

算法: 基于在线随机森林的统计翻译模型

输入: 双语短语对 $\langle s, t \rangle$, 源语言短语对应的随机森林 \mathcal{F} , \mathcal{F} 中决策树的数量 M

输入: 最小样本数阈值 α , 最大信息增益阈值 β , 动态加权最小样本数阈值 γ

输出: 更新后的随机森林 \mathcal{F}

```
1:  //针对所有森林中的所有决策树
2:  for  $t = 1 \rightarrow T$  do
3:       $k \leftarrow \text{Poisson}(\lambda)$ 
4:      if  $k > 0$  then
5:          //更新 $k$ 次
6:          for  $u = 1 \rightarrow k$  do
7:               $j = \text{FindLeaf}(s)$ 
8:               $\text{UpdateNode}(j, \langle s, t \rangle)$ 
9:              if  $|\mathcal{R}_j| > \alpha$  and  $\exists s \in \mathcal{S}: \Delta L(\mathcal{R}_j, s) > \beta$  then
10:                  //搜索最佳的决策因子
11:                   $s_j = \arg\max_{s \in \mathcal{S}} \Delta L(\mathcal{R}_j, s)$ 
12:                   $\text{CreateLeftChild}(\mathcal{R}_{jls})$ 
13:                   $\text{CreateRightChild}(\mathcal{R}_{jrs})$ 
14:              end if
15:          end for
16:      else
17:           $\text{OOBE}_t \leftarrow \text{UpdateOOBE}(\langle s, t \rangle)$ 
18:           $\text{WeightingTree}(f_t, \text{OOBE}_t, \gamma)$ 
19:      end if
20:  end for
21:
22:  function  $\text{WeightingTree}(f_t, \text{OOBE}_t, \gamma)$ 
23:       $\text{age}_t \leftarrow \text{NumberOfSamples}(f_t)$ 
24:      if  $\text{age}_t > \gamma$  and  $\text{OOBE}_t > \text{Rand}()$  then
25:          //移除该决策树
26:           $f_t = \text{NewTree}()$ 
27:      end if
28:  end function
```