

水滴石穿

探索，保持渴望，无所畏惧

 首页

 归档

 关于

 订阅

“

gbRank & logitRank自顶向下

📅 Sep 18, 2016 | 📁 机器学习 | 📖 1310 Hits



上一篇[Ranking Relevance in Yahoo Search](#)一文中提到的logitRank方法吃不太透，没展开。这两天刚好中秋，整理出来。

介绍思路如下：

1. gbrank : gbdn怎么用在排序，如何进行pair-wise训练
2. logitRank : logitRank和gbrank的区别，和关于scale因子的思考

本篇博文分两篇将从应用出发，往算法细节深入了解的顺序进行介绍。上篇：第1部分，介绍基础的gbrank，可以知道gbdn用在learning to rank领域的应用法子。第2部分，介绍logitRank，希望进一步理解前一篇笔记中作者提出改进。下篇：主要按提出者 Jerome H. Friedman 论文思路介绍gbdn

相关拓展：纵向可以往前看adboost及在learning to rank(ltr)应用[2]，或者往后看xgboost、FastBDT、LightGBM。横向可以看看bagging的一些算法，包括随机森林。ltr中常用的还有rankSVM（应用svm分类器）[3,4,5]、rankNet(应用神经网络)[6]、LambdaMART[9]

gbrank：如何进行pair-wise训练

learning to rank需要解决的问题是给定一个query，如何选择最相关的document。gbrank核心为将排序问题转化为一组回归问题，对于回归问题可以用gbdn进行求解，也可以用其他的回归函数[7]。

先介绍一组记号，对于所有的query-document pair，我们从pair抽取出一系列特征对其进行表示。例如query1-document1记为 x ，query1-document2记为 y 。记 $x \succ y$ 表示，用户发起查询query1时， x 比 y 更适合，更加满足query1的需求。记训练集合为

$$S = \{ \langle x_i, y_i \rangle \mid x_i \succ y_i, i = 1, \dots, N \}$$

给定排序函数空间 H ，我们希望得到一个排序函数 $h(h \in H)$ ，当 $x_i \succ y_i$ 时，我们有

