## Neural Collaborative Filtering: A Generalization of Biased Matrix Factorization

## 1 Preliminary

Let  $\boldsymbol{W} \in \mathbb{R}^{d \times (n+m)}, \boldsymbol{p} \in \mathbb{R}^n, \boldsymbol{q} \in \mathbb{R}^m$ , and then we have:

$$egin{aligned} oldsymbol{W}egin{bmatrix} oldsymbol{p} \ oldsymbol{q} \end{aligned} = oldsymbol{W}_poldsymbol{p} + oldsymbol{W}_qoldsymbol{q} \end{aligned}$$

where  $W_p, W_q$  are respectively the first n and last m columns of W.

## 2 Neural Collaborative Filtering (NCF)

Rating estimate  $\hat{y}_{ui}$  that user u gives item i is obtained by:

$$egin{aligned} \hat{y}_{ui} &= \sigma \left( oldsymbol{h}^{ op} egin{bmatrix} \phi^{ ext{GMF}} \ \phi^{ ext{MLP}} &= oldsymbol{p}_u^G \odot oldsymbol{q}_i^G \ \end{pmatrix} \ \phi^{ ext{MLP}} &= a_L \left( oldsymbol{W}_L^{ op} a_{L-1} \left( \dots a_2 \left( oldsymbol{W}_2^{ op} egin{bmatrix} oldsymbol{p}_u^M \ oldsymbol{q}_i^M \ \end{pmatrix} + oldsymbol{b}_2 
ight) \dots 
ight) + oldsymbol{b}_L 
ight) \end{aligned}$$

- $\sigma$ : sigmoid function (because NCF solves binary ratings  $\{1,0\}$ ).
- $\bullet$   $\odot$ : element-wise multiplication.
- $p_u^G, p_u^M$ : latent factor vectors of user u.
- $q_i^G, q_i^M$ : latent factor vectors of item i.
- h, W, b: weights in neurons.
- a: activation function.

## 3 Simplification of NCF

We limit MLP to one layer with linear activation  $a(\mathbf{p}) = \mathbf{p}$ :

$$\phi^{ ext{MLP}} = a \left( oldsymbol{W}^{ op} egin{bmatrix} oldsymbol{p}_u^M \ oldsymbol{q}_i^M \end{bmatrix} + oldsymbol{b} 
ight) \ = oldsymbol{W}_v^{ op} oldsymbol{p}_u^M + oldsymbol{W}_q^{ op} oldsymbol{q}_i^M + oldsymbol{b} 
ight)$$

Then we rewrite NCF rating estimate (without sigmoid function for clear presentations):

$$egin{aligned} \hat{y}_{ui} &= oldsymbol{h}^{ op} egin{bmatrix} \phi^{ ext{GMF}} \ \phi^{ ext{MLP}} \end{bmatrix} \ &= oldsymbol{h}_{G}^{ op} \phi^{ ext{GMF}} + oldsymbol{h}_{M}^{ op} \phi^{ ext{MLP}} \ &= oldsymbol{h}_{G}^{ op} \left( oldsymbol{p}_{u}^{G} \odot oldsymbol{q}_{i}^{G} 
ight) + oldsymbol{h}_{M}^{ op} \left( oldsymbol{W}_{p}^{ op} oldsymbol{p}_{u}^{M} + oldsymbol{W}_{q}^{ op} oldsymbol{q}_{i}^{M} + oldsymbol{b} 
ight) \end{aligned}$$

Let  $h_G = 1$  (vector of all elements 1's), and we have:

$$\hat{y}_{ui} = \underbrace{\boldsymbol{p}_{u}^{G\top}\boldsymbol{q}_{i}^{G}}_{\text{MF}} + \underbrace{\boldsymbol{h}_{M}^{\top}\boldsymbol{W}_{p}^{\top}\boldsymbol{p}_{u}^{M}}_{\text{user bias}} + \underbrace{\boldsymbol{h}_{M}^{\top}\boldsymbol{W}_{q}^{\top}\boldsymbol{q}_{i}^{M}}_{\text{item bias}} + \underbrace{\boldsymbol{h}_{M}^{\top}\boldsymbol{b}}_{\text{global bias}}$$

The last three scalar terms model biases in matrix factorization (MF) rating estimation. That is, the simplified version of NCF is equivalent to biased MF.