

068 | 基于隐变量的模型之三：分解机

2018-03-09 洪亮劼

AI技术内参

[进入课程 >](#)



讲述：初明明

时长 05:04 大小 2.33M



周三我们分享了“基于回归的隐变量模型”，这是在基本的矩阵分解基础上衍生出来的一类模型。这种模型把显式特性和隐变量结合起来，对解决“冷启动”问题有一定作用。

今天，我们来介绍一种叫作“**分解机**”（Factorization Machines）的推荐技术。这个模型是从基于回归的隐变量模型中衍生出来的，已成为了主流的推荐模型。

矩阵分解和基于回归的隐变量模型存在哪些问题？

在介绍分解机的基本原理之前，我们先来回顾一下从“矩阵分解”到“基于回归的隐变量模型”的一个发展脉络。

首先，矩阵分解主要解决了两个问题，那就是从一个大矩阵降维到两个小矩阵，并且寄希望这两个小矩阵能够抓住用户和物品的相关度。

然而，单纯的矩阵分解无法融入很多用户和物品的特性，这就引导我们开发出了基于回归的矩阵分解。所谓的回归部分，也就是从显式特性出发，建立从显式特性到隐变量之间关系的流程，从而使我们能够把更多的信号放进模型中。

在一定程度上，基于回归的隐变量模型实现了把显式变量和隐变量结合的目的，但是这类模型的学习过程非常麻烦。实际上，因为这类模型复杂的训练流程，其在实际应用中并不常见。

那么，有没有其他思路来统一显式变量和隐变量的处理方式呢？

分解机的基本原理

分解机 [1] 是学者斯特芬·润顿（Steffen Rendle）在德国康斯坦扎大学任教期间开发出来的推荐模型。斯特芬后来加入谷歌，分解机是他的代表作品。

分解机结合了“基于内容的推荐系统”和“基于回归的隐变量模型”的一些基本思想。

基于内容的推荐系统，其核心就是认为需要预测的变量（这里我们依然讨论评分）是所有显式变量的一个回归结果。分解机直接借鉴了这一点，也就是说，分解机的输入是所有的显式变量。

实际上，分解机在对待显式变量的手法上更进了一步，那就是不仅直接对显式变量进行建模，还对显示变量的两两关系进行建模。当然，在原始的论文中，分解机其实还可以对更高维的关系进行建模，我们这里局限在两两关系上。

什么意思呢？比如说我们有一个用户特性，是用户的年龄；我们有一个物品特性，是物品的种类。那么，普通的回归模型，就是把用户的年龄和物品的种类直接当作特性输入到模型中。而对于分解机来说，这只是第一步。

第二步，分解机是把这两个特性的数值进行乘积，当作一个新的特性，然后进一步处理这种两两配对的关系。把原始特性进行两两配对是构建模型的一种重要的方法，特别是对于非深度学习模型，需要自己做特征工程的模型。

两两配对的特征有什么好处呢？好处就是可以对两种特性的交互信息进行建模。举个例子，如果我们特别在意某个年龄段的用户在某种商品类别中的评分，那么，把这两个特性相乘，从而抓取到这个交互信息，是一个非常有效的手段。

但是，两两配对存在什么问题吗？一个问题就是**特性空间会急速增长**。如果我们有一个 100 维的用户特性向量，然后有一个 100 维的物品特性向量，那对于两两配对的特征，就是 100 乘 100 这个数量级的。另一个更严重的问题就是，如果我们的单独特性中，有一些是“类别特性”（Categorical Feature），那么在两两配对之后就会产生大量的 0，从而变成一个巨大的稀疏矩阵。

如何解决这个问题呢？**分解机利用了矩阵分解的降维思路**。就是说，我们不对一个稀疏矩阵直接建模，而是把这个稀疏矩阵分解之后再行建模。具体到上面这个例子，就是先假定，所有特性都对应一个隐变量向量，两个显式特性的乘积是两个特性的隐变量的点积。也就是说，我们把两个显式特性的乘积分解为了两个向量的乘积。这样，我们就不需要直接表示原来的稀疏矩阵。

在这样的思路下，分解机成功地把隐变量和显式变量结合到了一起。当我们的显式特性仅仅是用户 ID 和物品 ID 的时候，分解机的表达退回了最原始的矩阵分解。也就是说，矩阵分解其实可以表达成为特性的两两作用矩阵的分解。

在原始的论文中，作者还用分解机模拟了好几种流行的模型，我们这里就不复述了。

虽然也是为了建立从显式特性到隐变量的关系，但是对比基于回归的矩阵分解而言，分解机的训练过程大大简化了。在实际应用中，我们经常使用“**随机梯度下降**”（SGD, Stochastic Gradient Descent）来对分解机直接进行求解。

在最近几年的 Kaggle 比赛中以及一些工业级的应用中，分解机凭借其简单易用的特点，成为了很多产品的核心算法。

小结

今天我为你讲了隐变量模型中分解机的基本原理。

一起来回顾下要点：第一，我们简要介绍了矩阵分解的一些问题；第二，我们详细介绍了分解机的基本原理；第三，我们简要讲了如何求解分解机。

最后，给你留一个思考题，分解机能够解决“冷启动”的问题吗？

欢迎你给我留言，和我一起讨论。

参考文献

1. Steffen Rendle. Factorization Machines with libFM. ACM Trans. Intell. Syst. Technol. 3, 3, Article 57 (May 2012), 22 pages, 2012.

 极客时间

AI 技术内参

你的360度人工智能信息助理

洪亮劼
Etsy 数据科学主管
前雅虎研究院资深科学家



新版升级：点击「 请朋友读」，10位好友免费读，邀请订阅更有**现金**奖励。

© 版权归极客邦科技所有，未经许可不得传播售卖。页面已增加防盗追踪，如有侵权极客邦将依法追究其法律责任。

上一篇 067 | 基于隐变量的模型之二：基于回归的矩阵分解

下一篇 069 | 高级推荐模型之一：张量分解模型

精选留言 (2)

 写留言



林彦

2018-03-12

 2

分解机能比传统的方法更好地解决“冷启动”的问题。分解机是基于显式变量两两配对来建模，我的理解是只要对应的2个显式变量在所有的数据集中能计算出值即可。如果某个物品或某个用户的评分是缺失的，我们可以用显式变量的整体分布，如文中的例子的某个年龄段和某类商品的显式特性来计算。



Wesley

2019-02-27

👍 1

矩阵分解可以无需负样本就能训练. 分解机是否一定需要负样本来训练?

展开 ∨