

《数学之美》阅读笔记 (2)

西安交通大学

联系方式: williamyi96@gmail.com

易凯

2017 年 1 月 26 日

目录	2
----	---

目录

1 布隆过滤器	3
2 马尔科夫链的扩展—贝叶斯网络	3
3 条件随机场和句法分析	4
4 维特比和他的维特比算法	4
5 再谈文本自动分类问题—期望最大化算法	4
6 逻辑回归和搜索广告	4
7 各个击破算法和 Google 云计算基础	4

1 布隆过滤器

在日常生活中，经常要判断一个元素是否在一个集合中。

如果用哈希表进行存储，那么其所占据的空间在数据量较大时占用量较大。

如果使用布隆过滤器则可以将存储空间缩短到原来的四分之一或更多。

布隆过滤器原理

布隆过滤器是一个很长的二进制向量和一系列随机映射函数构成的。

当一个元素被加入集合时，通过 K 个散列函数将这个元素映射成一个位数组中的 K 个点，把他们置为 1。检索时，我们只要看看这些点是不是都是 1 就 (大约) 知道集合中有没有它了：如果这些点有任何一个 0，被检索的元素一定不再；如果都是 1，则被检索元素可能在。

优点：

1. 存储空间和插入/查询时间都是常数 $O(k)$;
2. 保密性能较好。不需要存储元素本身。

缺点：

1. 存在误差率。其他元素可能将待测位设置为 1(常见的方式是先建立一个小的白名单);
2. 不能够从布隆过滤器中删除元素。

2 马尔科夫链的扩展—贝叶斯网络

贝叶斯网络是一个加权有向图，是马尔科夫链的扩展。

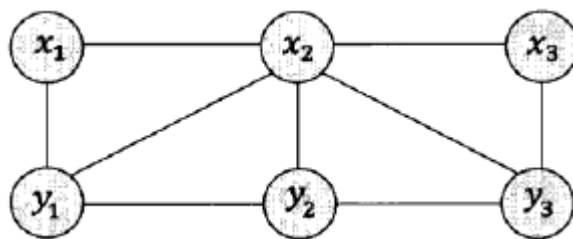
在有向图中，各个节点 (每个状态) 之和与它直接相连的状态有关，而和它间接相连的状态没有直接关系。

由于网络的每个弧都有一个可信度，贝叶斯网络也被称为信念网络 (Belief Networks)。

使用贝叶斯网络必须先确定这个网络的拓扑结构，然后还要知道各个状态之间相关的概率。得到拓扑结构和这些参数的过程分别叫做结构训练和参数训练。

3 条件随机场和句法分析

条件随机场是一个无向图，而贝叶斯网络是一个有向图。



4 维特比和他的维特比算法

5 再谈文本自动分类问题—期望最大化算法

6 逻辑回归和搜索广告

7 各个击破算法和 Google 云计算基础