# Recommender systems based on variational inference

Kunwoong Kim    Dongwuk Kim    Minwoo Kim

Team 10
Graduate Student, Seoul National Univeristy, Department of Statistics

June 2, 2020

# Current Section

1. Recommender system
2. Method
3. Evaluation
4. Applications
5. Discussion

# User ratings matrix

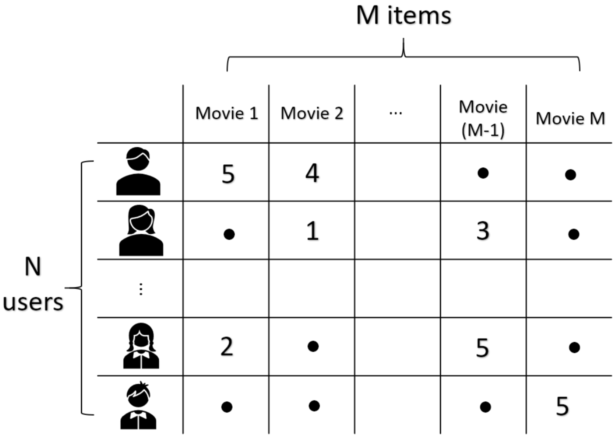
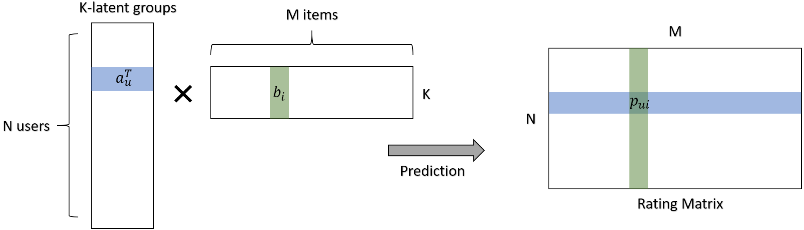



Figure: Example of user ratings

# Matrix Factorization




- $K \in \mathbb{N}$ latent factor: groups of users who share the same tastes.
- $\mathrm{a}_u \in \mathbb{R}^K$: $\mathrm{a}_{u,k} = \Pr(\text{user } u \in \text{Group } K)$, so $\sum_{k=1}^K \mathrm{a}_{u,k} = 1$.
- $\mathrm{b}_i \in \mathbb{R}^K$: $\mathrm{b}_{u,k} = \Pr(\text{users in the group } k \text{ like item } i)$.
- $\mathrm{p}_{u,i} = \mathrm{a}_u^T \mathrm{b}_i$ : estimated rating (up to scaling).

# Example

**Table 9**
Second example of rating matrix.

| | $I_1$ | $I_2$ | $I_3$ | $I_4$ | $I_5$ | $I_6$ | $I_7$ | $I_8$ | $I_9$ | $I_{10}$ | $I_{11}$ | $I_{12}$ | $I_{13}$ | $I_{14}$ | $I_{15}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $U_1$ | 5 | 5 | 5 | 5 | 5 | • | 1 | • | • | • | 1 | • | • | • | • |
| $U_2$ | 5 | 5 | 5 | 5 | 5 | • | • | 1 | • | • | • | • | • | • | • |
| $U_3$ | 5 | 5 | • | 5 | 5 | • | • | • | • | • | • | 3 | • | • | • |
| $U_4$ | • | • | 1 | • | • | 5 | 5 | 5 | 5 | 5 | • | 1 | • | 2 | • |
| $U_5$ | • | 3 | • | • | 2 | 5 | 5 | • | 5 | 5 | • | 2 | • | 4 | • |
| $U_6$ | • | 1 | • | • | • | 5 | 5 | 5 | 5 | 5 | • | • | • | • | • |
| $U_7$ | • | • | • | 4 | • | • | • | • | 1 | • | 5 | 4 | 5 | 5 | 5 |
| $U_8$ | • | 1 | • | • | • | • | 4 | • | • | • | 5 | 5 | 5 | 4 | 5 |
| $U_9$ | • | • | • | • | • | 3 | • | • | 3 | • | 5 | 4 | 5 | 5 | 5 |
| $U_{10}$ | 5 | 5 | 5 | 5 | 5 | 1 | 5 | 2 | 1 | 5 | • | • | • | • | • |
| $U_{11}$ | • | 1 | • | • | • | 5 | 5 | 5 | 5 | 5 | • | • | 5 | • | • |

# Example

**Table 13**
Predictions of the ratings according to our technique.

| | $I_1$ | $I_2$ | $I_3$ | $I_4$ | $I_5$ | $I_6$ | $I_7$ | $I_8$ | $I_9$ | $I_{10}$ | $I_{11}$ | $I_{12}$ | $I_{13}$ | $I_{14}$ | $I_{15}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $U_1$ | **5** | **5** | **5** | **5** | **5** | 2 | **2** | 2 | 2 | 4 | **2** | 3 | 3 | 3 | 3 |
| $U_2$ | **5** | **5** | **5** | **5** | **5** | 2 | 2 | **2** | 2 | 4 | 2 | 3 | 3 | 3 | 3 |
| $U_3$ | **5** | **5** | 5 | **5** | **5** | 2 | 2 | 2 | 2 | 4 | 2 | **3** | 3 | 3 | 3 |
| $U_4$ | 3 | 2 | **3** | 3 | 3 | **5** | **5** | **5** | **5** | **5** | 3 | **2** | 4 | **3** | 3 |
| $U_5$ | 3 | **2** | 3 | 3 | **3** | **5** | **5** | 5 | **5** | **5** | 3 | **2** | 4 | **3** | 3 |
| $U_6$ | 3 | **2** | 3 | 3 | 3 | **5** | **5** | **5** | **5** | **5** | 3 | 2 | 4 | 3 | 3 |
| $U_7$ | 3 | 2 | 3 | **4** | 3 | 3 | 4 | 3 | **3** | 3 | **5** | **4** | **5** | **5** | **5** |
| $U_8$ | 3 | **2** | 3 | 4 | 3 | 3 | **4** | 3 | 3 | 3 | **5** | **4** | **5** | **5** | **5** |
| $U_9$ | 3 | 2 | 3 | 4 | 3 | **3** | 4 | 3 | **3** | 3 | **5** | **4** | **5** | **5** | **5** |
| $U_{10}$ | **5** | **5** | **5** | **5** | **5** | **3** | **3** | **3** | **3** | **4** | 2 | 3 | 3 | 3 | 3 |
| $U_{11}$ | 3 | **2** | 3 | 3 | 3 | **5** | **5** | **5** | **5** | **5** | 3 | 2 | **4** | 3 | 3 |

- Make a prediction and recommend items to user.

# Proposed method

- Antonio Hernando et al., A non negative matrix factorization for collaborative filtering recommender systems based on a Bayesian probabilistic model, *Knowledge-Based Systems*, volume 97, 2016, 188–202
- Adopt a latent graphical model for $a_u, b_i$ and $p_{u,i}$.
- Use variational inference to estimate a ratings matrix.

# Structure of the presentation

1. Understanding the model and algorithm.
2. Evaluations of the model performance.
3. Applications to various domain: Steam games, Book recommendation.
4. Discussion.

# Current Section

# Graphical model

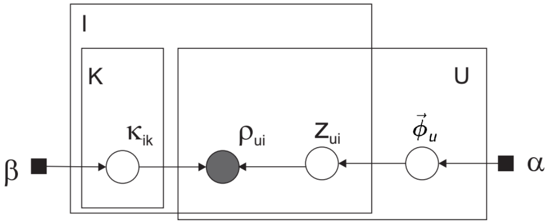



**Fig. 1.** Graphical model representation of our probabilistic approach.

- $\vec{\phi}_u \sim Dir(\alpha, \dots, \alpha)$: $\phi_{u,k} = \Pr(\text{user } u \in \text{ group } k)$.
- $\kappa_{i,k} \sim Beta(\beta, \beta)$: Pr(user in group $k$ likes the item $i$).
- $z_{u,i} \sim Cat(\vec{\phi}_u)$ and $\rho_{u,i} \sim Bin(R, \kappa_{i,z_{u,i}})$.
- The normalized rating: $r^*_{u,i} = \frac{\rho_{u,i}}{R}$. Usually, $R = 4$.

# Variational inference

- Let $\mathcal{M} = (\rho_{u,i}) \in \mathbb{R}^{N \times M}$ be the rating matrix that we ovserved.
- Goal: estimating the posterior distribution $p(\vec{\phi}_u, \kappa_{i,k}, z_{u,i} | \mathcal{M})$.
- Variational Inference
  : Approximate the real posterior distribution $p(\vec{\phi}_u, \kappa_{i,k}, z_{u,i} | \mathcal{M})$ by

$$q(\vec{\phi}_u, \kappa_{i,k}, z_{u,i}) = \prod_{u=1}^{N} q_{\vec{\phi}_u}(\vec{\phi}_u) \prod_{i=1}^{M} \prod_{k=1}^{K} q_{\kappa_{i,k}}(\kappa_{i,k}) \prod_{r_{u,i} \neq \bullet} q_{z_{u,i}}(z_{u,i}),$$

where

- $q_{\vec{\phi}_u}(\vec{\phi}_u) \sim Dir(\gamma_{u,1}, \ldots, \gamma_{u,K})$.
- $q_{\kappa_{i,k}}(\kappa_{i,k}) \sim Beta(\epsilon^{+}_{i,k}, \epsilon^{-}_{i,k})$.
- $q_{z_{u,i}}(z_{u,i}) \sim Cat(\lambda_{u,i,1}, \ldots, \lambda_{u,i,K})$.
- $\gamma_{u,1}, \ldots, \gamma_{u,K}, \epsilon^{+}_{i,k}, \epsilon^{-}_{i,k}$ and $\lambda_{u,i,k}$ are parameters to be learned.

The parameters must fulfill the following conditions:

1. $\gamma_{u,k} = \alpha + \sum_{\{i:r_{u,i} \neq \bullet\}} \lambda_{u,i,k}$
2. $\epsilon^{+}_{i,k} = \beta + \sum_{\{u:r_{u,i} \neq \bullet\}} \lambda_{u,i,k} \cdot r^{+}_{u,i}$
3. $\epsilon^{-}_{i,k} = \beta + \sum_{\{u:r_{u,i} \neq \bullet\}} \lambda_{u,i,k} \cdot r^{-}_{u,i}$
4. $\lambda'_{u,i,k} = \exp(\Psi(\gamma_{u,k}) + r^{+}_{u,i} \cdot \Psi(\epsilon^{+}_{i,k}) + r^{-}_{u,i} \cdot \Psi(\epsilon^{-}_{i,k}) - R \cdot \Psi(\epsilon^{+}_{i,k} + \epsilon^{-}_{i,k}))$
5. $\lambda_{u,i,k} = \frac{\lambda'_{u,i,k}}{\lambda'_{u,i,1} + \ldots \lambda'_{u,i,K}}$

where

- $\Psi(\mathrm{x}) = \frac{\Gamma'(\mathrm{x})}{\Gamma(\mathrm{x})}$
- $r^{+}_{u,i} = \rho_{u,i} = R \cdot r^{*}_{u,i}$
- $r^{-}_{u,i} = R - \rho_{u,i} = R(1 - \cdot r^{*}_{u,i})$

# Algorithm

*Antonio Hernando et al. (2016)* used a coodrinate ascent algorithm to update the parameters sequantially.

- Input: $\mathcal{M} = (r^*_{u,i})$, $\alpha$, $\beta$
- Output: $(\gamma_{u,\,k})$, $(\epsilon^-_{i,k})$, $(\epsilon^+_{i,k})$
- Steps:
  - Initialize randomly $\gamma_{u,\,k}$.
  - Initialize randomly $\epsilon^+_{i,k}$.
  - Initialize randomly $\epsilon^-_{i,k}$.
  - Repeat until changes are not significant
    - For each user $u$:
    - For each item $i$ rated by the user $u$ in the training set:
    - For each factor $k$: update $\lambda_{u,\,i,\,k}$ according to Eq. (9).
    - For each user $u$:
    - For each factor $k$: update $\gamma_{u,\,k}$ according to Eq. (6).
    - For each item $i$ rated by the user $u$ in the training set:
    - For each factor $k$: update $\epsilon^+_{i,k}$ according to Eq. (7).
    - For each factor $k$: update $\epsilon^-_{i,k}$ according to Eq. (8).
- Output $a_{u,\,k}$ according to Eq. (10).
- Output $b_{k,\,i}$ according to Eq. (11).

# Remarks

- $a_{u,k}, b_{u,k} \in [0, 1]$, while $a_{u,k}, b_{u,k}$ can take arbitrary real values in other classical matrix factorization techinqes.
- The proposed model has good probabilistic interpretations.
- $K$ may be low and $a_{u,k}, b_{u,k}$ are sparse.
- **Proposition.** If the recommender system predicts that a user $u$ will like the item $i$, then there are some users with the same taste as $u$ who have very positively rated the item $i$.

# Current Section

# Implementation

- 실제 implementation 할 때 어떠한 방식으로 했는지? 주요하게 언급할만한 것 적기.

# Evaluation

- papaer에 나온 것처럼 영화 데이터에 대해 재현에 방점 두고 작성.

# Accuracy in predictions

- different types of errors: MAE, CMAE, 0-1-Loss 소개 및 간단히 의미 설명
- 각각에서 hyper parameter tuning 정해진 거 report
- quality of recommendations
- PMF(classical methods)랑 our approach 사용
- reproducing Table 19, 20.

# Accuracy in recommendations

- 각각에서 hyper parameter tuning 정해진 거 report
- quality of recommendations
- PMF(classical methods)랑 our approach 사용
- reproducing Table 19, 20.

# Current Section

# Data description

- 우리가 사용한 data 정리 출처 및 간단한 요약통계량
- 먼저 movie 로 이야기하고(논문이랑 동일), steam game and book data

# Data transformation

- steam game / book data 변환 과정 설명

# Results

- accruacy and results of game/book applications.

# Current Section

# Discussion

- 다양한 데이터들에 variational inference based 추천시스템을 적용해봄.
- 영화 이외의 domain data에 대해서도 어떤 의미가 있는지?
- $K$ group이 어떻게 생겼는지, 우리 직관 혹은 알려진 사실과 맞는지 확인해보는 것도 재밌을 듯.
- 등등 필요한 내용...