

039 | 机器学习排序算法经典模型：RankSVM

2018-01-01 洪亮劼

AI技术内参

[进入课程 >](#)



讲述：初明明

时长 07:54 大小 3.62M



到目前为止，我们在专栏里已经讨论了关于搜索引擎方方面面的很多话题，包括经典的信息检索技术、查询关键字理解、文档理解以及现代搜索引擎的架构等等。同时，我们也从机器学习角度出发对搜索引擎的最核心部分，也就是排序算法进行了最基本的分享，囊括了单点法排序学习（Pointwise Learning to Rank）、配对法排序学习（Pairwise Learning to Rank）以及列表法排序学习（Listwise Learning to Rank），相信你应该对这类算法的大概内容有所掌握。

那么，这周我们就来看看机器学习排序算法中几个经典的模型，希望能够通过这几个经典的算法为你深入学习和研究排序算法指明方向。

今天，我就来分享配对法排序中最有价值一个算法，**排序支持向量机（RankSVM）**。这个算法的核心思想是**应用支持向量机到序列数据中，试图对数据间的顺序直接进行建模。**

排序支持向量机的历史

20 世纪 90 年代中后期，受统计学习理论（Statistical Learning Theory）思想和风险最小化框架（Risk Minimization Framework）趋于成熟的影响，支持向量机逐渐成为当时机器学习界的主流模型。一时间，各个应用领域的学者和工程师都在思考如何把支持向量机利用到自己的问题领域上，从而获得更好的效果。

拉夫·赫博里奇（Ralf Herbrich）发表于 1999 年 [1] 和 2000 年 [2] 的论文中讨论了如何把支持向量机和有序回归（Ordinal Regression）结合起来。赫博里奇当时在柏林科技大学（Technical University of Berlin）攻读博士学位。2000 年到 2011 年，他在微软研究院和 Bing 任职，从事机器学习，特别是贝叶斯方法（Bayesian method）的研究。2011 年到 2012 年，他在 Facebook 短暂任职后，于 2012 年加入了亚马逊负责机器学习的研发工作，并且担任在柏林的研发中心主管经理（Managing Director）。尽管赫博里奇很早提出了把有序回归和支持向量机结合的思路，但是当时的论文并没有真正地把这个新模型用于大规模搜索系统的验证。

更加完整地对排序支持向量机在搜索中的应用进行论述来自于康奈尔大学教授索斯滕·乔基姆斯（Thorsten Joachims）以及他和合作者们发表的一系列论文（见参考文献 [3]、[4]、[5] 和 [6]）。索斯滕我们前面介绍过，他是机器学习界享有盛誉的学者，是 ACM 和 AAAI 的双料院士；他所有论文的引用数超过 4 万次；他获得过一系列奖项，包括我们前面讲的 2017 年 ACM KDD 的时间检验奖等等。

排序支持向量机模型

在说明排序支持向量机之前，我们先来简要地回顾一下支持向量机的基本思想。

在二分分类问题中（Binary Classification），线性支持向量机的核心思想是找到一个“超平面”（Hyperplane）把正例和负例完美分割开。在诸多可能的超平面中，支持向量机尝试找到距离两部分数据点边界距离最远的那一个。这也就是为什么有时候支持向量机又被称作是“**边界最大化**”（Large Margin）分类器。

如果问题并不是线性可分的情况，支持向量机还可以借助“**核技巧**”（Kernel Trick）来把输入特性通过非线性变换转化到一个线性可分的情况。关于支持向量机的具体内容你可以参考各类机器学习教科书的论述。

要把支持向量机运用到排序场景下，必须改变一下原来的问题设置。我们假设每个数据点由特性 X 和标签 Y 组成。这里的 X 代表当前文档的信息、文档与查询关键字的相关度、查询关键字的信息等方方面面关于文档以及查询关键字的属性。 Y 是一个代表相关度的整数，通常情况下大于 1。

那么，在这样的设置下，我们针对不同的 X ，需要学习到一个模型能够准确地预测出 Y 的顺序。意思是说，如果有两个数据点 X_1 和 X_2 ，他们对应的 Y_1 是 3， Y_2 是 5。因为 Y_2 大于 Y_1 （在这里，“大于”表明一个顺序），因此，一个合理的排序模型需要把 X_1 通过某种转换，使得到的结果小于同样的转换作用于 X_2 上。这里的转换，就是排序支持向量机需要学习到的模型。

具体说来，在线性假设下，排序支持向量机就是要学习到一组线性系数 W ，使得在上面这个例子中， X_2 点积 W 之后的结果要大于 X_1 点积 W 的结果。当然，对于整个数据集而言，我们不仅仅需要对 X_1 和 X_2 这两个数据点进行合理预测，还需要对所有的点，以及他们之间所有的顺序关系进行建模。也就是说，模型的参数 W 需要使得数据集上所有数据点的顺序关系的预测都准确。

很明显，上述模型是非常严格的。而实际中，很可能并不存在这样的 W 可以完全使得所有的 X 都满足这样的条件。这也就是我们之前说的线性不可分在排序中的情况。那么，更加现实的一个定义是，**在允许有一定误差的情况下，如何使得 W 可以准确预测所有数据之间的顺序关系，并且 W 所确定的超平面到达两边数据的边界最大化**，这就是线性排序向量机的定义。

实际上，在线性分类器的情境下，线性排序向量机是针对数据配对（Pair）的差值进行建模的。回到刚才我们所说的例子，线性排序向量机是把 X_2 减去 X_1 的差值当做新的特性向量，然后学习 W 。也就是说，原理上说，整个支持向量机的所有理论和方法都可以不加改变地应用到这个新的特征向量空间中。当然，这个情况仅仅针对线性分类器。

因为是针对两个数据点之间的关系进行建模，排序支持向量机也就成为配对法排序学习的一个经典模型。

排序支持向量机的难点

我们刚刚提到的排序支持向量机的定义方法虽然很直观，但是有一个非常大的问题，那就是**复杂度是 N 的平方级，这里的 N 是数据点的数目**。原因是我们需要对数据点与点之间的所

有配对进行建模。当我们要对上万，甚至上百万的文档建模的时候，直接利用排序支持向量机的定义来求解模型参数显然是不可行的。

于是，针对排序支持向量机的研究和应用就集中在了**如何能够降低计算复杂度**这一难点上，使得算法可以在大规模数据上得以使用。

比较实用的算法是索斯腾在 2006 年发表的论文 [6] 中提出的，这篇论文就是我们前面讲的 2017 年 KDD 时间检验奖，建议你回去复习一下。这里，我再简要地梳理一下要点。

这个算法的核心是重新思考了对排序支持向量机整个问题的设置，把解决结构化向量机 (Structural SVM) 的一种算法，CP 算法 (Cutting-Plane)，使用到了排序支持向量机上。简单来说，这个算法就是保持一个工作集合 (Working Set) 来存放当前循环时依然被违反的约束条件 (Constraints)，然后在下一轮中集中优化这部分工作集合的约束条件。整个流程开始于一个空的工作集合，每一轮优化的是一个基于当前工作集合的支持向量机子问题。算法直到所有约束条件的误差小于一个全局的参数误差为止。

索斯腾在文章中详细证明了该算法的有效性和时间复杂度。相同的方法也使得排序支持向量机的算法能够转换成为更加计算有效的优化过程，在线性计算复杂度的情况下完成。

小结

今天我为你讲了利用机器学习技术来学习排序算法的一个基础的算法，排序支持向量机的基本原理。作为配对法排序学习的一个经典算法，排序支持向量机有着广泛的应用。一起来回顾下要点：第一，我们简要介绍了排序支持向量机提出的历史背景。第二，我们详细介绍了排序支持向量机的问题设置。第三，我们简要提及了排序支持向量机的难点和一个实用的算法。

最后，给你留一个思考题，排序支持向量机是否给了你一些启发，让你可以把更加简单的对数几率分类器 (Logistic Regression) 应用到排序问题上呢？

欢迎你给我留言，和我一起讨论。

参考文献

1. Herbrich, R.; Graepel, T. & Obermayer, K. Support vector learning for ordinal regression. The Ninth International Conference on Artificial Neural Networks

- (ICANN 99), 1, 97-102 vol.1, 1999.
2. Herbrich, R.; Graepel, T. & Obermayer, K. Smola; Bartlett; Schoelkopf & Schuurmans (Eds.). Large margin rank boundaries for ordinal regression. Advances in Large Margin Classifiers, MIT Press, Cambridge, MA, 2000.
 3. Tsochantaridis, I.; Hofmann, T.; Joachims, T. & Altun, Y. Support Vector Machine Learning for Interdependent and Structured Output Spaces. Proceedings of the Twenty-first International Conference on Machine Learning, ACM, 2004.
 4. Joachims, T. A Support Vector Method for Multivariate Performance Measures. Proceedings of the 22Nd International Conference on Machine Learning, ACM, 377-384, 2005.
 5. Tsochantaridis, I.; Joachims, T.; Hofmann, T. & Altun, Y. Large Margin Methods for Structured and Interdependent Output Variables. The Journal of Machine Learning Research, 6, 1453-1484, 2005.
 6. Joachims, T. Training Linear SVMs in Linear Time. Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 217-226, 2006.

论文链接

[Support vector learning for ordinal regression](#)

[Support Vector Machine Learning for Interdependent and Structured Output Spaces](#)

[A Support Vector Method for Multivariate Performance Measures](#)

[Large Margin Methods for Structured and Interdependent Output Variables](#)

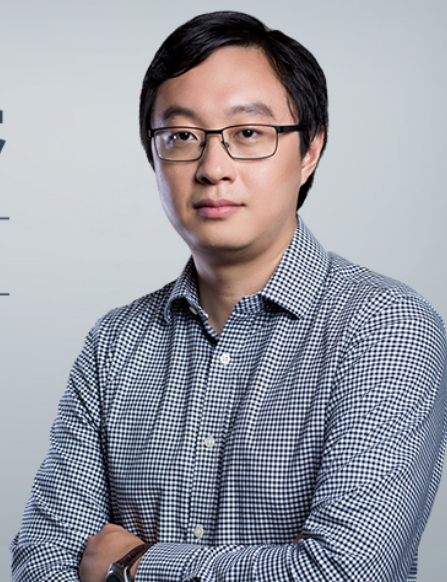
[Training Linear SVMs in Linear Time](#)

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 038 | 社区检测算法之“模块最大化”

下一篇 040 | 机器学习排序算法经典模型：GBDT

精选留言

写留言

由作者筛选后的优质留言将会公开显示，欢迎踊跃留言。