VISHWAKARMA INSTITUTE OF INFORMATION TECHNOLOGY, PUNE

COMPUTER ENGINEERING DEPARTMENT

APRIL-MAY 2018

Synopsis

Group number:

Group Members :

1. Sameer Khan   -421002   sameer.khan@viit.ac.in

2. Abhishek Kumar -421072   abhishek.kumar@viit.ac.in

3. Mayur Kharbas   -421064   mayur.kharbas@viit.ac.in

4. Rahul Gupta   -421004   rahul.gupta@viit.ac.in

Mobile no : Sameer Khan – 8805662445

Title :

Traffic Prediction and Anomaly detection in wireless cellular networks from Call Detail Record Analysis

Objective :

To predict cellular traffic in wireless mobile networks and anomaly detection to improve Quality of experience for mobile users using call detail records.

Abstract :

With rising capacity demand in mobile networks, the infrastructure is also becoming increasingly denser and complex. This results in collection of a larger amount of raw data(big data) that is generated at different levels of network architecture and is typically underutilized. To unleash it's full value, valuable insights need to be extracted for improving overall network performance. Additionally, a major challenge for network operators is to cope up with increasing number of call drop rates, managing heavy network usage, resolving cell outages while simultaneously reducing operational expenditure. This project aims to provide solutions to the aforementioned problems by exploiting the call detail records generated by the network operators. These call records are primarily used for billing purposes however, they can provide meaningful insights to detect cell outages and help to predict future cellular traffic for a region which will help telecom operators to better manage and operate their network.

Working :

- Predict cellular traffic of users based on ground truth data

- To detect anomaly in cellular network. Anomalies result in Network outage or are a result of network outage. Anomaly could be a sudden spike in usage leading to network congestion or an unusually low usage due to poor network

- Providing various infographics from data analysis to provide actionable insights to users

Tehnical Details :

Dataset Description :

The Spatio-temporal data was made public by Telecom Italia as a part of their Open Big data Initiative.

The data is geo-referenced and is processed from CDRs of their subscribers

For each grid, following information is present -

1) Grid ID,
2) Time stamp (in milliseconds) of 10 minutes duration,
3) Country code,
4) Activity in terms of received SMS,
5) Activity in terms of sent SMS,
6) Activity in terms of inbound calls,
7) Activity in terms of outbound calls, and
8) Internet usage.

Algorithms and Techniques :

1) K means clustering for anomaly detection

K-means can be applied among "total activity and activity hours" to find the usage pattern with respect to the activity hours.

By using this clustering mechanism, we can find the clusters making more traffic to the telecom network in the measure of total activity.

2) Arima model for cellular traffic prediction

ARIMA stands for Auto-Regressive Integrated Moving Averages.
It can be used for forecasting a stationary time series data.
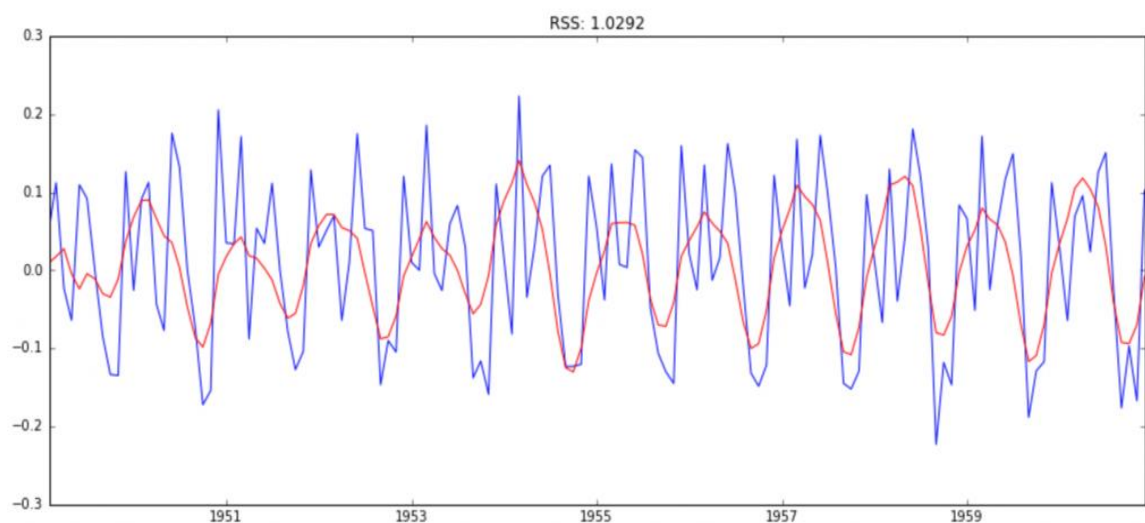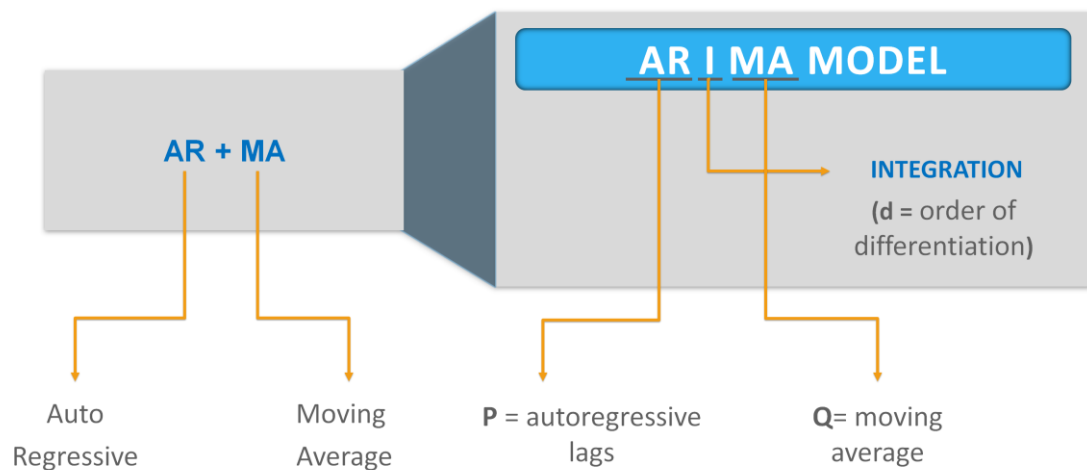
Applications :

- The application proposed is aimed for providing solutions to telecom operators.

- It will help them gain an insight into their consumer base and analyze customer pattern and behaviour to come up with efficient sales strategy and optimize operations.

- It will also help to improve the quality of service and hence reduce the customer churn.

References :

1. Bilal Hussain, Qinghe Du, Pinyi Ren, "Semi-Supervised Learning Based Big Data-Driven Anomaly Detection in Mobile Wireless Networks", April 2018

2. Kashif Sultan, Hazrat Ali, Zhongshan Zhang, "Call Detail Records Driven Anomaly Detection and Traffic Prediction in Mobile Cellular Networks", DOI 10.1109/ACCESS.2018.2859756, IEEE Access

3. Dataset link : https://dandelion.eu/datagems/SpazioDati/telecom-sms-call-internet-mi/description/