Factorization meets the Item Embedding : Regularizing Matrix Factorization with Item Co-occurrence

Main Idea

- 다른 사용자가 연달아 소비한 아이템들의 쌍이 유사할 것임
- simultaneously factorization
  - the click matrix and the item co-occurrence matrix
  - Matrix factorization + item embedding
  - Cofactor model 제안

Matrix factorization

rating

item factor

$$\mathcal{L}_{\mathrm{mf}} = \sum_{u,i} c_{ui} (y_{ui} - \theta_u^\top \beta_i)^2 + \lambda_\theta \sum_u \|\theta_u\|_2^2 + \lambda_\beta \sum_i \|\beta_i\|_2^2. \quad (1)$$

scaling parameter

user factor

Regularization term

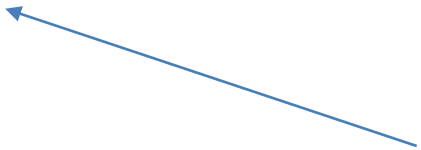
Word embedding

- #(i,j) is the number of times word j appears in the cont ext of word I
- D is the total number of word-context pairs

$$\mathrm{PMI}(i,j) = \log \frac{\#(i,j) \cdot D}{\#(i) \cdot \#(j)}. \qquad (2)$$

$$\#(i) = \sum_j \#(i,j) \text{ and } \#(j) = \sum_i \#(i,j)$$

$$\mathrm{SPPMI}(i,j) = \max\{\mathrm{PMI}(i,j) - \log k, 0\}$$



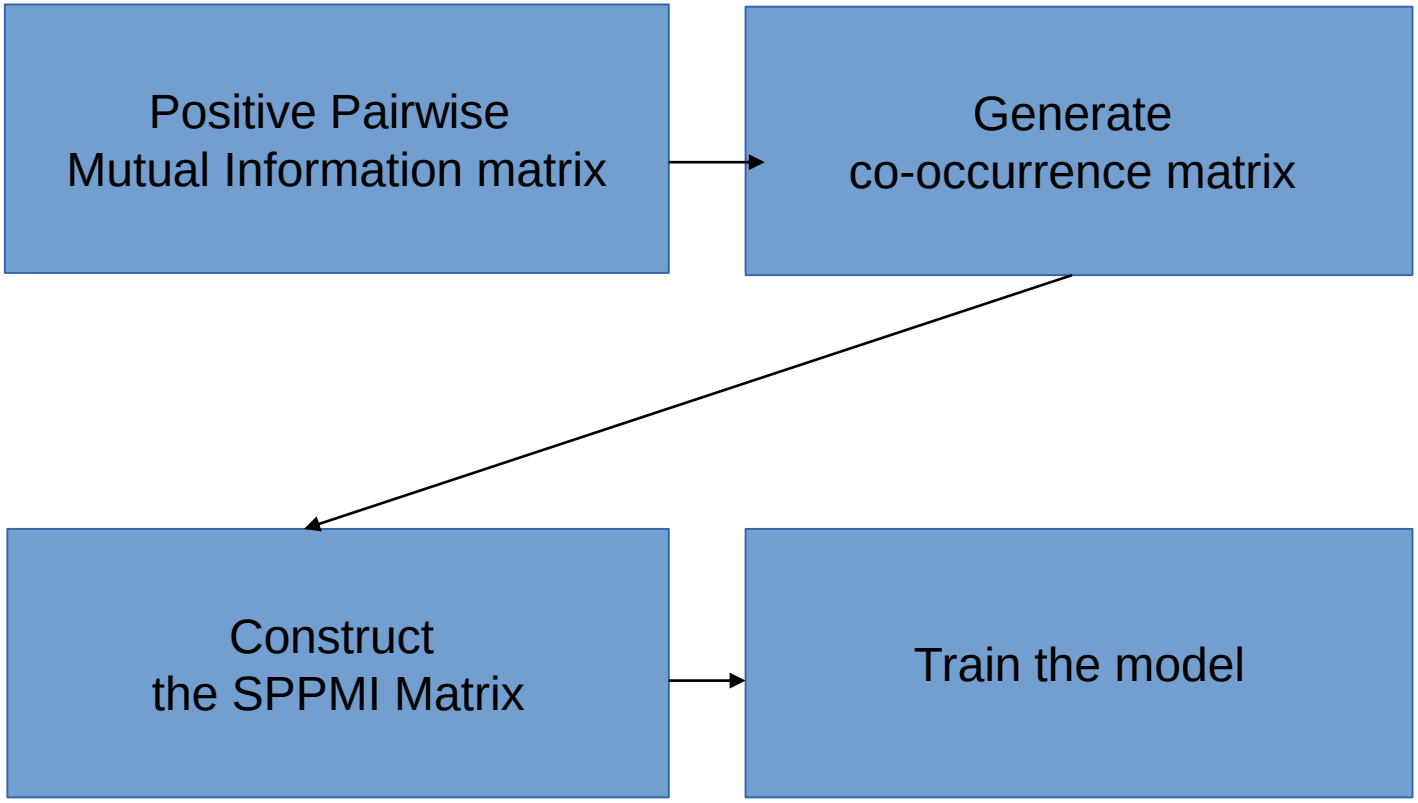
SGNS와 동치

Cofactor model

- MF + item embedding

$$\mathcal{L}_{\text{co}} = \overbrace{\sum_{u,i} c_{ui}\left(y_{ui} - \theta_u^\top \beta_i\right)^2}^{\text{MF}} + \overbrace{\sum_{m_{ij}\neq 0} \left(m_{ij} - \beta_i^\top \gamma_j - w_i - c_j\right)^2}^{\text{item embedding}}$$
$$+ \lambda_\theta \sum_u \|\theta_u\|_2^2 + \lambda_\beta \sum_i \|\beta_i\|_2^2 + \lambda_\gamma \sum_j \|\gamma_j\|_2^2 \qquad (3)$$

Process


Positive Pairwise
Mutual Information matrix
Generate
co-occurrence matrix
Construct
the SPPMI Matrix
Train the model

Example

	4.5	**3.0**	
4.0		3.5	
	5.0		2.0
	3.5	4.0	2.0

0	**1**	**1**	**0**
1	0	1	0
0	1	0	1
0	1	1	1

Example

V

0	**1**	**1**	**0**
1	0	1	0
0	1	0	1
0	1	1	1

PPMI matrix

```
(0, 1)        1
(0, 2)        1
(1, 0)        1
(1, 2)        1
(2, 1)        1
(2, 3)        1
(3, 1)        1
(3, 2)        1
(3, 3)        1
```

```
[[0 1 1 0]
 [1 0 1 0]
 [0 1 0 1]
 [0 1 1 1]]
```

Co-occurrence matrix

```
(1, 3)        1.0
(2, 3)        2.0
(2, 4)        2.0
(3, 1)        1.0
(3, 2)        2.0
(3, 4)        1.0
(4, 2)        2.0
(4, 3)        1.0
```

```
[[ 0.  0.  1.  0.]
 [ 0.  0.  2.  2.]
 [ 1.  2.  0.  1.]
 [ 0.  2.  1.  0.]]
```

SPPMI matrix

```
[ 1.09861231  0.4054651   0.69314718  1.09861231  0.4054651   0.69314718]
```

```
[[ 0.          0.          1.09861231  0.        ]
 [ 0.          0.          0.4054651   0.69314718]
 [ 1.09861231  0.4054651   0.          0.        ]
 [ 0.          0.69314718  0.          0.        ]]
```

Experiment

- movielens의 20M dataset
  - 138,000 user
  - 20million ratings
  - 27,000 movies
  -

```
Test Recall@20: 0.1448
Test Recall@50: 0.1765
```