### Recommender Systems

**Purpose**
- Survey state of the art Matrix Factorization solutions to the Cold Start problem.
- Suggest avenues of further research.

**User-Item Utility Matrix**
Ratings for m users and n items in \( m \times n \) matrix, \( R \).
\[ R_{u,i} = u \text{’s rating of item } i \]

### Cold Start Problem

How to give ratings for new users and new items for which no ratings are recorded in the dataset?

### Modeling Global Behavior

\[ R_{u,i} = b_{u} + p_{u} \cdot q_{i} + \epsilon_{u,i} \]

\( b_{u} = \mu + \sigma_{u} \) and \( \sigma_{i} = \) observed deviations of user \( u \) and item \( i \) from \( \mu \).

**Objective Function**
\[ L(Q, P) = \sum_{(u,i)} (R_{u,i} - q_{i}^T p_{u})^2 + \lambda (\|q_{i}\|^2 + \|p_{u}\|^2) \]

**Gradient Descent Step**
\[ \frac{\partial L}{\partial p_{u}} = 2 \sum_{i \in u} \sum_{h \in i} R_{h,i} (q_{h}^T p_{u}) - 2 \lambda p_{u} \]

### Functional Matrix Factorization

- Hybrid MF-Decision Tree Recommender System
- Uses MF to find best quotas to estimate user profiles

#### Item Cold Start Solution

For new item \( z \), use decision tree and quotas to derive user profile \( p_{z} \). Use \( z \)'s estimated rating for \( i \) then \( q_{i}^T p_{z} + \mu \).

#### Local Collective Embeddings

- Hybrid content-based MF recommender
- TF-IDF

**Objective Function**
\[ \min_{(t_j, d_j)} (f_{k,j} - \max_{t_j \in d_j} \log N) \frac{f_{k,j}}{n_{d,j}} \]

**Collective Factorization**
\[ D_{k,j} = TF-IDF(t_j, d_j) \]

### User Cold Start

**Objective Function**
\[ \frac{1}{2} \sum_{(u,i)} (R_{u,i} - q_{i}^T p_{u})^2 \]

**Closed Form**
\[ h_{t_j} = \arg \min_{h} \sum_{(u,i)} \sum_{w \in u \cap (h \cup z)} (R_{u,i} - q_{i}^T p_{h}) \]

### Item Cold Start Solution

For new item \( z \) and TF-IDF vector \( d_{z} \), derive item profile \( q_{z} \) using common factors \( f_{k,j} \) by solving \( W Q = d_{z} \). Use \( z \)'s estimated rating for \( i \) is then \( q_{z}^T p_{i} + \mu \).

### General Cold Start

#### Further Research
- Benchmark state of the art methods using Amazon dataset
- Explore cold start metric based on new user/item frequency and user/item ratio which leverages performance of both user and item cold start scenarios
- Develop method which addresses user and item cold start
- Use product descriptions to generate topic factorization \( W Q^T \) and topic keywords to generate decision tree \( T \).

**User Cold Start Solution**

**Collective Factorization**
\[ D_{k,j} = TF-IDF(t_j, d_j) \]

**Local Collective Embeddings**

- Hybrid content-based MF recommender
- TF-IDF

**Objective Function**
\[ \min_{(t_j, d_j)} (f_{k,j} - \max_{t_j \in d_j} \log N) \frac{f_{k,j}}{n_{d,j}} \]

**Collective Factorization**
\[ D_{k,j} = TF-IDF(t_j, d_j) \]

### Amazon Dataset

- 6,643,669 users
- 28 product categories & descriptions for all products

### References