Code Review - Shabbir-Navya-YuWen

Ayushi-Balaraj-Shikha

October 23, 2017

The Code Review is as follows:

- GNUParser seems to be deprecated, needs to be replaced with appropriate expressions.

```
CommandLine cmd = new GnuParser().parse(options, args, true);
```

- Readme could have been more informative. Information on Job Flow is needed.

- No Justification for why they considered training data for just one year

```
// by default we only extract on year's range as training data
```

- Code could have been better formatted in some places (ex: in line 96 (main.java))

- No description on the functionality of the class file. It would be better if the description given in the code.

- Jobs could be more efficient and utilize the functionality of Map Reduce by using more reducers for the system.

```
job.setNumReduceTasks(1);
```

- The following code

```
    // AirportID,  AirportSeqID, CityMarketID, StateFips, Wac should be larger than 0
    if ((float) m.get("ORIGIN_AIRPORT_ID") <= 0)              return false;
    if ((float) m.get("ORIGIN_AIRPORT_SEQ_ID") <= 0)         return false;
    if ((float) m.get("ORIGIN_CITY_MARKET_ID") <= 0)         return false;
    if ((float) m.get("ORIGIN_STATE_FIPS") <= 0)             return false;
    if ((float) m.get("ORIGIN_WAC") <= 0)                    return false;
```

could be written as:

```
if ((float) m.get("ORIGIN_AIRPORT_ID") <= 0)   ||
      ((float) m.get("ORIGIN_AIRPORT_SEQ_ID") <= 0)  ||
      ((float) m.get("ORIGIN_CITY_MARKET_ID") <= 0)  ||
      ((float) m.get("ORIGIN_STATE_FIPS") <= 0)             ||
      ((float) m.get("ORIGIN_WAC") <= 0)
return false;
```

- There are typos in the report.

- Unable to understand why s1,s2,s3 is being used and there is no documentation for the same.

```
// flightDate >= journeyDateMs && flightDate <= journeyDateMs + X_DAYS
      context.getCounter("MyMap", "s1").increment(1);
      if (!(flightDateMs >= q.dateMs)) return;
      context.getCounter("MyMap", "s2").increment(1);
      if (!(flightDateMs <= q.dateMs + THREE_DAYS)) return;
      context.getCounter("MyMap", "s3").increment(1);
```

- We were unable to run the code since it was giving heap space issues.

- Multiple conversions are used

```
new IntWritable((int) ((float) f.get("YEAR")))
```

could have been written as

```
new IntWritable((int) f.get("YEAR"))
```

- Report is incomplete. Results are missing:

```
Results
Random forest with Depth 10
Correct Prediction : 41.6 %
Error Prediction :%
Precision (proportion of the predicted flight delays that were correct) : %
```

- The report has mentioned that the the ML model was running out of heap space and taking too much time, it would have been better if they would have tested the system using R.

```
Results
Random forest with Depth 10
Correct Prediction : 41.6 %
Error Prediction :%
Precision (proportion of the predicted flight delays that were correct) : %
```