

The *eighth* Master of Computing and innovation Group Project weekly meeting will be held in **Room 464 at 3.00pm on Tuesday 24 Apr 2018.**

## Meeting minutes on 24th Apr

**Attendance: Zheng Xu, Junjie Guo**

### 1 Apologies

None

### 2 Progress

After downloading all the related jar files and set up the running environment, we could use the xgboost package installed in the Java to build the training model. However, we meet the issues that we are hard to continent to develop the system.

#### 2.1 Findings

1. Xgboost algorithm could not handle the non-numeric values.

Answer: We could use a hash function to replace the name to specific number such as tcp -i 001, udp -i 002.

2. The recursive tree which was produced by the default xgboost package is hard to understand. Its parameters are not similar to decision tree, but we want it could be read as the later one.

Answer: It is related to the question 3.

3. We just implement the training model without Hadoop platform, there are few resources about combining xgboost and Map-reduce function. It is more likely to utilized in Yarn mode.

Answer: We should learn the algorithm not just use it. Maybe we could rewrite the code to build a understandable tree by ourselves. The tree will be recorded to the XML file. In addition, the xgboost algorithm need to be implemented on Hadoop platform. Here are two choices:

1. Rewriting the code in our logic without using the default xgboost package.
2. Implementing the xgboost algorithm in Spark, because Spark is also a HDFS and dealing with huge dataset platform that is suitable for our project.

## **2.2 About the next steps**

1. How could we develop this system as plan? Or we could just use this algorithm without Hadoop?

Answer: Rewrite the code or using the Spark platform to implement this algorithm. We will do some researches first.

Note: Next meeting to be held on 1 May 2018.