Using Explainability for Constrained Matrix Factorization

Behnoush Abdollahi, Olfa Nasraoui

Knowledge Discovery & Web Mining Lab, Dept. of Computer Engineering & Computer Science, University of Louisville

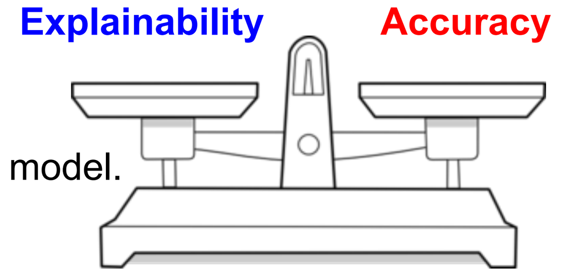
Background & Introduction

Explainable Model

- Black Box (opaque) predictors such as Deep learning and Matrix Factorization are accurate,
- …. but lack interpretability and ability to give explanations.
- White Box models such as rules and decision trees are interpretable (explainable),
- … but lack accuracy.

Why explanation?

➔ Explanations provide a rationale behind predictions,
➔ help the user gauge the validity of a prediction,
➔ may reveal prediction errors and reasons behind errors,
➔ increase trust between human and machine.



Tradeoff between Accuracy and Explainability

- Using Explanations, we can increase the transparency of the model.
- However, explainable models should also remain accurate!

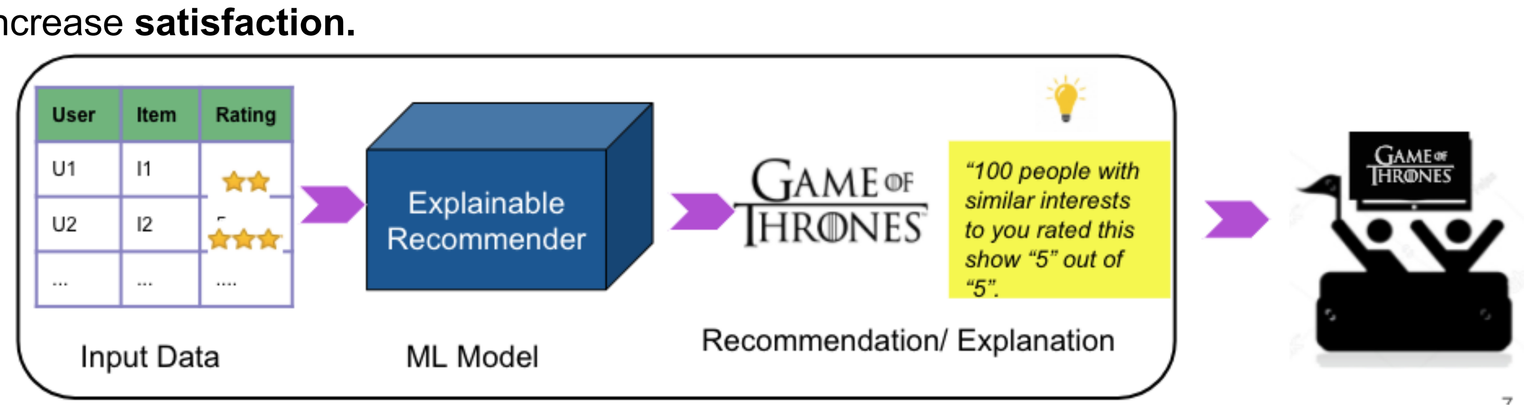
Our Focus: Explanations in Recommender Systems

Research Questions

- Can we measure/quantify explainability in recommender systems?
- Can we build accurate recommender systems that can recommend explainable items?

Why explain recommendations?

- Improve **acceptance** and **trust** by adding **justification.**
- Increase **effectiveness.**
- Increase **transparency.**
- Increase **satisfaction.**



Explainability

NSE-based Explainability

- For a user-item pair, *(u,i)*:
  - probability of item *i* having rating *k*, given the set of similar users for user *u*:

$$\mathbf{Pr}(r_{u,i} = k|N_u) = \frac{|N_u \cap U_{i,k}|}{|N_u|}$$

  - Where N_u is the set of neighbors of user *u, and* $U_{i,k}$ is the set of users who have given rating *k* to item *i.*
  - **Explainability**: expected value of the ratings given by the similar users of user *u* to the item *i*:

$$\mathbf{E}(r_{u,i}|N_u) = \sum_{k\in\kappa} k \times \mathbf{P}(r_{u,i} = k|N_u)$$

Rating NSE-based Explanation

"8 out of 10 people with similar interests to you have rated this movie 6 and higher, out of 10."

Review NSE-based Explanation

"8 out of 10 people with similar interests to you have reviewed this movie positively."

ISE-based Explainability

- For a user-item pair, *(u,i)*:
  - probability of item *i* having rating *k*, by user *u* given the set of similar items for item *i:*

$$\mathbf{Pr}(r_{u,i} = k|N_i) = \frac{|N_i \cap I_{u,k}|}{|N_i|}$$

  - Where N_i is the set of neighbors of item *i*, and $I_{u,k}$ is the set of items that user *u* has given rating *k* to.
  - **Explainability**: expected value of the ratings given by user *u* to similar items to item *i*:

$$\mathbf{E}(r_{u,i}|N_i) = \sum_{k\in\kappa} k \times \mathbf{P}(r_{u,i} = k|N_i)$$

ISE-based Explanation

Our recommendation is "Pulp Fiction", because you rated similar movies:

Movie	Your Rating
From Dusk Till Dawn (1996)	2
Seven (Se7en) (1995)	4
Usual Suspects The (1995)	4

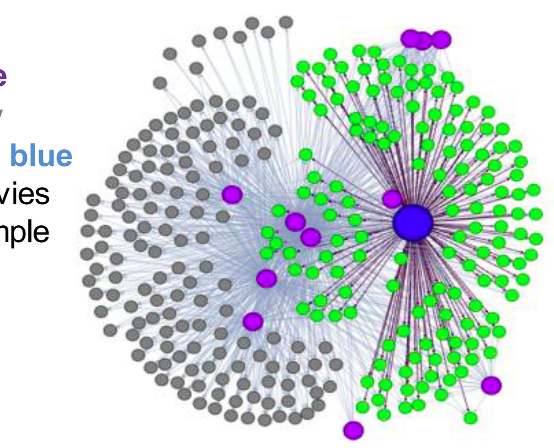
Explainable Matrix Factorization

Explainability Graph

- Explainability Matrix, *W*, between user-item pairs in the Explainability Graph:

$$W_{u,i} = \begin{cases} Expl_{u,i} & if\ Expl_{u,i} \geq \theta \\ 0 & otherwise \end{cases}$$

- users ⇒ purple
- movies ⇒ gray
- sample user ⇒ blue
- explainable movies for the blue sample user⇒ green



Explainable Matrix Factorization

- Extension of Matrix Factorization (MF).
- Add Soft Explainability Constraints to bring users closer to their explainable items in Latent Space!

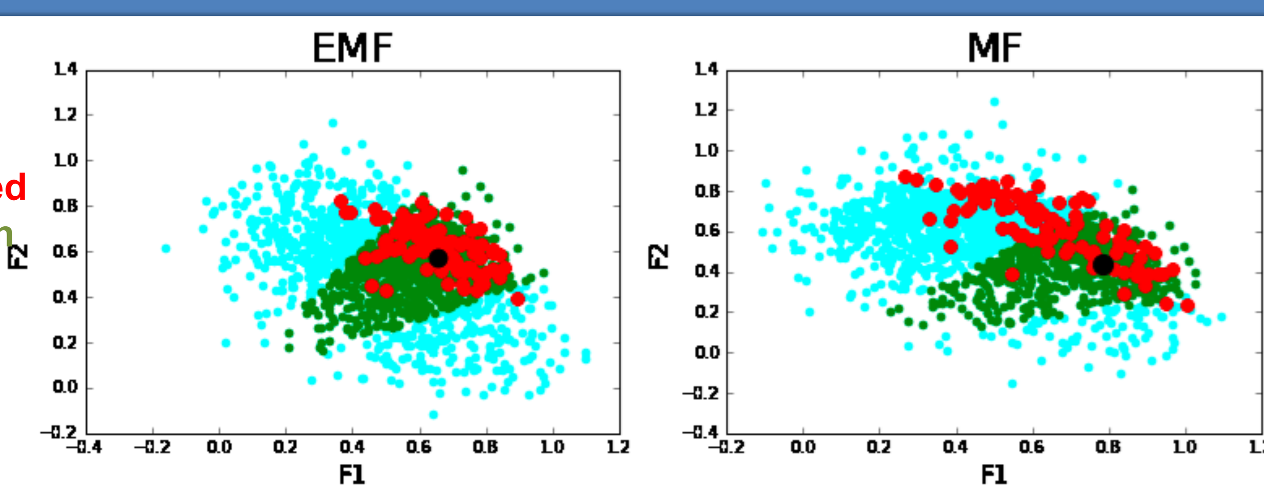
$$J = \underbrace{\sum_{u,i\in R}(r_{u,i} - p_u q_i^T)^2 + \frac{\beta}{2}(\| p_u \|^2 + \| q_i \|^2)}_{\text{Rating-based optimization (MF objective function)}} + \underbrace{\frac{\lambda}{2} \| p_u - q_i \|^2 W_{u,i}}_{\text{Explainability soft constraint}}$$

- EMF Update Rules:

$$p_u^{(t+1)} \leftarrow p_u^{(t)} + \alpha(2(r_{u,i} - p_u q_i^T)q_i - \beta p_u - \lambda(p_u - q_i)W_{u,i})$$
$$q_i^{(t+1)} \leftarrow q_i^{(t)} + \alpha(2(r_{u,i} - p_u q_i^T)p_u - \beta q_i + \lambda(p_u - q_i)W_{u,i})$$

Explainability Effect in Latent Space

- sample user ⇒ black
- explainable items ⇒ red
- relevant items ⇒ green
- other items ⇒ cyan



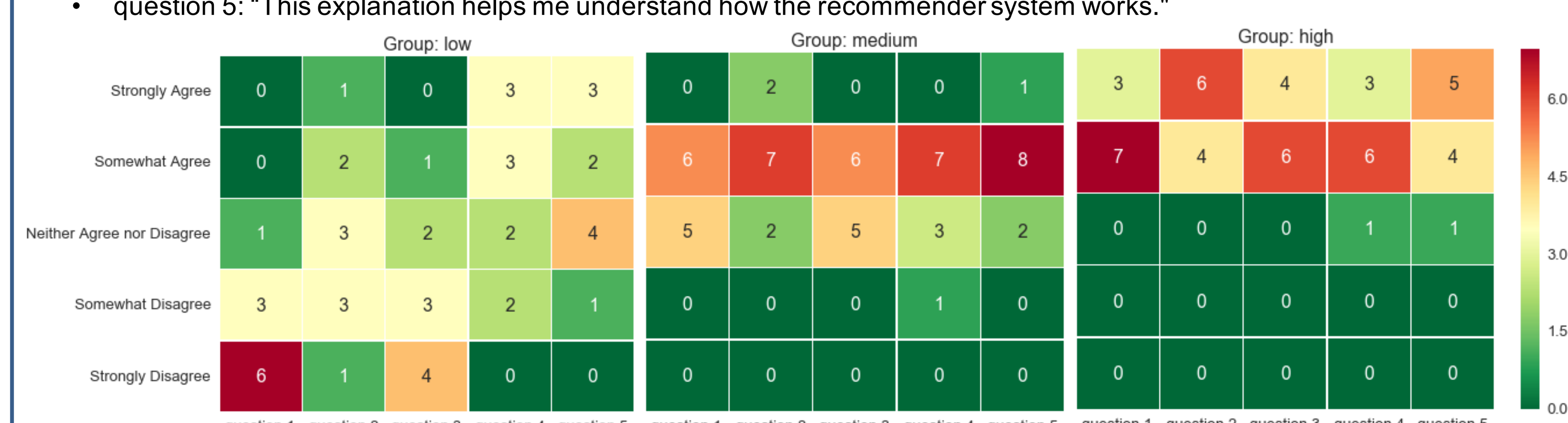
User Study

Does the explainability value of the explanation have an impact on user satisfaction?

- 3 groups:
  - **low**: explainability value < 2.
  - **medium**: 2 <= explainability value between < 4.
  - **high**: explainability value >= 4.

p-value	High	Low
Low	$4.0e-11$	-
Medium	0.016	$2.2e-08$

- Likert scale survey questions:
  - question 1: "Based on the ratings of people with similar interest to mine, this is a good recommendation."
  - question 2: "This explanation helps me understand why this movie was recommended."
  - question 3: "Based on the ratings of people with similar interests to mine, I will watch this movie."
  - question 4: "Based on the ratings of people with similar interests to mine, I can determine how well I will like this movie."
  - question 5: "This explanation helps me understand how the recommender system works."



- Results showed that **there was <u>significant difference between the explainability in the three groups.</u>**

Experimental Results

Data

- MovieLens ratings data which consists of 100,000 ratings, on a scale of 1 to 5, for 1700 movies and 1000 users.
- 10% of the latest ratings from each user are selected for the test set and the remaining 90% of the ratings are used in the training set.

Baseline methods

- Standard latent factor model based on Matrix Factorization(MF).
- Probabilistic Matrix Factorization(PMF).
- Hybrid technique - Content boosted Collaborative filtering.
- User-based top-n CF.
- Item-based top-n CF.

Accuracy Metrics

- Mean Average Precision.
- Area Under Curve (AUC): area under the true positive rate against the fallout (false positive rate) plot.
- Mean Explainability Precision: $\frac{|\{i : i \in top-n, Expl_{u,i} > \theta\}|}{|top-n|}$
- Mean Explainability Recall: $\frac{|\{i : i \in top-n, Expl_{u,i} > \theta\}|}{|Expl_{u,i} > \theta|}$

MAP@50

f	*UB*	*IB*	*PMF*	*MF*	EMF_{UB}	EMF_{IB}
5	0.009	0.0064	0.0113	0.0149*	0.0108	0.011
10	0.009	0.0064	0.0108	0.0145	0.0157*	0.0112
20	0.009	0.0064	0.0116	0.0143	0.0146*	0.0118
50	0.009	0.0064	0.0126	0.015	0.0165*	0.0138
$\|N_.\|$	*UB*	*IB*	*PMF*	*MF*	EMF_{UB}	EMF_{IB}
5	0.009	0.0065	0.0108	0.0145*	0.0102	0.0138
10	0.0087	0.0064	0.0108	0.0145	0.0101	0.0197*
20	0.0085	0.0071	0.0108	0.0145	0.009	0.0272*
50	0.0081	0.0077	0.0108	0.0145	0.0105	0.0328*

top: MAP vs #factors, @ 50 neighbors
bottom: MAP vs #neighbors, @ 10 factors

AUC

f	*UB*	*IB*	*PMF*	*MF*	EMF_{UB}	EMF_{IB}
5	0.4988	0.4982	0.5743	0.7129*	0.5616	0.5745
10	0.4988	0.4982	0.5629	0.7033	0.7115*	0.5791
20	0.4988	0.4982	0.563	0.6843	0.6873*	0.5791
50	0.4988	0.4982	0.54	0.5697	0.5984*	0.5019
$\|N_.\|$	*UB*	*IB*	*PMF*	*MF*	EMF_{UB}	EMF_{IB}
5	0.4759	0.4711	0.563	0.7011	0.7131	0.7707*
10	0.4851	0.4835	0.563	0.7011	0.6787	0.7821*
20	0.489	0.4826	0.563	0.7011	0.6522	0.7872*
50	0.4905	0.4991	0.563	0.7011	0.6855	0.7463*

top: AUC vs #factors, @ 50 neighbors
bottom: AUC vs #neighbors, @ 10 factors

MEP@50

f	*UB*	*IB*	*PMF*	*MF*	EMF_{UB}	EMF_{IB}
5	0.449	0.551	0.6284	0.7079	0.7080	0.7090*
10	0.449	0.551	0.5412	0.7085	0.7089*	0.7187
20	0.449	0.551	0.3617	0.7187	0.7224	0.7242*
50	0.449	0.551	0.0843	0.5502	0.5845*	0.4011
$\|N_.\|$	*UB*	*IB*	*PMF*	*MF*	EMF_{UB}	EMF_{IB}
5	0.4831	0.5895	0.5412	0.708	0.7081*	0.708
10	0.4489	0.5516	0.5412	0.708	0.7083	0.7099*
20	0.4195	0.5423	0.5412	0.708	0.7082	0.7087*
50	0.4124	0.5416	0.5412	0.708	0.7083	0.7096*

top: MEP vs #factors, @ 50 neighbors
bottom: MEP vs #neighbors, @ 10 factors

MER@50

f	*UB*	*IB*	*PMF*	*MF*	EMF_{UB}	EMF_{IB}
5	0.054	0.07	0.0706	0.0756	0.0757*	0.073
10	0.054	0.07	0.0622	0.0757	0.0758*	0.0748
20	0.054	0.07	0.0399	0.0778	0.0785*	0.0755
50	0.054	0.07	0.0085	0.0564	0.0569*	0.0362
$\|N_.\|$	*UB*	*IB*	*PMF*	*MF*	EMF_{UB}	EMF_{IB}
5	0.0583	0.0708	0.062	0.075	0.0756*	0.0729
10	0.0534	0.0701	0.062	0.075	0.0757*	0.0732
20	0.0496	0.0668	0.062	0.075	0.0756*	0.073
50	0.0485	0.0652	0.062	0.075	0.0757*	0.0731

top: MER vs #factors, @ 50 neighbors
bottom: MER vs #neighbors, @ 10 factors

Conclusion and Future Directions

- We proposed a probabilistic formulation for measuring **explainability** for recommendations.
- We proposed an **Explainable-Matrix Factorization** (EMF) model for providing explainable recommendations that are accurate.
- We proposed **offline metrics to evaluate the explainability** of recommender systems.
- Improved Explainability without significant sacrifice in Accuracy.

Why is Explainability so Important?

- We are relying on Machine learning algorithms in critical activities:
  - Credit Scoring, Criminal investigation, justice, Healthcare, education, insurance risk modeling, etc.
- Real life data can include biases that will affect the predictions.
  - May result in unfair models (discriminative, unreasonable, opaque ..)
- Transparency is crucial to avoid or at least scrutinize biased predictions and to have more trust in ML models!

Future Directions

- Utilize different domains of data.
- Incorporate other explanation generation techniques.
- Apply EMF to other machine learning areas.

Acknowledgement

- This research was partially supported by KSEF Award KSEF-3113-RDE-017