Optimizing resource allocation for government services in New York City

Devashish Khulbe
NetID: dk3596
Center for Urban Science + Progress
New York University
Email: dk3596@nyu.edu

Innovation review:

The paper presents an approach to optimize resource allocation for various government agencies by building a predictive approach with ARMA model [1] based on looking at the autocorrelation values of number government services used for every zip code in New York City.

Innovation proposal:

Introduction and Idea

One of the key parameters by which a city can be judged is how it performs in providing various kinds of services to its citizens. Comparing around various cities in the world, New York City has a fairly good mechanism of providing governmental services by 311 calls [2]. The data of these call, thus, creates a huge potential for understanding and improving how the city agencies function and respond to different types of service requests. One way the city agencies can improve resource allocation for services could be if they create a tool to know where are the areas that with the highest number of services currently and in the near future. This can be thought of as a way to predict for an area the number of services that are going to come up for a particular agency. Knowing that, the agency can optimize its resource allocation to that area to decrease the response time for the services provided.

Amazon's cache algorithms

The idea on which this work is based is inspired from Amazon.com's way of optimally distributing orders for minimum delivery time. For a neighborhood, Amazon predicts the kind of orders that would be coming up and ships those items to the local distribution center based on the past 30 days of orders for that area. As an analogy to computer memory, the local distribution center behaves as a cache memory for fast retrieval of orders for that particular area.

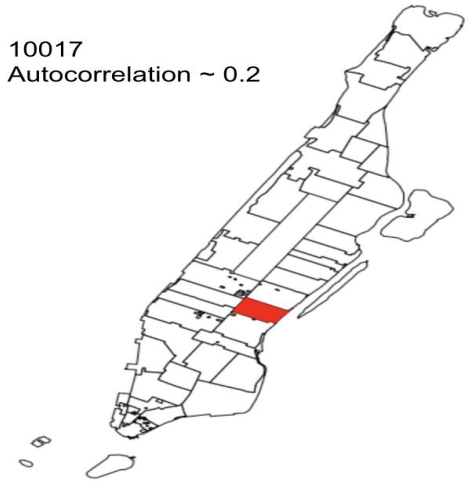
This paper proposes a similar solution for optimally allocating resources for the services used by citizens in New York City (NYC).

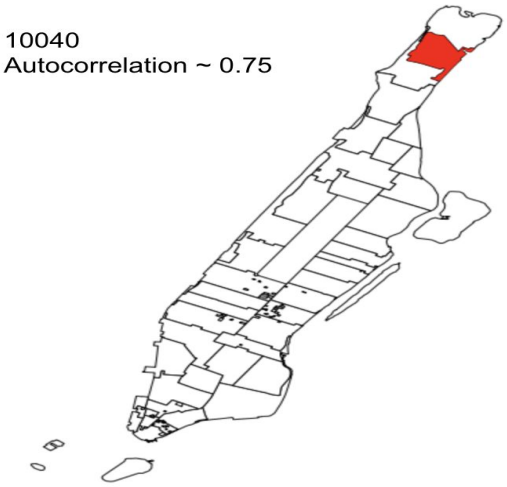
Innovation roadmap:

How could this work?

One way to think about implementing this technique is to answer the question "If we look at the recently used services by people for an area, are those services likely to be used again?" To tackle this question, we need to find the areas, if any, where higher number of complaints/services requested are followed number of the same. This would mean the concerned agency handling those requests may need to focus more on that area. Conversely, we also need to find the areas where low numbers of services used are followed by low numbers of the same.

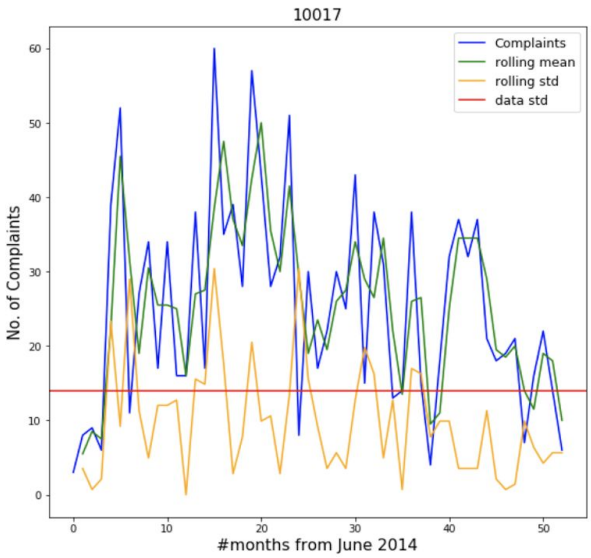
Essentially, looking at autocorrelation [3] of the time series of services of each area would give an idea about the nature of that area.

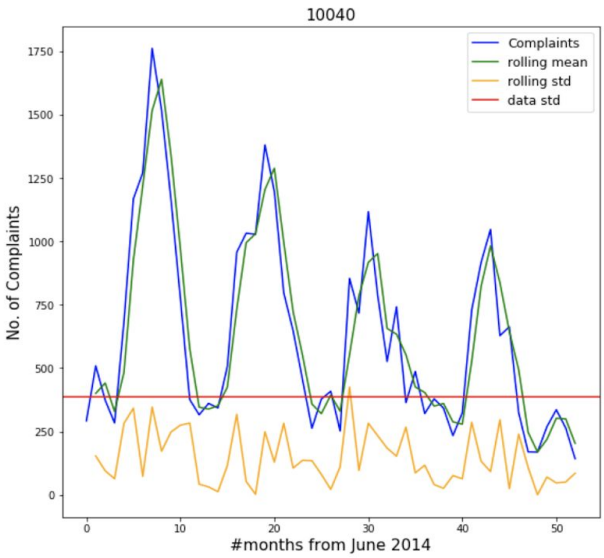




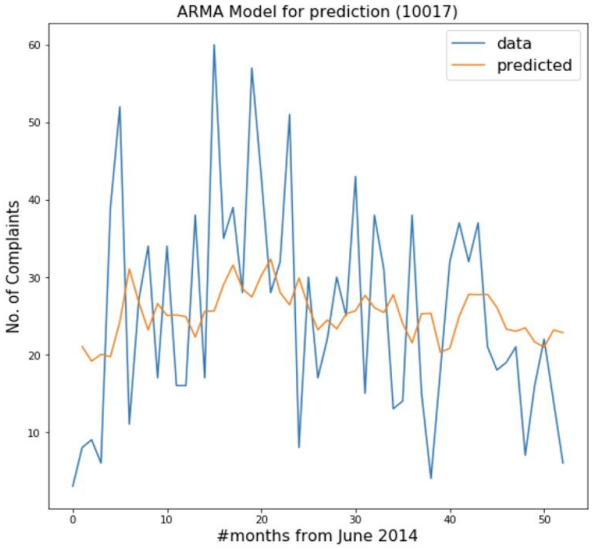
Comparing the neighborhoods and analyzing results

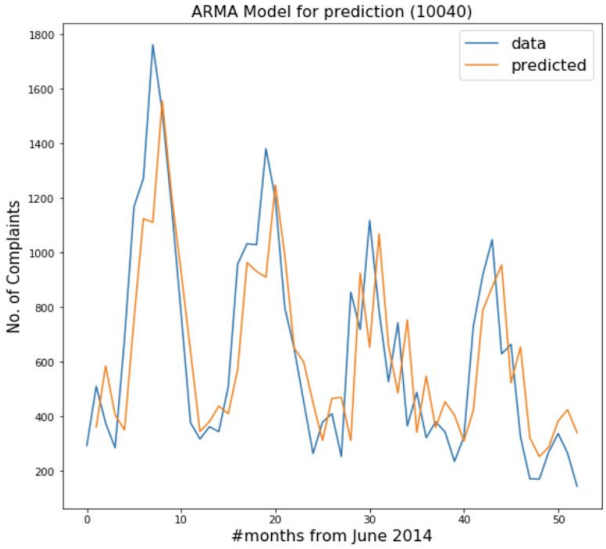
Looking at the autocorrelation of number of services used for Housing Preservation and Development (HPD) of NYC from June 2014 for the Manhattan borough on a monthly basis, we find that almost all areas have positive autocorrelation values. High positive autocorrelation values would mean that HPD experiences higher complaints/services followed by months of higher services. To build and interpret a model for prediction purposes, we looked at the statistics which differentiate the neighborhoods with different values of autocorrelation.





From the analyses, it is clear that for a region with high autocorrelation value (~0.75), the rolling mean [4] is much more aligned with the time series as compared to the other region (~0.2). Also, another important statistic that could be help for predictive purposes is the rolling standard deviation (rSTD). For the zip 10040, the rSTD value crosses the standard deviation of the time series very rarely as compared to the region with low autocorrelation (10017). This would imply that if we make a prediction of the number of services for a month based on looking at the previous month, we would know beforehand that the difference between the two would not be greater than the standard deviation of the time series for the high autocorrelation region. This cannot be said for low autocorrelation region as the time series for it is much more volatile.





The ARMA model gives a much better prediction for high autocorrelation region where the predicted model aligns almost perfectly with the data. This confirms our belief that the same ARMA model cannot be used for the regions with low autocorrelation regions.

Innovation Constituents:

Agencies need to involved

For this tool to be implemented, the agencies that handle the requests and complaints of citizens of NYC would need to involved to achieve the goal of optimization of allocation of their resources. Since we looked at only the 311 data for building this proposal, the city agencies that address the complaints from 311 calls would benefit from this implementation. The main concerned organizations are New York Police Department (NYPD), Housing Preservation and Department (HPD), Department of Environment Protection (DEP) and the Department of Sanitation (DSNY).

Innovation Impact

One important factor where this implementation could help is in reducing the response time for the services requested and complaints resolved for the citizens. HPD, for example, has an average response time of 2.3 days for 311 calls. While it is hard to analyze what would be the response time if this tool is implemented, we believe that the response time should reduce by 20-30% based on the data of the previous policy changes by the agencies. Apart from the reduction in response time, the model would also help to better understand the demographics and the nature of complaints. Also, one key thing worth noticing is that the most regions with high number of complaints also have high autocorrelation. This might be worth looking into further and could be a possibility for future work in this area.

Conclusion

The ARMA model proposed in the work for prediction purposes performs well for high autocorrelation areas. However, as we saw it cannot be applied to the areas with autocorrelation values < 0.6. Another limitation of the model could be its simplicity. The complaints can have a relationship with a variety of other variables and just looking at the previous month's number

could be very simplistic. That said, we believe the model would perform well for the relevant areas and would benefit the agencies in allocating resources better.

References

[1] https://en.wikipedia.org/wiki/Autoregressive%E2%80%93moving-average_model

[2]https://nycopendata.socrata.com/Social-Services/311-Service-Requests-from-2010-to-Present/erm2-nwe9

[3] https://en.wikipedia.org/wiki/Autocorrelation

[4] https://en.wikipedia.org/wiki/Moving_average

[5] Hydrologic Variability of the Cosumnes River Floodplain(Booth et al., San Francisco Estuary and Watershed Science, Volume 4, Issue 2, 2006)