

CHAPTER 6

CONCLUSION

Based on all data throughout this study, overall, it can be said that to some degree, neural network can be used as prediction for full database in days. The pattern in space usage and how the oracle database adding extent depends on how much data received in range of each time the program fetch data. In case of this study, the simulation database really depends on the day it fetched data.

The sample data are taken from simulated database that emulate production database from retail company. This simulation database total insert rows every fetch based on the week of month, and day of week. Day 23rd have the highest amount of rows that get smaller everyday until it reached day 23rd again. Weekend (Saturday and Sunday) have higher amount of rows than weekdays.

The most optimal structure of deep neural network from this study is using 3 input neurons, 3 hidden layer with shape of diamond ($2 \times \text{total input} \rightarrow \text{total input} + (\text{total input}/2) \rightarrow \text{total input}$), and 1 output neurons. For hyperparameter, adam optimizer default learning rate of 0.001 could reach loss below 2 before 20 epoch passed. This in turn to train the network, it only need small amount of epoch, which in this study, it use 150 epoch. For batch size, 10 batch size is the most optimal value since 5 makes the neural network overfit, and more than 10 makes the neural network accuracy goes down.

The day that is counted as accurate is correct day and one day difference. This is because the way float number get rounded (if the predict is 34.7 rounded to 35 and the real value is 34, it still counted as accurate since the difference is only one day).

Best way to split the dataset to training set and validating set is 80% training set and 20% validating. The average accuracy all across tablespace for this settings is 55.5%.

The train cycle average accuracy can drop to 36.49%. This possibly because the pattern (how it add extent inside datafile or decreasing datafile free space) from one cycle to other is different and the usage of adam as optimizer can't generalize that well. But if the pattern from one cycle and another cycle is similiar, the accuracy goes up significantly. Further research regarding the cause this accuracy drop is needed.

Another thing to keep in mind is if the predict using the cycle dataset from last trained cycle dataset usually have accuracy above 80% (for example train until cycle 2, the if it try predict using cycle 2, it usually have high accuracy since the last trained dataset cycle is from cycle 2).

Based on the result of this study, database administrator shouldn't rely too much of the neural network prediction for only that day, because the cycle pattern could be different based on the trained cycle, and that can make the day difference with real and predicted days left goes more than 5 days. Instead, compare prediction values of different point in time. The alarming system also can be used as a preventive measure if the prediction goes more than 2 day and administrator doesn't expect it. The threshold also act as prevention to give enough room in prediction before it completely full.

Database administrator also can consider training until cycle 1 only. This is because based on the testing result, train cycle average accuracy using only cycle 1 goes above 60%.

The suggestion for further research is:

- a) Analyze the cause of accuracy drop (for example implementing other optimization algorithm)

- b) Use other type of Neural Network model, for example Long-Short Term Memory to compare the difference in performance with this study.
- c) Compare the accuracy of this model in real world production database.

