Predicting Familiarity in Recommender Systems

Group 7
Tavvi Taijala, Haiwei Ma, Wei Chen

Problem

RQ: How does familiarity impact the recommender system experience?

- Do users prefer to be recommended more familiar items?
- Do users prefer to be recommended less familiar items?
- Does this preference vary between users?
- Does this preference vary over time?

To answer this question, we need a way to manipulate familiarity experimentally.

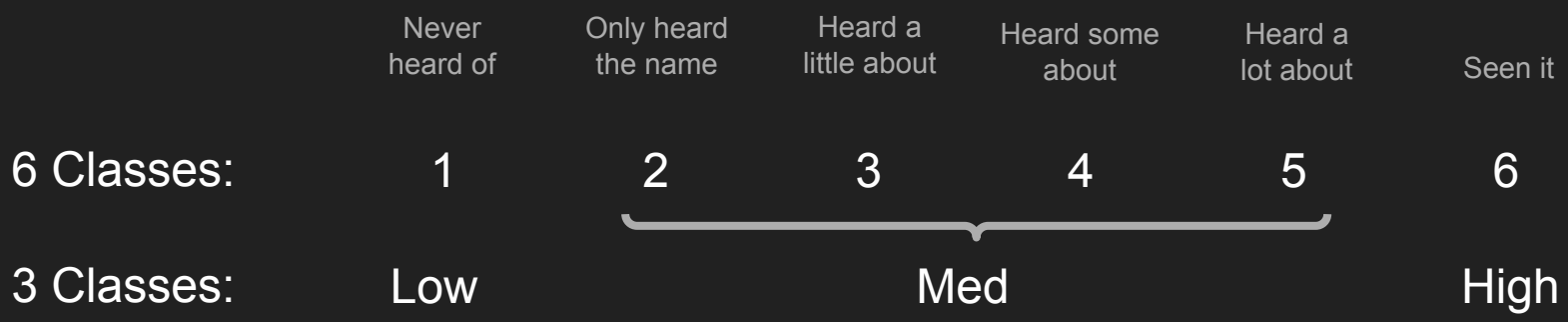
Our approach was to build a model that could predict familiarity (accurately).

Dataset

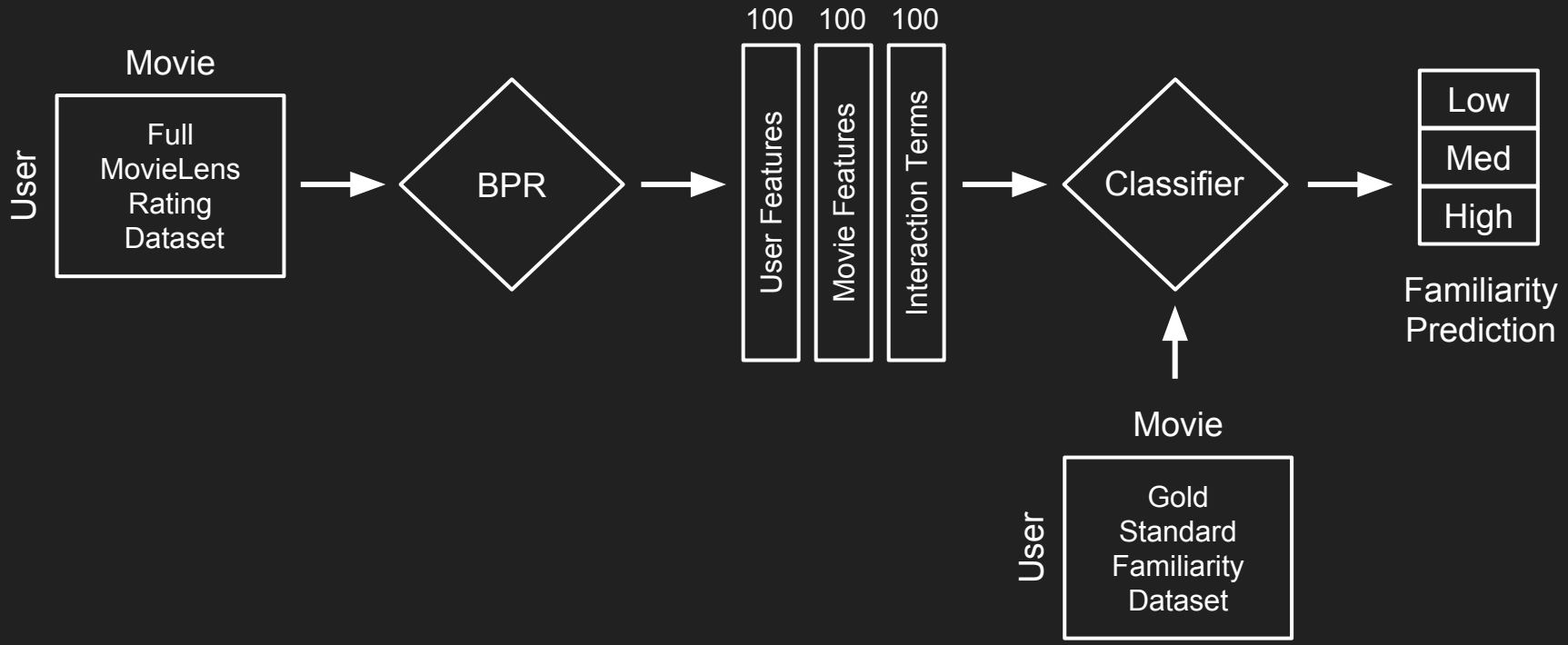
Collected a gold standard familiarity dataset on MovieLens users.

- 1,000 MovieLens users
- 40,000 (user, movie, familiarity) data points

Familiarity was measured on a 6-point ordinal scale, condensed to a 3-point scale.



Approach


Movie
User
Full MovieLens Rating Dataset
BPR
100
100
100
User Features
Movie Features
Interaction Terms
Classifier
Low
Med
High
Familiarity Prediction
Movie
User
Gold Standard Familiarity Dataset

Models and Algorithms

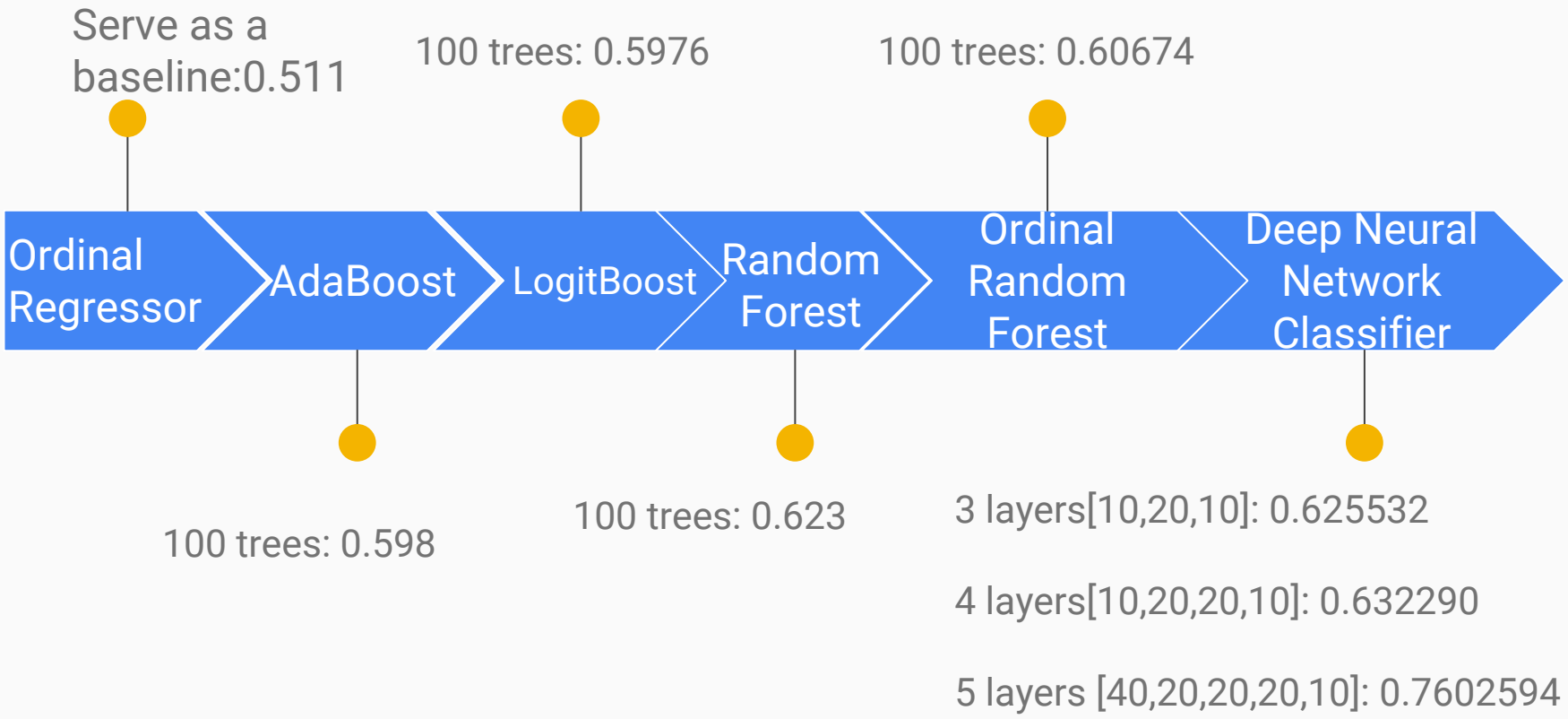
Matrix Factorization Models

- BPR
- AdaptiveBPR
- WRMF
- FunkSVD

Classification Models

- OrdRec
- AdaBoost
- LogitBoost
- RandomForest
- Ordinal Random Forest
- Deep Neural Networks

K-folder Cross Validation Results



Random Forest Vs Ordinal Random Forest

- Accuray
  - Non-ordinal random forest: 0.61454
  - Ordinal random forest: 0.60674
- Confusion matrix
  - Non-ordinal random forest:

predict/true	1	2	3
1	13795	4848	2581
2	1510	4425	2298
3	1055	3107	6331

Ordinal random forest:

predict/true	1	2	3
1	13789	4947	2473
2	1762	3912	2199
3	809	3521	6538

Experimental Setup

Evaluation Metrics

- Accuracy
- Per-class accuracy
- Confusion matrices

Crossfold Validation Approaches

- K folds
- K stratified folds
- K out of sample folds

Out of Sample Validation Results

Algorithm	Parameters	Mean (Std Dev) Accuracy
Baseline	Popularity	47.65% (1.36%)
OrdRec		54.97% (1.92%)
Random Forest	300 Trees	53.03% (0.88%)
AdaBoost	300 Trees	53.53% (1.51%)
LogitBoost	300 Trees	56.55% (1.10%)
Neural Network	4 Layers	57.35% (2.22%)
Neural Network	5 Layers	57.92% (1.07%)
Neural Network	6 Layers	57.34% (2.40%)

Next Steps

Neural Networks

- Pretraining
- Ordinal Classification

Synthetic Training Data

Thanks!