朱爵的做法：assign a uniform weight to the missing data to reduce computational complexity
即负样本采样是正样本的倍

本文：1. weight the missing data based on item popularity，根据物品流行度来采样
↓ 有随机数年问题
2. element-wise Alternating Least Square 算法 (eALS)，不需要设置学习率

sampled based learning，samples negative instances from missing data
whole-data based learning，treate all missing data as negative
都有优缺点

sampled based learning：提高了训练的效率，但降低了模型的鲁棒性
whole-data based learning：相反

本文仍然采用 whole-data based learning，但和 ALS-based algorithm 来提升效率

MF：$\hat{r}_{ui} = \langle \vec{p}_u, \vec{q}_i \rangle$，如果将 p_u 看成 $\vec{p}_u = [\vec{p}_u, b_u, 1]$，$q_i = [\vec{q}_i, 1, b_i]$
那么 $\langle p_u, q_i \rangle = b_u + b_i + \vec{p}_u^T \cdot \vec{q}_i$ ⇒ bias MF

$$J = \sum_{u=1}^{M} \sum_{i=1}^{N} w_{ui} (r_{ui} - \hat{r}_{ui})^2 + \lambda \left(\sum_{u=1}^{M} \|\vec{p}_u\|^2 + \sum_{i=1}^{N} \|\vec{q}_i\|^2 \right)$$

w_{ui} 是实际 r_{ui} 的权重，$W = [w_{ui}]_{M \times N}$，missing entries 的 $r_{ui} = 0$，$w_{ui} \neq 0$

ALS：一次只优化一个参数，将其余参数固定（理解为常数），
前提条件是优化子问题能够解析求解

对于 p_u 而言，$J_u = \|W^u (r_u - Q p_u)\|^2 + \lambda \|p_u\|^2$
$r_u = (N, 1)$　$Q = (N, k)$　$p_u = (k, 1)$　$W^u = (N, N)$

$2W^u (r_u - Qp_u) \cdot (-Q)$
$2W^u r_u (-Q) + 2W^u Q p_u Q$
$-2Q^T W^u r_u + 2Q^T W^u Q p_u$

我们希望 $J_u = 0$，$\therefore \frac{\partial J_u}{\partial p_u} = 0 = 2\lambda p_u + 2W^u Q p_u Q - 2W^u r_u Q$
矩阵交换相乘顺序，必须转置

$= 2\lambda p_u + 2Q^T W^u Q p_u - 2Q^T W^u r_u = 0$

$\therefore (Q^T W^u Q + \boxed{\lambda I}) p_u = Q^T W^u r_u$

$\therefore \vec{p}_u = (Q^T W^u Q + \lambda I)^{-1} \cdot Q^T W^u r_u$，问题可以转化到 $\vec{q}_u$

↓

要对 Q 求转置，时间复杂度为 $O(Nk^2)$

∴ 更新一个用户的隐向量的时间复杂度为 $O(Nk^2 + Nk^2)$

更新所有参数的时间复杂度为 $O((M+N)k^3 + MNk^2)$

全量，非采样

$$L = \sum_{u=1}^{M} \sum_{i \in I_u} w_{ui} (\hat{r}_{ui} - r_{ui})^2 + \sum_{u=1}^{M} \boxed{\sum_{k \notin I_u}} s_k \hat{r}_{uk}^2 + \lambda \left(\sum_{u=1}^{M} \|U_u\|^2 + \sum_{i=1}^{N} \|V_i\|^2 \right)$$

point-wise

↓ $s_k (\hat{r}_{uk} - 0)^2$

如果 s_k 比较大，为了 minimize loss，会让 $\hat{r}_{uk}$ 尽可能趋近于 0

$s_k = s_0 \cdot \frac{f_k^{\alpha}}{\sum_{i=1}^{N} f_i^{\alpha}}$，$f_k$ 是物品 k 的流行度

一个物品越流行，但用户 u 很喜欢，$\hat{r}_{uk}$ 是可能为 0 → 对于强假设（所有物品 k 和 u 有关联）的削弱

令 $L_1 = \sum_{u=1}^{M} \sum_{i \in I_u} w_{ui} (\hat{r}_{ui} - r_{ui})^2$，$L_2 = \sum_{u=1}^{M} \sum_{k \notin I_u} s_k \hat{r}_{uk}^2$，$L_{reg} = \lambda \left(\sum_{u=1}^{M} \|U_u\|^2 + \sum_{i=1}^{N} \|V_i\|^2 \right)$

$\therefore \frac{\partial L_{eALS}}{\partial U_{uf}} = \frac{\partial (L_1 + L_2 + L_{reg})}{\partial U_{uf}}$　U_{uf} 为 U_u 的第 f 维度

$\hat{r}_{ui}^{-f} = \hat{r}_{ui} - U_{uf} V_{if}$

$$\frac{\partial L_1}{\partial U_{uf}} = \frac{\partial \sum_{i \in I_u} w_{ui} (\hat{r}_{ui} - r_{ui})^2}{\partial U_{uf}}$$

$\hat{r}_{ui} = \sum_{f=1}^{k} U_{uf} \cdot V_{if}$

$$= \sum_{i \in I_u} 2 w_{ui} (\hat{r}_{ui} - r_{ui}) V_{if}$$

$$= \sum_{i \in I_u} 2 w_{ui} (\hat{r}_{ui}^{-f} + U_{uf} V_{if} - r_{ui}) V_{if}$$

$$= \sum_{i \in I_u} 2 w_{ui} (\hat{r}_{ui}^{-f} - r_{ui}) V_{if} + \sum_{i \in I_u} 2 w_{ui} U_{uf} V_{if}^2$$

$$\frac{\partial L_2}{\partial U_{uf}} = \frac{\partial \sum_{k \notin I_u} s_k \hat{r}_{uk}^2}{\partial U_{uf}}$$

$\hat{r}_{uk} = U_u \cdot V_k = \sum_{f=1}^{k} U_{uf} \cdot V_{kf}$

$$= \sum_{k \notin I_u} 2 s_k \hat{r}_{uk} \cdot V_{kf}$$

$$= \sum_{k \notin I_u} 2 s_k (\hat{r}_{uk}^{-f} + U_{uf} V_{kf}) V_{kf}$$

$$= \sum_{k \notin I_u} 2 s_k \hat{r}_{uk}^{-f} V_{kf} + \sum_{k \notin I_u} 2 s_k U_{uf} V_{kf}^2$$

令 $\frac{\partial L_{eALS}}{\partial U_{uf}} = 0$，可得 $U_{uf} = \frac{-\sum_{i \in I_u} w_{ui} (\hat{r}_{ui}^{-f} - r_{ui}) V_{if} + \boxed{\sum_{k \notin I_u} s_k \hat{r}_{uk}^{-f} V_{kf}}}{\sum_{i \in I_u} w_{ui} V_{if}^2 + \boxed{\sum_{k \notin I_u} s_k V_{kf}^2} + \lambda}$

missing data 会连续大量的计算量，
使用 memorization strategy 来优化

先集合成为常量的项

$$\sum_{k \notin I_u} s_k \hat{r}_{uk}^{-f} V_{kf} = \sum_{k=1}^{m} s_k U_{uf} \sum_{e \neq f} U_{ue} V_{ke} - \sum_{k \in I_u} s_k \hat{r}_{uk}^{-f} V_{kf}$$

item-wise → element-wise
将 s_k 和 V_{kf} 作 distribute

$s_1 V_{1f} \cdot (U_{u1} V_{11} + \dots + U_{uk} V_{1k}) +$
$s_2 V_{2f} (U_{u1} V_{21} + \dots + U_{uk} V_{2k}) +$
$\vdots$
$s_m V_{mf} (U_{u1} V_{m1} + \dots + U_{uk} V_{mk})$

发现的整合成一次

$$= \sum_{e \neq f} U_{ue} \sum_{k=1}^{m} s_k V_{kf} V_{ke} - \sum_{k \in I_u} s_k \hat{r}_{uk}^{-f} V_{kf}$$

矩阵 C^q 中第 q 列的大小

$$= \sum_{e \neq f} U_{ue} C_{fe}^{q} - \sum_{k \in I_u} s_k \hat{r}_{uk}^{-f} V_{kf}$$

计算量小

转置 × 本身 → 方阵

$C^q = \sum_{k=1}^{m} s_k \boxed{V_k^T V_k} \in R^{d \times d}$　可以提前计算好

$$\sum_{k \notin I_u} s_k V_{kf}^2 = \sum_{k=1}^{m} s_k V_{kf}^2 - \sum_{k \in I_u} s_k V_{kf}^2$$

同理

$$= C_{ff}^{q} - \sum_{k \in I_u} s_k V_{kf}^2$$