

双塔模型：模型和训练

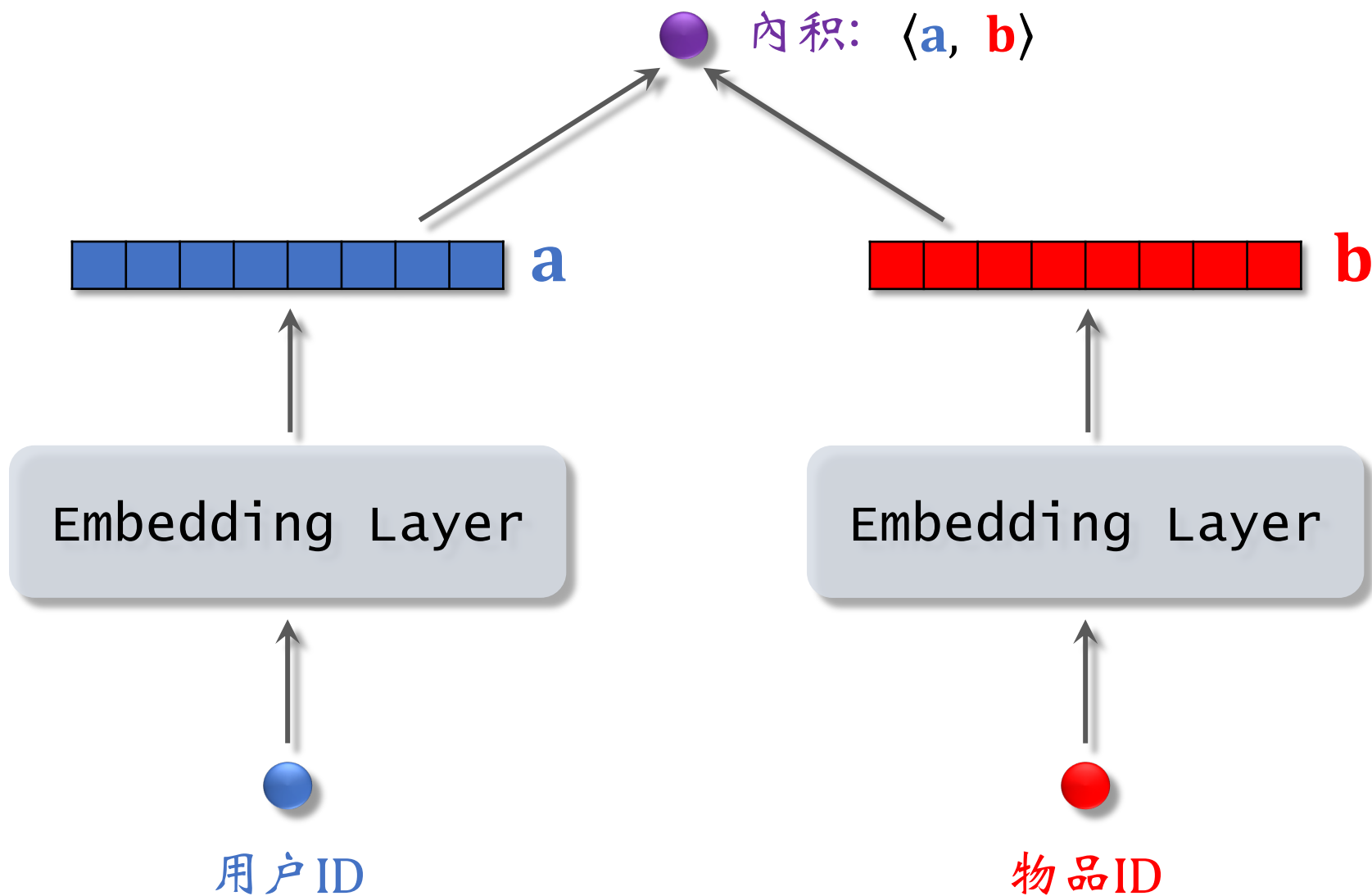
王树森

ShusenWang@xiaohongshu.com

<http://wangshusen.github.io/>



矩阵补充模型



双塔模型



Embedding Layer



用户 ID



Embedding Layers



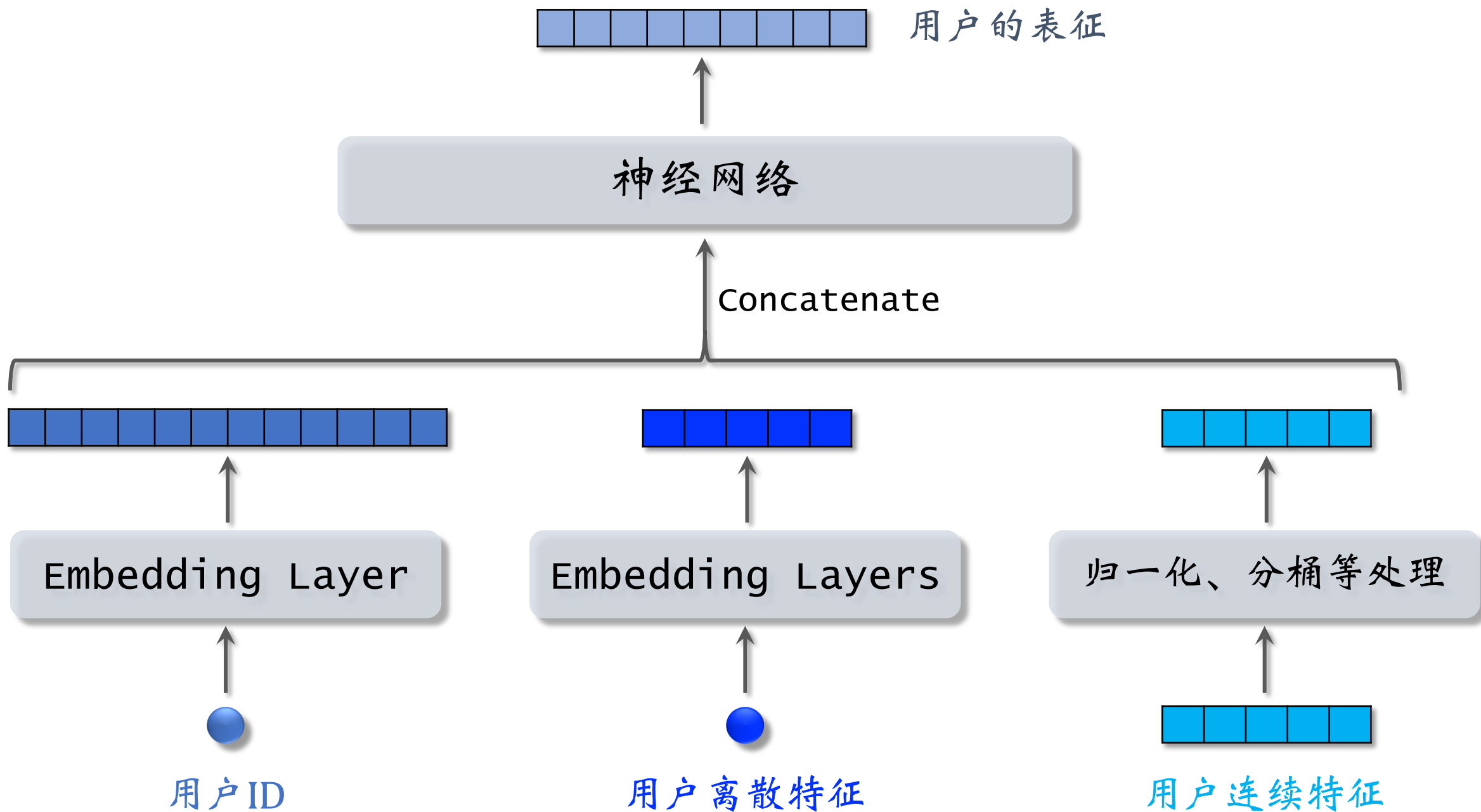
用户离散特征

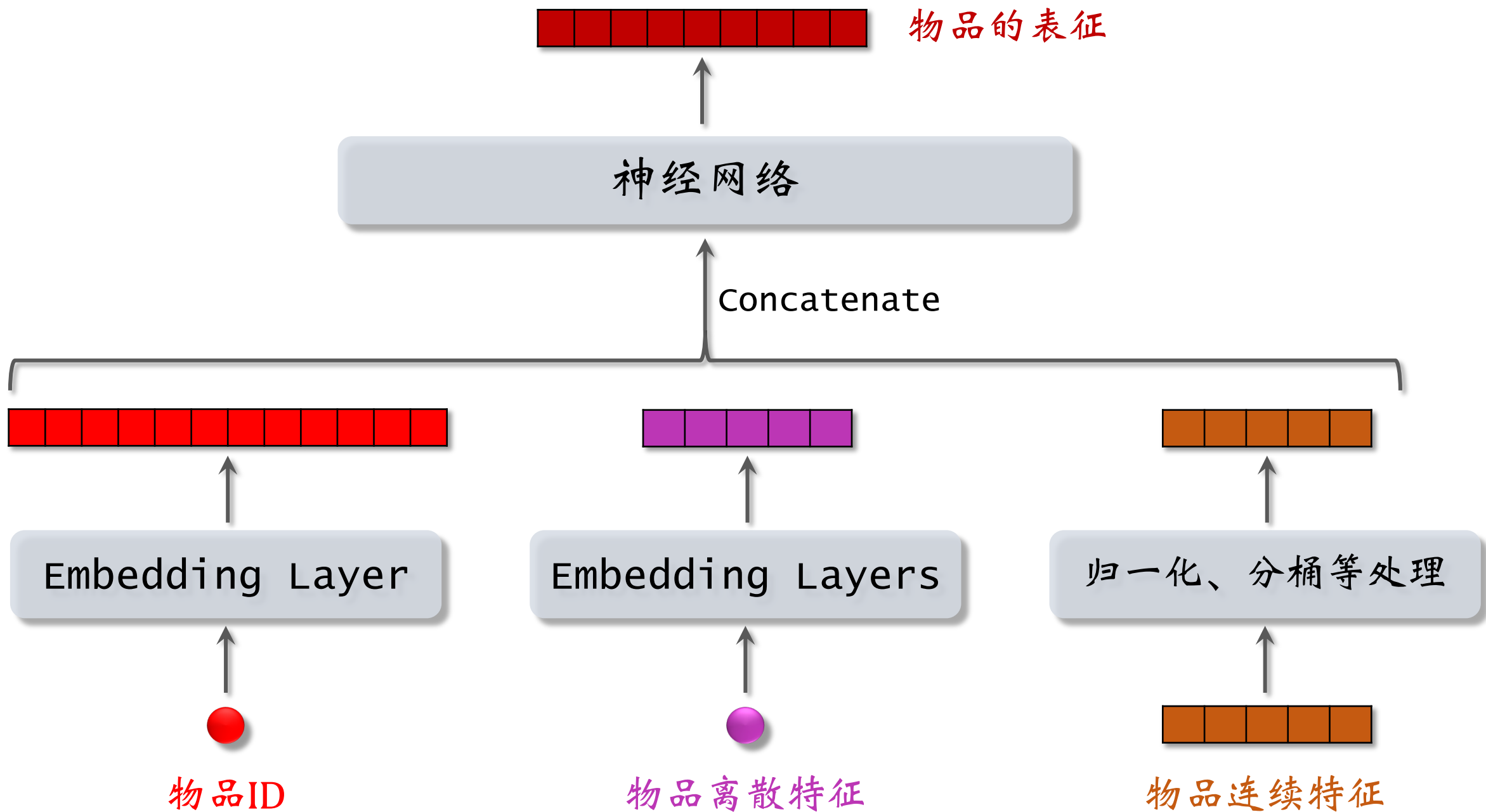


归一化、分桶等处理

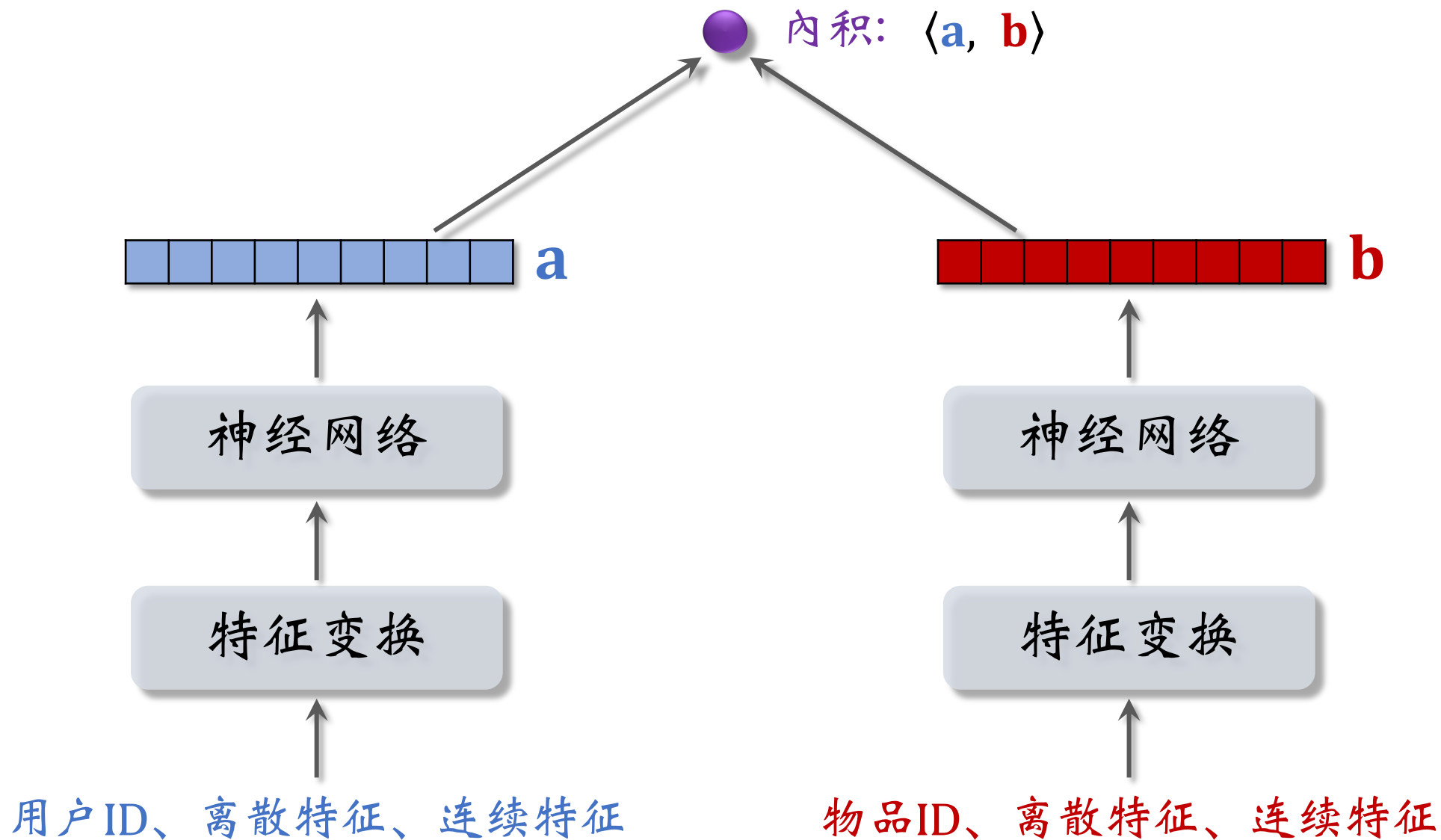


用户连续特征

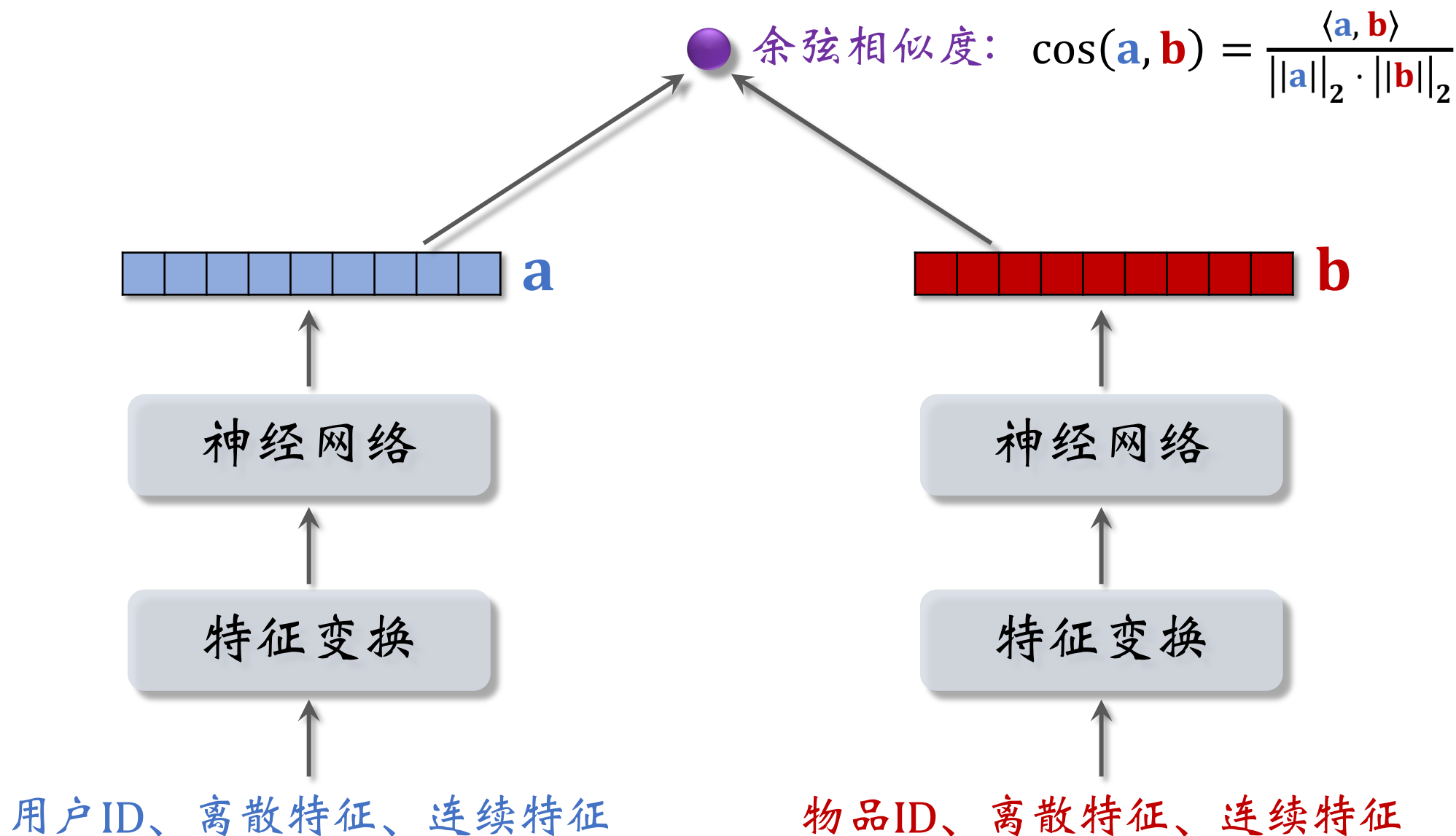




双塔模型



双塔模型



双塔模型的训练

- Pointwise：独立看待每个正样本、负样本，做简单的二元分类。
- Pairwise：每次取一个正样本、一个负样本 [1]。
- Listwise：每次取一个正样本、多个负样本 [2]。

参考文献：

1. Jui-Ting Huang et al. [Embedding-based Retrieval in Facebook Search](#). In *KDD*, 2020.
2. Xinyang Yi et al. [Sampling-Bias-Corrected Neural Modeling for Large Corpus Item Recommendations](#). In *RecSys*, 2019.

正负样本的选择

- 正样本：用户点击的物品。
- 负样本 [1, 2]：
 - 没有被召回的？
 - 召回但是被粗排、精排淘汰的？
 - 曝光但是未点击的？

参考文献：

1. Jui-Ting Huang et al. [Embedding-based Retrieval in Facebook Search](#). In *KDD*, 2020.
2. Xinyang Yi et al. [Sampling-Bias-Corrected Neural Modeling for Large Corpus Item Recommendations](#). In *RecSys*, 2019.

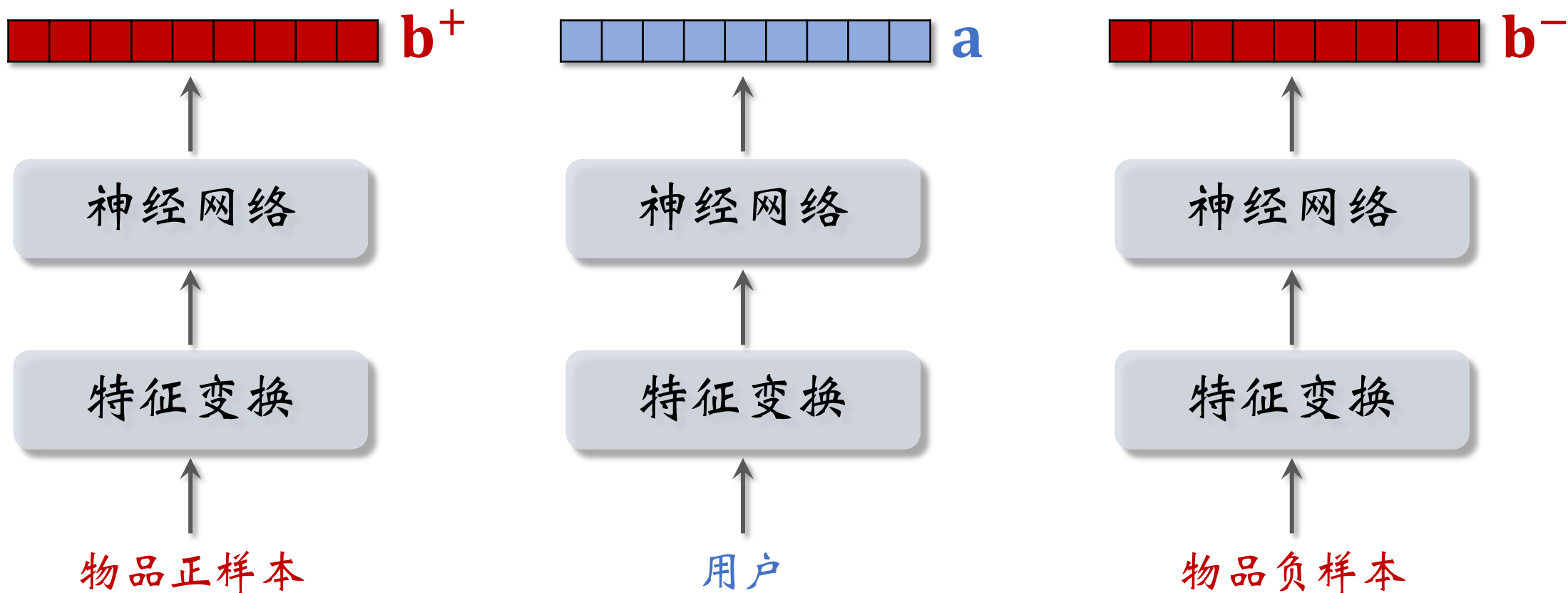
Pointwise训练

Pointwise训练

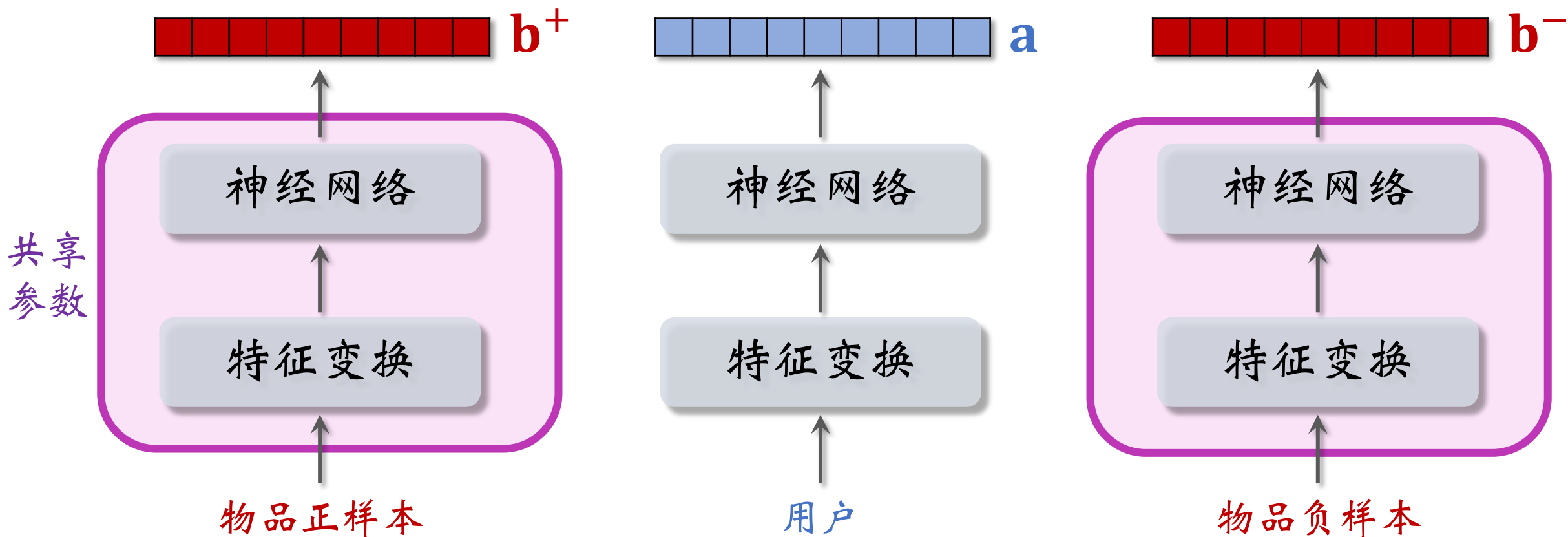
- 把召回看做二元分类任务。
- 对于正样本，鼓励 $\cos(\mathbf{a}, \mathbf{b})$ 接近 $+1$ 。
- 对于负样本，鼓励 $\cos(\mathbf{a}, \mathbf{b})$ 接近 -1 。
- 控制正负样本数量为 1:2 或者 1:3。

Pairwise训练

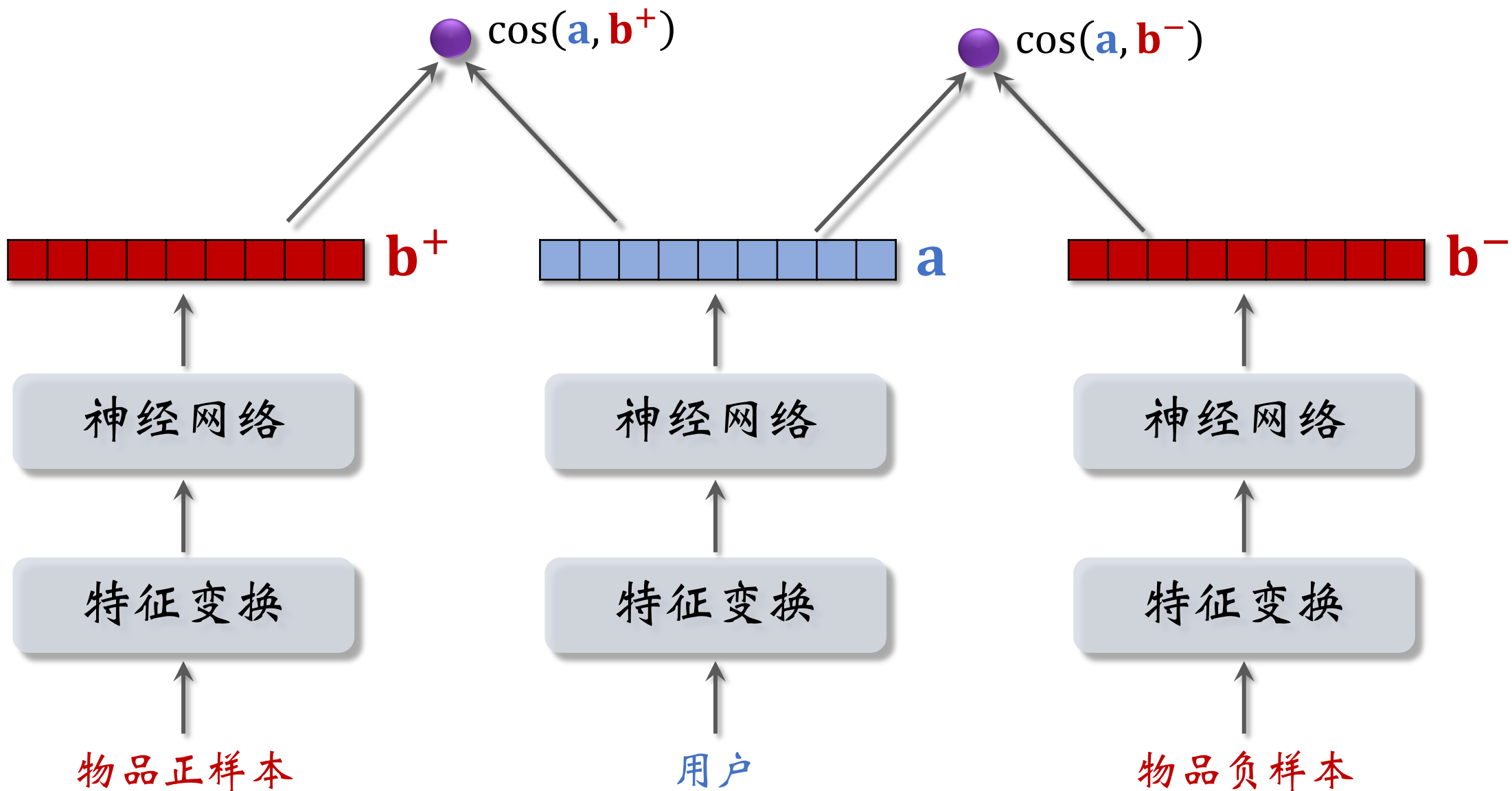
Pairwise训练



Pairwise训练



Pairwise训练



Pairwise训练

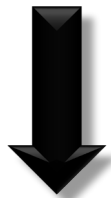
基本想法：鼓励 $\cos(\mathbf{a}, \mathbf{b}^+)$ 大于 $\cos(\mathbf{a}, \mathbf{b}^-)$

- 如果 $\cos(\mathbf{a}, \mathbf{b}^+)$ 大于 $\cos(\mathbf{a}, \mathbf{b}^-) + m$ ，则没有损失。
- 否则，损失等于 $\cos(\mathbf{a}, \mathbf{b}^-) + m - \cos(\mathbf{a}, \mathbf{b}^+)$ 。

Pairwise训练

基本想法：鼓励 $\cos(\mathbf{a}, \mathbf{b}^+)$ 大于 $\cos(\mathbf{a}, \mathbf{b}^-)$

- 如果 $\cos(\mathbf{a}, \mathbf{b}^+)$ 大于 $\cos(\mathbf{a}, \mathbf{b}^-) + m$ ，则没有损失。
- 否则，损失等于 $\cos(\mathbf{a}, \mathbf{b}^-) + m - \cos(\mathbf{a}, \mathbf{b}^+)$ 。



Triplet hinge loss:

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

Pairwise训练

基本想法：鼓励 $\cos(\mathbf{a}, \mathbf{b}^+)$ 大于 $\cos(\mathbf{a}, \mathbf{b}^-)$

Triplet hinge loss:

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

Pairwise训练

基本想法：鼓励 $\cos(\mathbf{a}, \mathbf{b}^+)$ 大于 $\cos(\mathbf{a}, \mathbf{b}^-)$

Triplet hinge loss:

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

Triplet logistic loss:

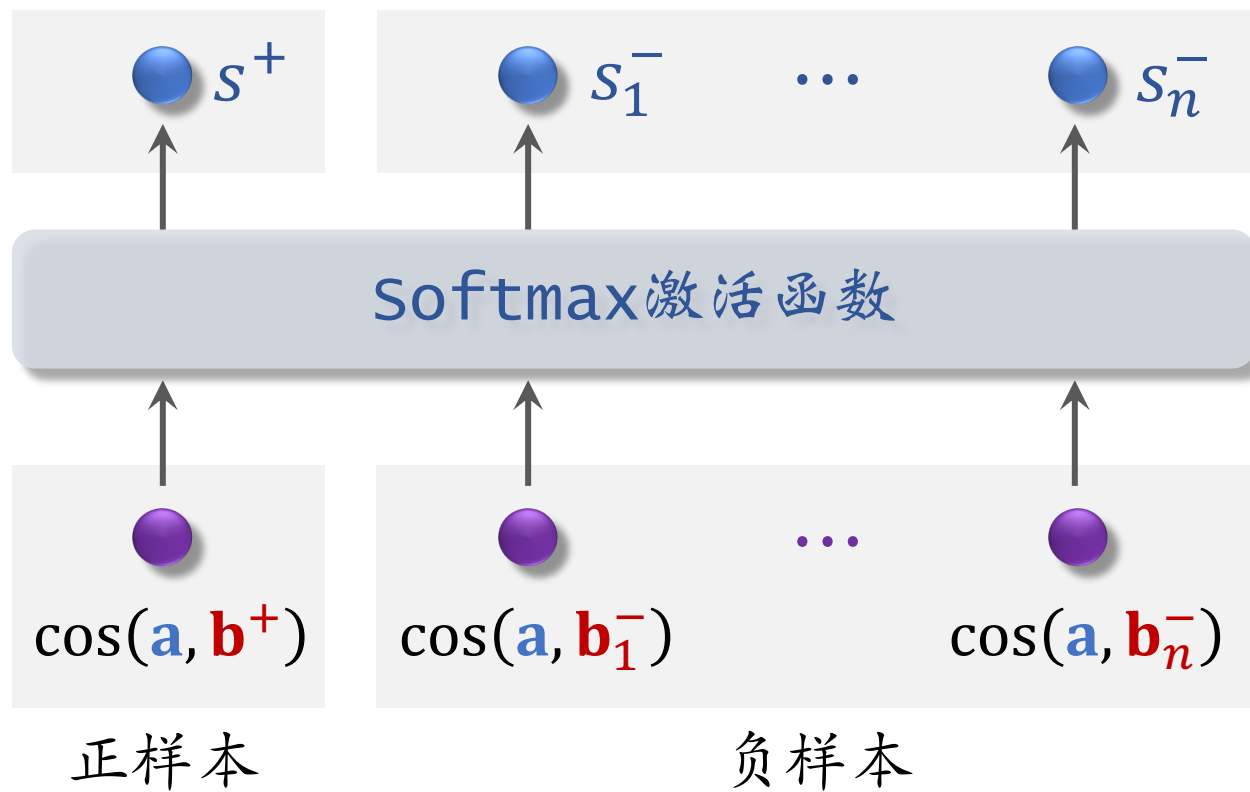
$$L(\mathbf{a}, \mathbf{b}^+, \mathbf{b}^-) = \log(1 + \exp[\sigma \cdot (\cos(\mathbf{a}, \mathbf{b}^-) - \cos(\mathbf{a}, \mathbf{b}^+))]).$$

Listwise训练

Listwise训练

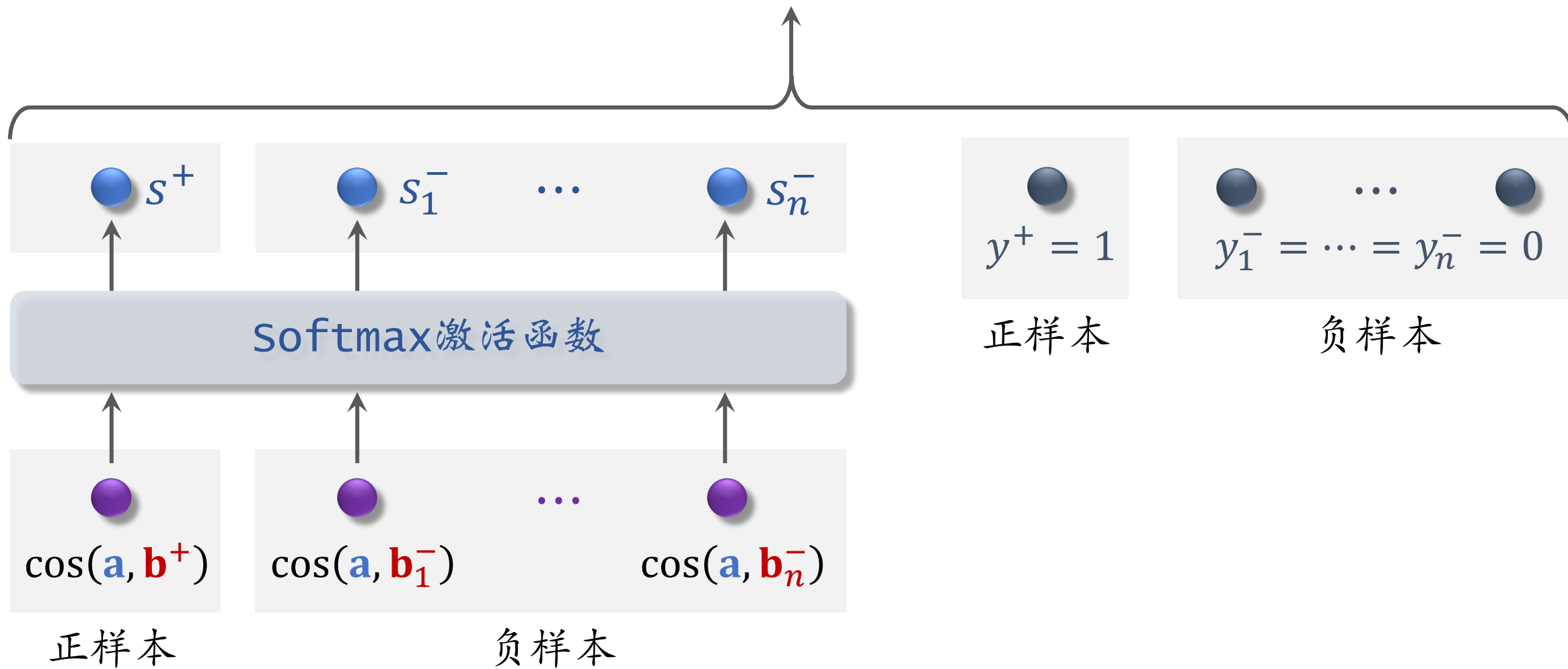
- 一条数据包含：
 - 一个用户，特征向量记作 \mathbf{a} 。
 - 一个正样本，特征向量记作 \mathbf{b}^+ 。
 - 多个负样本，特征向量记作 $\mathbf{b}_1^-, \dots, \mathbf{b}_n^-$ 。
- 鼓励 $\cos(\mathbf{a}, \mathbf{b}^+)$ 尽量大。
- 鼓励 $\cos(\mathbf{a}, \mathbf{b}_1^-), \dots, \cos(\mathbf{a}, \mathbf{b}_n^-)$ 尽量小。

Listwise训练



Listwise训练

$$\text{CrossEntropyLoss}(\mathbf{y}, \mathbf{s}) = -\log s^+$$

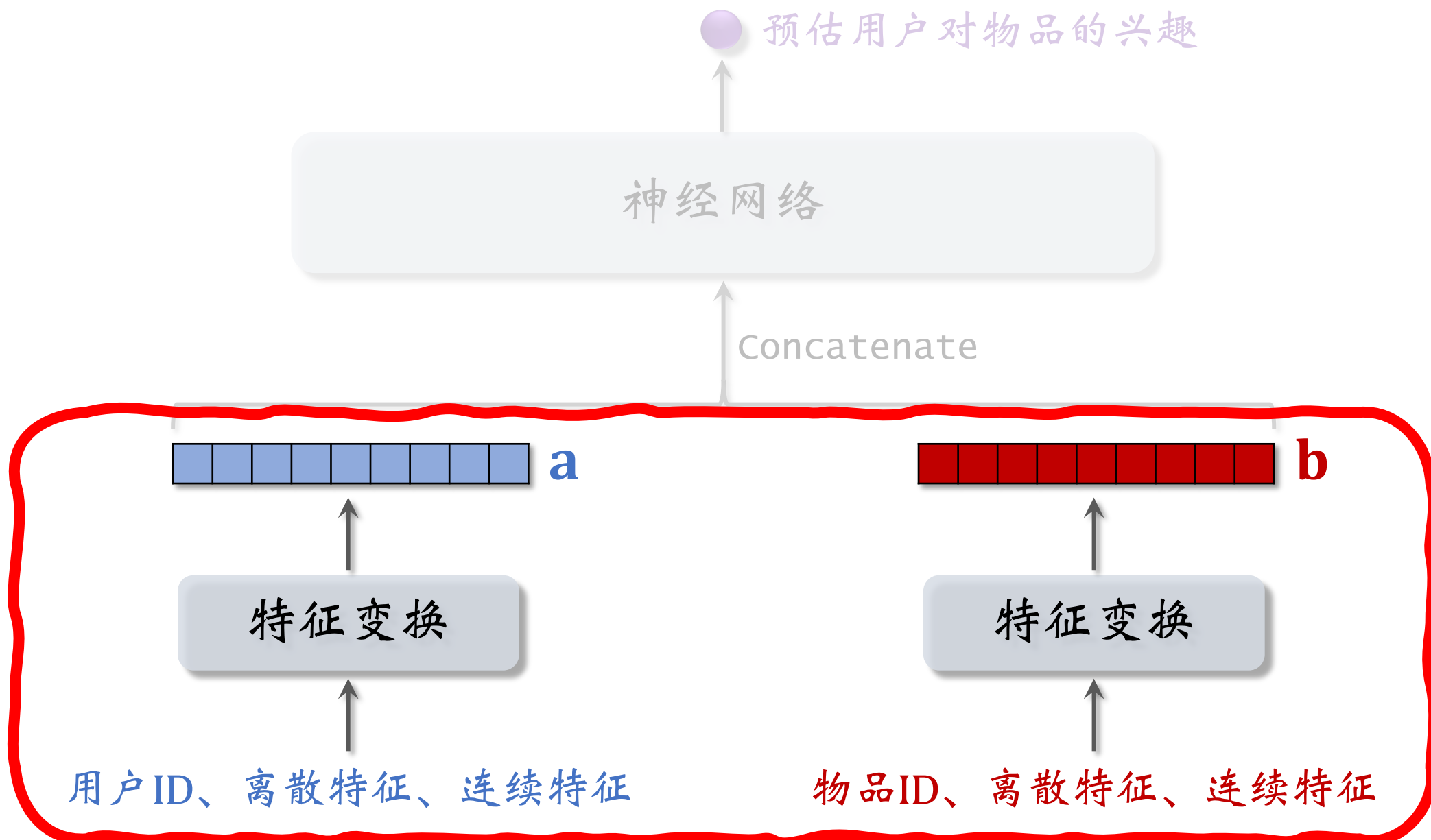


总结

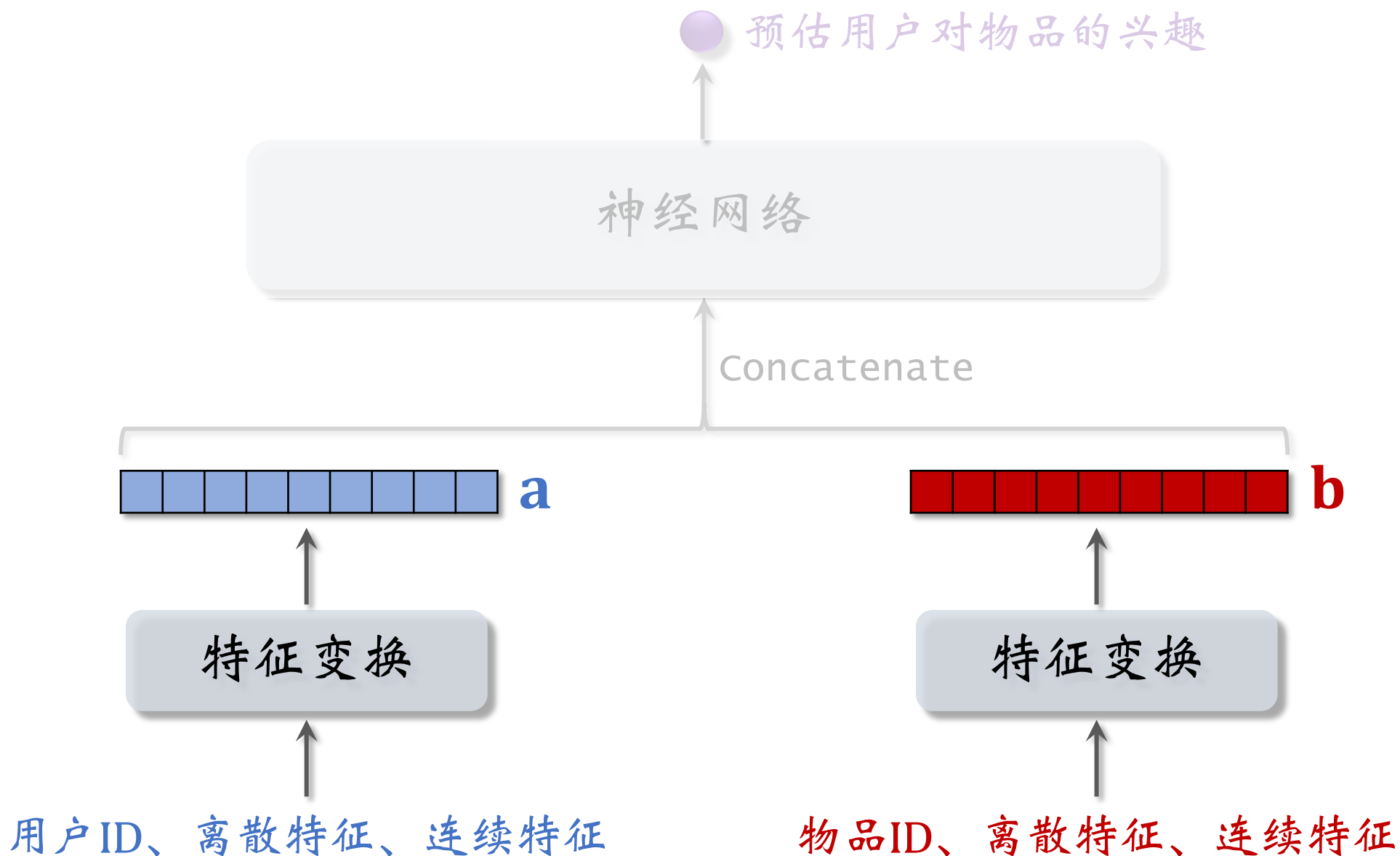
双塔模型

- 用户塔、物品塔各输出一个向量。
- 两个向量的余弦相似度作为兴趣的预估值。
- 三种训练方式：
 - Pointwise：每次用一个用户、一个物品（可正可负）。
 - Pairwise：每次用一个用户、一个正样本、一个负样本。
 - Listwise：每次用一个用户、一个正样本、多个负样本。

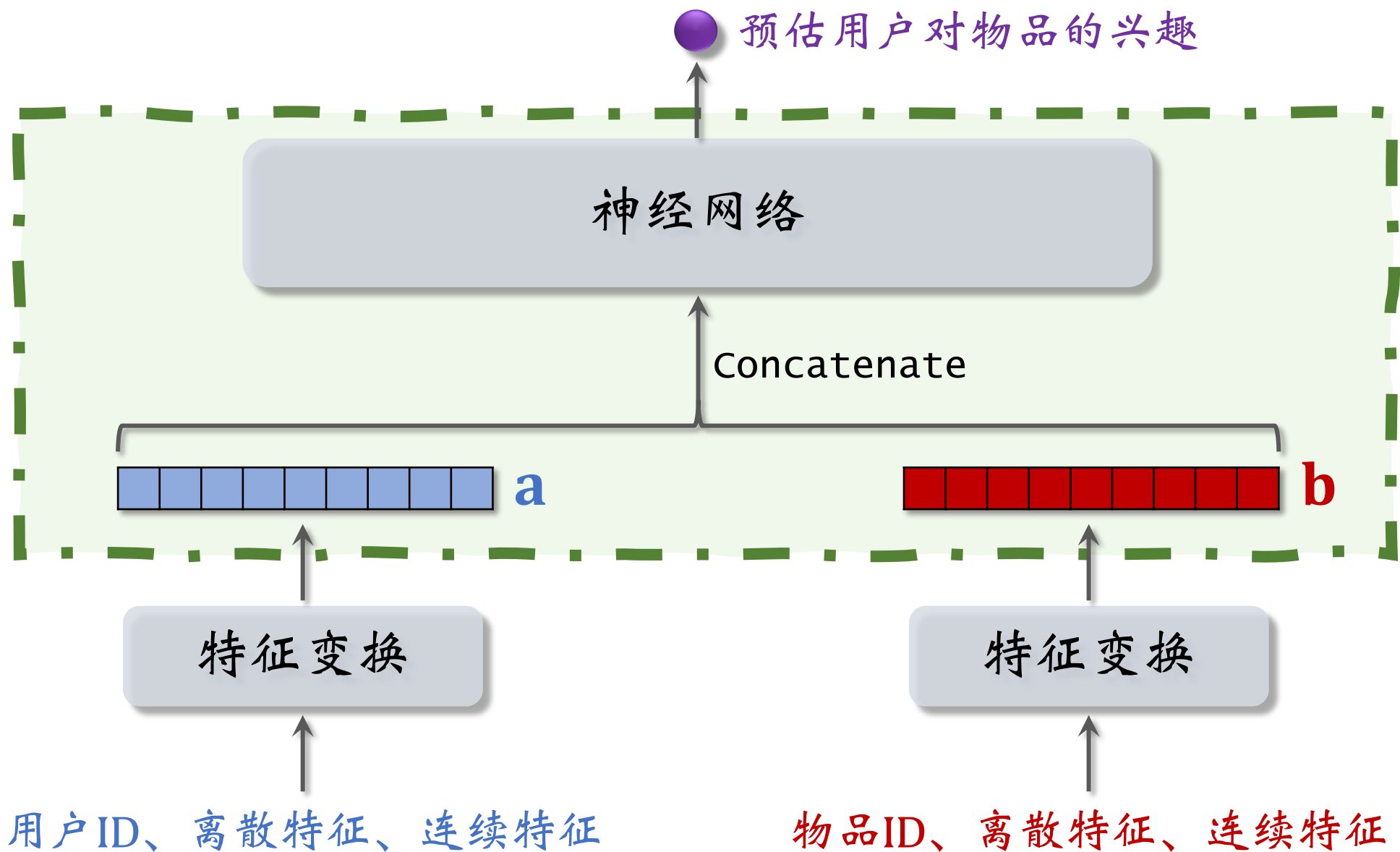
不适用于召回的模型



不适用于召回的模型



不适用于召回的模型



Thank You!

<http://wangshusen.github.io/>

长期招聘优秀的算法工程师

- 部门：小红书社区技术部。
- 方向：搜索、推荐。
- 职位：校招、社招、实习。
- 地点：上海、北京。
- 联系方式：`ShusenWang@xiaohongshu.com`