

023 | 机器学习排序算法：列表法排序学习

2017-11-24 洪亮劼

AI技术内参

[进入课程 >](#)



讲述：初明明

时长 08:06 大小 3.71M



本周我们已经分别讨论了最基本的单点法排序学习（Pointwise Learning to Rank）和配对法排序学习（Pairwise Learning to Rank）两种思路。单点法排序学习思路简单实用，目的就是经典的检索问题转化成机器学习问题。配对法排序学习则是把排序的问题转化成针对某个查询关键字每两个文档之间的相对相关性的建模问题。不过，这两种思路也都有很明显的问题，需要进一步对算法进行优化，以实现我们需要的最终目标。

今天我就来讲直接优化排序问题的“终极方法”：列表法排序学习（Listwise Learning to Rank）。相对于尝试学习每一个样本是否相关或者两个文档的相对比较关系，列表法排序学习的基本思路是尝试直接优化像 NDCG（Normalized Discounted Cumulative Gain）这样的指标，从而能够学习到最佳排序结果。

列表法排序学习的历史

2000 年后，学术界和工业界都开始研究如何用机器学习来解决最优排序问题，五六年之后，研究者们才开始尝试直接优化整个排序列表。

这方面的研究工作很多都来自微软研究院。比如 2007 年左右的 AdaRank，就来自微软亚洲研究院的徐君和李航。这篇论文算是较早提出列表法排序观点的研究工作。同一年在国际机器学习大会 ICML 2007 (International Conference on Machine Learning) 上发表的 ListNet 算是从理论上开启了列表法的大门。这篇论文也来自微软亚洲研究院，是刘铁岩等人的重要工作。类似的研究工作在这一年里如雨后春笋般涌现。

另外一个方向，接下来我会提到，LambdaRank 出现稍早，而 LambdaMART 则稍微晚一点。这方面的工作是在微软西雅图的研究院开发的。主导人是克里斯托弗·博格斯 (Christopher J.C. Burges)。博格斯 2016 年退休，在微软工作了 16 年，可以说，他领导的团队发明了微软的搜索引擎 Bing 的算法。

列表法排序学习详解

列表法排序学习有两种基本思路。**第一种，就是直接针对 NDCG 这样的指标进行优化。**目的简单明了，用什么做衡量标准，就优化什么目标。**第二种，则是根据一个已经知道的最优排序，尝试重建这个顺序，然后来衡量这中间的差异。**

我先来说一下第一大思路，直接针对 NDCG 这样的指标进行优化。

首先，重温一下排序测试集的测试原理。总体来说，所有的基于排序的指标都要考察测试集上，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。有两种比较通用的做法。第一个方法主要适用于二分的相关信息，对于某一个查询关键字，针对排序产生的“顶部的 K”个文档进行评估，查看精度 (Precision)、召回 (Recall) 等。第二种方法，利用五级相关信息定义，在这样的情况下，就可以利用类似于 NDCG 这样的评价指标。具体解读你可以回到本周前面两期我们讲解过的内容进行复习。

那么，**直接优化排序指标的难点和核心在什么地方呢？**

难点在于，希望能够优化 NDCG 指标这样的“理想”很美好，但是现实却很残酷。NDCG 以及我之前说过的基于“顶部的 K”的精度，都是在数学的形式上的“非连续” (Non-Continuous) 和“非可微分” (Non-Differentiable)。而绝大多数的优化算法都是基于“连续” (Continuous) 和“可微分” (Differentiable) 函数的。因此，直接优化难度比较大。

针对这种情况，主要有这么几种方法。

第一种方法是，既然直接优化有难度，那就找一个近似 NDCG 的另外一种指标。而这种替代的指标是“连续”和“可微分”的。只要我们建立这个替代指标和 NDCG 之间的近似关系，那么就能够通过优化这个替代指标达到逼近优化 NDCG 的目的。这类的代表性算法的有 SoftRank 和 AppRank。

第二种方法是，尝试从数学的形式上写出一个 NDCG 等指标的“边界” (Bound)，然后优化这个边界。比如，如果推导出一个上界，那就可以通过最小化这个上界来优化 NDCG。这类的代表性算法有 SVM-MAP 和 SVM-NDCG。

第三种方法则是，希望从优化算法上下手，看是否能够设计出复杂的优化算法来达到优化 NDCG 等指标的目的。对于这类算法来说，算法要求的目标函数可以是“非连续”和“非可微分”的。这类的代表性算法有 AdaRank 和 RankGP。

说完了第一大思路后，我们再来看看第二大思路。这种思路的主要假设是，已经知道了针对某个搜索关键字的完美排序，那么怎么通过学习算法来逼近这个完美排序。**我们希望缩小预测排序和完美排序之间的差距。**值得注意的是，在这种思路的讨论中，优化 NDCG 等排序的指标并不是主要目的。这里面的代表有 ListNet 和 ListMLE。

讲了这两大思路以后，最后我再来提一下第三类思路。**这类思路的特点是在纯列表法和配对法之间寻求一种中间解法。**具体来说，这类思路的核心思想，是从 NDCG 等指标中受到启发，设计出一种替代的目标函数。这一步还和我刚才介绍的第一大思路中的第一个方向有异曲同工之妙，都是希望能够找到替代品。

这第三类思路更进一步的则是**找到替代品以后，把直接优化列表的想法退化成优化某种配对。**这第二步就更进一步简化了问题。这个方向的代表方法就是微软发明的 LambdaRank 以及后来的 LambdaMART。微软发明的这个系列算法成了微软的搜索引擎 Bing 的核心算法之一。

我这里简单提一下 LambdaRank 这个系列模型的基本思想。

首先，微软的学者们注意到，一个排序算法是否达到最优的情况，简单来看，就是查看当前的排序中，相比于最优的情况，有哪些两两文档的关系搞错了。**学习最优排序的问题就被转**

化成了减小这些两两排错的关系。更进一步，在设计这个优化过程中，我们其实并不需要知道真正的目标函数的形式，而仅仅需要某种形式的梯度（Gradient）。

这里有这样一个洞察，对于绝大多数的优化过程来说，目标函数很多时候仅仅是为了推导梯度而存在的。而如果我们直接就得到了梯度，那自然就不需要目标函数了。最后，通过实验，微软的学者们把这个 NDCG 通过梯度变化的差值再乘以这个梯度，这样就达到了增强效果的目的。

早期的 LambdaRank，特别是 RankNet 是采用了神经网络来进行模型训练，而 LambdaMART 则采用了“集成决策树”的思想，更换到了基于决策树的方法。**后来实践证明，基于决策树的方法对于排序问题非常有效果，也就成了很多类似方法的标准配置。**

最后，有一点需要你注意，我们讨论了不同的列表法思路，列表法从理论上和研究情况来看，都是比较理想的排序学习方法。因为列表法尝试统一排序学习的测试指标和学习目标。尽管在学术研究中，纯列表法表现优异，**但是在实际中，类似于 LambdaRank 这类思路，也就是基于配对法和列表法之间的混合方法更受欢迎。**因为从总体上看，列表法的运算复杂度都比较高，而在工业级的实际应用中，真正的优势并不是特别大，因此列表法的主要贡献目前还多是学术价值。

小结

今天我为你讲了列表法排序学习。你可以看到，列表法排序有很多种思路，在 2000 年到 2010 年之间是一个非常活跃的研究领域，积累了大量的成果。

一起来回顾下要点：第一，简要介绍了列表法排序学习的历史。第二，详细介绍了列表法排序学习的三大思路以及每个思路里的主要细节和方法。

最后，给你留一个思考题，列表法是不是就完全解决了排序算法的问题呢？

欢迎你给我留言，和我一起讨论。

参考文献

1. Jun Xu and Hang Li. AdaRank: a boosting algorithm for information retrieval. Proceedings of the 30th annual international ACM SIGIR conference on research and development in information retrieval, 391-398, 2007.

2. Zhe Cao, Tao Qin, Tie-Yan Liu, Ming-Feng Tsai, Hang Li. Learning to rank: from pairwise approach to listwise approach. ICML, 129-136, 2017.
 3. Q. Wu, C.J.C. Burges, K. Svore and J. Gao. Adapting boosting for information retrieval measures. Journal of Information Retrieval, 2007.
 4. C.J.C. Burges, R. Ragno and Q.V. Le. Learning to rank with non-smooth cost functions. Advances in Neural Information Processing Systems, 2006.
 5. C.J.C. Burges, T. Shaked, E. Renshaw, A. Lazier, M. Deeds, N. Hamilton and G. Hullender. Learning to rank using gradient descent. Proceedings of the twenty second international conference on machine learning, 2005.
 6. F. Xia, T.-Y. Liu, J. Wang, W. Zhang, and H. Li. Listwise approach to learning to rank — Theorem and algorithm. ICML, 1192–1199, 2008.
 7. S. Chakrabarti, R. Khanna, U. Sawant, and C. Bhattacharyya. Structured learning for non-smooth ranking losses. SIGKDD, 88–96, 2008.
 8. T. Qin, T.-Y. Liu, and H. Li. A general approximation framework for direct optimization of information retrieval measures. Technical Report, Microsoft Research, MSR-TR-2008-164, 2008.
 9. M. Taylor, J. Guiver, S. Robertson, and T. Minka. SoftRank: Optimising non-smooth rank metrics. WSDM, 77–86, 2008.
 10. J.-Y. Yeh and J.-Y. Lin, and etc. Learning to rank for information retrieval using genetic programming. SIGIR 2007 Workshop in Learning to Rank for Information Retrieval, 2007.
 11. Y. Yue, T. Finley, F. Radlinski, and T. Joachims. A support vector method for optimizing average precision. SIGIR, 271–278, 2007.
-

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 022 | 机器学习排序算法：配对法排序学习

下一篇 024 | “查询关键字理解”三部曲之分类

精选留言 (2)

 写留言



sky

2018-05-22

 2

列表法的假设是，已经知道了针对某个搜索关键字的完美排序，但是在并不知道完美排序或者不存在完美排序的时候就不好用了



幻大米

2018-01-28

 1

洪老师，列表法只是在训练阶段的运算复杂度高吧，在线应用时运算复杂度不跟单点法一样吗？

