $\hat{r}_{ui} = \frac{\sum_{j \in \mathrm{S}^k(i;u)} s_{ij} r_{uj}}{\sum_{j \in \mathrm{S}^k(i;u)} s_{ij}}$

- Ours: Factor
  - $k = 100$ if not mentioned otherwise

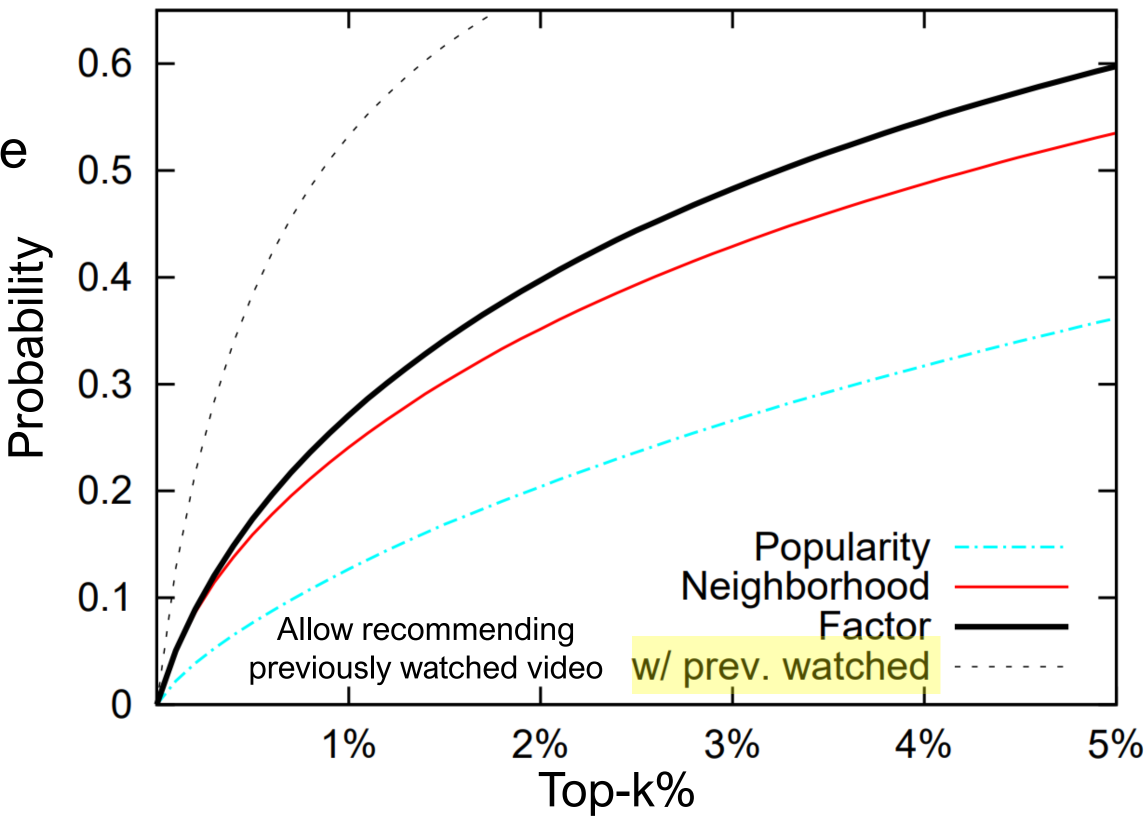
Experiments

- Result
  - Popularity (16.46%) > Neighborhood (10.74%) > Factor (8.35%)
  - Analysis
    - Popularity is a good indicator
    - More factors the better



Experiments

- Result
  - 27% of top-1% recommendations are in GT
  - Analysis
    - Factor performs slightly better
    - 50% of top-1 recommendations are in GT if recommending previously watched videos is allowed
      - Good indicator & reminder
      - …though less serendipitous



Allow recommending previously watched video

Experiments

- Ablations $(\lambda_1 = 500, \lambda_2 = 150)$
  - Direct factorization of observation r_{ui} (w/o preference, confidence)

$$\min_{x_\star, y_\star} \sum_{u,i} (r_{ui} - x_u^T y_i)^2 + \lambda_1 \left(\sum_u \|x_u\|^2 + \sum_i \|y_i\|^2 \right)$$

  - Direct factorization of preference p_{ui} (w/o confidence)

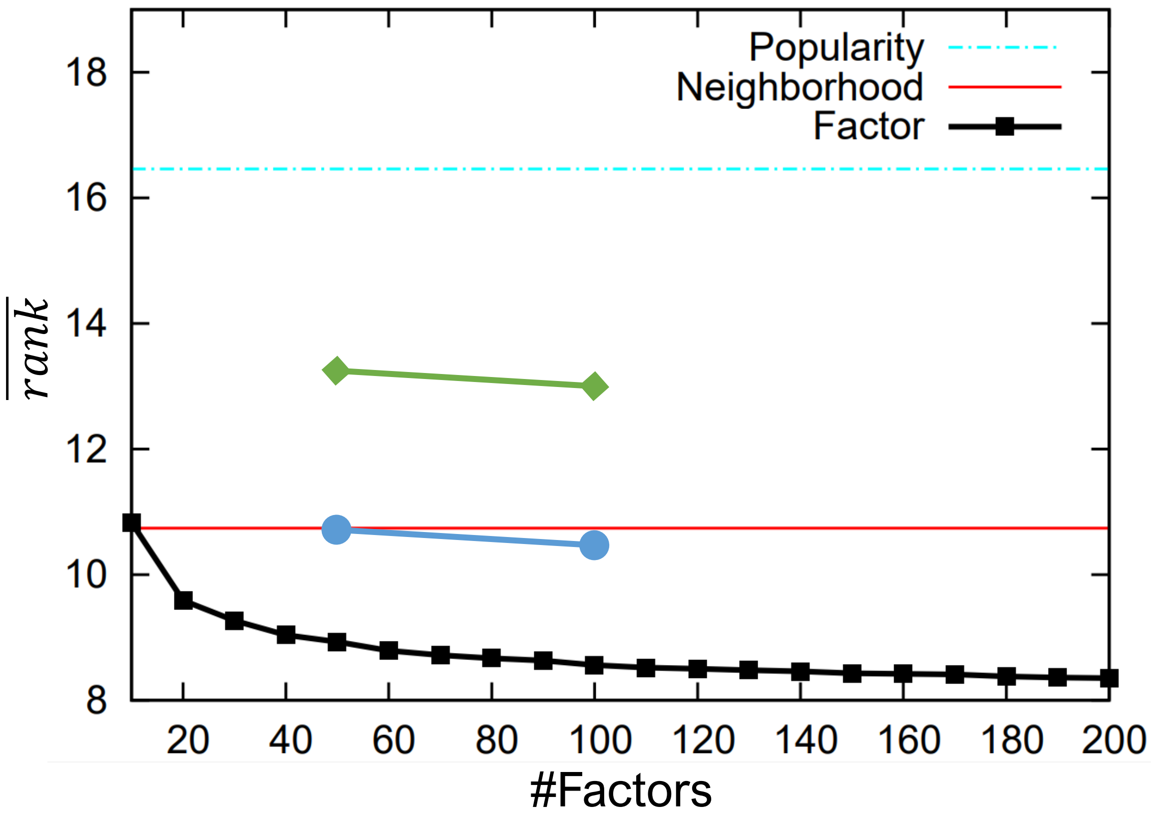
$$\min_{x_\star, y_\star} \sum_{u,i} (p_{ui} - x_u^T y_i)^2 + \lambda_2 \left(\sum_u \|x_u\|^2 + \sum_i \|y_i\|^2 \right)$$

  - cf) Proposed model

$$\min_{x_\star, y_\star} \sum_{u,i} c_{ui}(p_{ui} - x_u^T y_i)^2 + \lambda \left(\sum_u \|x_u\|^2 + \sum_i \|y_i\|^2 \right)$$

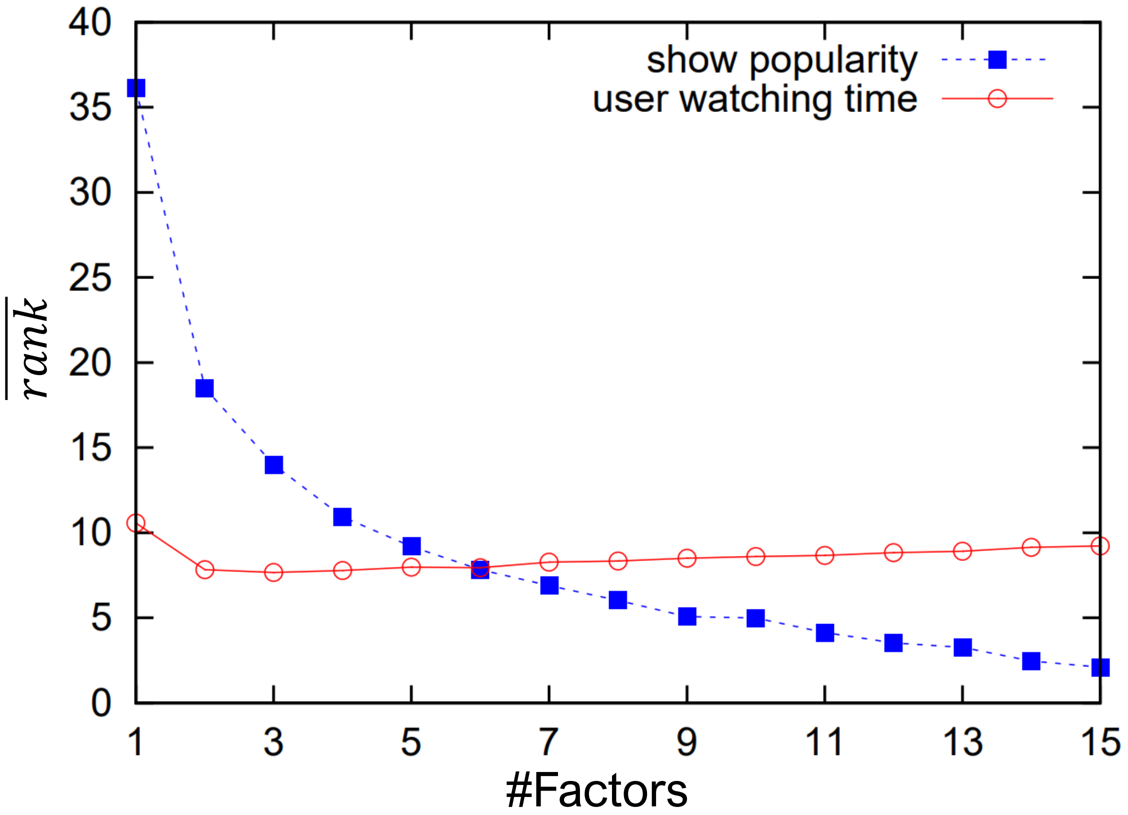
Experiments

- Result
  - Factor w/o p, c ◆—◆
    - 13.63% (50) > 13.40% (100)
  - Factor w/o c ●—●
    - 10.72% (50) > 10.49% (100)
  - Analysis
    - Naïve SVD as in explicit feedback does not work
    - Confidence as well as preference does count



Experiments

- User & item binning
  - 15 bins w.r.t. watching time
    - Bin 1 being least popular
    - Blue (item bin), Red (user bin)
  - Analysis
    - Hard to predict non-popular shows
    - "More user data, more accuracy" rule in explicit feedback does not hold
      - Heterogeneous accounts (multiple watchers on same TV)



Experiments

- Explainability
  - No previous works on explainability for factorization-based methods
  - Top-5 shows accounts for 35-40%

$$\hat{p}_{ui} = \sum_{j:r_{uj}>0} s^u_{ij} c_{uj}$$

So You Think You Can Dance	**Spider-Man**	**Life In The E.R.**
Hell's Kitchen	Batman: The Series	Adoption Stories
Access Hollywood	Superman: The Series	Deliver Me
Judge Judy	Pinky and The Brain	Baby Diaries
Moment of Truth	Power Rangers	I Lost It!
Don't Forget the Lyrics	The Legend of Tarzan	Bringing Home Baby
Total Rec = 36%	Total Rec = 40%	Total Rec = 35%

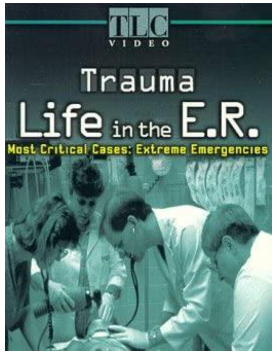
Recommendation

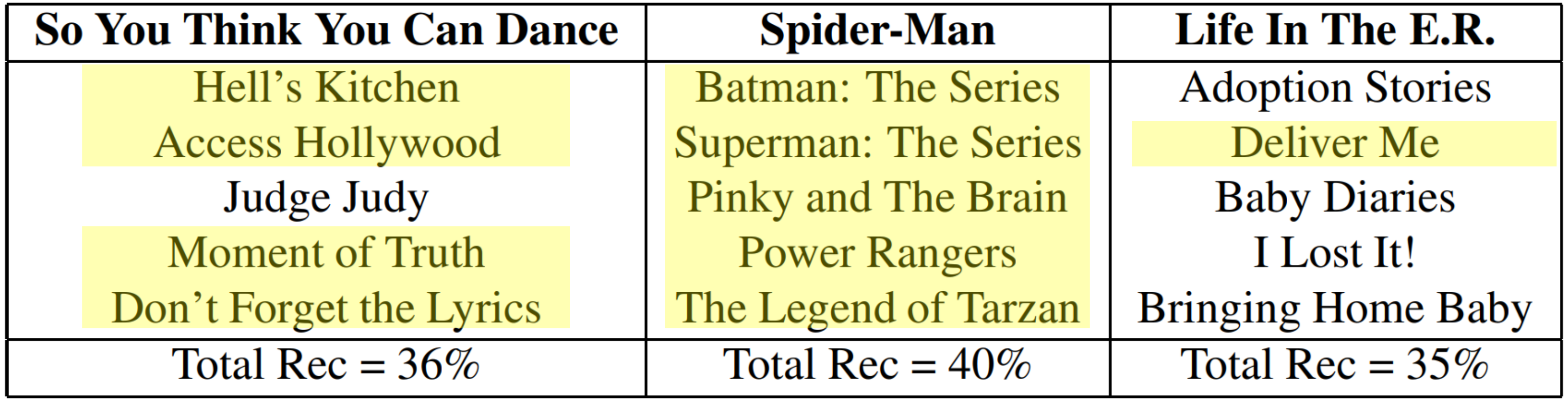
Explanation

Accountability

Experiments

- Explainability





So You Think You Can Dance	**Spider-Man**	**Life In The E.R.**	Recommendation
Hell's Kitchen	Batman: The Series	Adoption Stories	Explanation
Access Hollywood	Superman: The Series	Deliver Me	
Judge Judy	Pinky and The Brain	Baby Diaries	
Moment of Truth	Power Rangers	I Lost It!	
Don't Forget the Lyrics	The Legend of Tarzan	Bringing Home Baby	
Total Rec = 36%	Total Rec = 40%	Total Rec = 35%	Accountability

Conclusion

Conclusion

- Contribution
  - RS Framework with implicit feedback factorization
    - Decomposition of implicit observation = preference + confidence
    - Scalability via alternaternating least squares
    - Explainability
    - Novel dataset

- Extension
  - Zero preference with non-uniform confidence level
    - Varying reasons: unawareness, schedule conflict, not interested, etc.
  - Time-dependent popularity of specific genres

Conclusion

- Additional Remarks
  - (+) Implicit feedback-specific properties, modeling, optimization, explainability, dataset, etc.
  - (–) test configuration of dataset, measuring explainability (user study)

Thank You

Yifan Hu, Yehuda Koren and Chris Volinsky. Collaborative Filtering for Implicit Feedback Datasets. In ICDM 2008.

Appendix

Alternating Least Squares Proof

- $\min \sum_{u,i} c_{ui}(p_{ui} - x_u^T y_i)^2 + \lambda(\sum|x_u|^2 + \sum|y_i|^2)$
- $\frac{\Delta Loss}{\Delta x_u} = \sum_i c_{ui}\left(-2p_{ui}y_i + 2x_u y_i^T y_i\right) + 2\lambda x_u = 0$
- $-\sum_i c_{ui}\left(p_{ui} - x_u y_i^T\right)y_i + \lambda x_u = 0$
- $\sum_i c_{ui} x_u y_i^T y_i + \lambda x_u = \sum_i c_{ui} p_{ui} y_i = Y^T C^u p(u)$
    - $Y: n \times f, C^u: n \times n, p(u): n$-D
- $(\sum_i c_{ui} y_i y_i^T + \lambda I)x_u = (Y^T C^u Y + \lambda I)x_u = Y^T C^u p(u)$
- $\therefore\ x_u = (Y^T C^u Y + \lambda I)^{-1} Y^T C^u p(u)$