Sequential Recommendation with User Memory Networks

11/13/2018

Xiaosong Jia

Motivation

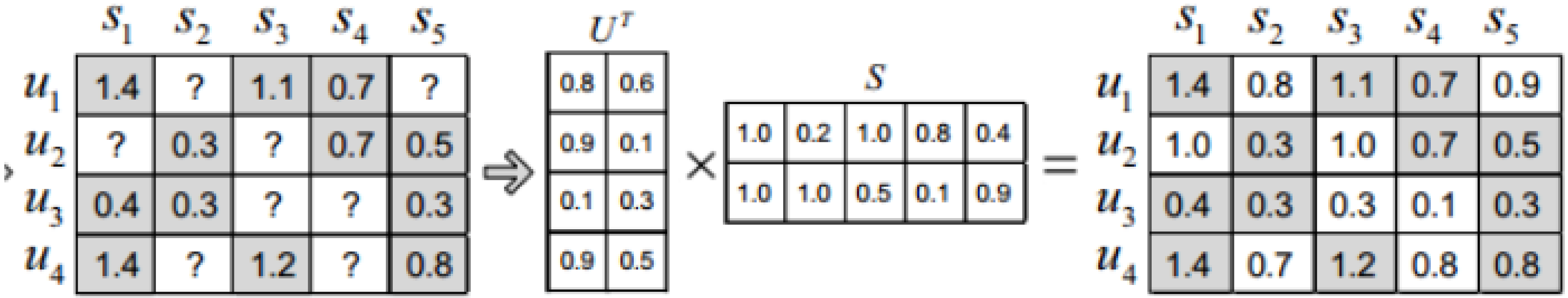
- Compress all of a user's previous records into a fixed hidden representation?

- Weaken the signal of highly correlated items for sequential recommendation
- Overlooking such signal makes it difficult for us to understand and explain the sequential recommendations.

General Framework

- N users, M item
- p_u-user u's vector, q_i-item I's vector
- M^u the personalized memory matrix
- $p_u^m = READ(M^u, q_i)$
- $p_u = MERGE(p_u^*, p_u^m)$
- MERGE(x,y)=$x + ay$
- Prediction： PREDICT(p_u, q_i)= $\delta(p_u^T q_i)$
- $M^u < -WRITE(M^u, q_i)$

Matrix Factorization



Item-level method

- Read:

$$w_{ik} = (\boldsymbol{q}_{v_i^u})^T \cdot \boldsymbol{m}_k^u,\ z_{ik} = \frac{\exp(\beta w_{ik})}{\sum_j \exp(\beta w_{ij})},\ \forall k = 1, 2, \cdots, K \quad (7)$$

$$\boldsymbol{p}_u^m = \sum_{k=1}^{K} z_{ik} \cdot \boldsymbol{m}_k^u$$

- Write： Recent K items.

Feature-level method

- Global latent feature matrix F= $\{\boldsymbol{f}_1, \boldsymbol{f}_2, \ldots \boldsymbol{f}_k\}$
- User preference matrix $\boldsymbol{M}^u = \{\boldsymbol{m}_1^u, \ldots, \boldsymbol{m}_k^u\}$
- Read：

$$w_{ik} = \boldsymbol{q}_i^T \cdot \boldsymbol{f}_k,\ z_{ik} = \frac{\exp(\beta w_{ik})}{\sum_j \exp(\beta w_{ij})},\ \forall k = 1, 2, \cdots, K$$

$$\boldsymbol{p}_u^m = \sum_{k=1}^{K} z_{ik} \cdot \boldsymbol{m}_k^u$$

- Write：

$$\boldsymbol{erase}_i = \sigma(E^T \boldsymbol{q}_i + \boldsymbol{b}_e) \quad \boldsymbol{m}_k^u \leftarrow \boldsymbol{m}_k^u \odot (1 - z_{ik} \cdot \boldsymbol{erase}_i)$$

$$\boldsymbol{add}_i = \tanh(A^T \boldsymbol{q}_i + \boldsymbol{b}_a),\ \boldsymbol{m}_k^u \leftarrow \boldsymbol{m}_k^u + z_{ik} \cdot \boldsymbol{add}_i$$