ITU ML Different models and approaches with results

Saturday, June 27, 2020 6:24 PM

ITU ML 5g Challenge
K Venkat Ramnan
PES1201801319
4th sem ECE DEPT

Prior 24th june 2020: Understanding basics of ML and keras,scikitlearn and tensorflow

BASIC APPROACHES

24th june 2020

Approach: Neural Network with 2 hidden layers and 128 nodes. ReLU activation. Adam optimizer
Test Dataset: Scenario 1a – 100 CSV
Columns included: node type, x, y, primary channel, min channel, max channel, Throughput
RMSE: 10.54
R2: 0.7796

Approach: Neural Network with 2 hidden layers and 128 nodes. ReLU activation. Adam optimizer
Test Dataset: Scenario 1b – 100 CSV
Columns included: node type, x, y, primary channel, min channel, max channel, Throughput
RMSE: 16.65
R2: 0.39

Approach: Neural Network with 7 hidden layers and 512 nodes. ReLU activation. Adam optimizer. Batch size 250
Test Dataset: 600 csv All scenarios
Columns included: node type, x, y, primary channel, min channel, max channel, Throughput
RMSE: 12.91
R2: 0.759

Approach: Neural Network with 7 hidden layers and 512 nodes. ReLU activation. RMSEprop optimizer. Batch size 250
Test Dataset: 600 csv All scenarios
Columns included: node type, x, y, primary channel, min channel, max channel, Throughput
RMSE: 13.19
R2: 0.749

25th june 2020

Approach: ElasticNet
Test Dataset: 600 csv All scenarios
Columns included: node type, x, y, primary channel, min channel, max channel,

Throughput
RMSE: 17.93
Mean Absolute error:8.644

Approach: Random Forest Regressor, 300 estimators
Test Dataset: 600 csv All scenarios
Columns included: node type, x, y, primary channel, min channel, max channel, Throughput
RMSE: 12.71
Mean Absolute error:6.47

Approach: PLS Regression
Test Dataset: 600 csv All scenarios
Columns included: node type, x, y, primary channel, min channel, max channel, Throughput
RMSE: 16.36

Approach: Basic Linear Regression (no neural network)
Test Dataset: 600 csv All scenarios
Columns included: node type, x, y, primary channel, min channel, max channel, Throughput
RMSE: 16.35
R2: 0.615

26th june 2020

Approach: Neural Network with 2 hidden layers and 128 nodes. ReLU activation. Adam optimizer. Batch size 250
Test Dataset: 600 csv All scenarios
Columns included: node type, x, y, primary channel, min channel, max channel,RSSI Throughput
Note: Inf in RSSI replaced by value of 20
RMSE: 13.012
R2: 0.75

Approach: Neural Network with 7 hidden layers and 512 nodes. ReLU activation. Adam optimizer. Batch size 250
Test Dataset: 600 csv All scenarios
Columns included: node type, x, y, primary channel, min channel, max channel,RSSI Throughput
Note: Inf in RSSI replaced by value of 20
RMSE: 13.18
R2: 0.749

Approach: Random Forests , 300 estimators
Test Dataset: 600 csv All scenarios
Columns included: node type, x, y, primary channel, min channel, max channel,RSSI Throughput
Note: Inf in RSSI replaced by value of 20
RMSE: 13.18
MAE:6.4

Approach: Linear Regression (no neural network)
Test Dataset: 600 csv All scenarios
Columns included: node type, x, y, primary channel, min channel, max channel,RSSI Throughput

Note: Inf in RSSI replaced by value of 20
RMSE: 15.94
R2: 0.625

27th june 2020

Approach: Neural Network with 2 hidden layers and 128 nodes. ReLU activation. Adam optimizer. Batch size 250
Test Dataset: 600 csv All scenarios
Columns included: node type, x, y, primary channel, min channel, max channel, RSSI,Channel-bonding_available(Max-Min), Throughput
Note: Inf in RSSI replaced by value of 20
RMSE: 13.13
R2: 0.749

Approach: Neural Network with 2 hidden layers and 128 nodes. ReLU activation. Adam optimizer. Batch size 250
Test Dataset: 600 csv All scenarios
Columns included: node type, x, y, primary channel, RSSI, Channel-bonding_available(Max-Min), Throughput
Note: Inf in RSSI replaced by value of 20
RMSE: 13.12
R2: 0.749

Approach: Neural Network with 7 hidden layers and 512 nodes. ReLU activation. Adam optimizer. Batch size 250
Test Dataset: 600 csv All scenarios
Columns included: node type, x, y, primary channel,RSSI,Channel-bonding_available(Max-Min), Throughput
Note: Inf in RSSI replaced by value of 20
RMSE: 12.99
R2: 0.75

28th june 2020

Approach: Neural Network with 7 hidden layers and 512 nodes. ReLU activation. Adam optimizer. Batch size 250
Test Dataset: 600 csv All scenarios
Columns included: node type, x, y, primary channel, min channel, max channel, RSSI,Channel-bonding_available(Max-Min), Throughput
Note: Inf in RSSI replaced by value of 20
RMSE: 12.9
R2: 0.755

Approach: Random forest regressor , 300 estimators
Test Dataset: 600 csv All scenarios
Columns included: node type, x, y, primary channel, min channel, max channel, RSSI,Channel-bonding_available(Max-Min), Throughput
Note: Inf in RSSI replaced by value of 20
RMSE: 13.18
MAE:6.4

Approach: Random forest regressor , 300 estimators
Test Dataset: 600 csv All scenarios
Columns included: node type, x, y, primary channel, RSSI,Channel-bonding_available(Max-Min), Throughput
Note: Inf in RSSI replaced by value of 20

RMSE: 13.17
MAE:6.39

Approach: Linear Regression (no neural network)
Test Dataset: 600 csv All scenarios
Columns included: node type, x, y, primary channel, min channel, max channel, RSSI,Channel-bonding_available(Max-Min), Throughput
Note: Inf in RSSI replaced by value of 20
RMSE: 15.94
R2:0.625

Approach: Linear Regression (no neural network)
Test Dataset: 600 csv All scenarios
Columns included: node type, x, y, primary channel, RSSI,Channel-bonding_available(Max-Min), Throughput
Note: Inf in RSSI replaced by value of 20
RMSE: 15.94
R2: 0.625

Approach: Neural Network with 2 hidden layers and 128 nodes. ReLU activation. Adam optimizer. Batch size 250
Test Dataset: 600 csv All scenarios
Columns included: node type, x, y, RSSI, Channel-bonding_available(Max-Min), Throughput
Note: Inf in RSSI replaced by value of 20
RMSE: 13.16
R2: 0.744

Approach: Neural Network with 27hidden layers and 512 nodes. ReLU activation. Adam optimizer. Batch size 250
Test Dataset: 600 csv All scenarios
Columns included: node type, x, y, RSSI, Channel-bonding_available(Max-Min), Throughput
Note: Inf in RSSI replaced by value of 20
RMSE:12.99
R2: 0.76

Approach: Neural Network with 2 hidden layers and 128 nodes. ReLU activation. Adam optimizer. Batch size 250
Test Dataset: 600 csv All scenarios
Columns included: node type, x, y, min channel, max channel, RSSI,Channel-bonding_available(Max-Min), Throughput
Note: Inf in RSSI replaced by value of 20
RMSE: 13.28
R2: 0.745

Approach: Neural Network with 7 hidden layers and 512 nodes. ReLU activation. Adam optimizer. Batch size 250
Test Dataset: 600 csv All scenarios
Columns included: node type, x, y, min channel, max channel, RSSI,Channel-bonding_available(Max-Min), Throughput
Note: Inf in RSSI replaced by value of 20
RMSE: 13.27
R2: 0.74

Approach: Random forest regressor , 300 estimators
Test Dataset: 600 csv All scenarios
Columns included: node type, x, y, min channel, max channel, RSSI,Channel-bonding_available(Max-Min), Throughput
Note: Inf in RSSI replaced by value of 20
RMSE: 13.18
MAE:6.4

Approach: Random forest regressor , 300 estimators
Test Dataset: 600 csv All scenarios
Columns included: node type, x, y, RSSI,Channel-bonding_available(Max-Min), Throughput
Note: Inf in RSSI replaced by value of 20
RMSE: 13.29
MAE:6.49

29th June 2020

Approach: XGBoost
Test Dataset: 600 csv All scenarios
Columns included: node type, x, y, RSSI,Channel-bonding_available(Max-Min), Throughput
Note: Inf in RSSI replaced by value of 20
RMSE: 12.64
R2:0.764

Approach: XGBoost
Test Dataset: 600 csv All scenarios
Columns included: node type, x, y, min channel, max channel, RSSI,Channel-bonding_available(Max-Min), Throughput
Note: Inf in RSSI replaced by value of 20
RMSE: 12.68
R2:0.7628

2nd July 2020

Approach: Random forest regressor , 300 estimators
Test Dataset: sce1a
Columns included: node type, x, y,primary,Channel-bonding_available(Max-Min), Interference, Throughput
RMSE: 11.74
MAE:4.79

Approach: Neural Network with 2 hidden layers and 128 nodes. ReLU activation. Adam optimizer. Batch size 250
Test Dataset: sce1a
Columns included: node type, x, y,primary,RSSI,Channel-bonding_available(Max-Min),
Interference, Throughput
RMSE: 11.6
R2: 0.81

Approach: Neural Network with 2 hidden layers and 128 nodes. ReLU activation. Adam optimizer. Batch size 250
Test Dataset: sce1a
Columns included: node type, x, y,RSSI,Channel-bonding_available(Max-Min), Interference, Throughput

RMSE: 11.5
R2: 0.80

Approach: Neural Network with 7 hidden layers and 1024 nodes. ReLU activation. Adam optimizer. Batch size 250
Test Dataset: sce1a
Columns included: node type, x, y,RSSI,Channel-bonding_available(Max-Min), Interference, Throughput
RMSE: 10.89
R2: 0.80

Approach: Random forest regressor , 300 estimators
Test Dataset: sce1a
Columns included: node type, x, y,Channel-bonding_available(Max-Min), Interference, Throughput
RMSE: 10.85
MAE:4.76

KNN Method

Test data: all scenarios (Aps and STAs included)
Data contained 12 nearest neighbors of each entity
Columns included as features: Node type,x,y primary channel, min channel, max channel
Label column: Throughput
Splitting : 80 for training and 20 for testing

Approach: ANN 7 layers with 1024 nodes each and ReLU activated
RMSE: 25.44

Approach: Random Forest Regressor
RMSE:26.15

Approach:SVR (Support Vector Regressor)
RMSE: 26.49

Approach: XGBoost Regressor
RMSE:25.96

When the trial was done with 30 neighbors the RMSE was in the range of 27 in the ANN method with 7 layers of 1024 node ReLU activation functions.

Other trial was done with 30 neighbors but the testing data was the whole of sce2c which gave an RMSE of 37.

FIXED RSSI VALUE TRIALS

The RSSI values were found to be missing in some of the files.
So after fixing it in sce1a, sce2a, sce2b and sce2c , all these scenarios were combined to test for RMSE and MAE.

BOTH AP and STAs

Training: 80-20 split
Columns included: Node type, x , y ,primary channel, min channel, max channel, RSSI, throughput
Approach 1: K nearest neighbors , 18 neighbors gave best RMSE of 14.2 and MAE

of 7.22
Approach 2: Random Forest Regressor gave RMSE of 14.6 and MAE of 7.34 with max depth of 10 and after iteration till max depth of 200, the RMSE reduced to 14.53.
Approach 3: SVR gave RMSE of 17 and MAE of 8.1
Approach 4: Gradient Boosting Regression gave RMSE of 15.7 and MAE of 7.85
Approach 5: Stacking Regressor comprising Ridge Regressor, SVR, KNN, and random forest put together gave 14.99 as RMSE.

ONLY STAs
Similar Approaches were tried for STAs separately. Note that all the columns as above are included except for the node type.
Approach 1: KNN gave RMSE of 6.88 and MAE of 4.67
Approach 2: Random forest Regressor gave RMSE of 6.91 and MAE of 4.88
Approach 3: SVR gave RMSE of 7.45 and MAE of 4.86
Approach 4:Gradient Boosting Vector gave RMSE of 6.99 and MAE of 4.96

ONLY APs
Similar approaches but RMSE was very bad in the range of 45.

ANN with only X, Y , RSSI and Throughput
ANN was used on only these features. It used 4 layers of 128 nodes each. The RMSE turned out to be 15.55.

PROPOSED METHOD==> CNN
The CNN method first converts the data into array format.
The features array are 800 x 600 x 5 arrays. The 800 x 600 is the dimension of the total area and 5 is the number of features encoded in each of the array.
So at total there are 600 Feature arrays , each array pertaining to one single deployment and the features are encoded in the x-y location wrt the grid.
Similarly throughput arrays were built with 800 x 600 x 1 as only throughput was encoded.
This model was taken into consideration in order to take into account the spatial information.
The CNN model had to perform regression and give an array as output.
The Conv2D and Conv2DTranspose layers are built with various kernel sizes.
Only 2 layers of each were used for trial number 1.
The number of epochs for training were 5 and the RMSE on the first trial was about 45 to 50.
More work has to be put in this method.

New RSSI Values
Data: All scenarios
Features :node_type, x , y , primary , min , max, RSSI
Label: throughput
ANN with 7 layers and 1024 nodes and 500 epochs
STAs and Aps ==> RMSE=13.2; R2=0.74; MAE=6.22; Adj R2=0.74;
STAs ==> RMSE=6.25; R2=0.8; MAE=4.2; Adj R2=0.8;

26 September 2020
SINR Values Combined
Data: All Scenarios
Features: node type, x , y ,primary , min , max , RSSI , SINR
Label: throughput
Note: sinr values replacements: nan = -9999, inf = 10000
Data split : 80 percent training and 20 percent testing

ANN with 7 layers and 1024 nodes and 500 epochs
STAs and Aps ==> RMSE=13.54 ; MAE = 5.67
STAs ==> Testing data :: RMSE = 5.2 ; MAE =3.89
Training data :: RMSE = 1.32 ; MAE = 0.667