

本章重难点梳理

本章我们学习了推荐系统中举足轻重的召回层，下面让我们来快速回顾一下本章的内容，温故而知新。

一、召回层的作用

推荐系统本身就是一个信息过滤系统，召回层作为其中的第一层筛选器，承担着将候选集的规模从数百万缩小至几百或几千的作用。

召回层通常具有如下特点：模型复杂度比较低，处理速度快，且所用特征较少。

在实践中，通常会使用多策略同时进行召回，从而保证用户所喜欢的物品尽可能地被全面捕获。比如平台内最热门的物品、近期点击次数激增的物品等策略都不失为一个优秀的召回策略。

二、Embedding技术

Embedding正如其字面含义，是希望对于每一个物品都找到一个对其的数字“表示”。使得相似的物品的“表示”比较接近，甚至互相之间还可以做代数运算。这样我们就可以用这组数字来代表候选集中的图片，从而对其进行召回了。

一个经典的Embedding的例子就是word2vec，如果我们将“国王”，“男人”，“女人”分别表示为下面的向量，那他们就可以通过计算得出“女王”的表示，是不是很神奇呢？

king - man + woman ≈ queen

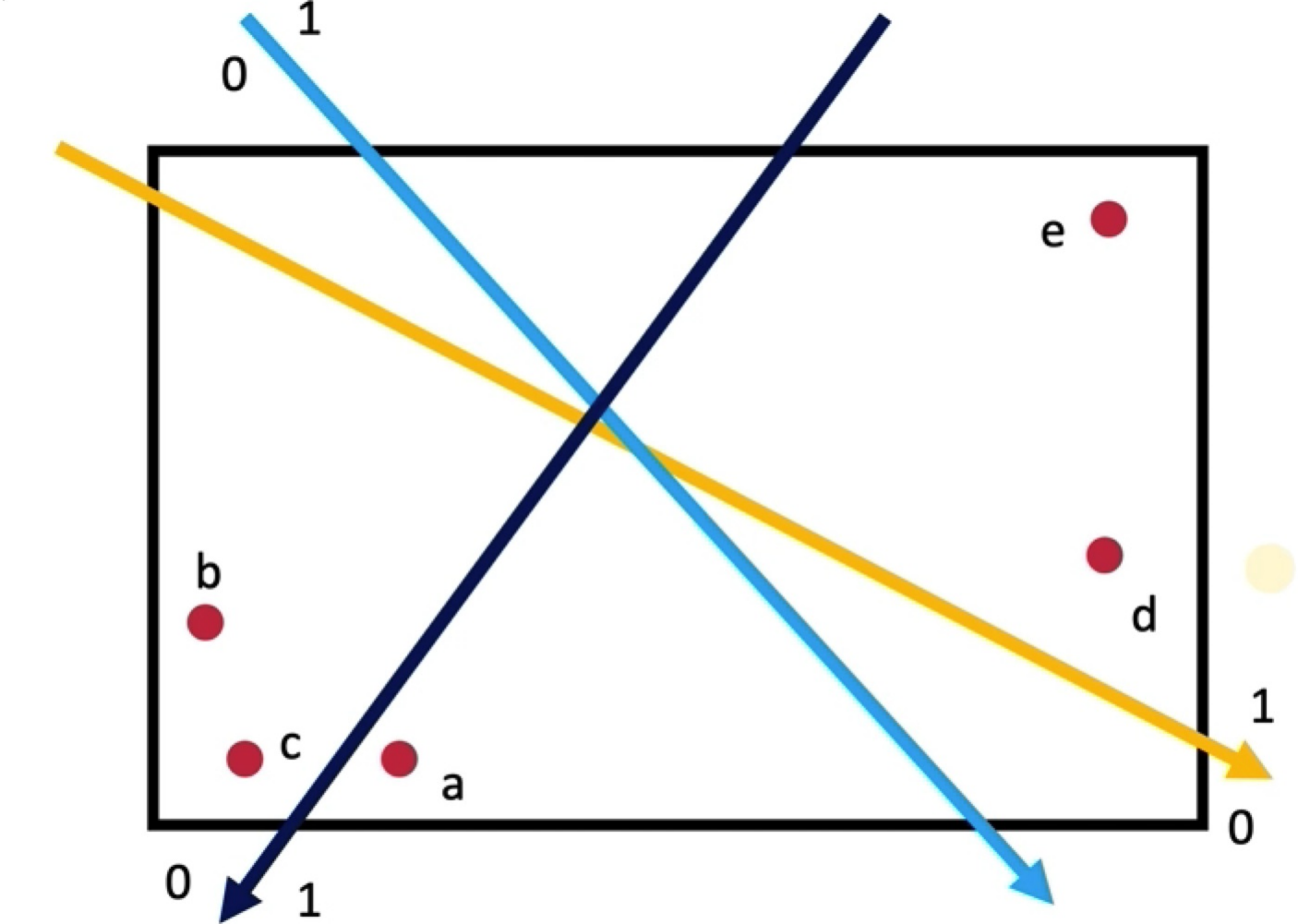


三、最近邻搜索

当我们通过计算获得了每个物品的Embedding表示之后，最常遇到的一个问题就是在给定某个目标物品时，如何找出何其最类似的n个物品？

这一类问题被称为最近邻搜索问题，经过多年的发展，这个问题在业界也有诸多的解决方案。

目前最常用在推荐系统中的是近似最近邻检索算法（ANN）。由于我们并不需要精确的知道某个物品的n个最相近的物品是什么，只要大致结果正确就行，因此可以采用许多办法来加速计算、减少时间和空间复杂度。在这方面Facebook团队开源了FAISS，提供了诸多开箱即用的模型。



在学习了如上的知识后，我们利用Embedding技术和ANN算法，结合其他若干高效的召回策略共同实现了课程项目中的排序模块。当然这还远远没有结束，在后续课程中，我们还会学习在实践中如何增强召回层的实时性、如何解决物品冷启动等问题。