Assignment5 report

Chengze Li
cs6240

pseudo-code

```
distArray = LoadMutltiStack.load(distFile) // generate a Array[Array[Byte]]
imageArray = LoadMultiStack.load(imageFile) // generate a Array[Array[Byte]]

distRdd =
    distArray
        .map(layerArray) // make a long array to a 3d array(z, x, y)
        .flatMap(extractLabel) // each row in rdd is (axis, label)

imageRdd =
    imageArray
        .map(layerArray) // make a long array to a 3d array(z, x, y)
        .flatMap(extractNeighbors) // each row in rdd is (axis, neighbors)

originalJoinedRdd
    = imageRdd.join(distRdd) // each row is(axis, (label, neighbors))
```

//then add the diversity of training data

rotated90DistArray = distArray.map(rotate90)

rotated90ImageArray = imageArray.map(rotate90)

// do the same thing as we do for original dist and image rdds

// then rotate the rotated90 data 90 degree to get rotated180 data and so on

// then load other images and do the same thing

how you represent image and distance data in your Spark Scala program

(1) Are you using (pair) RDDs or DataSets?

I use pair Rdd to represent image and distance data

(2) What is stored in each row of the RDD/DataSet—a single pixel, an entire image, a 2D layer of an image, a matrix or array containing some part of an image stack, etc.?

in my case, for the data returned from java file, I put the full image and dist file in one row, then after processing, each row represent a single pixel

(3) Do you store each image in a separate RDD/DataSet?

yes, I put each image in a separate Rdd

Performance and results

6m-100000

```
INFO total process run time: 496 seconds
2017-11-17T14:25:41.342Z INFO Step created jobs:
2017-11-17T14:25:41.342Z INFO Step succeeded with exitCode 0 and took 496
seconds
```

6m-1000000

```
INFO total process run time: 896 seconds
2017-11-17T17:40:47.045Z INFO Step created jobs:
2017-11-17T17:40:47.045Z INFO Step succeeded with exitCode 0 and took 896
seconds
```

11m-100000

```
INFO total process run time: 534 seconds
2017-11-17T20:15:34.249Z INFO Step created jobs:
2017-11-17T20:15:34.249Z INFO Step succeeded with exitCode 0 and took 534
seconds
```

11m-1000000

```
INFO total process run time: 558 seconds
2017-11-17T20:36:46.703Z INFO Step created jobs:
2017-11-17T20:36:46.703Z INFO Step succeeded with exitCode 0 and took 558
seconds
```