PROJECT SYNOPSIS

Project Guide: Sulekha Aloorravi

Project Leader: Sanku Lahiri

Team Members: Amol Ravatale, Devendra Bharadwaj, Kushal Thakral

Project Title: Predict Ambulation Locations to serve accident hotspots

**Problem Definition**

Road accidents contribute to approximately 1.35 million deaths [*source: WHO]* worldwide annually. A major factor in road accidents which decide mortality is post trauma care. This is dependent on how fast medical attention is provided through ambulances after accident. The accessibility of ambulance service is fragmented based on the development index of a nation. Wealthier and developed economies have very fast and reliable ambulance care services while developing and weaker economies lag drastically. Research shows that the survival rate of accident victim increases if an ambulance is in the vicinity of 3-4 miles of accident site to provide post trauma care. This means that if an ambulance could be placed beforehand in the vicinity of all such accident locations it would be able to serve victims faster. Our paper aims to develop a Machine Learning based model to predict locations of 6 ambulances for 8 timeslots in a day (24 Hrs.) based on the history of accident locations. For our analysis we have taken data set from “Uber Nairobi Ambulation Pre-ambulation” challenge which provides accident history, locations etc. for Nairobi.

**Introduction and Literature Survey**

With an average mortality rate of 17.4 per 100,000 people road accidents by far are one of the major (top 10) contributors in human deaths every year. While developed countries fare better, this rate is significantly high for low-and middle-income countries comprising majority of global population. One of the major factors contributing to high mortality is inefficient pre-hospitalization medical attention [**2**]. Most of these injury deaths could be prevented with the timely access/arrival of medical emergency team/ambulances. As per the trauma experts the first 60 min in road accidents or “golden hours” are a matter of life or death for the victim [**4**]. This means that an ambulance’s presence in the vicinity of an accident location could drastically change the equation in favor of accident victims.

As per WHO report African nations have the highest death rate for accident road victims to the extent that some countries have more than double the average global rate [**1**]. We have chosen our problem statement for Nairobi, Kenya which struggles with high rate of road accident victims due to high response time of up to 162 min for an ambulance to reach accident site [**6**].

This huge time taken by ambulances is mainly due to traffic issues and lack of availability of ambulances at all the trauma or hospitals. One way to improve the quick response times of the ambulances is that they could be placed at/near locations which are more accident prone. Eventually, the framework could be developed to identify large areas of city where sets of ambulances could be placed to cater to each respective area instead of all ambulances servicing the entire city.

The problem of predicting different aspects of future (road/marine/air) accidents like location, timing and co-related factors like traffic flow etc. is studied by different researchers at academia and institutes associated with traffic management in area like roads, ports, marine transport etc. Studies have been done to predict accidents using ‘Grey System’ model consisting of radial basis function networks have been proposed to [8][9]. Such models are designed to model complex system featured by nonlinearity, uncertainty, and time-varying dynamics line in case of road/marine traffic system and to predict accidents, corelated factors. Some researchers have done study of predicting accidents at ports and co-related factors using particle swarm optimization-based RBF neural network (PSO-RBF neural network) and RBF-neural network and found that PSO-RBF neural network models are better [10][11]. Literature established the model of the relationship between traffic volume, lane number, turning lane number, bicycle facilities, control mode and traffic safety at intersections [13]. Literature [12] shows work done to predict highway traffic accident potential by identifying hazardous traffic conditions and normal traffic conditions with real-time traffic data. Literature [14] attempts to use spatial and temporal data to predict accidents.

Thus, literatures show that researchers have applied different machine learning models like liner/logistic regression, Bayesian classifier, KNN clustering, deep learning methods, LSTM, ConvLSTM to predict various aspects of accident event like timing, location and co-related factors like traffic flow, speed of vehicles, road features like intersection, cycle tracks etc.

**Tentative List of Algorithms & Approach**

The “Uber Nairobi Ambulation Pre-Ambulation” challenge provides data set with accident locations across Nairobi for the period Jan 2018 to June 2019. A supplemental data comprising weather parameters e.g., precipitation, humidity, temperature, and wind speed is also provided for this duration. Moreover, road conditions comprising pavement, zebra crossing availability, gradient etc. are also provided for entire city.

The main objective of our paper is to provide a solution which could be applied to predict the locations where ambulances could be placed across the city during a certain time slot in a day to cater to accident victims promptly. These locations would change of course for any given timeslot and ambulances will have to travel to the new locations. These locations will be derived based on the minimum distance an ambulance will have to travel to reach accident spot. Further, we have used certain assumptions like, no one set of ambulance will be available throughout the day to be placed at different locations; for a certain timeslot only the given set of ambulances will cater to all the accidents through multiple trips if required and finally the minimum distance calculated for ambulance location does not guarantee minimum response time as traffic parameters etc. have not been considered.

We have attempted to find the solutions through two approaches. The first approach considers the time-series approach. For this the date/timestamp is considered as input while accident locations are considered as output or targets. A custom solution is designed to segregate all the accident locations based on time stamps and convert to a unique number. Eventually, using historical values of locations the future locations of ambulances locations are derived using time-series analysis techniques. This approach does not consider weather or road parameters as input and simply relies on the behavior (trend, seasonality etc.) of accident locations.

The second approach is much more inclusive where it considers all weather parameters, road conditions, average vehicle speeds for a period and accident locations to predict the feasibility of an accident at a particular location. This approach uses conventional machine Learning methods like Logistic Regression, SVC etc to determine the possibility of accident and applies a clustering method to group locations with accident possibilities. Further, it utilizes tailored solution to derive locations based on various inputs like weather, road condition, speed etc. Eventually both the approaches will determine 6 locations where when ambulances are placed will be at minimum distance from accident spots.

Approach-(1)

The first approach using time series analysis is formulated in the following steps:

1. Segregate data as per respective 8 time slots for entire duration of provided data.
2. Derive a location centroid for each daily timeslot.
3. Using Bayesian Structural Time Series /Continuous Time Series Markov Model to predict the next location (centroid).
4. Validation whether the prediction locations are at optimum distance.
5. Identify clusters using DBSCAN in a particular time slot in similar period (date-time).
6. Apply RMSE or Euclidean measure to derive distance on test set.

As mentioned previously, the first approach uses time-series as the core solution approach. The accident locations provided in the data are of continuous nature and the location-based service model is the best fit for such scenario. Moreover, since this is a changing or dynamic location, we decided to attempt Continuous Time Series – Markov Model and Gaussian Mixed Model [**3**]. A brief snapshot of the time series “Location Centroid” approach is depicted in Figures below.

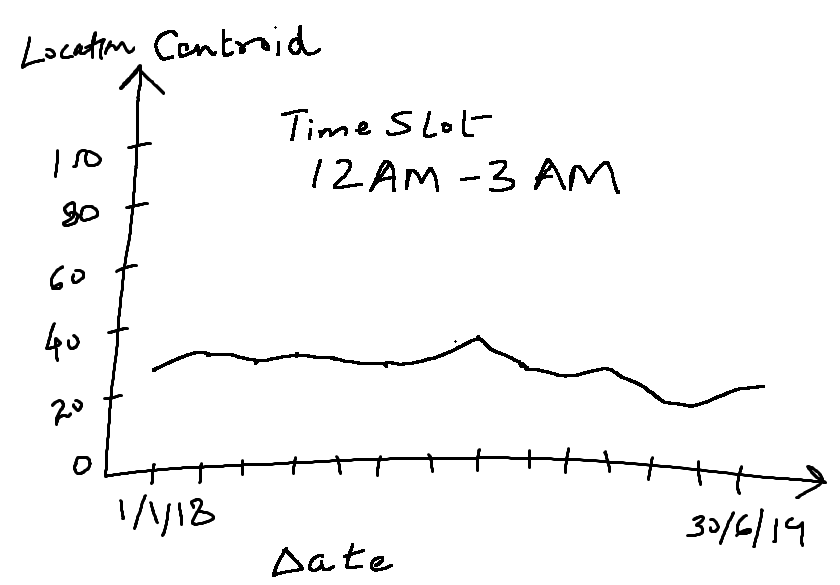


Figure 1: Accident Location Centroid – Training Data Trend

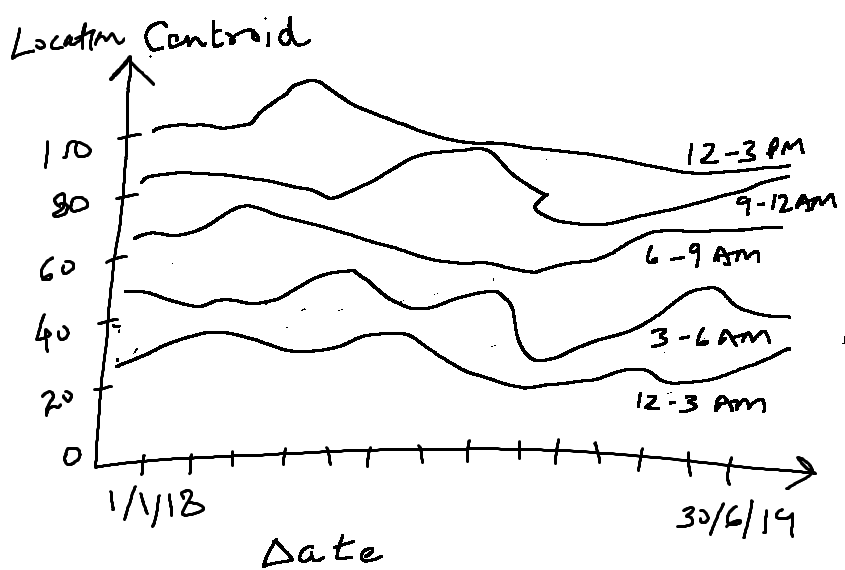


Figure 2: Accident Locations Centroid: All Timeslots

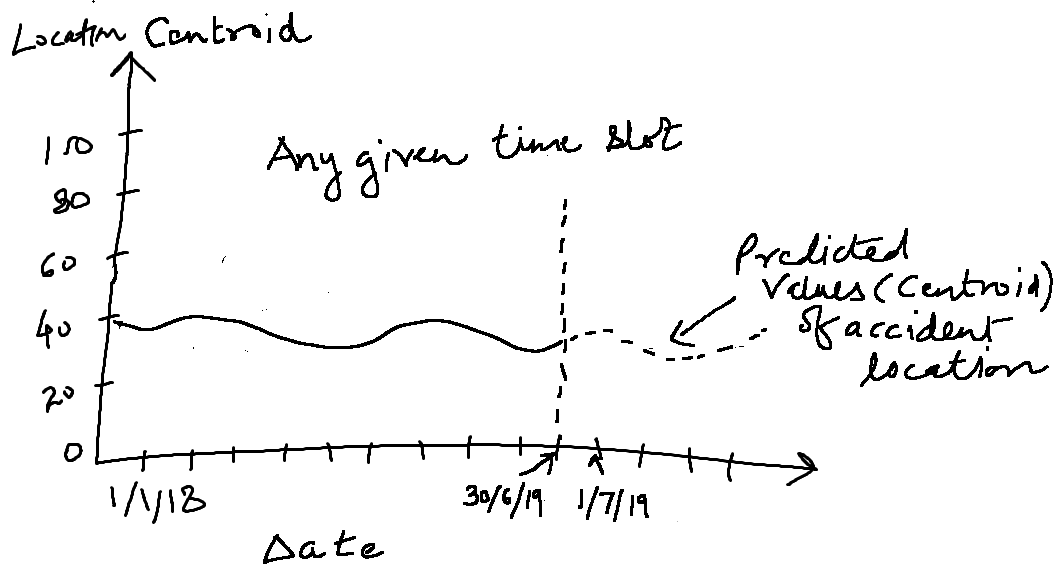


Figure 3: Predicted Accident Locations Centroids

Approach-(2)

## Prediction of Future Accidents

### Feature Engineering - Create below fautres using the data set menitoned in section (II)

### Date & Time Features – Create feature for combination of Day of the month, Day of the week, Hour and minute.

* 1. Segment Features – Create feature for each road characteristic like existence of crosswalk, presence of obstacle, behavioral characteristics like people walking along the side of road etc. create feature.
  2. Weather Features – Create feature for each weather attribute like humidty, temprature etc.
  3. Traffic Features – Create fautre like travel time between zone, average street speed etc.

### Build Model - Create classification model to classify if accidents happens or not due to input features described above.

1. Prediction
   1. In order to predict future crash locations, randomly identify multiple geospatial points (lattitude and longitude) on different roads/segments. Along with other input details such as future day of month, day of week, weather input, segment features of those segments, traffic features pass these details to the model to predict if accident will happen or not.
   2. Using above step predict multiple crash locations for future ‘n’ days which will give multiple crash locations where accident may happen.

## Placement of six virtual ambulances : Use crash locations predicted in step 3(b) and idetify 6 clusters using clustering methods. The centroid of the cluster will be the location of ambulance placment.

**References:**

1. Brodsky, H. (n.d.). Emergency medical service rescue time in fatal road accidents. Retrieved April 30, 2021, from Trb.org website: http://onlinepubs.trb.org/Onlinepubs/trr/1990/1270/1270-011.pdf
2. Coats, T. J., & Davies, G. (2002). Prehospital care for road traffic casualties. *BMJ (Clinical Research Ed.)*, *324*(7346), 1135–1138.
3. Du, Y., Wang, C., Qiao, Y., Zhao, D., & Guo, W. (2018). A geographical location prediction method based on continuous time series Markov model. *PloS One*, *13*(11), e0207063.
4. F. Sam, E., K. Blay, D., Antwi, S., Anaafi, C., & A. Adoma, J. (2019). Pre-hospital and trauma care to road traffic accident victims: Experiences of residents living along accident-prone highways in Ghana. In O. Karcioglu & M. G. Eneyli (Eds.), *Emergency Medicine and Trauma*. London, England: IntechOpen.
5. Dong, C., Shao, C., Li, J., & Xiong, Z. (2018). An improved deep learning model for traffic crash prediction. Journal of Advanced Transportation, 2018, 1–13.
6. <https://zindi.africa/competitions/uber-nairobi-ambulance-perambulation-challenge/data>
7. Yu, R., Xia, X., Liao, S., & Wang, X. (2015). A location prediction algorithm with daily routines in location-based participatory sensing systems. International Journal of Distributed Sensor Networks, 2015, 1–12.
8. Yang, Z.-Q. (2011). Marine traffic accident prediction based on particle swarm optimization-based RBF neural network. 2011 3rd International Conference on Computer Research and Development, 1, 435–438. IEEE.
9. Su, H., Zhang, W., & Li, Z. (2018). Analysis and prediction of water traffic accidents in jingtang port based on improved GM (1,1) model. 2018 37th Chinese Control Conference (CCC), 2212–2217. IEEE.
10. J. C. Yin, Z. J. Zou, F. Xu and N. N. Wang, "Online ship roll motion prediction based on grey sequential extreme learning machine", Neurocomputing, vol. 129, pp. 168-174, 2014
11. J. C. Yin and N. N. Wang, "Online grey prediction of ship roll motion using variable RBFN", Applied Artificial Intelligence, vol. 27, no. 10, pp. 941-960, 2013
12. Lv, Y., Tang, S., & Zhao, H. (2009). Real-time highway traffic accident prediction based on the k-nearest neighbor method. 2009 International Conference on Measuring Technology and Mechatronics Automation, 3, 547–550. IEEE.
13. Literature established the model of the relationship between traffic volume, lane number, turning lane number, bicycle facilities, control mode and traffic safety at intersections

Greibe Poul, "Accident prediction models for urban roads", Accident Analysis and Prevention, vol. 35, pp. 273-285, 2003.

1. Zhou, Z. (2019). Attention based stack ResNet for citywide traffic accident prediction. 2019 20th IEEE International Conference on Mobile Data Management (MDM), 369–370. IEEE.

**Glossary**

ACF – Auto Correlation Function

AIC - Akaike Information Criterion

AR – Auto Regression

ARIMA - Auto Regressive Integrated Moving Average

BIC - Bayesian Information Criterion

MA – Moving Average

MAE – Mean Average Error

PACF – Partial Correlation Function

RMSE – Root Mean Square Error

WHO – World Health Organization