
SCTE®
CABLE-TEC EXPO®
FAST FORWARD 2021
ATLANTA, GA
OCTOBER 11-14
SCTE
a subsidiary of CableLabs®
UNLEASH THE POWER OF LIMITLESS CONNECTIVITY
2021 Fall Technical Forum
SCTE® • NCTA • CABLELABS®

SCTE
a subsidiary of CableLabs®

Cloud & Virtualization

Cluster-Based Network Traffic Prediction Pipeline For Big Data Time Series

WEI CAI

Network Planning Engineer
Cox Communication

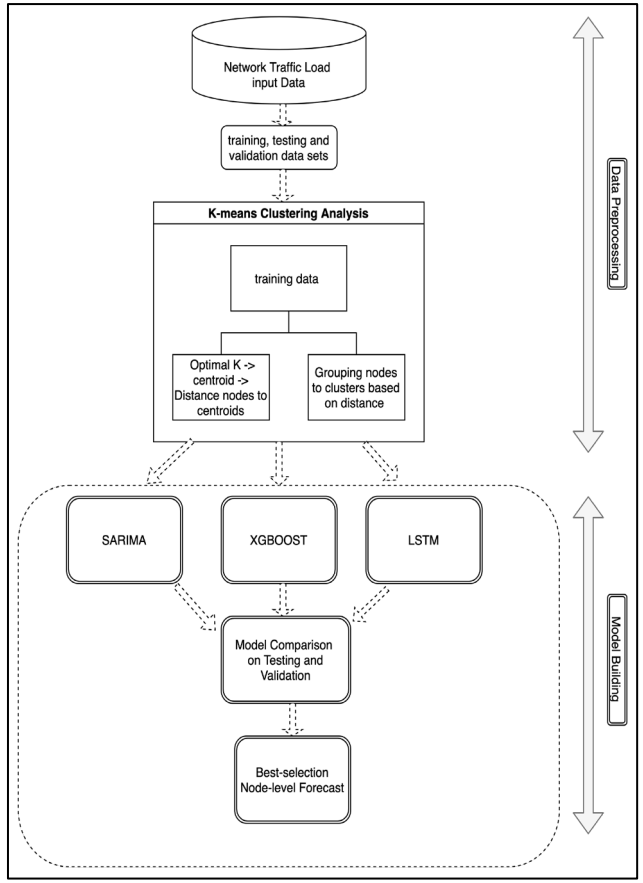
2021 Fall Technical Forum
SCTE • NCTA • CABLELABS

SCTE CABLE-TEC EXPO FAST FORWARD 2021
VIRTUAL EXPERIENCE
OCTOBER 11-14

Rationales

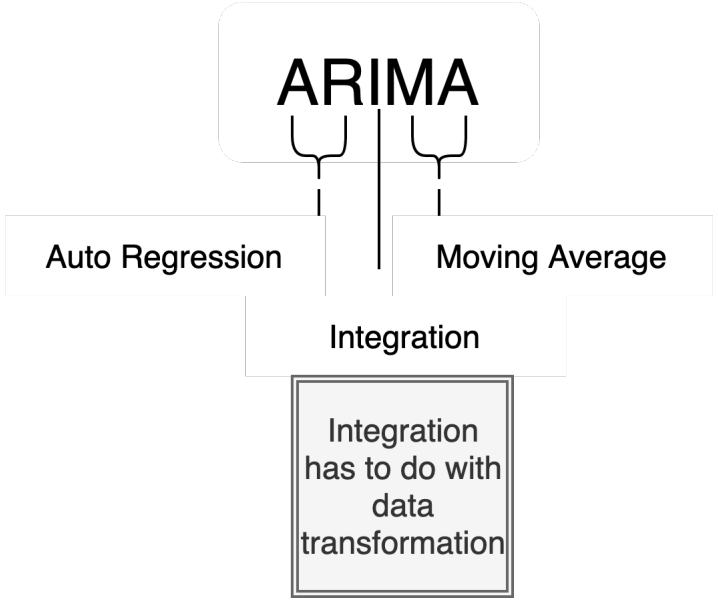
- Irregularity and more fluctuations
- Big data
- Similarity in traffic patterns across time series

Proposed clustering-based best-selection forecast pipeline



ARIMA

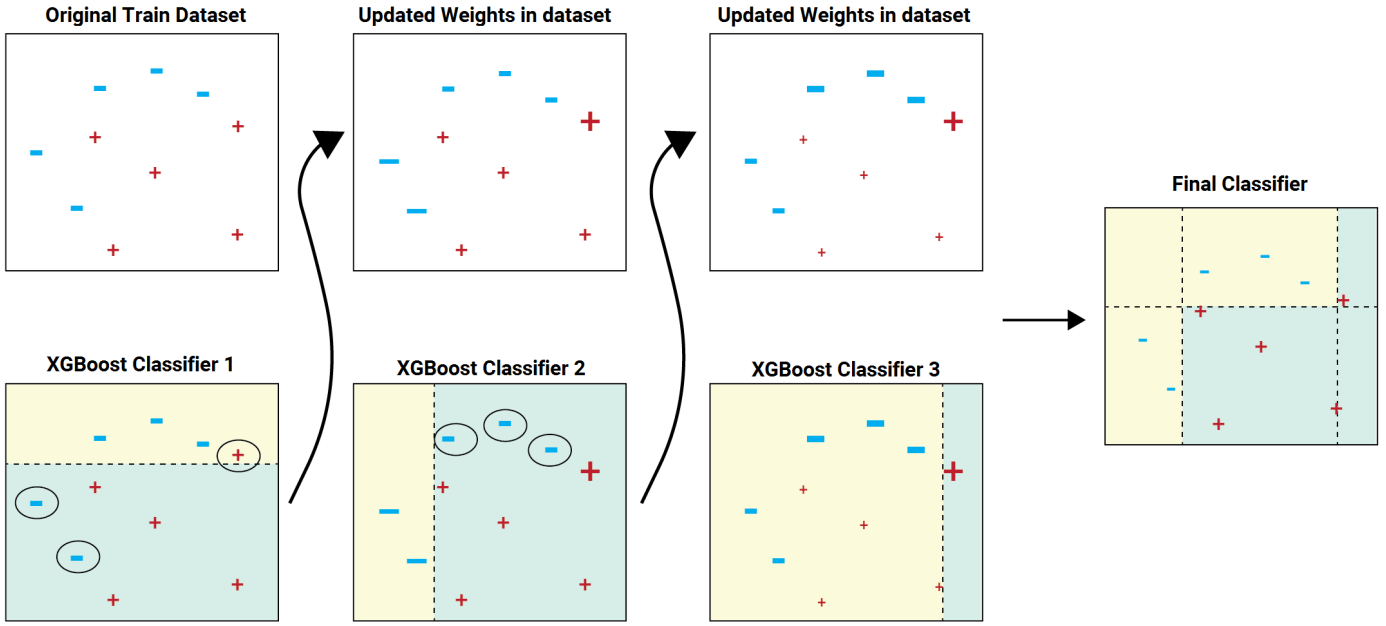
ARIMA models are designated by Autoregression, Integration and Moving average



ARIMA structure

XGBoost

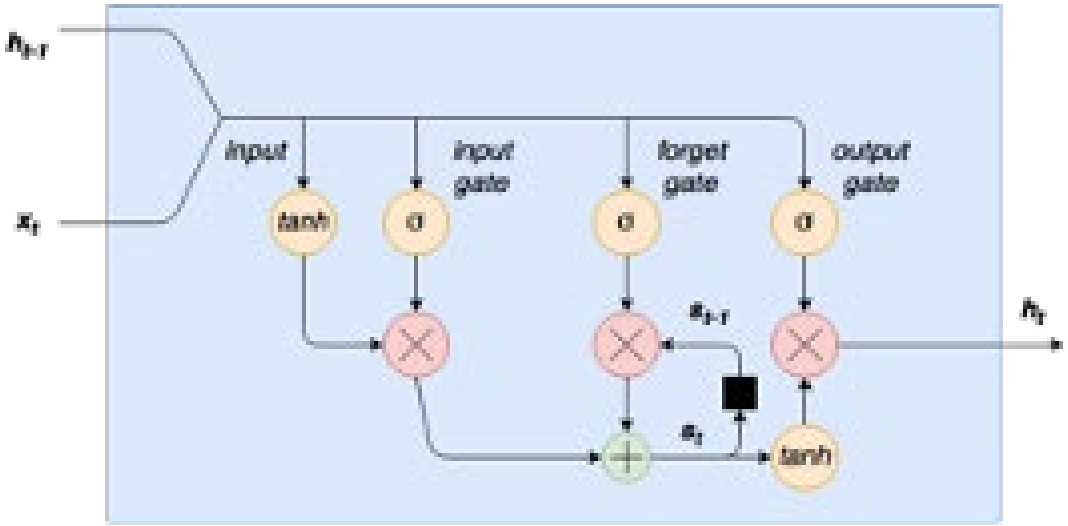
XGBoost models combines weak learners to form a strong model through iterations



XGBoost structure (https://blog.quantinsti.com/xgboost-python/)

LSTM

XGBoost models combines weak learners to form a strong model through iterations



LSTM structure (https://adventuresinmachinelearning.com/keras-lstm-tutorial/)

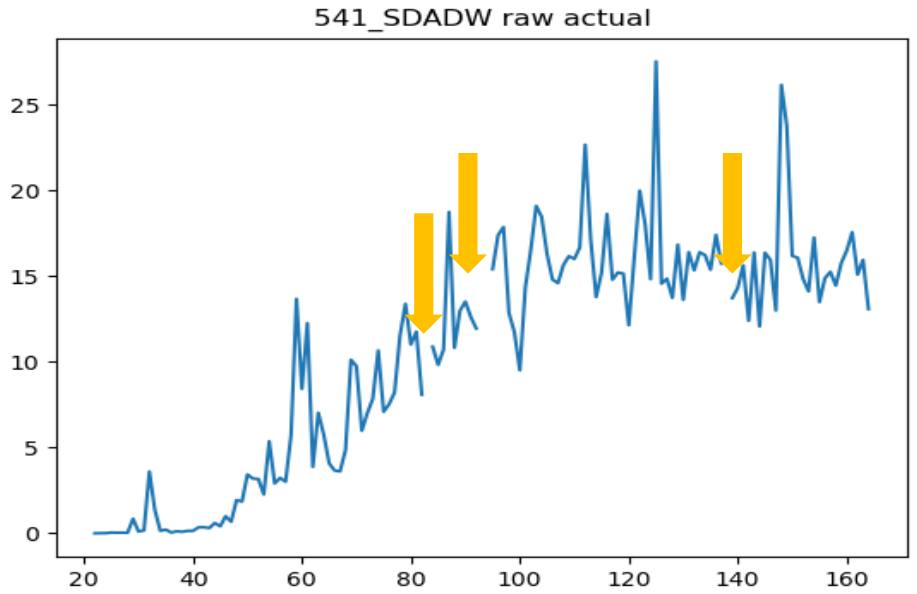
Two major gaps

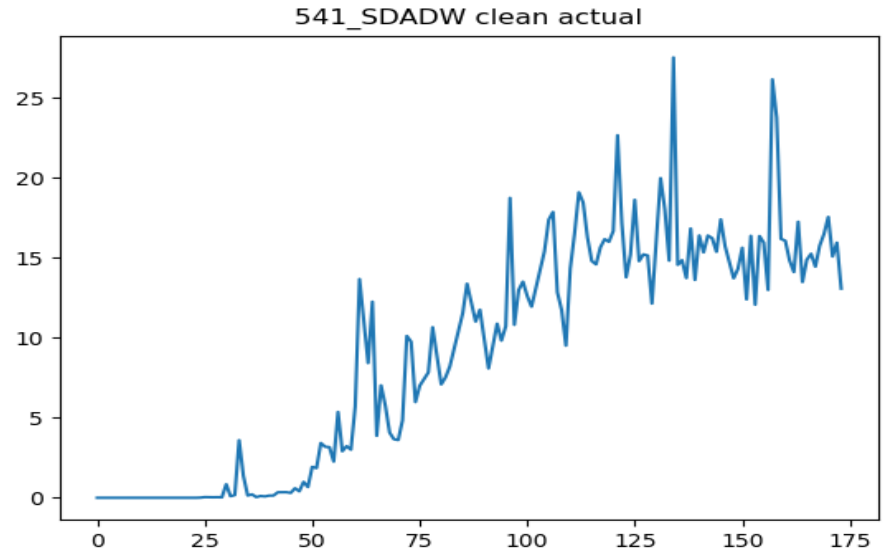
1. Very few research papers have employed a clustering approach into large scale network traffic forecasting workflow

2. Best selection from different models performs better than averaging the results from different models, particularly for network time series with more volatilities

Raw data with missing polls

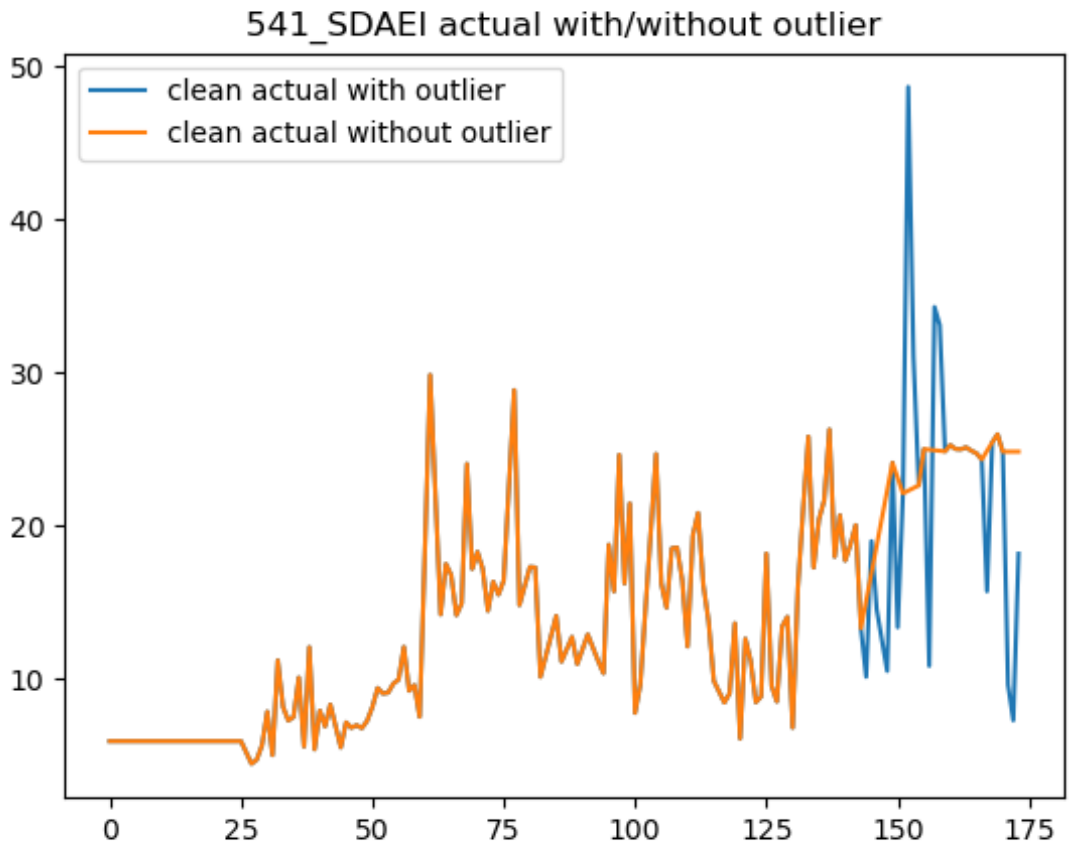
Imputed clean data





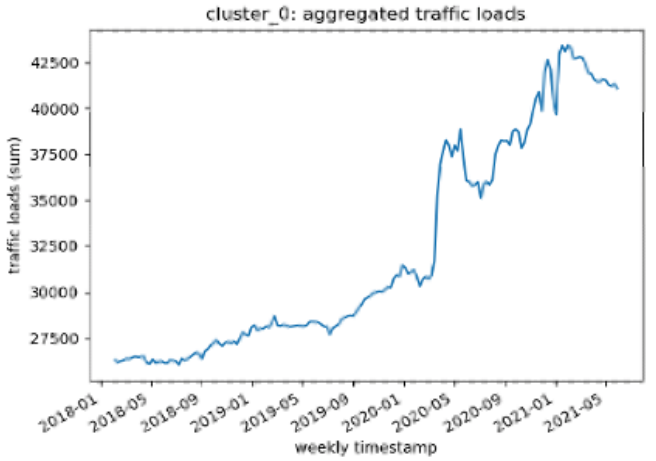
Clean data with outliers

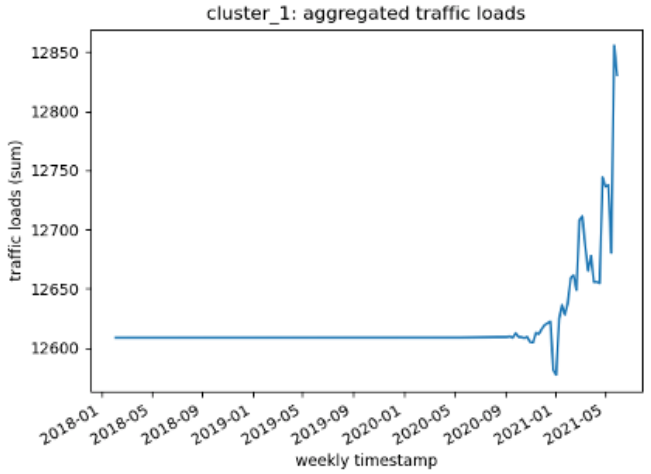
Clean data with no outliers

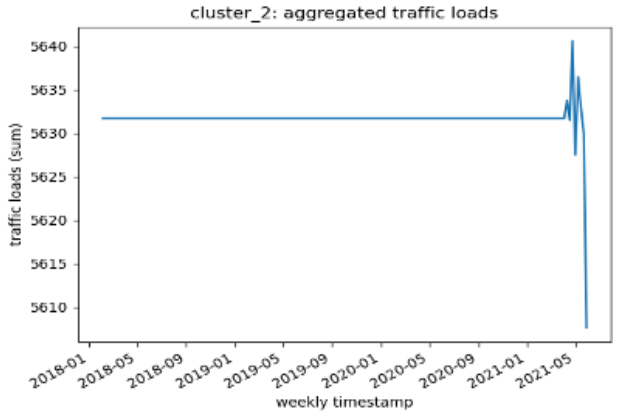


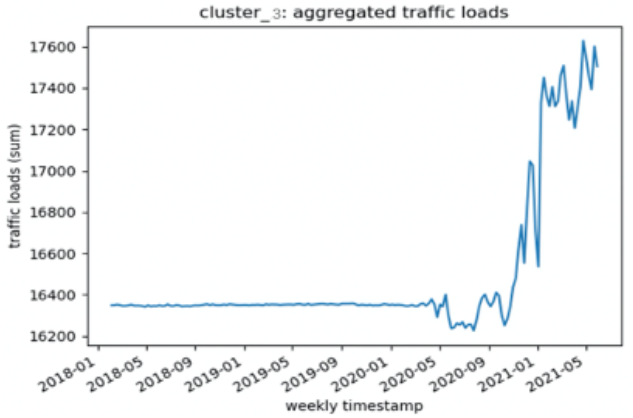
Clustering nodes to 4 clusters

Cluster Label	Node Counts
Cluster 0	10,082
Cluster 1	3,92
Cluster 2	71
Cluster 3	1,193









METHODOLOGY: FORECASTING MODEL SET-UP

- Grid search was used to automatically discover the optimal order of non-seasonal and seasonable parameters at a cluster level.
- Dataset is split into 70%, 20%, 10% training, test and validation sets, respectively.
- Stationarity of the series are checked.
- SARIMAX models are trained and obtained to make estimation on fresh test data.
- MAPE values for both training dataset and testing data set are calculated.

SARIMA forecasting Procedures

METHODOLOGY: FORECASTING MODEL SET-UP

- Grid search algorithm is used to optimize the parameters at a cluster level.
- Dataset is split into 70%, 20%, 10% training, test and validation sets, respectively.
- XBGoost models are trained and obtained to make estimation on fresh test data.
- MAPE values for both training dataset and testing data set are calculated.

XGBoost forecasting Procedures

METHODOLOGY: FORECASTING MODEL SET-UP

- Scale data using MinMaxScaler to speed up the learning process and help model.
- Hyperparameter such as number of layers, layer depths, activation functions, dropout coefficients are repeatedly tuned at a cluster.
- Dataset is split into 70%, 20%, 10% training, test and validation sets, respectively.
- LSTM models are trained and obtained to make estimation on fresh test data.
- MAPE values for both training dataset and testing data set are calculated.

LSTM forecasting Procedures

RESULTS AND DISCUSSION

Table 1 Training MAPE by Models

Model	Training MAPE				Total
MAPE Range	<= 5%	>=6% & <= 10%	>=11% & <= 15%	>=16%	
SARIMA	1,360	2,828	2,059	5,295	11,542 nodes
XGBoost	11,333	55	20	134	11,542 nodes
LSTM	3,573	2,979	2,909	2,081	11,542 nodes

Table 2 Test MAPE by Models

Model	Test MAPE				Total
MAPE Range	<= 5%	>=6% & <= 10%	>=11% & <= 15%	>=16%	
SARIMA	982	746	1,175	8,639	11,542 nodes
XGBoost	2,244	2,228	2,870	4,200	11,542 nodes
LSTM	2,299	3,996	2,753	2,494	11,542 nodes

Table 3 Node Counts with <= 10% MAPE by Models

Model	Node counts with good fit (Mape <= 10%)	Node counts with overfit (Mape <= 10%)
SARIMA	956	586
XGBoost	1,195	3,248
LSTM	2,129	2,278

Table 4 Node Counts MAPE by Models

Model	Node counts with Mape <= 10%)	Node counts with Mape >10%)
SARIMA	101	495
XGBoost	139	457
LSTM	346	250

SCTE® CABLE-TEC EXPO® FAST FORWARD 2021

ATLANTA, GA
OCTOBER 11-14

SCTE
a subsidiary of CableLabs®

Thank You!

WEI CAI

Network Planning Engineer
Cox Communication

2021 Fall Technical Forum
SCTE® • NCTA • CABLELABS®