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:

$$\boldsymbol{W}\begin{bmatrix}\boldsymbol{p}\\ \boldsymbol{q}\end{bmatrix} = \begin{bmatrix}\boldsymbol{W}_p \boldsymbol{W}_q\end{bmatrix}\begin{bmatrix}\boldsymbol{p}\\ \boldsymbol{q}\end{bmatrix} = \boldsymbol{W}_p\boldsymbol{p} + \boldsymbol{W}_q\boldsymbol{q}$$

where $\boldsymbol{W}_p, \boldsymbol{W}_q$ are respectively the first n and last m columns of $\boldsymbol{W}$.

2 Neural Collaborative Filtering (NCF)

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

$$\begin{aligned}
\hat{y}_{ui} &= \sigma\left(\boldsymbol{h}^\top \begin{bmatrix}\phi^{\text{GMF}}\\ \phi^{\text{MLP}}\end{bmatrix}\right)\\
\phi^{\text{GMF}} &= \boldsymbol{p}_u^G \odot \boldsymbol{q}_i^G\\
\phi^{\text{MLP}} &= a_L\left(\boldsymbol{W}_L^\top a_{L-1}\left(\dots a_2\left(\boldsymbol{W}_2^\top \begin{bmatrix}\boldsymbol{p}_u^M\\ \boldsymbol{q}_i^M\end{bmatrix} + \boldsymbol{b}_2\right)\dots\right) + \boldsymbol{b}_L\right)
\end{aligned}$$

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

3 Simplification of NCF

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

$$\begin{aligned}
\phi^{\text{MLP}} &= a\left(\boldsymbol{W}^\top \begin{bmatrix}\boldsymbol{p}_u^M\\ \boldsymbol{q}_i^M\end{bmatrix} + \boldsymbol{b}\right)\\
&= \boldsymbol{W}_p^\top \boldsymbol{p}_u^M + \boldsymbol{W}_q^\top \boldsymbol{q}_i^M + \boldsymbol{b}
\end{aligned}$$

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

$$\begin{aligned}
\hat{y}_{ui} &= \boldsymbol{h}^\top \begin{bmatrix} \phi^{\text{GMF}} \\ \phi^{\text{MLP}} \end{bmatrix} \\
&= \boldsymbol{h}_G^\top \phi^{\text{GMF}} + \boldsymbol{h}_M^\top \phi^{\text{MLP}} \\
&= \boldsymbol{h}_G^\top \left(\boldsymbol{p}_u^G \odot \boldsymbol{q}_i^G \right) + \boldsymbol{h}_M^\top \left(\boldsymbol{W}_p^\top \boldsymbol{p}_u^M + \boldsymbol{W}_q^\top \boldsymbol{q}_i^M + \boldsymbol{b} \right)
\end{aligned}$$

Let $\boldsymbol{h}_G = \mathbf{1}$ (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.