# 实验结果：

完全匹配：

interval：20；basic：300（A）；

1.对经过SAX的图例建树；

2.对流数据直接进行SAX，流入suffix-tree判断；

3.对于匹配长度大于n（n设为3）的结果输出；

4.在第三步中会选择最长的子序列，即当a1a2a3a4符合条件时，不会再输出a1a2a3等等。

子树的匹配结果：

DFEDCDEF…

4\*20,DEF

50\*20,DFED

27\*20,CDEF

30\*20,FEDCDEF

# 问题思考与解决：

真实图谱：

1. 主要问题为：
   1. index中各个图例的区分/归一化：

实验得到的图例有些过于平坦，特征意义不大；

图例的各个分支要做合并

* 1. stream的偏移问题

仍旧存在切错的问题，而且目前没有证明无遗漏

* 1. 距离度量方式：只选用匹配维度（或者说匹配字符长度）不能反应“近似程度”；采用(Σdi)/M?

1. suffix-tree的效率较低
   1. 自己实现的suffix-tree较为原始，空间消耗没有优化；下周的任务：根据Ukkonen的论文优化suffix树；
   2. 原有的字符型结构并不一定符合要求，下周尝试修改为[a,b]型key
2. 优化思考：
   1. 分割方式：如果用PAA，则不能假设所有的拐点等间距或近似等间距。那么使用加了长度限制的（PICA？）
   2. 不同图例之间垂直高度差别很大，又不能归一化，是否抛弃线性划分sax，试一下gauss分布
   3. 【模糊】对流数据，根据设定的参数进行模糊；对图例，根据参数进行模糊，增加各个分支的融合度？
3. **下周目标：**
   1. 用优化过的suffix-tree对数据进行流式分析，评估效率；
   2. 尝试优化思考c
   3. 并实现一种距离度量方式。.

通过

思考:

不处理

利用方差处理宽度

map-reduce？

王鹏的论文；

间断的时间序列的相关度；

为什么老的方法不work；

bitmap？a,b之间边界上的模糊。