Predicting Call Drop Quality based on Data collected using TRAI MyCall App

Kandula Mohan Krishna

CS & IS Department  
BITS Pilani Hyderabad CampusIndia Maddimsetti Trinadh

CS & IS Department  
BITS Pilani Hyderabad CampusIndia

Uday Shahapur

CS & IS Department  
BITS Pilani Hyderabad CampusIndia

*Abstract*—This paper summarizes the observations obtained by using various data mining techniques in visualizing and analyzing call record patterns across various states in India which will lead to finding insights about the efficiency of different Network Operators in different locations.

Keywords—Call record patterns, Network Operator, Data mining, Network efficiency

# Introduction

India has the one of the fastest growing economies in the world. This is the age of wireless networks and as taxpayers, we have every right to enjoy Quality of Service (QoS). But unfortunately, now-a-days it is very common to hear from customers that they aren’t satisfied with the services provided and suffer from sudden or frequent termination of their calls. The sudden and undesirable termination of successfully established calls from Telecom Service Provider’s (TSP’s) side are known as call drops. A call drop is a situation where a call on a wireless network is disconnected before the caller ends the call. Some of the main reasons for call drops:[1]

* Lack of tower infrastructure
* Improper network planning
* Non-optimisation of network.
* Points of Interconnection

# Data PreProcessing

## Defining a Problem

* To Predict the Call Drop category base on Network operator, Network Type, State name and mobility state (Indoor/Outdoor/Travelling.)
* State of Call Drop Categories (‘Poor Call Quality’, ‘Call Dropped’, ‘Satisfactory’).
* Analysing Data for Best Network provider in a specific state.
* Best TSP for a given mobility type in State.

Python Libraries are used for analysing the Data.

## Data Description

Birla Institute of Technology and Sciences, Hyderabad Campus

The dataset which we have selected (Voice call quality Customer experience) contains the records over 277000 call recorded in different states all-over India by TRAI.

* **Network Operator**​: The dataset contains 5 classes namely ​Rjio, Idea, Airtel, Vodafone, BSNL. ​It is a nominal and discrete attribute.
* **In Out Travelling**​ ​: ​This attribute consists of 3 classes namely ​Indoor, Outdoor, Travelling. ​It is a nominal and discrete attribute.
* **Network type​** ​: ​ This attribute consists of 3 types namely ​2G,3G,4G. ​It also holds numerous instances of missing values. It is a nominal and a discrete attribute.
* **Rating**​ ​: ​The calls are rated as an integer between 1 to 5, 5 being the best. It is an ordinal and discrete attribute.
* **Latitude**​ ​: ​ It is a ratio and continuous attribute.
* **Longitude**​ ​: ​ It is a ratio and continuous attribute
* **State name**​ ​: ​ It is a nominal, discrete attribute.
* **Call drop Quality**​ ​: ​ This attribute consists of 3 classes namely ​Poor Voice Quality, Call Dropped, Satisfactory. ​It is an ordinal and discrete attribute.

## Exploratory Data Analysis

**Visualization**:

Libraries like Matplotlib and Seaborn are used for visualization. As it is mostly nominal Data, Visual representations like Bar graph and Pie chart are used to visualise the Data. Bar graph showing frequency distribution –

1. In a state between different call drop categories (Poor Call Quality, Satisfactory, Call Dropped).
2. In a Network operator between different call drop categories.
3. In a Network type.

Also visualise using pie-charts showing frequency distribution of call drop category in all states.

**Test of Independence**: **(chi-square test)**

It is a test used to check the dependence of two attributes. It considers two hypotheses.

Null Hypothesis: Two attributes are independent.

Alternate Hypothesis: Two attributes are not independent.

X2=(observed−expected)2**/**(expected)

D.O.F = (no. of rows - 1) (no. of columns -1)

Using these, find p-value, if p < 0.05, two attributes are statistically dependent. For this test between call drop quality and Network operator, Travelling mode and state.

## Data Cleaning

* **Missing Values:** The data objects whose values in Operator attribute or when it has ‘-1’ in latitude or longitude column are removed.
* The data objects which has State name missing with latitude and longitude are handled by reverse geocoding using geopy library.
* And for resulting states, Original state names are replaced using python dictionary data type.

## Data Transformation

Network operator which is a nominal attribute is converted into binary attributes with each of its unique element as attribute by one-hot encoding. First each of operator will be assigned a label which later converted into binary attribute.

## Data Reduction

The operator attribute having 8 distinct nominal elements are converted into 4 numeric discrete attributes to ease application of Machine Learning Algorithms.

##### References

1. Puru Gaur, “A Review of Menace of Call Drops in India and Possible Ways to Minimize It,” International Journal of Mathematical, Engineering and Management Sciences Vol. 1, No. 3, 130–138, 2016 ISSN: 2455-7749.

**IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove template text from your paper may result in your paper not being published.**