(1) 计算相对熵和变化率的变异系数：

**1) relative\_entropy\_cv.py**

a.程序名：relative\_entropy\_cv.py

b.程序处理描述：计算2个节点之间某段时间内某项指标的相对熵值；计算2个节点之间该段时间内该项指标的微分变化的变异系数。

C．传递参数：

|  |  |  |  |
| --- | --- | --- | --- |
| 参数名称 | 是否必须 | 类型 | 描述 |
| node1 | 是 | String | 节点1 |
| node2 | 是 | String | 节点2 |
| indexname | 是 | String | 某一个指标名 |
| tablename | 是 | String | 该指标所属的table名称 |
| starttime | 是 | String | 某个stage的开始时间 |
| endtime | 是 | String | 某个stage的结束时间 |
| bins | 是 | int | 分段数，可以设置10，100，这个值越大，计算的相对熵值越精确 |

d.使用方法（控制台）：python relative\_entropy\_cv.py node1 node2 indexname tablename starttime endtime bins

e.程序介绍：

|  |
| --- |
| 举例：  python relative\_entropy\_cv.py hw004 hw073 cpu\_usage os 1467298880 1467308880 100  程序输出：  0.167729238953 (相对熵)  coefficient of variation: 0.166014299557 0.133761088016 (两个节点微分变化的变异系数) |

(2) K-means算法对节点进行聚类分析：

**1) node\_kmeans.py node\_sim\_linear.py**

a.程序名：node\_kmeans.py node\_sim\_linear.py

b.程序处理描述：对某段时间内各节点之间的某个Spark stage阶段各项指标的余弦相似度进行聚类分析，默认分成两类。

C．传递参数：

|  |  |  |  |
| --- | --- | --- | --- |
| 参数名称 | 是否必须 | 类型 | 描述 |
| stage\_id | 是 | String | Spark stage |

d.使用方法（控制台）：python node\_kmeans.py stage\_id

e.程序介绍：

|  |
| --- |
| 举例：  python node\_kmeans.py spark\_stage\_app-20160719212517-0001\_2  程序输出：  关于余弦相似度K-means聚类分析后的结果（png图片）  E:\3W\Internship\02_Institute of Computing Technology, Chinese Academy of Sciences\Task\08_diff Coefficient of Variation & Kmeans of Node Similarity\codes\kmeans.png |