**统计学习方法读书笔记**

61518122 丁自民

**读书内容**

第21章 PageRank算法

**读书收获**

1. PageRank算法是图的链接分析的代表性算法，属于图数据上的无监督学习算法
2. 最早用于互联网网页重要度的计算方法
3. 基本思路：在有向图上定义一个随机游走模型，即一阶马尔可夫链。描述随机游走者沿着有向图随机访问各个节点的行为。在一定条件下，极限情况访问每个节点的概率收敛到平稳分布。此时，各个节点值就是其PageRank值，表示节点的重要程度。
4. PageRank的是递归定义的，其计算也可以由迭代算法给出。

**读书疑问**

PageRank和马尔可夫链有什么不同？为什么要单独定义一个PageRank的定义？

**疑问解答**

王贵涛问题1：

那只是平滑因子。只是对平滑因子要求一个节点到任意节点概率相等。

王贵涛问题2:

多步转移相当于概率转移矩阵的高次方。用特征值和特征向量做就好了

郑健雄问题2

阻尼因子的作用之一，是保证图的连通性。阻尼因子越接近1，转移矩阵越接近自身。阻尼因子越接近0，越接近等概率转移。