Data Mining Final Presentation

P3 - Adoption Prediction in Social Inflence

R03921048 - 廖宜修
R03942071 - 沈昇勳
R03942039 - 呂相弘

Outline

- **Problem Description**
- **Motivation**
- **Proposed Method**
- **Experiments**
- **Future Work**

Problem Description

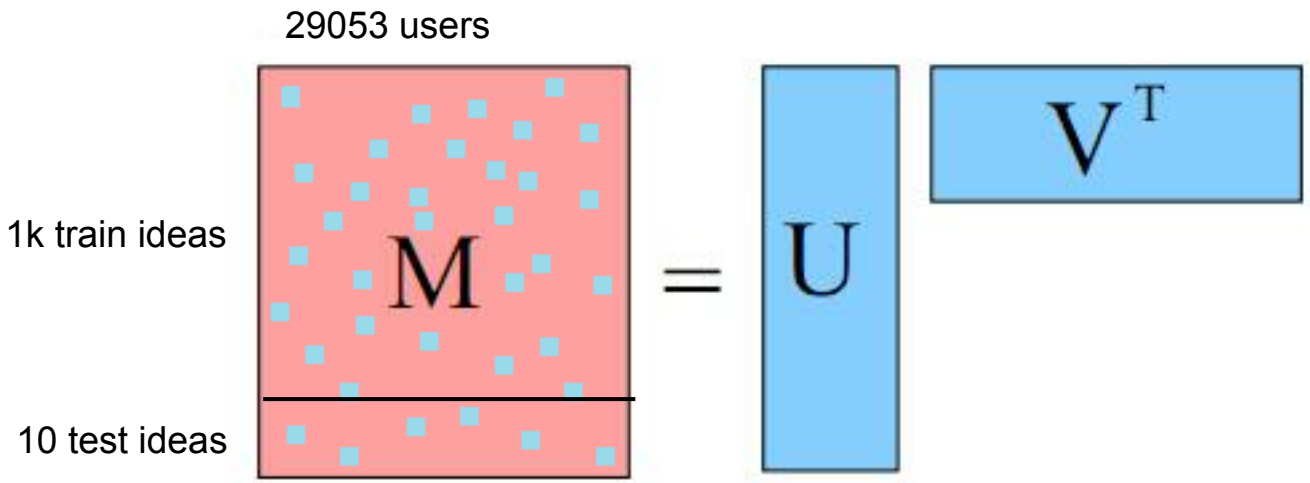
- **Given :**
  - Social Network (vertice, edges)
  - Degree toward ideas with temporal information
  - Initial adopters w.r.t an idea.
- **Goal:**
  - Retrieve following adopters w.r.t and idea.
- **Evaluation:**
  - F- score

Motivation

- **Recommendation System**
  - Given previous music listening logs, the system would recommend you with songs you may like too.
- **Find latent topics in the User-Idea distribution.**
  - Users prefer similar ideas.
  - Ideas are adopted by similar users.
  - Like word - documents relation.

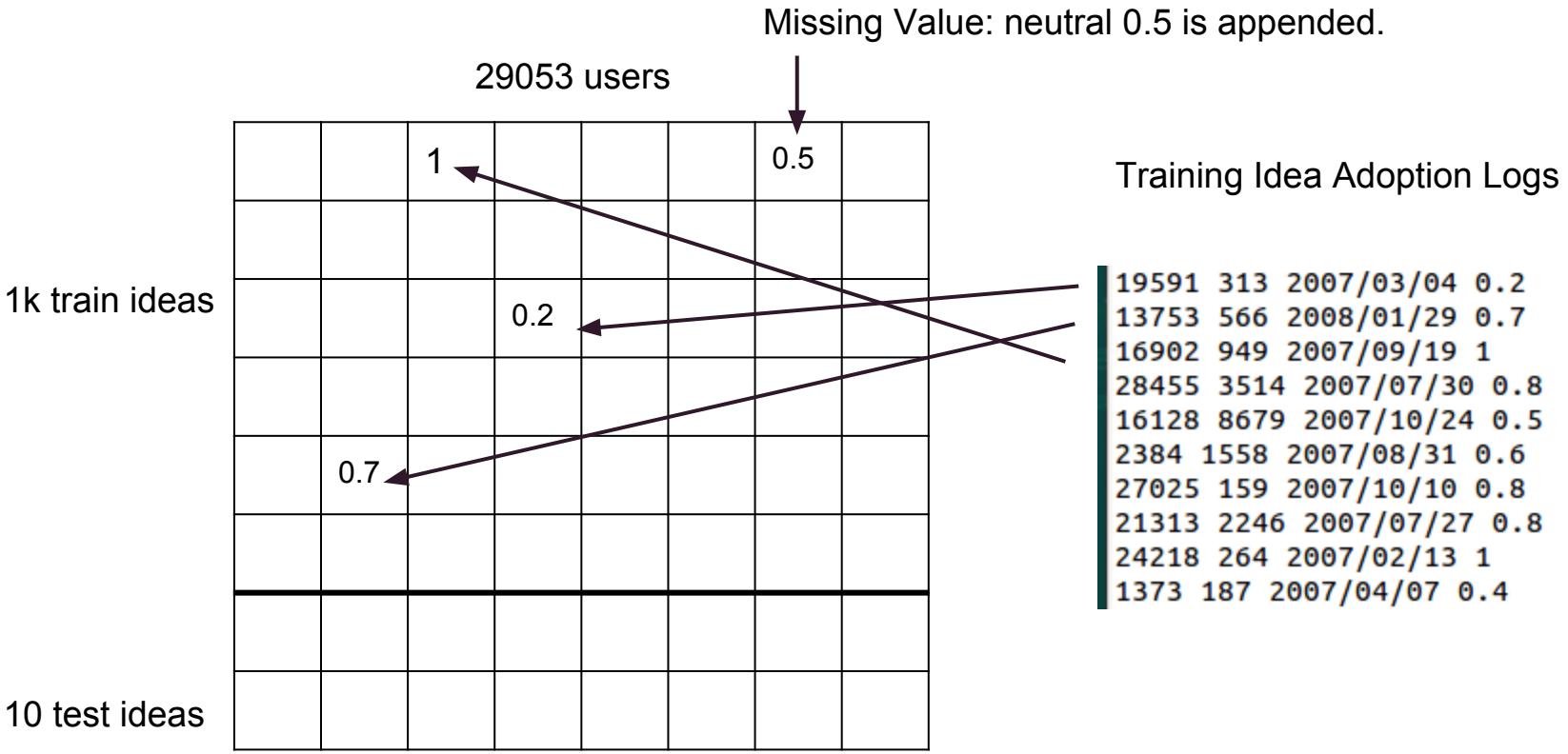
Proposed Methods

- Matrix Factorization based prediction.
- toolkit: libmf



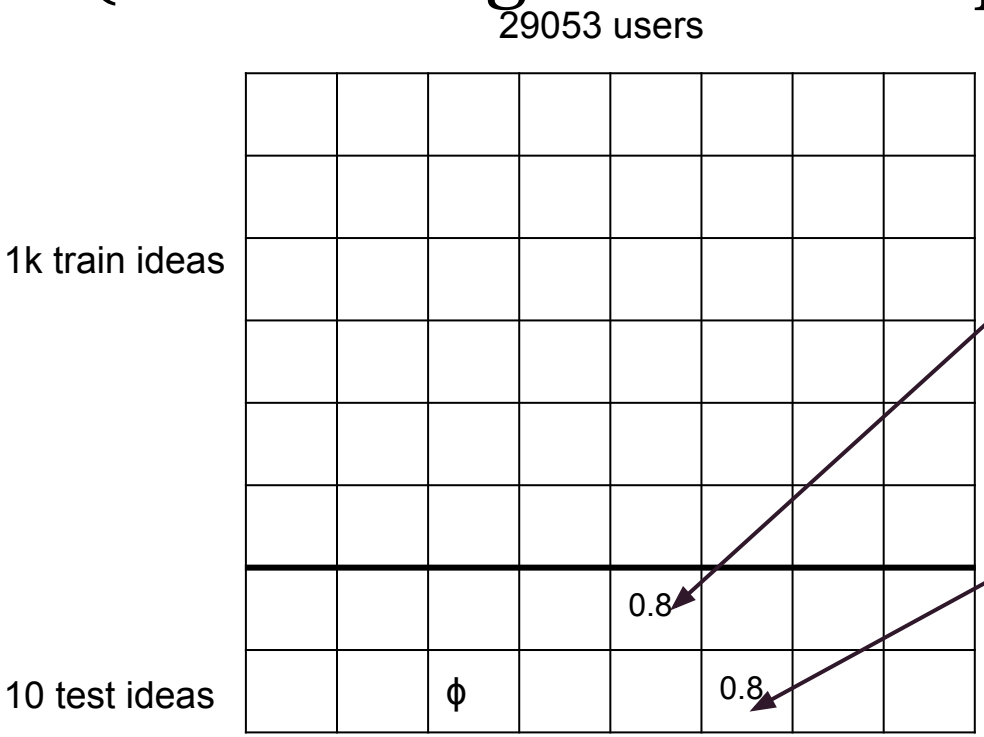
Stage 1:

Generate matrix M of size (1000+10) x 29053



Stage 2:

Append testing ideas with initial adopters (assume degree = 0.8 empirical number.)



Testing Idea Initial Adopters

18815 9068 2516 14186 2844 17678
27753 14955 21439 12145 7252 205
18815 9415 21454 15057 14317 149
58 11781 12232 15390 13321 20867
4865 17678 17640 9415 13847 2858
5312 17266 10464 25907 23232 282
29003 23579 1185 7398 7860 27066
2 8513 5800 3805 6539 6995 10701
6 28352 5487 8485 8190 5574 2602
40 8630 4513 10104 2361 10292 12
5452 19222 18551 27464 12926 282
520 21868 5222 6038 22997 19165
712 19970 4499

idea 1

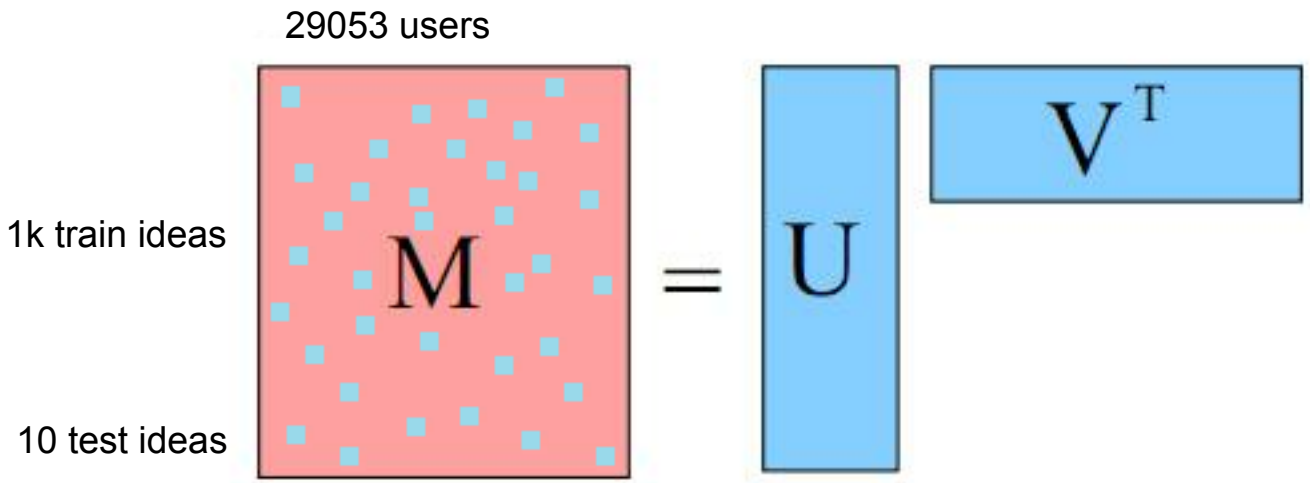
idea 2

Missing Value: leave it blank. (ϕ)

Stage 3:

Objective: L2 - MF reconstruction error.

$$\min_{U,V} O(U,V) = \sum_{m_{ij} \neq \phi} \left(m_{ij} - \sum_{k=1}^{K} u_{ik} v_{jk} \right)^2 + p \sum_{i,k} (u_{ik})^2 + q \sum_{j,k} (v_{jk})^2$$



Stage 4:

- Predict by matrix reconstruction. $M^* = UV^T$
- Sort predicted degree.
- Report first 100 adopters w.r.t an idea.
- Evaluate it by F-score

Experiments

- SGD iteration = 1000
- hidden topics K = 50
- regularization coefficient = 0.05
- learning rate = 0.003

MF	q1	q2	q3
precision	0.332	0.456	0.428
recall	0.035	0.037	0.052
f-score	0.063	0.069	0.093

baseline	q1	q2	q3
precision	0.086	0.076	0.066
recall	0.009	0.006	0.008
f-score	0.016	0.011	0.014

Oracle	q1	q2	q3
precision	1.000	1.000	1.000
recall	0.105	0.082	0.122
f-score	0.191	0.152	0.218

Future Work

- **Social MF**
  - integrate social graph
- **Temporal information**
  - weight training bias in MF
  - different initialization schemes
- **Explore other models.**
  - NN based