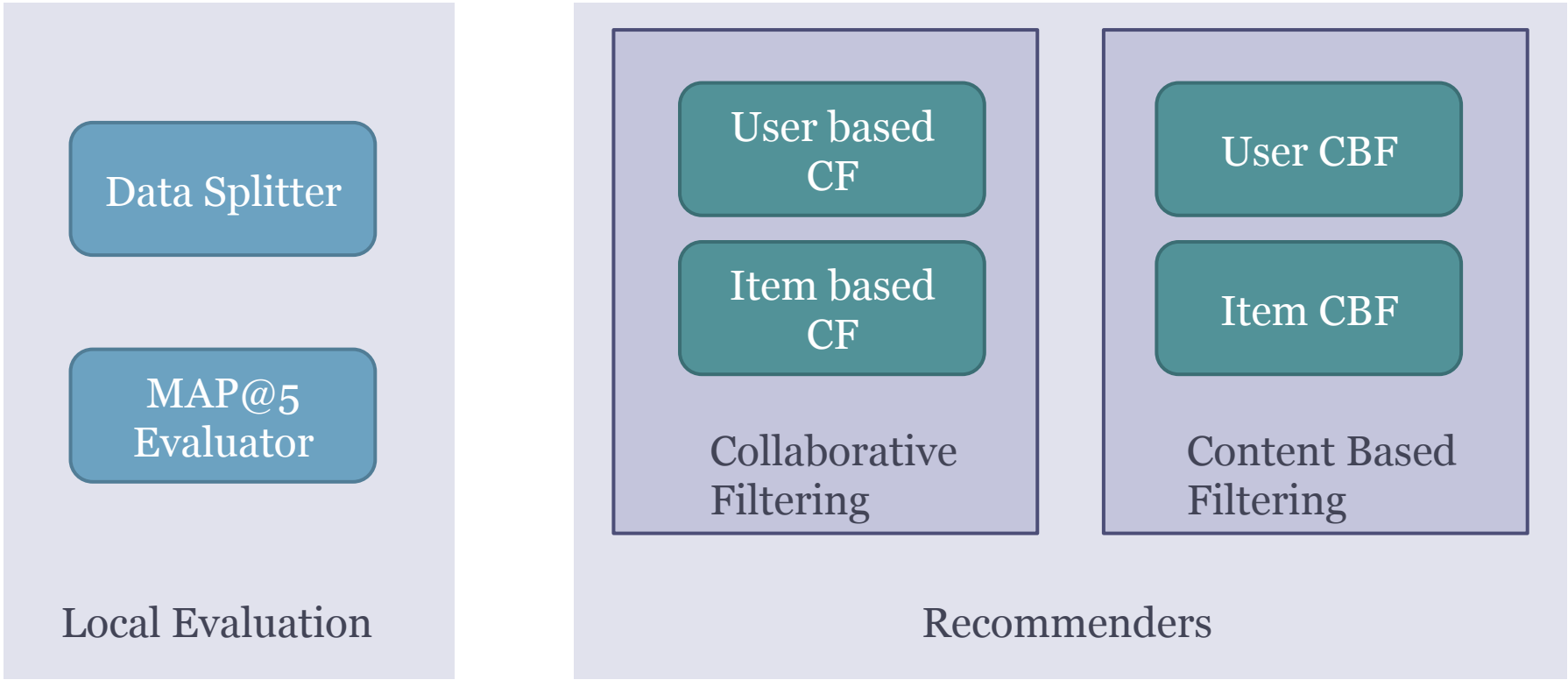
POLIMI RecSys Challenge
Team NOONE

KANG Shuwen

Feb 13, 2018

System Architecture

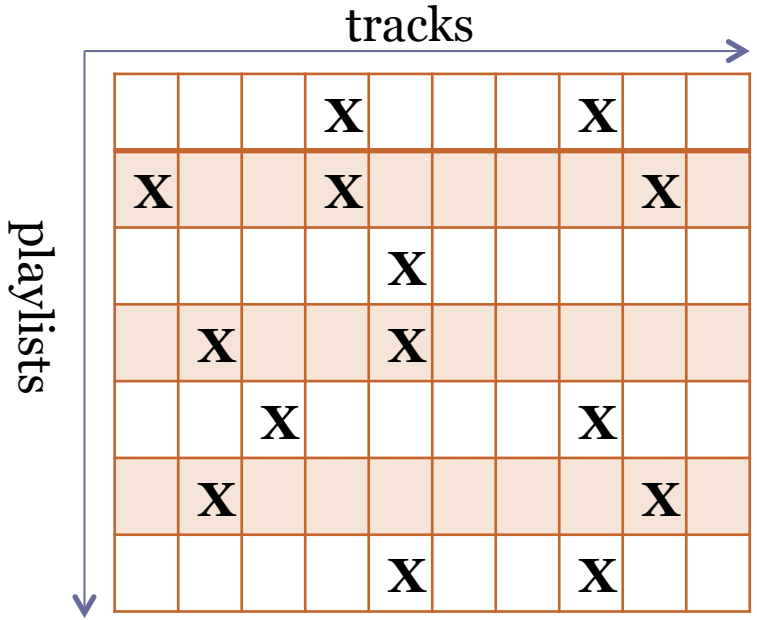
Playlist -> User
Track -> Item



Data Preprocessing

Playlist -> User
Track -> Item

From “train_final.csv” to user rating matrix (**URM**)

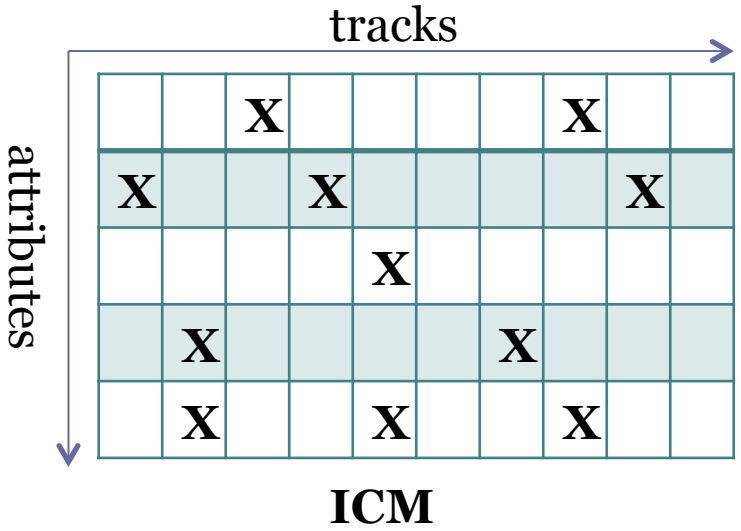


Data Preprocessing

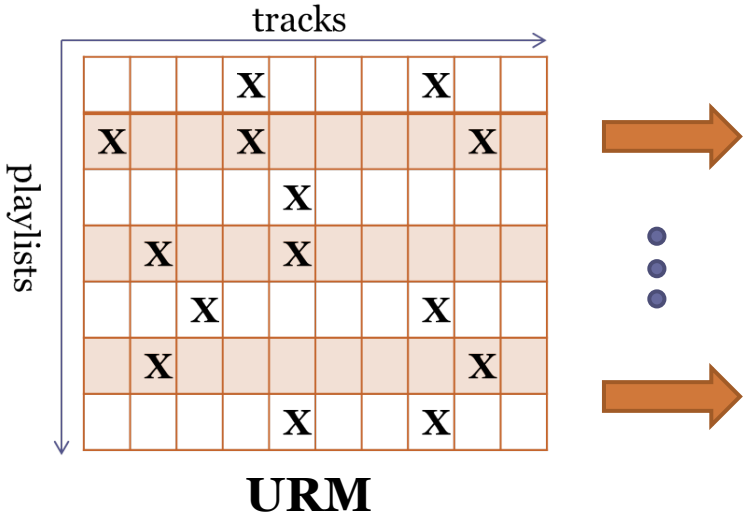
Playlist -> User
Track -> Item

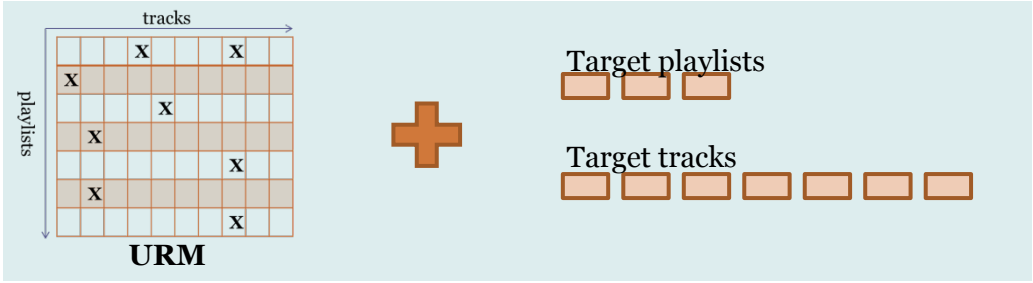
From “tracks_final.csv” to item content matrix (**ICM**)

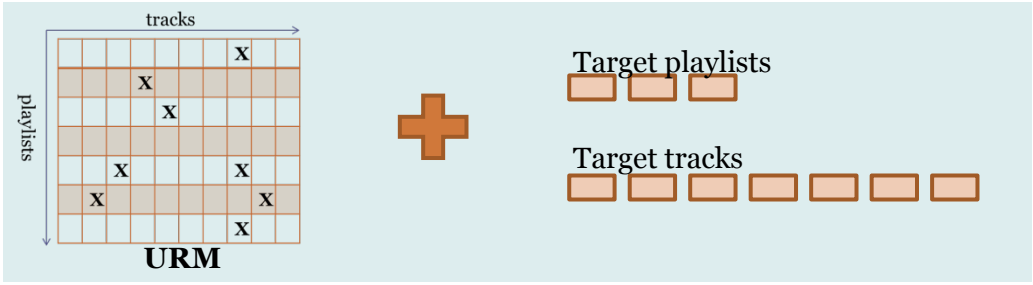
Similarly , from “playlists_final.csv” to user content matrix(**UCM**)



Data Preprocessing







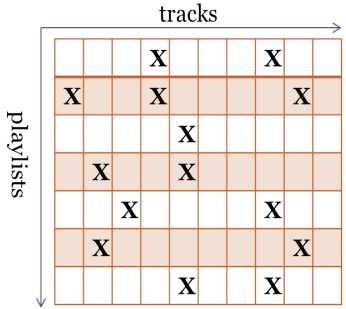
Algorithms

Playlist -> User
Track -> Item

User based CF

Item based CF

Collaborative Filtering



URM

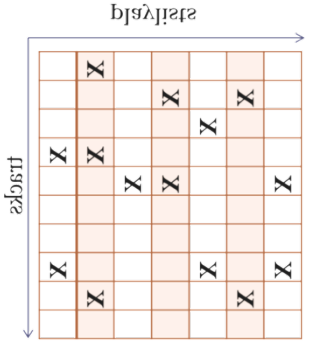
Cosine similarity

Shrinkage

User Similarity Matrix

K nearest neighbor

Su



URM_t

Cosine similarity

Shrinkage

Item Similarity Matrix

K nearest neighbor

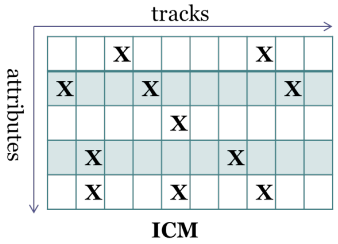
Si

Algorithms

User CBF

Item CBF

Content Based Filtering



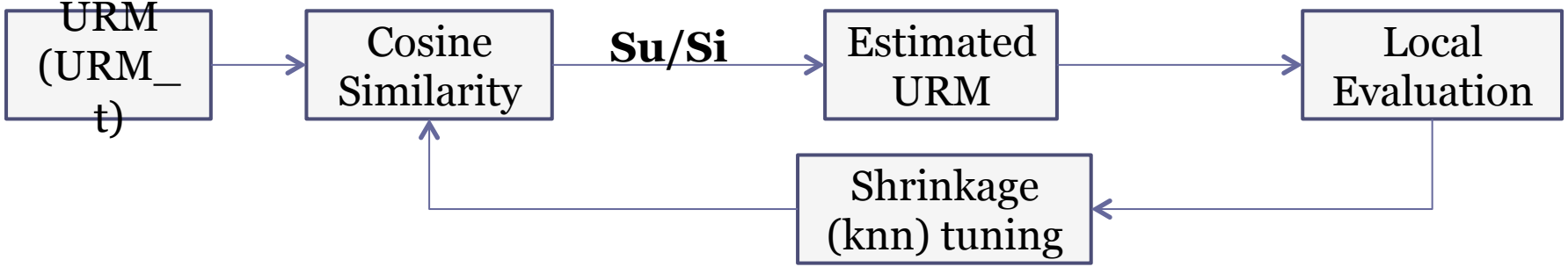
Playlist -> User
Track -> Item

UCM_t → Cosine similarity / Shrinkage → User Similarity Matrix → K nearest neighbor → **Su**

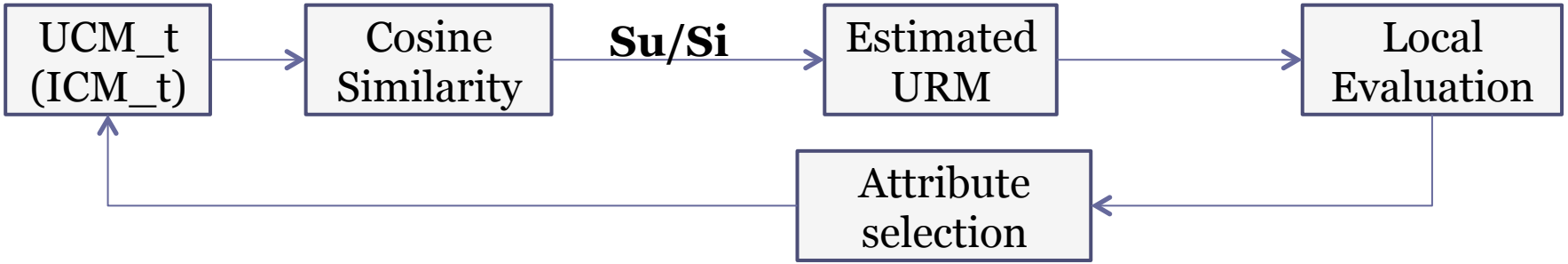
ICM_t → Cosine similarity / Shrinkage → Item Similarity Matrix → K nearest neighbor → **Si**

Parameter Tuning

Collaborative Filtering: shrinkage and knn tuning

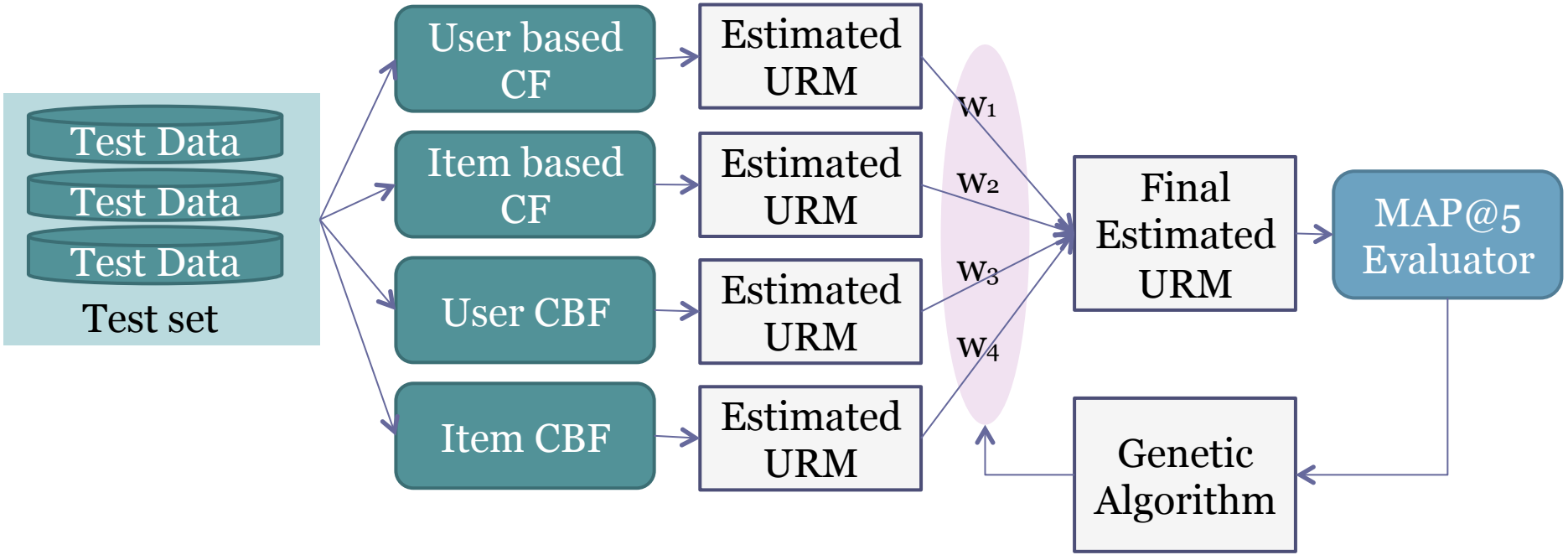


Content Based Filtering: attribute selection



Parameter Tuning

Hybrid System: weight training with *Genetic Algorithm*



Useful Tricks

- Partial UCM and URM_t combination : consider owner, title and tracks together as playlist attributes
- Partial ICM : only artist_id and album are considered as track attributes
- TF-IDF for tags and title
- Shrinkage and K nearest neighbor for similarity matrix

Thank you for your attention!