Luan Nguyen
11-01-2020

Predictive Police Patrolling

1 Problem statement

The goal of this work is to forecast daily *police attention level needed* for each neighborhoods in Detroit city, Michigan. The forecast is aim at 24 hours with 3-hour increments.

2 Introduction

Given the increased in number of violent activities recent and how public safety has always been a concern in the community, police patrol is crucial in ensuring public security. In the past, police patrol with random routes. However, this is highly inefficient because the number of available police officers is limited. More recent approach is to use crime hot spots to plan patrol route. However, crime hot spots are based on static information, such as previous year crime record or demographics. This does not address the suggested dynamic nature of emergency need.

The work in this report aim at providing a more dynamic forecast (daily in 3-hour intervals) to help police office plan their routes effectively.

3 Data

3.1 Obtain

The data was obtained from the City of Detroit Open Data Portal using their url. The data sets downloaded (saved as csv files) include: 911 calls since Sep 2016, 911 calls of the last 30 days, and neighborhood bounding box.

The 911 calls records includes over 800 thousands 911 calls (initiated by police officers and citizen) for the year 2020 alone. Each entry (row) is a 911 call. There are 27 features that were recorded for each 911 calls. For example the features that are important to this work includes: *time of the call*, *nearest intersection gps of the call*, *who initiated the call, category of the 911 call,* and *priority of the 911 call.*

The neighborhood bounding boxes includes gps coordinates that draw out the boundary of each neighborhood in Detroit City. There are 208 neighborhoods in the city.

3.2 Processing

For the features that are important for this work (see section above), there are no missing values that needed to be dealt with. The data processing process involves 4 main steps.

3.2.1 Filter

The call records were filtered such that only calls which are not related to public safety / disturbance are excluded. These, for examples, are: officer shift change information calls, traffic stops, officer remarks, bus boarding, vehicles towing, animal bite, etc…

In addition, calls that are outside of Detroit city boundary are also discarded.

3.2.2 Label 911 calls

Each 911 calls are labeled with the appropriate neighborhood FID and name from which the calls came from. This was done using neighborhood bounding box data.

3.2.3 Grouping

The calls record was grouped by each *neighborhood* and by *3-hour intervals*. This is to have the data equally spaced out in time. Within this grouping, the number of 911 calls was total and the calls priority was averaged.

3.2.4 Feature engineering

The target of this work is the police attention level needed by each neighborhoods. In this case, that is defined as:

$$Attention\ level = \left(\frac{\#\ of\ calls}{avg.\ \ priority + 1} \right) + 1$$

Note that adding one on the denominator is to avoid division by zero. The value is also offset by 1 to avoid working with 0 which can cause mathematical problem for some loss-metric calculation later on.

4 Modeling

Several different models were used for comparison. These include: naive forecasting, ARIMA, Random Forest Regressor, Dense Neural Network, and Recurrent Neural Network. For neural network models, both univariate and multivariate approaches were used. The models were used to make 8 forecast points (3-hour intervals) which equals to 24 hours

forecast. Mean absolute percentage error was used to judge each model forecasting performance.

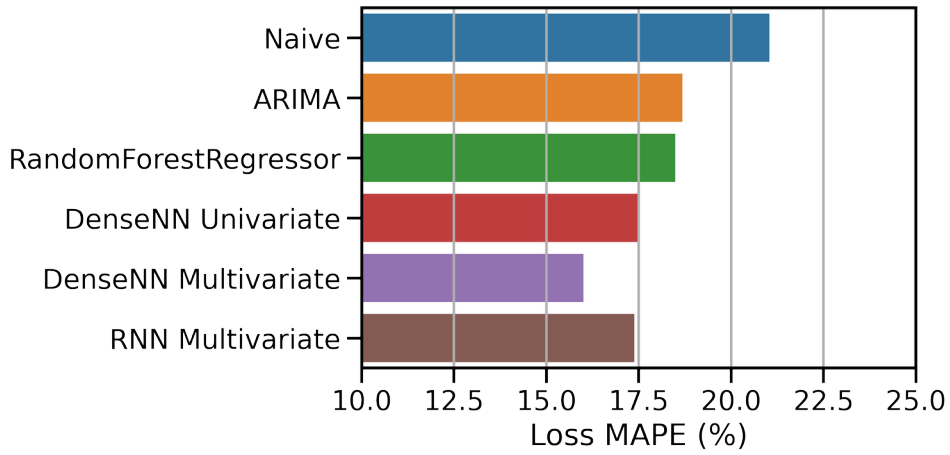
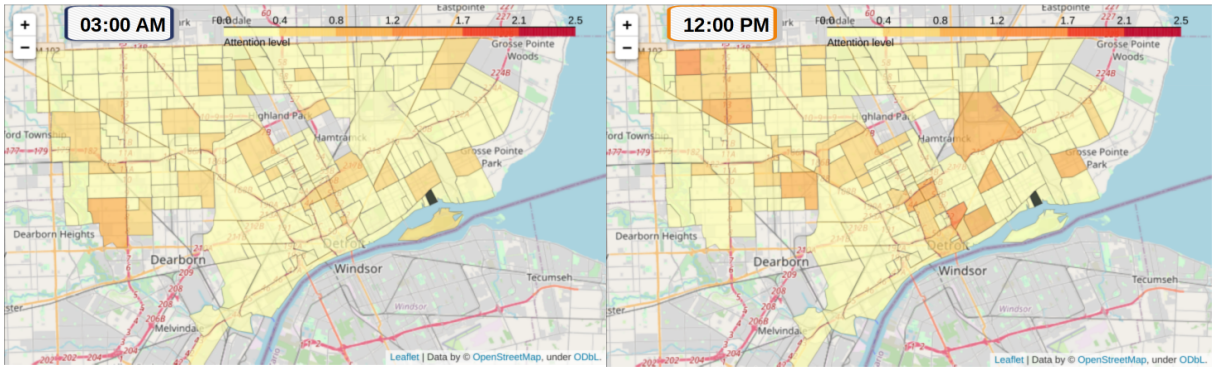


Figure 1. Loss metric MAPE for different models used.

From Figure 1, the dense neural network with multivariate approach yielded the best result with the lowest amount of forecasting error of 16%.

In addition, there appears to be a performance improve when all neighborhoods data were fitted in a multivariate fashion into one model, as shown between the Dense Univariate and Dense Multivariate results. This indicates that there is possible influence between neighborhoods.

5 Results



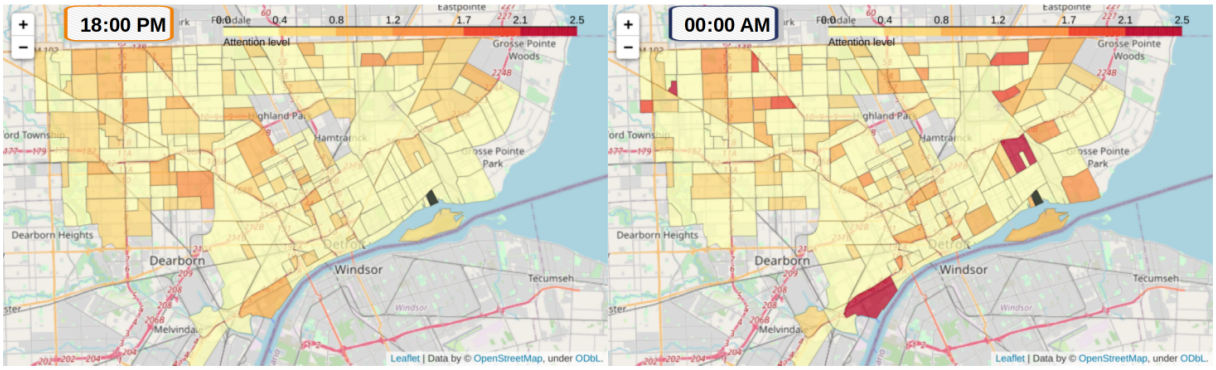


Figure 2. Heatmap of police attention level needed for neighborhoods in Detroit. Data is forecast for Oct 6th 2020.

Using the best model, the police attention level needed by each neighborhoods can be forecast. For example, forecast attention level needed for Oct 6th, 2020 is shown in Figure 2. Notice that the needed attention level is dynamic throughout the day.

6 Outlook

Since the number of available police officers is limited, effective assignment of officers for each area in the city is crucial for the public safety. One way their effectiveness can be improved is by predictive patrolling. Police officers can use the forecast results to plan ahead their patrol route such that more officers are assigned to neighborhoods that require more attention at that particular period of the day. The officers can then reassigned to other neighborhoods where their level of attention needed are expected to increase in the next hour period.

In addition, this work can be further extended to include human mobility tracking within the city to help improve the predictive power. My initial approach was to use geo-tagged tweets to track human mobility. However, our preliminary results show that there is not enough geo-tagged tweets generated. It is due to the fact that most Twitter users don't enable the geo-tag feature in the account setting. Other mobility tracking options such as Four Square or Google Map remain available and were not investigated.