New Explainable Active Learning Approach for Recommender Systems

Sami KHENISSI, Behnoush Abdollahi, Wenlong Sun, Pegah Sagheb, Olfa Nasraoui

Knowledge Discovery and Web Mining Lab – University of Louisville

Introduction and Motivations

- Recommender Systems are intelligent programs that analyze patterns between items and users to predict the user's taste.

Problems for Recommender Systems

Lack of interpretability for accurate models

Missing Data and sparse input

Explainable Matrix Factorization

Idea: Combine both solutions into a unified approach

Active Learning: Select most useful data

Objective

- Design an efficient Active Learning Strategy to increase the **explainability** and the **accuracy** of an "Explainable Matrix Factorization" model.

Matrix Factorization

Input: Rating matrix

	item v	
user u	r_{uv}	

Rating from user u to item v

Idea: Learn p and q to predict all values of the rating matrix

- *p* and *q* are the representation of the user *u* and item *v* in a latent space.

$$r_{uv} = q_v^T * p_u$$

Learning process: $\min_{P,Q} \sum_{(u,v)\in R} (r_{uv} - q_v^T p_u)^2 + \lambda(\|q_v^2\| + \|p_u^2\|)$

Main Problem: Matrix Factorization is a **Black Box Model**

Explainable Matrix Factorization

Idea: Provide neighborhood style explanations with recommendations

Recommendation:

Justification:

80% of users who have similar interests to you have liked this video

New objective function:

$$J = \sum_{(u,v)\in R} (r_{uv} - q_v^T p_u)^2 + \frac{\beta}{2}(\|p_u^2\| + \|q_v^2\|) + \boxed{\frac{\lambda}{2}(p_u - q_v)^2 W_{uv}}$$

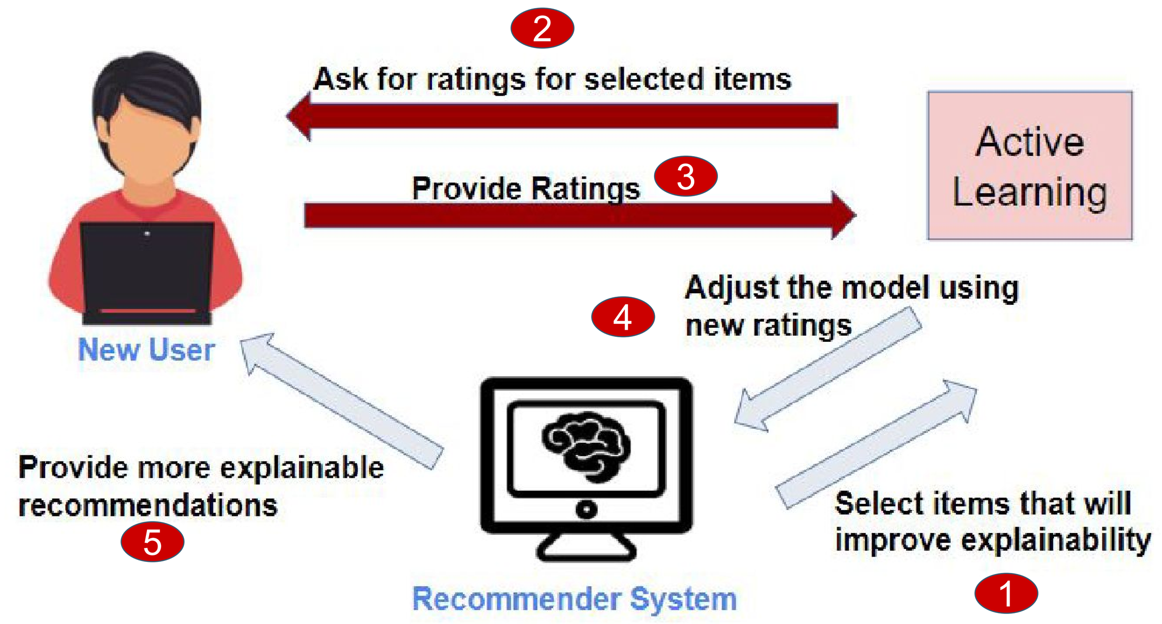
Explainability term to favor users and items with similar p and q

- W_{uv} is the explainability score calculated for the user *u* and item *v*.

$$W_{uv} \begin{cases} \frac{|N'(u)|}{|N'_k(u)|} \; if \; \frac{|N'(u)|}{|N'_k(u)|} > \theta; \\ 0 \quad Otherwise; \end{cases}$$

N' is the total number of neighbors who rated item *v* and N'_k is the total number of neighbors

ExAL algorithm



1. Select items from an unlabeled pool set of items using a **selection strategy**
2. Get the true ratings of the selected item from the new user
3. Adjust the parameters of the model using the new ratings
4. Repeat process until meeting a stopping criterion

ExAL-min Strategy

- Select items using the following criterion:

$$\boxed{i_u^*} \simeq \underset{i\in I^u_{pool}}{\operatorname{argmin}} \sum_{j\in I^u_{test}} \left| \boxed{1 - r_{uj} + 2\alpha((r_{ui} - \bar{R}_i)\sum_{f=1}^{k} q_{if}q_{jf}} + \boxed{\lambda W_{ui}(r_{uj} - \sum_{f=1}^{k} q_{if}q_{jf}))} \right|$$

Selected item

Predicted change in the Estimated test error

Regularization term to take into consideration the explainability

- The strategy picks the item that is expected to have the least estimated Mean Absolute Error with taking into consideration explainability

Pros	Good explainability increase
Cons	Slow accuracy increase

ExAL-max Strategy

- Select items using the following criterion:

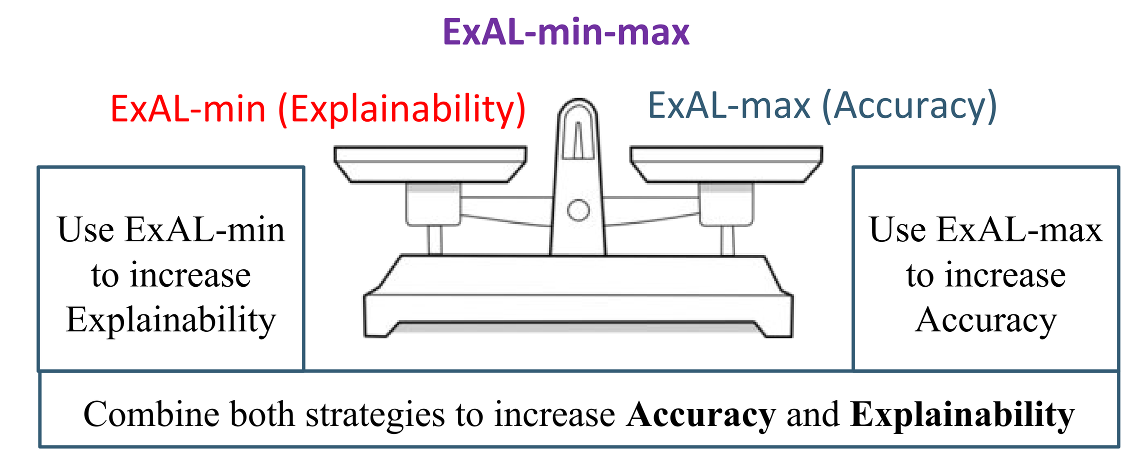
$$i_u^* \simeq \boxed{\underset{i\in I^u_{pool}}{\operatorname{argmax}}} \sum_{j\in I^u_{test}} \left| 1 - r_{uj} + 2\alpha((r_{ui} - \bar{R}_i)\sum_{f=1}^{k} q_{if}q_{jf} + \lambda W_{ui}(r_{uj} - \sum_{f=1}^{k} q_{if}q_{jf})) \right|$$

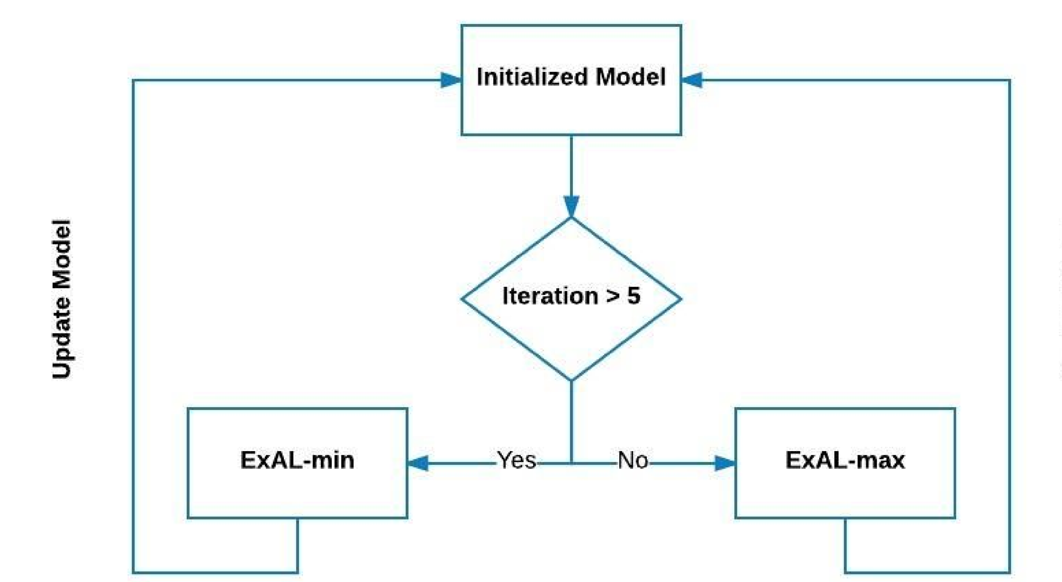
Learn the true rating of items that are expected to increase the test error

The newly selected items will provide new information to help ameliorate the predictions of the model

Pros	Good accuracy increase
Cons	Low Explainability improvement

ExAL-min-max Strategy





ExAL-min-max can control the tradeoff Explainability and Accuracy

Experimental Results

Data set: Movie-lens

- Users: 943
- items: 1680

Splitting strategy

- 343 test users (new users)
- 600 train users

Explainability Evaluation metrics:

Explainability Precision: $xP = \frac{1}{|U_{test}|}\sum_{u\in U_{test}} \frac{|N^u_{rec} \cap N^u_{exp}|}{|N^u_{rec}|}$

Explainability Recall: $xR = \frac{1}{|U_{test}|}\sum_{u\in U_{test}} \frac{|N^u_{rec} \cap N^u_{exp}|}{|N^u_{exp}|}$

Explainability F-score: $xF - score = 2\frac{xP \times xR}{xP + xR}$

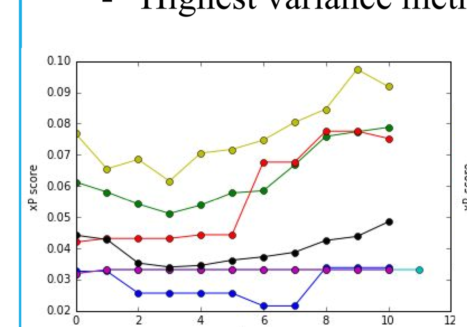
- N^u_{rec} is the number of recommended items to the user u, and N^u_{exp} is the number of recommended and explainable items to the user u.

Accuracy Evaluation metrics:

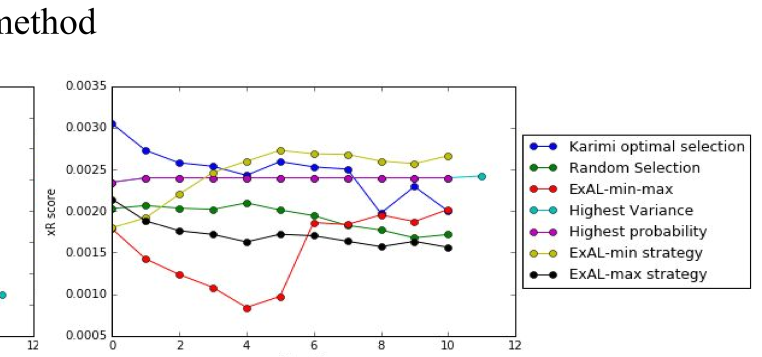
Mean Absolute Error $MAE_u = \frac{1}{|I^u_{test}|}\sum_{j\in I^u_{test}} |r^*_{uj} - r_{uj}|$

Baseline Methods

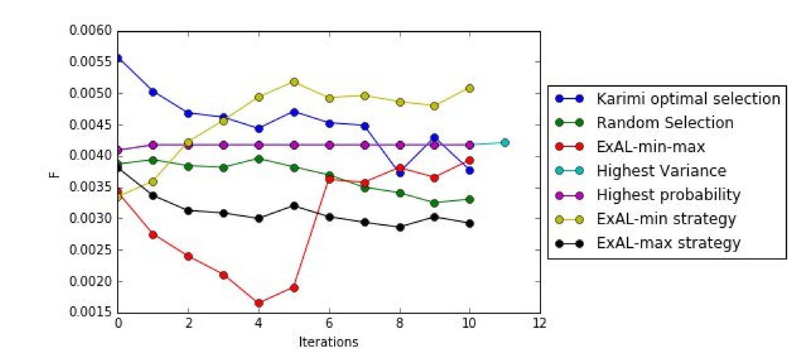
- Random selection
- Karimi's Optimal Selection
- Highest rating probability
- Highest variance method



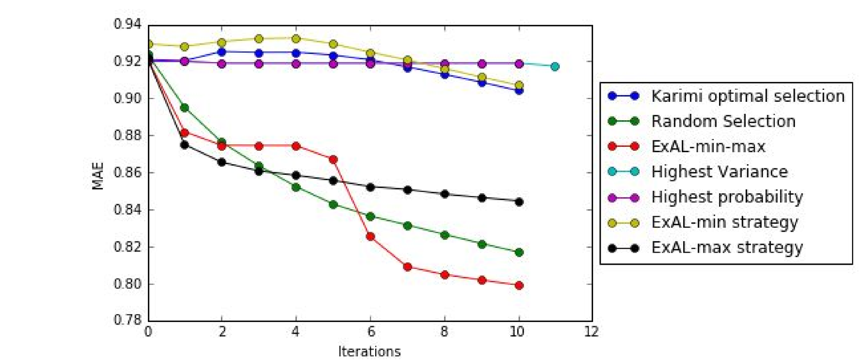
Explainability Precision



Explainability Recall



Explainability F-score



Mean Absolute Error

Method	Explainability	Accuracy
ExAL-min	Optimal Explainability increase	Very slow Accuracy increase
ExAL-max	Poor Explainability increase	Effective Accuracy increase
ExAL-min-max	Boost the Explainability starting from the 5th iteration (when we switch from ExAL-max to ExAL-min)	Optimal Accuracy increase especially after the 5th iteration

Conclusion

- **ExAL-min**: chooses items that are predicted to reduce the test error, takes into account an explainability term to favor the interpretability of the model.
  - **Optimal method for increasing explainability**.
- **ExAL-max**, choose items with the highest predicted test error.
  - Helps to increase the accuracy at a faster pace
- **ExAL-min-max**: Mixed strategy that uses both ExAL-max and ExAL-min
  - Can control the tradeoff between Accuracy and Explainability

Future Work

- Apply our strategy to other black box algorithms (Deep Learning)
- Generalize it to other machine learning areas (supervised learning)

References

(1) Abdollahi, B. and Nasraoui, O. (2016). Explainable matrix factorization for collaborative filtering. In Proceedings of the 25th International Conference Companion on World Wide Web

(2) Karimi, R., Freudenthaler, C., Nanopoulos, A., and Schmidt-Thieme, L. (2011b). Towards optimal active learning for matrix factorization in recommender systems. In Proceedings of the 2011 IEEE 23rd International Conference on Tools with Artificial Intelligence

Acknowledgements

This work was partially funded by KSEF Award KSEF-3113-RDE-017 and NSF INSPIRE (IIS)- Grant #1549981