DataSplitter

2019年3月28日

20:48

* 引用了工具包；拆分包；配置包；（？）master 包（//

master.app.AMContext//提供方法声明，该AMContext为接口

）

* final AMContext context ； //定义上下文
* 定义了Map类 Map<Integer, SplitClassification> splitClassifications；拆分类化
* 指定相应API
* publicMap<Integer, SplitClassification> generateSplits**(**) // 粗略拆分数据，放置 workergroup
  + 获取配置信息，判断训练数据（目录）
  + 获取工作组和任务组 成员量（//~worker,~number）
  + 判断API与debug等信息，（//调整以适应环境）
* List<org.apache.hadoop.mapreduce.InputSplit> generateSplitsUseNewAPI(Configuration conf,int expectedSplitNum)
  + 获取上下文
  + 获得最小数目和最大数目的拆分配置jobConf.getConfiguration()
  + 获得获得其配置信息类型 赋值给String 变量 inputFormatClassName
  + 返回拆分数据 input.getSplits（jobConf）
* List<org.apache.hadoop.mapreduce.InputSplit> generateSplitsUseOldAPI(Configuration conf,int expectedSplitNum)
  + （//逻辑同上，拆分的是old 数据）
  + 返回拆分数据
* input.getSplits(jobConf, splitNum)

dispatchSplitsUseLocation(List<org.apache.hadoop.mapreduce.InputSplit> splitsNewAPI,int groupNumber, int groupItemNumber)

\* 获得拆分数，拆分组数

\* 循环便利，为locationList （！string 属性）赋值，数量为拆分组数，值为获得的拆分API的location属性值

\* 为拆分类化属性变量对象 赋值为对应的位置赋其值，（locationList列表键值）

原理同上，操作属性为groupItemNumber

* · String[] getSplitLocations(int index)仅仅返回拆分类化属性值
* 定义了方法将其对外接口，输入输出流进行了序列化

|  |  |
| --- | --- |
|  |  |