

# 涨指标的方法：召回

王树森

# 召回模型 & 召回通道

- 推荐系统有几十条召回通道，它们的召回总量是固定的。总量越大，指标越好，粗排计算量越大。
- 双塔模型（two-tower）和 item-to-item（I2I）是最重要的两类召回模型，占据召回的大部分配额。
- 有很多小众的模型，占据的配额很少。在召回总量不变的前提下，添加某些召回模型可以提升核心指标。
- 有很多内容池，比如30天物品、1天物品、6小时物品、新用户优质内容池、分人群内容池。
- 同一个模型可以用于多个内容池，得到多条召回通道。

# 双塔模型

# 改进双塔模型

方向1：优化正样本、负样本。

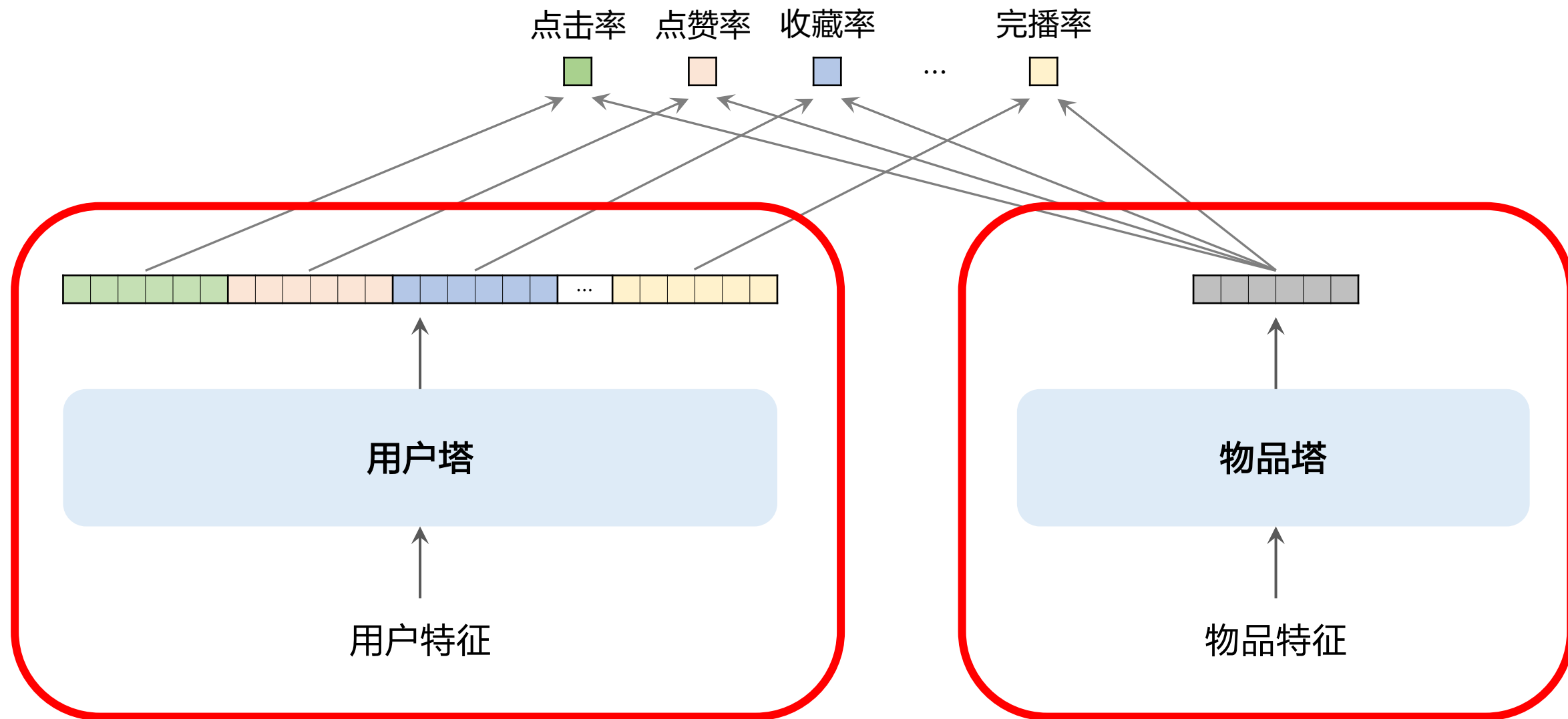
- 简单正样本：有点击的（用户，物品）二元组。
- 简单负样本：随机组合的（用户，物品）二元组。
- 困难负样本：排序靠后的（用户，物品）二元组。

# 改进双塔模型

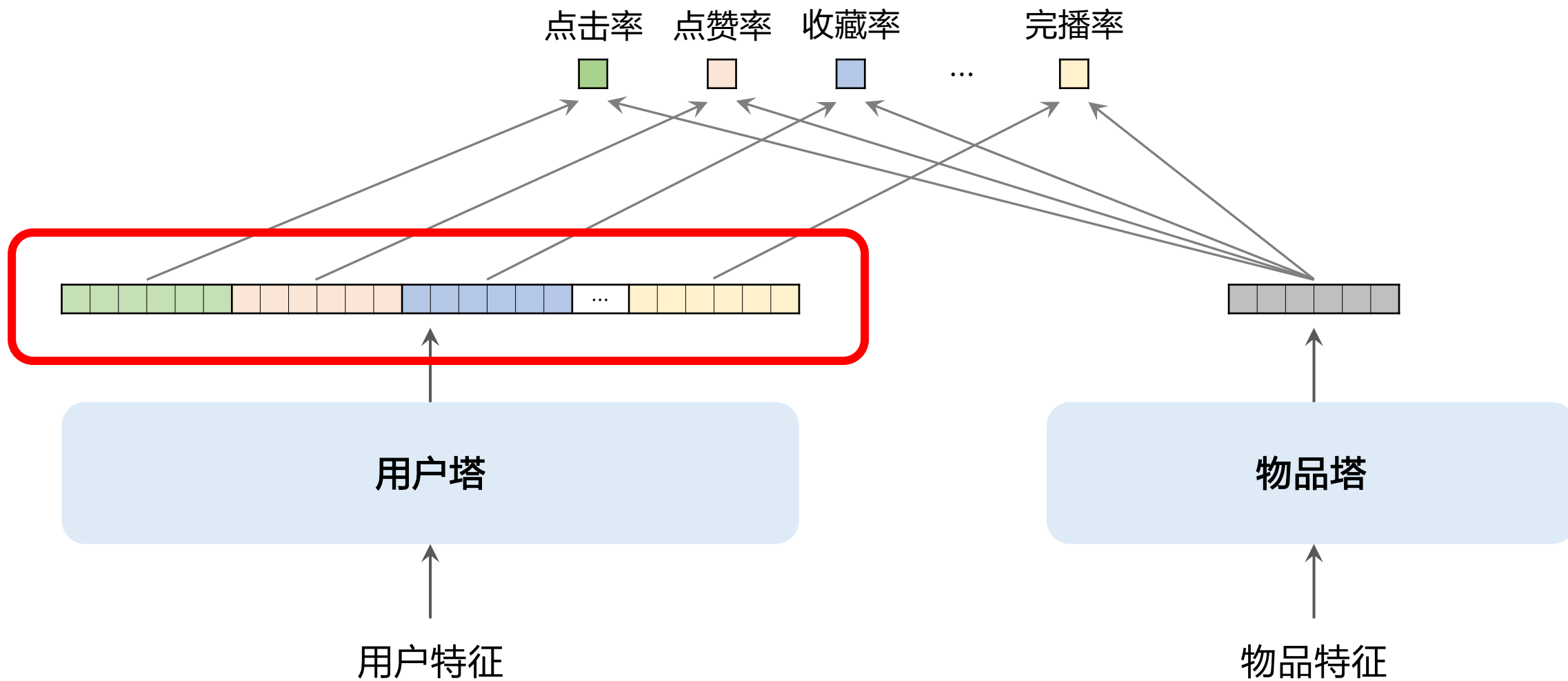
## 方向2：改进神经网络结构。

- Baseline：用户塔、物品塔分别是全连接网络，各输出一个向量，分别作为用户、物品的表征。
- 改进：用户塔、物品塔分别用 DCN 代替全连接网络。
- 改进：在用户塔中使用用户行为序列 (last-n) 。
- 改进：使用多向量模型代替单向量模型。（标准的双塔模型也叫单向量模型。）

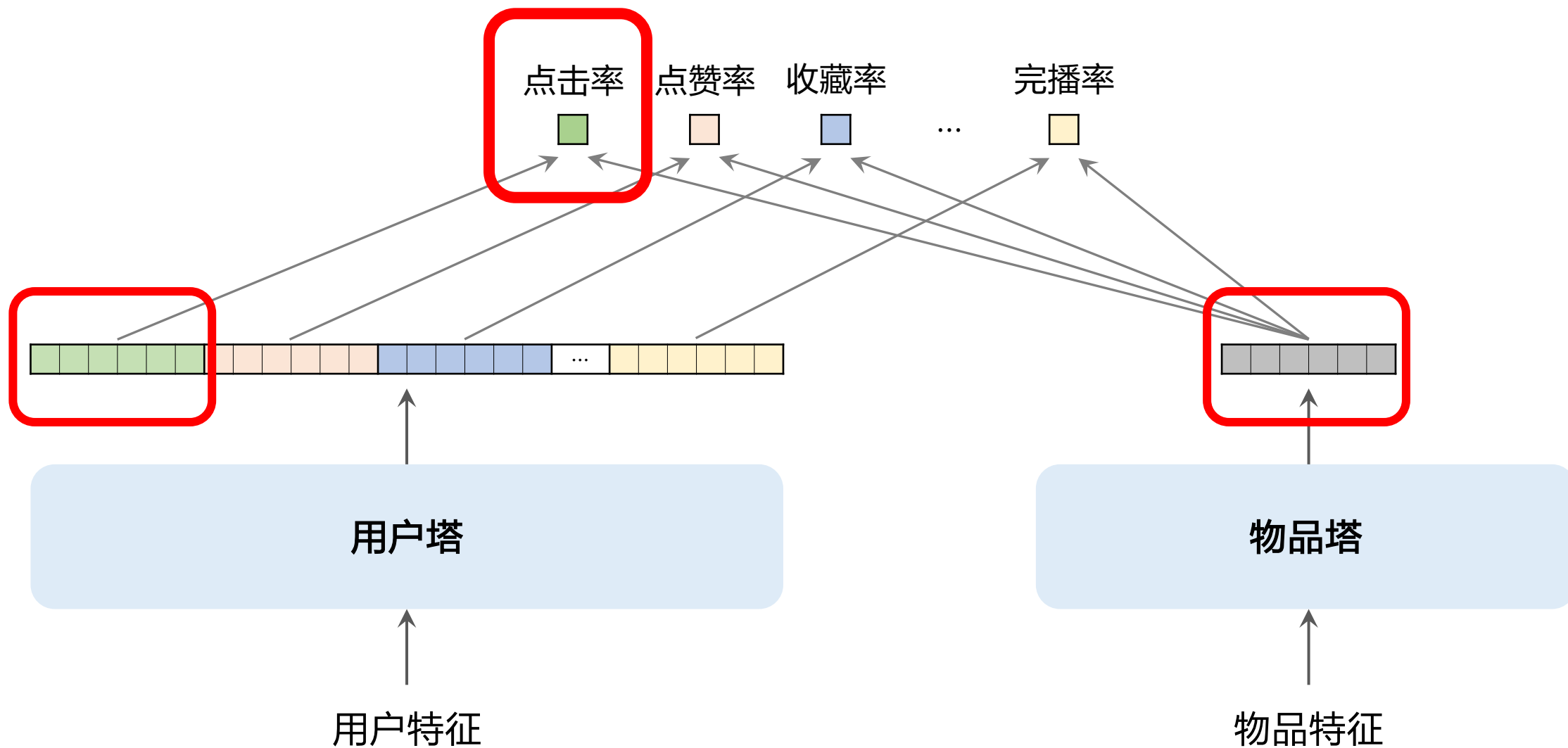
# 改进双塔模型



# 改进双塔模型

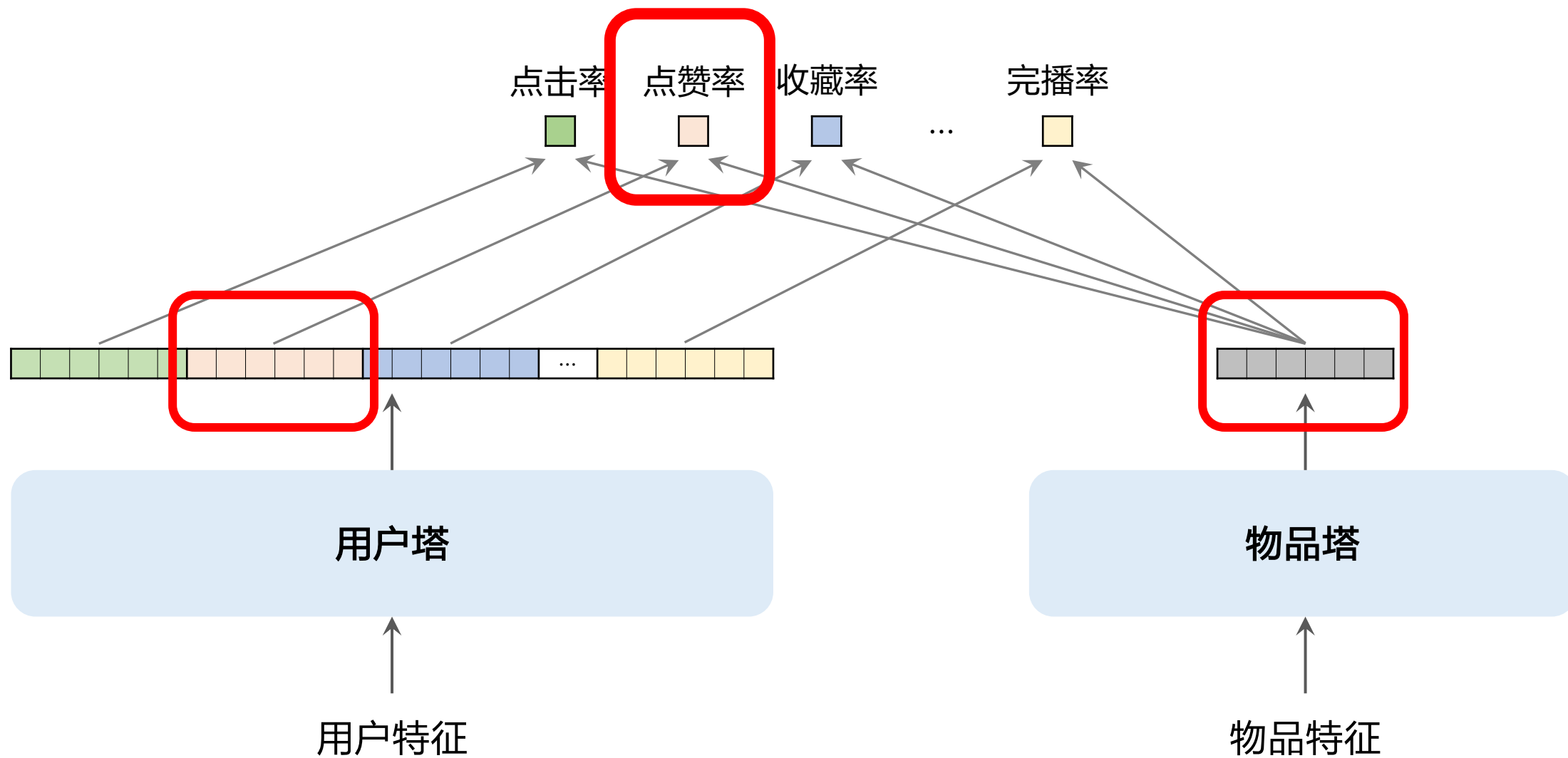


# 改进双塔模型

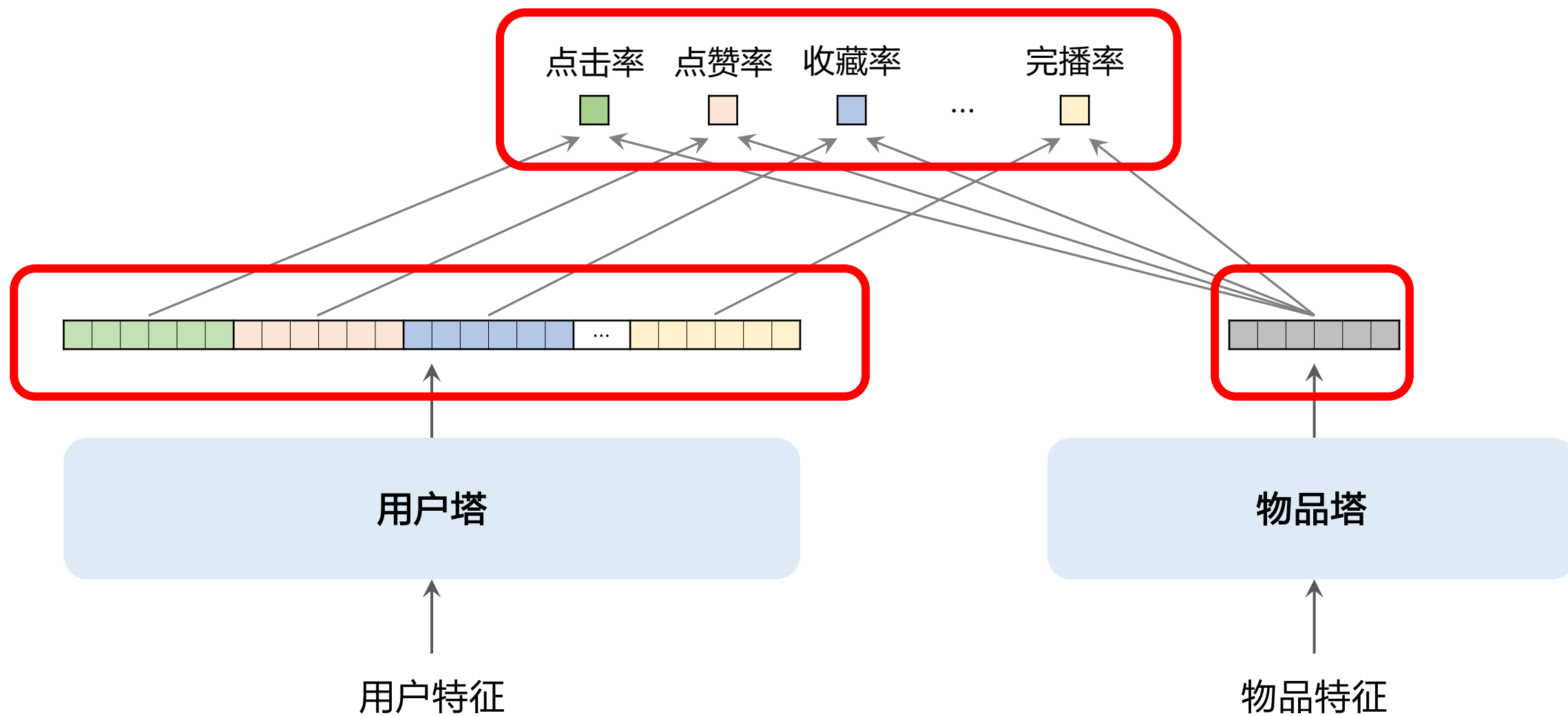




# 改进双塔模型



# 改进双塔模型



# 改进双塔模型

方向3：改进模型的训练方法。

- Baseline：做二分类，让模型学会区分正样本和负样本。
- 改进：结合二分类、batch 内负采样。（对于 batch 内负采样，需要做纠偏。）
- 改进：使用自监督学习方法，让冷门物品的 embedding 学得更好。

Item-to-Item (I2I)

# Item-to-Item (I2I)

- I2I 是一大类模型，基于相似物品做召回。
- 最常见的用法是 U2I2I ( $\text{user} \rightarrow \text{item} \rightarrow \text{item}$ )。
  - 用户  $u$  喜欢物品  $i_1$ （用户历史上交互过的物品）。
  - 寻找  $i_1$  的相似物品  $i_2$ （即 I2I）。
  - 将  $i_2$  推荐给  $u$ 。

# Item-to-Item (I2I)

- I2I 是一大类模型，基于相似物品做召回。
- 最常见的用法是 U2I2I ( $\text{user} \rightarrow \text{item} \rightarrow \text{item}$ )。
- 如何计算物品相似度？
- 方法1：ItemCF 及其变体。
  - 一些用户同时喜欢物品  $i_1$  和  $i_2$ ，则认为  $i_1$  和  $i_2$  相似。
  - ItemCF、Online ItemCF、Swing、Online Swing 都是基于相同的思想。
  - 线上同时使用上述 4 种 I2I 模型，各分配一定配额。

# Item-to-Item (I2I)

- I2I 是一大类模型，基于相似物品做召回。
- 最常见的用法是 U2I2I ( $\text{user} \rightarrow \text{item} \rightarrow \text{item}$ )。
- 如何计算物品相似度？
- 方法1：ItemCF 及其变体。
- 方法2：基于物品向量表征，计算向量相似度。（双塔模型、图神经网络均可计算物品向量表征。）

# 小众的召回模型



# 类似 I2I 的模型

- U2U2I (user  $\rightarrow$  user  $\rightarrow$  item) : 已知用户  $u_1$  与  $u_2$  相似，且  $u_2$  喜欢物品  $i$ ，那么给用户  $u_1$  推荐物品  $i$ 。
- U2A2I (user  $\rightarrow$  author  $\rightarrow$  item) : 已知用户  $u$  喜欢作者  $a$ ，且  $a$  发布物品  $i$ ，那么给用户  $u$  推荐物品  $i$ 。
- U2A2A2I (user  $\rightarrow$  author  $\rightarrow$  author  $\rightarrow$  item) : 已知用户  $u$  喜欢作者  $a_1$ ，且  $a_1$  与  $a_2$  相似， $a_2$  发布物品  $i$ ，那么给用户  $u$  推荐物品  $i$ 。

# 更复杂的模型

- Path-based Deep Network (PDN) [1]
- Deep Retrieval [2]
- Sparse-Interest Network (SINE) [3]
- Multi-task Multi-view Graph Representation Learning (M2GRL) [4]

## 参考文献

1. Li et al. [Path-based Deep Network for Candidate Item Matching in Recommenders](#). In *SIGIR*, 2021.
2. Gao et al. [Learning an end-to-end structure for retrieval in large-scale recommendations](#). In *CIKM*, 2021.
3. Tan et al. [Sparse-interest network for sequential recommendation](#). In *WSDM*, 2021.
4. Wang et al. [M2GRL: A multitask multi-view graph representation learning framework for web-scale recommender systems](#). In *KDD*, 2020.

# 总结：改进召回模型

- 双塔模型：优化正负样本、改进神经网络结构、改进训练的方法。
- I2I 模型：同时使用 ItemCF 及其变体、使用物品向量表征计算物品相似度。
- 添加小众的召回模型，比如 PDN、Deep Retrieval、SINE、M2GRL 等模型。
- 在召回总量不变的前提下，调整各召回通道的配额。（可以让各用户群体用不同的配额。）

**Thank You!**

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