区间树中区间重叠检测算法正确性的证明

jianjunwu@tecent.com

April 1, 2017

最近遇到一个版本区间冲检测问题，这个问题描述如下：

> 给定一个版本区间[x,y]，试问他是否与一个版本区间集合中的区间有重叠。

在网上收集了一下相关资料，发现一个叫区间树的数据结构，网址

`http://www.geeksforgeeks.org/interval-tree/`

对其有一个简单的介绍。

首先在这里简单介绍下区间树：区间树每个节点按区间的左端点(所以区间的左端点不能重复)形成一颗红黑树，每个节点中存储着一个区间$[u, v]$以及以本节点为根的子树中所有区间的最大右端点m。在区间树上进行区间查找的时间为$O(lgn)$。区间树如下图所示：

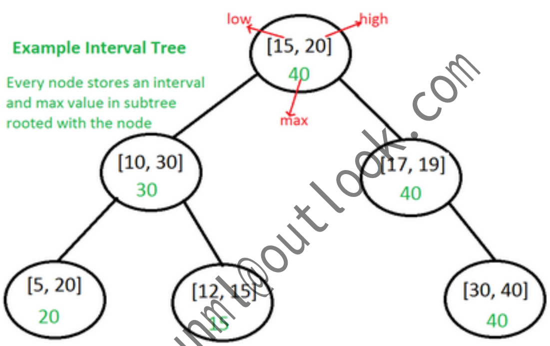

Figure 1: 区间树示例

树中每个节点的定义如下：

```
struct Interval
{
    int low, high;
};
struct ITNode
{
    Interval *i;
    int max;
    ITNode *left, *right;
};
```

在区间树root中查找是否有与区间i重叠的区间的算法为：

```
bool doOVerlap(Interval i1, Interval i2)
{
    if (i1.low <= i2.high && i2.low <= i1.high)
        return true;
    return false;
}
Interval *overlapSearch(ITNode *root, Interval i)
{
    if (root == NULL) return NULL;

    if (doOVerlap(*(root->i), i))
        return root->i;

    if (root->left != NULL && root->left->max >= i.low)
        return overlapSearch(root->left, i);
    return overlapSearch(root->right, i);
}
```

这个算法的正确性据我所知没有资料提及，而我个人觉得其正确性不是那么明显和直观。所以，我试着自己证明了一下。

我们首先形式化这个问题：有一颗区间树，根节点为$\{[a,b],m1\}$，根节点的左孩子为$\{[c,d],m2\}$，右孩子为$\{[e,f],m3\}$。

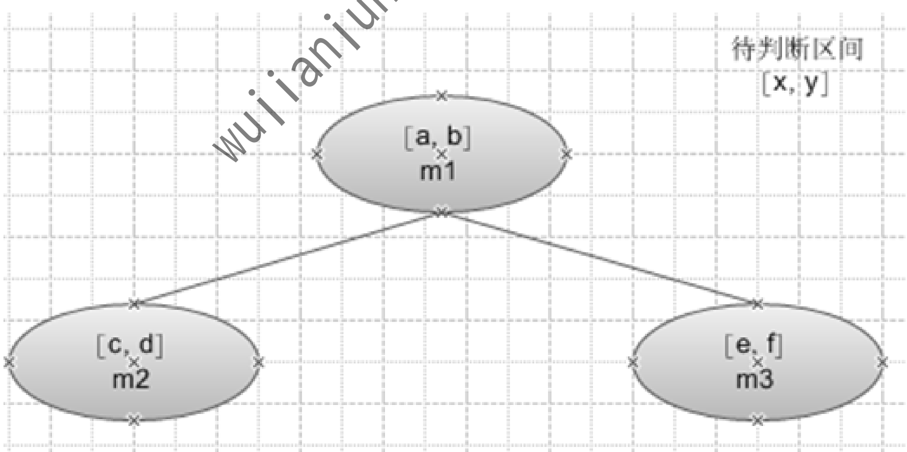

Figure 2: 区间树查找示意图

不失一般性，假设用上述算法判断在这颗区间树(图2)中是否有区间与

待判断区间$[x, y]$重叠。在开始证明前，需要说明的是：

L1 证明中出现的任何数都是正实数。

L2 $*$ 代表任意正实数。

L3 所有区间都是非0的开区间。

L4 $[u, v]$是一个区间， $\Rightarrow, u < v$

L5 我们采用集合运算来表示区间重叠与否，$[u_1, v_1] \cap [u_2, v_2] \neq \emptyset$ 表示两个区间相交。

L6 $[u_1, v_1] \cap [u_2, v_2] \neq \emptyset \Leftrightarrow v_1 \geq u_2 \;\&\&\; v_2 \geq u_1$ 。

下面我们就开始证明，我尽量给出每一个证明细节。

定理1. *若$[a, b] \cap [x, y] = \emptyset$并且$m_2 > x$，那么：*
1).要么，在左子树存在一个区间与$[x, y]$重叠，
2).要么，整颗树中都没有区间与$[x, y]$重叠。

Proof. 我们只需要证明逆定理不可能成立。也就是证明下面的情况在定理条件满足时不可能成立：

有重叠子区间且重叠子区间不在左子树而在右子树。

当左子树不存在与$[x, y]$重叠区间的时候，左子树可以分成两种情况讨论：

1 左子树的区间全部形如$[*, x - \varepsilon]$。

2 左子树至少有一个区间形如$[y + \varepsilon, *]$。

其实我们可以马上得出情况1是不可能的，因为此时左子树的最大右端点是$x - \varepsilon = m_2$，而定理条件有$m_2 > x$，所以得出矛盾$x - \varepsilon = m_2 > x$。

我们现在再证明情况2也是不可能的。假设右子树中有一个区间$[u, v]$和$[x, y]$重叠，也就是:

$$y \geq u \;\&\&\; x \leq v \tag{1}$$

因为区间树是以区间左端点为key的二叉排序树，所以：

$$y + \varepsilon < u \tag{2}$$

马上，我们由方程1和方程2便得到，下面的矛盾：

$$y + \varepsilon < u, y \geq u \;\&\&\; x \leq v\text{同时成立} \Rightarrow y + \varepsilon < u \leq y$$

证毕

$\square$

References