022 | 机器学习排序算法：配对法排序学习

2017-11-22 洪亮劼

AI技术内参 进入课程 >

讲述：初明明
时长 08:37 大小 3.95M

周一的文章里我分享了最基本的单点法排序学习（Pointwise Learning to Rank）。这个思路简单实用，是把经典的信息检索问题转化为机器学习问题的第一个关键步骤。简单回顾一下，我们介绍了在测试集里使用 NDCG（Normalized Discounted Cumulative Gain），在某个 K 的位置评价“精度”（Precision）和“召回”（Recall），以这些形式来评估排序算法。

你可以看到，单点法排序学习算法的模式和我们最终需要的结果中间还存在明显差距。这个差距并不是算法好坏能够决定的，而是算法所要优化的目标，也就是单个数据点是否相关，和我们的最终目的，一组结果的 NDCG 排序最优之间的结构化区别。这个结构化区别激发研究者们不断思考，是不是有其他的方法来优化排序算法。

今天我就来讲从单点法引申出来的“配对法”排序学习（Pairwise Learning to Rank）。相对于尝试学习每一个样本是否相关，配对法的基本思路是对样本进行两两比较，从比较中学习排序，离真正目标又近了一步。

配对法排序学习的历史

当人们意识到用机器学习来对排序进行学习，从文档与文档之间的相对关系入手，也就是配对法，就成了一个非常火热的研究方向。机器学习排序这个领域持续活跃了 10 多年，在此期间很多配对法排序算法被提出，下面我就说几个非常热门的算法。

2000 年左右，研究人员开始利用支持向量机（SVM）来训练排序算法，来自康奈尔的索斯藤·乔基姆斯（Thorsten Joachims）就构建了基于特征差值的**RankSVM**，一度成为配对法排序学习的经典算法。索斯藤我们前面讲过，他获得了今年的 KDD 时间检验奖。

2005 年，当时在雅虎任职的研究人员郑朝晖等人，开始尝试用**GBDT**（Gradient Boosting Decision Tree，梯度提升决策树）这样的树模型来对文档之间的两两关系进行建模。郑朝晖后来成为一点资讯的联合创始人。

2005 年，微软的学者克里斯·博格斯（Chris Burges）等人，开始使用神经网络训练**RankNet**文档之间两两关系的排序模型。这是最早使用深度学习模型进行工业级应用的尝试。这篇论文在 2015 年获得了 ICML 2015（International Conference on Machine Learning，国际机器学习大会）的 10 年“经典论文奖”。

配对法排序学习详解

在介绍配对法排序学习的中心思路之前，我们先来重温一下测试集的测试原理。总体来说，测试的原理和单点法一样，都是要考察测试集上，对于某一个查询关键字来说，某一组文档所组成的排序是否是最优的。

比如，对于某一个查询关键字，我们针对排序产生的“顶部的 K”个文档进行评估，首先查看精度（Precision），即在所有算法已经判断是相关的文档中，究竟有多少是真正相关的；其次看召回（Recall），即所有真正相关的文档究竟有多少被提取了出来。当然，还有 F1 值，也就是精度和召回“和谐平均”（Harmonic Mean）的取值，一个平衡精度和召回的重要指标。需要再次说明的是， 精度、召回以及 F1 值都是在二元相关信息的标签基础上定义的。

如果需要利用五级相关信息定义，也就是通常所说的“最相关”、“相关”、“不能确定”到“不相关”、“最不相关”，那么就需要用类似于 NDCG 这样的评价指标。NDCG 的假设是，在一个排序结果里，相关信息要比不相关信息排得更高，最相关信息需要排在最上面，最不相关信息需要排在最下面。任何排序结果一旦偏离了这样的假设，就会受到“扣分”或者“惩罚”。

在清楚了测试集的情况后，再回过头来看一看训练集的设置问题。在今天文章一开篇的时候，我就提到了单点法对于排序学习的“目标不明确”的问题。其实从 NDCG 的角度来看也好，基于顶部 K 的精度或者召回的角度来看也好，都可以看出，**对于一个查询关键字来说，最重要的其实不是针对某一个文档的相关性是否估计得准确，而是要能够正确估计一组文档之间的“相对关系”**。只要相对关系估计正确了，那么从排序这个角度来说，最后的结果也就准确了。理解这一个观点，对于深入理解排序和普通的分类之间的区别至关重要。

那么，如何从单点建模再进一步呢?

很显然，在排序关系中，一个关键关系就是每两个文档之间的比较，也就是我们通常所说的两两关系。试想一下，如果针对某一个查询关键字而言，有一个完美的排序关系，然后通过这个完美的排序关系，可以推导出文档之间的两两相对关系，再从这些相对关系中进行学习，从而可以进一步对其他查询关键字进行排序。

注意，**在这样的架构下，训练集的样本从每一个“关键字文档对”变成了“关键字文档文档配对”**。也就是说，每一个数据样本其实是一个比较关系。试想，有三个文档：A、B 和 C。完美的排序是“B>C>A”。我们希望通过学习两两关
系“B>C”、“B>A”和“C>A”来重构“B>C>A”。

这里面有几个非常关键的假设。

第一，我们可以针对某一个关键字得到一个完美的排序关系。在实际操作中，这个关系可以通过五级相关标签来获得，也可以通过其他信息获得，比如点击率等信息。然而，这个完美的排序关系并不是永远都存在的。试想在电子商务网站中，对于查询关键字“哈利波特”，有的用户希望购买书籍，有的用户则希望购买含有哈利波特图案的 T 恤，显然，这里面就不存在一个完美排序。

第二，我们寄希望能够学习文档之间的两两配对关系从而“重构”这个完美排序。然而，这也不是一个有“保证”的思路。用刚才的例子，希望学习两两关

系“B>C”、“B>A”和“C>A”来重构完美排序“B>C>A”。然而，实际中，这三个两两关系之间是独立的。特别是在预测的时候，即使模型能够正确判
断“B>C”和“C>A”，也不代表模型就一定能得到“B>A”。注意，这里的关键是“一定”，也就是模型有可能得到也有可能得不到。两两配对关系不能“一定”得到完美排序，这个结论其实就揭示了这种方法的不一致性。也就是说，我们并不能真正保证可以得到最优的排序。

第三，我们能够构建样本来描述这样的两两相对的比较关系。一个相对比较简单的情况，认为文档之间的两两关系来自于文档特征（Feature）之间的差异。也就是说，可以利用样本之间特征的差值当做新的特征，从而学习到差值到相关性差异这样的一组对应关系。

我前面提到的 RankSVM 就是这样的思路。RankSVM 从本质上来说其实还是 SVM，也就是支持向量机，只不过建模的对象从单一文档变成了文档的配对。更加复杂的模型，比如 GBRank，就是通过树的聚合模型 GBDT 来对文档之间的关系直接建模，希望通过函数值的差值来表达文档的相关性差异。

需要注意的是，**配对法排序学习特别是在测试集预测的时候，可能会有计算复杂度的问题。**因为原则上，必须要对所有的两两关系都进行预测。现实中，如果是基于线性特征的差值来进行样本构造的话，那么测试还可以回归到线性复杂度的情况。而用其他方法，就没那么幸运了。有很多计算提速或者是逼近算法为两两比较排序在实际应用中提供了可能性。

小结

今天我为你讲了文档检索领域基于机器学习的配对法排序学习。你可以看到，和单点法一样，整个问题的设置和传统的文字搜索技术有本质的区别，但在对文档之间关系的建模上，又比单点法前进了一大步 。

一起来回顾下要点：第一，在火热的机器学习排序研究中，提出了很多配对法排序算法，比如 RankSVM、GBDT 和 RankNet。第二，配对法排序学习测试集的测试原理和单点法一致，我们可以查看精度、召回和 F1 值，或者利用五级相关信息。第三，针对单点法对于排序学习的“目标不明确”问题，配对法排序学习有不一样的训练集设置，在这个基础上，我介绍了三个关键假设。

最后，给你留一个思考题，有没有什么办法可以把单点法和配对法结合起来呢？

欢迎你给我留言，和我一起讨论。

参考文献

1. Zhaohui Zheng, Keke Chen, Gordon Sun, and Hongyuan Zha. A regression framework for learning ranking functions using relative relevance judgments. Proceedings of the 30th annual international ACM SIGIR conference on research and development in information retrieval, 287-294，2007.
2. Thorsten Joachims. Optimizing search engines using clickthrough data. Proceedings of the eighth ACM SIGKDD international conference on knowledge discovery and data mining，133-142，2002.

上一篇　021 | 机器学习排序算法：单点法排序学习

下一篇　023 | 机器学习排序算法：列表法排序学习

精选留言 (2)

写留言

yaolixu
2018-11-08

1

洪老师，以配对法为基础，把单点法的特征作为配对法输入的一部分。但是，感觉应该有更高大上的结合方法？？

白杨

2018-05-16

1

可以把单点的输出，作为作为配对法的输入特征，两两之间作差。

展开