Two Tower Model

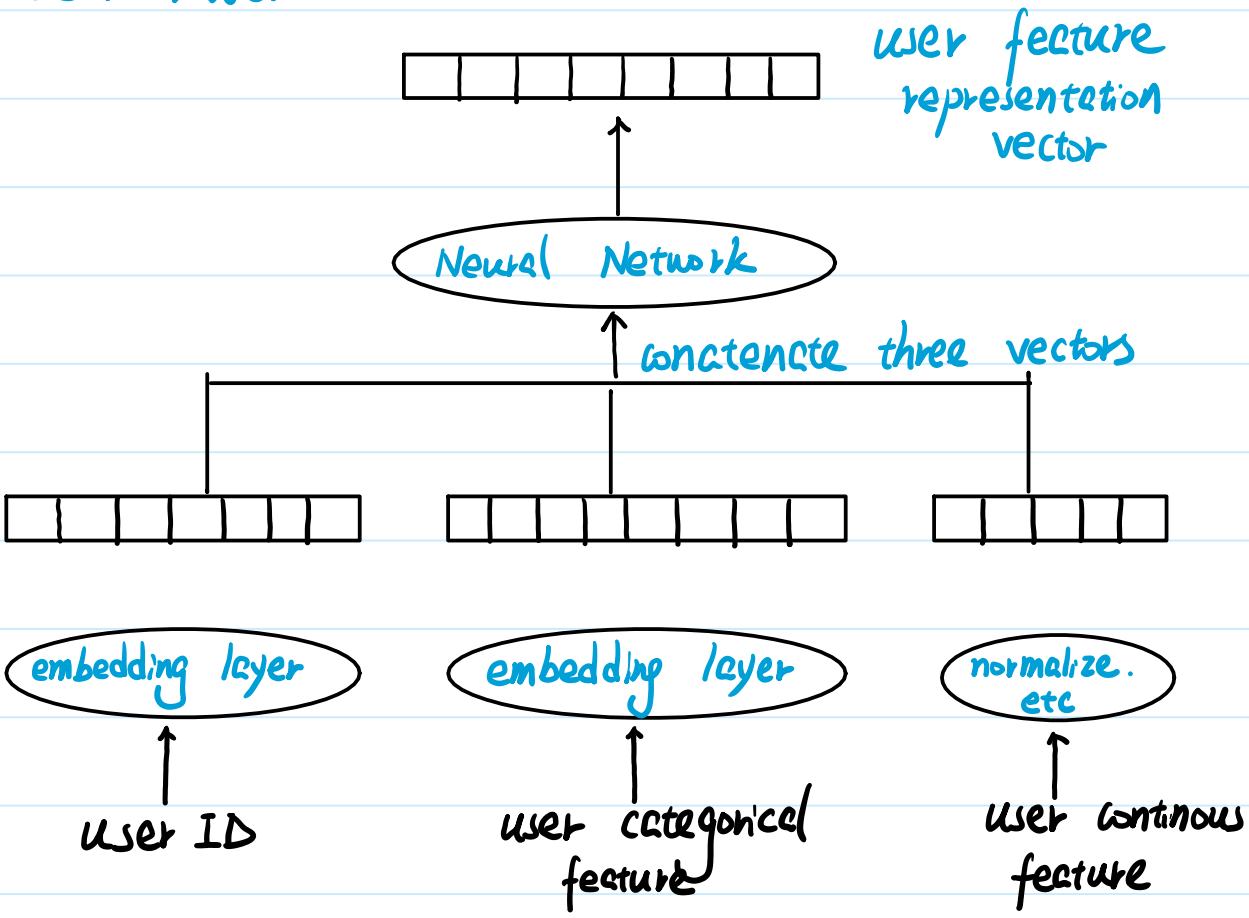
Thursday, March 7, 2024 7:45 AM

limitation of matrix completion:

Do NOT use user and item features.

Two Tower Model:

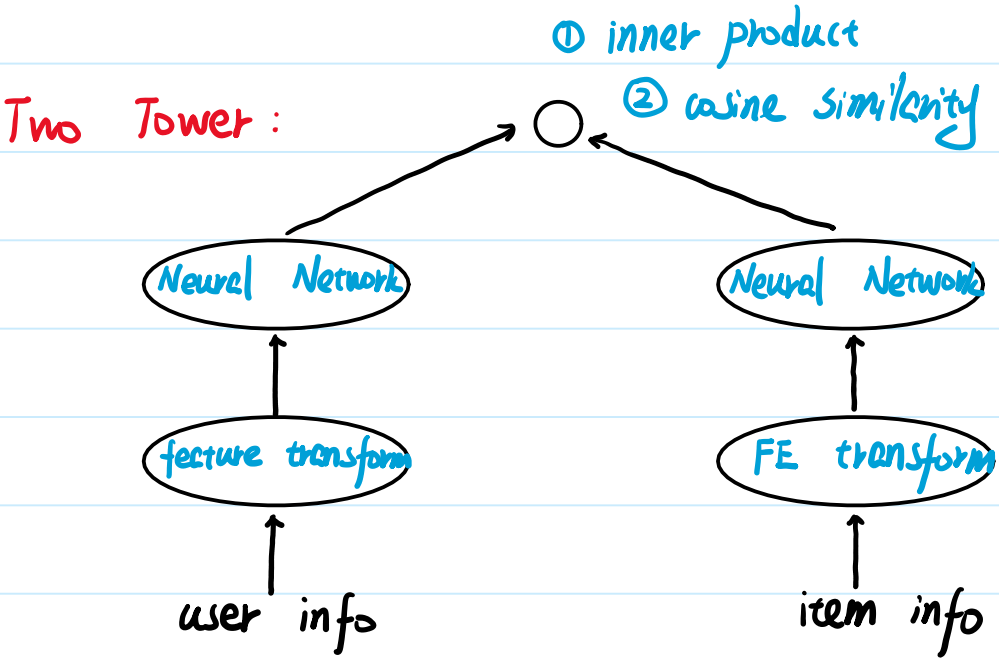
user tower:



item tower has the similar structure but consumes the item ID and features.

Two Tower Model

Thursday, March 7, 2024 7:45 AM



Training :

① point wise : (user FE, item FE, click)

⋮

(user FE, item FE, No click)

② pair wise : one positive + one negative sample

③ list wise : one positive + multiple negative sample

Positive Sample : the item clicked by users

Negative Sample : ① items NOT retrieved ?

② items retrieved but removed in rank ?

③ items exposed but not clicked ?

Two Tower Model

Thursday, March 7, 2024 7:45 AM

Pointwise Training:

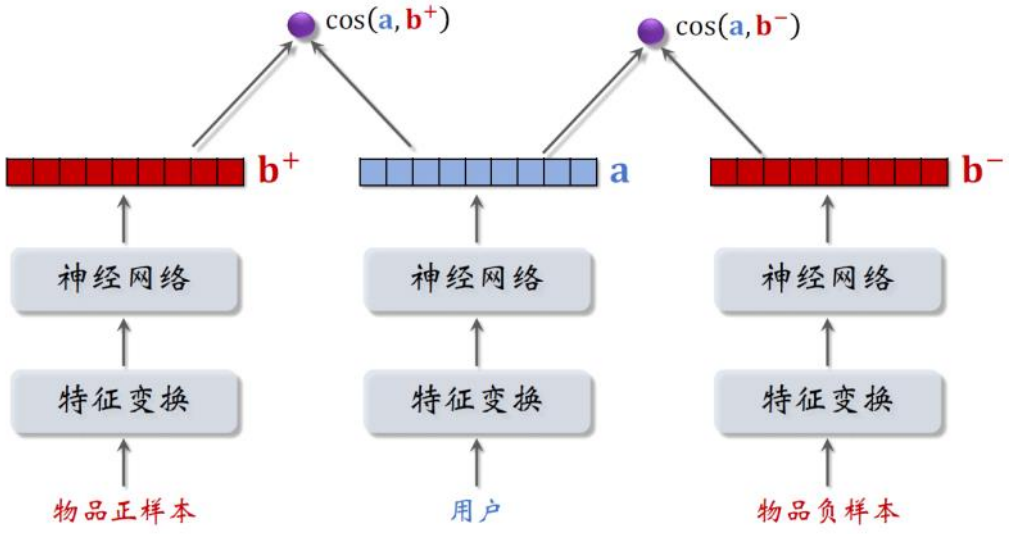
① view "retrieval" as binary classification

② positive sample: $\cos\langle a, b\rangle \rightarrow 1.0$

③ negative sample: $\cos\langle a, b\rangle \rightarrow -1.0$

④ # of positive to negative samples: 1:2 or 1:3

Pairwise Training:



(Picture from Shusen Wang on Youtube/Bilibili)

make $\cos(a, b^+) \gg \cos(a, b^-)$ hyperparameter

① if $\cos(a, b^+) > \cos(a, b^-) + m$, no loss

② else loss $= \cos(a, b^-) + m - \cos(a, b^+)$

Two Tower Model

Thursday, March 7, 2024 7:45 AM

Triplet hinge loss:

$$L(a, b^+, b^-) = \max(0,\ \cos(a, b^-) + m - \cos(a, b^+))$$

Triplet logistic loss:

$$L(a, b^+, b^-) = \log\left(1 + \exp\left[\sigma(\cos(a, b^-) - \cos(a, b^+))\right]\right)$$

list wise Training:

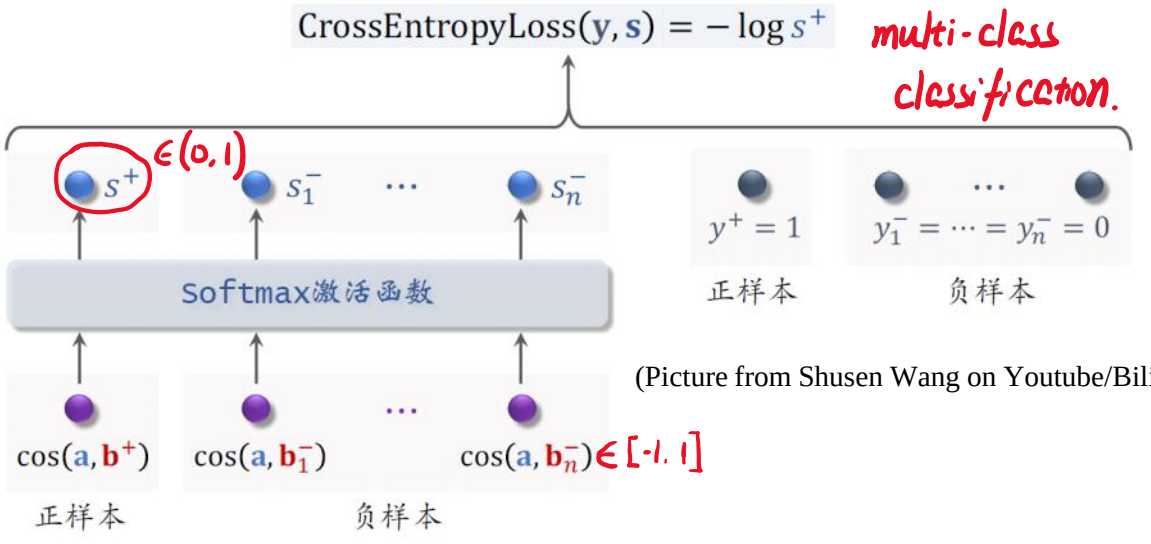
① one record:

user a, positive sample b^+

multiple negative samples: $b_1^-, \cdots b_n^-$

② make $\cos(a, b^+)$ large

③ make $\cos(a, b_1^-), \cdots \cos(a, b_n^-)$ small



(Picture from Shusen Wang on Youtube/Bilibili)