Profiling and Prediction of Non-Emergency Calls in New York City

Yilong Zha, Manuela Veloso*
* On leave from Carnegie Mellon University
CUSP (Center for Urban Science and Progress)
New York University
1 MetroTech Center, 19th Floor
Brooklyn, NY 11201

Abstract

Non-emergency calls, namely the 311 calls, capture different complaints of city residents and visitors about a variety of experienced problems in a city. The 311 calls in New York City (NYC) are publicly available and can provide an interesting status of the city. In this paper, we share a summary of an extensive analysis that we are performing in the 311 data of NYC, as well as a data-based prediction of the number of 311 calls. We present information about the 311 data files and content along multiple dimensions, and then proceed to present prediction results, in which we show that several semantic features affect the different types of complaints differently.

Introduction

Several cities, New York City in particular for this paper, have a 311 24-hour hot line and online service, which allows anyone, residents and tourists, to report a non-emergency problem. Reported 311 problems are passed along to government services, who address and solve the problem. The records of 311 calls are publicly open and updated daily.

Analysis of 311 calls can clearly be of great use for a wide variety of purposes, ranging from a rich understanding of the status of a city to the effectiveness of the government services in addressing such calls. Ideally, the analysis can also support a prediction of future 311 calls, which would enable the assignment of service resources by the city government.

We have been extensively analyzing 311 calls in NYC. In this paper, we profile the data set and highlight a few interesting facts. We provide statistics along complaint types, geolocation, and temporal patterns and show the diversity of the big 311 data along those dimensions. We then discuss the prediction problem of number of calls, where we experiment with different sets of semantic features. We show that the prediction error for different complaint types can significantly vary if some features are not considered.

We believe that the 311 data offer a compelling source to understand how cities work, what is influencing them, and the underlying relationships between different environments and temporal information to the reported incidents.

Profiling 311 Request Data

The 311 service request data is publicly available from the NYC Open Data Portal. It has been updated daily since year 2010. In our study, we use data from four complete years in the period of Jan 1, 2010 to Dec 31, 2013. These data come in a table with 52 columns and 6,588,519 rows, where each row is a record of the request and each column refers to one descriptor of the record.

Basically, the 52 descriptors of the 311 data can be classified into five main categories:

Time:
Descriptors of important time points of requests. There are 4 of them: *Created Date*, *Closed Date*, *Due Date*, and *Resolution Action Updated Date*.

Location:
Descriptors related to the geo-location of the requests. There are 31 descriptors in the category. Many of the descriptors provide redundant information about locations of requests, because they are designed only for some certain types of requests. Examples are *Incident Zip*,*Incident Address*, *X Coordinate (State Plane)*, and *Y Coordinate (State Plane)*.

Type:
Semantic descriptors of the requests. There are 2 members in this category, *Complaint Type* and *Descriptor*. The *Complaint Type* contains categories of the requests according to its content, while the information in *Descriptor* are more detailed subcategories inside *Complaint Type*s.

Agency:
Descriptors indicating which agency handled the request. There are 2 of them: *Agency* and *Agency Names*. The *Agency* is basically the abbreviation of the *Agency Name*.

Other:
13 other varied descriptors including *Unique Key*, *Status*,*Facility Type*, *Garage Lot Name*. Like the **Location**, many of the descriptors are designed to support only a few types of requests.

In this work, we will focus on selected descriptors that are applicable to all types of requests. These descriptors are *Created Date*, *Closed Date*, *Complaint Type*, *Agency*, and *Incident Zip*.

The complaint types of a 311 request are manually-chosen semantic tags that describe the basic purpose of each request. They are also closely related to the procedure for handling the requests. In our data set, there are 230 different complaint types. Figure 1 illustrates the distribution of volumes of calls in different complaint types as a pie plot. The type with the largest number of calls is *HEATING* that accounts for about 11.99% of the total request volumes. Note that, 204 types have fewer than 1% shares of requests volumes, but the total share of those are 17.99%, which is larger than the largest complaint type *HEATING*.

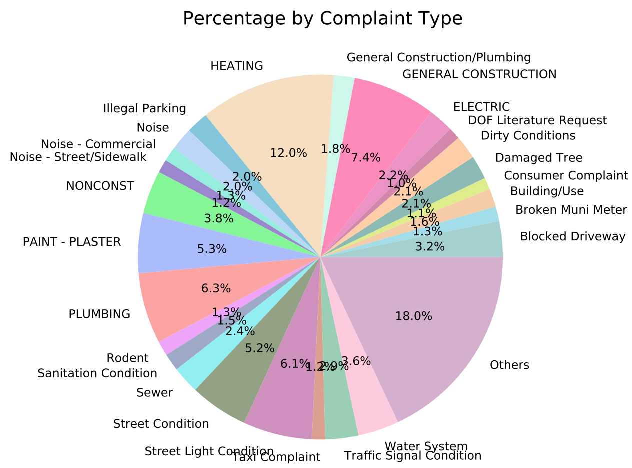

Figure 1: The distribution of volumes of calls in selected types of complaints. Note that only the 26 types that have request volumes larger than 1% of the total volume are displayed, the sum of the rest types is displayed as *Others*.

Figure 2 gives the distribution of requests by *Agency*. This distribution is greatly polarized. *HPD* (Department of Housing Preservation & Development) alone handles 37.50% of all the requests, which is the largest among all the 59 agencies. In comparison with that, many agencies only process a tiny share of requests, the smallest 48 of them account for only 1.07% of the total request volume altogether, which is represented as the *Others* in Figure 2.

Temporal Distribution

On average there are about $4,500$ new requests per day. However, the request volumes are periodically fluctuating over time. There are patterns in different scales, the daily cycle and weekly cycle being the most obvious ones.

Figure 3 compares how request volumes are distributed over the days of the week in different complaint types. It is observed that most of the types of complaints have their own distribution over a week. While a majority of the them show a decrease in complaints during weekends, the volumes of requests related to noise increases during weekends.

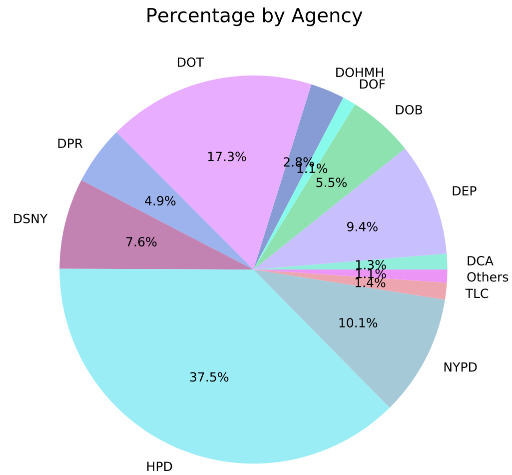

Figure 2: The distribution of volumes of calls in selected agencies. There are 59 different agencies, only 11 agencies which have request volumes larger than 1% of the total volume are displayed, the sum of the rest are displayed as *Others*.

In our prediction later, the weekly cycle of request volumes will be applied. And a noticeable improvement in predicting is observed before and after considering the difference in weekly cycle in given complaint types.

Another interesting aspect of temporal patterns is related to the processing times (denoted as dt) of requests. We use $Closed\ Date - Created\ Date$ to calculate dt. However, in the data set, about 15.89% of the records don't have a *Closed Date* (These requests may either still be under processing or may not require a *Closed Date*).

The dt we calculate will be based on those request records that have both the *Created Date* and the *Closed Date*. Since in 88.39% of the request records, the *Created Date*s describes only the date rather than the time of the day, the dt will be measured by the number of days. For example, if the request is closed at the same day it is created, $dt = 1$, if this happens the second day, $dt = 2$, etc..

According to this definition, the distribution of dt is shown in Figure 4 in a log-log plot. If we ignore the NYPD, the distributions can be approximated by a power-law distribution with a slope close to -2. This kind of distribution is also reported in many other human activities (Oliveira and Barabási 2005; Clauset, Shalizi, and Newman 2009). This also indicates that there may be different dynamics in the processing of emergency and non-emergency services.

Geo-correlations to Population

There are lots of descriptors in the category **Location**. We can locate about 91.25% of requests with latitude and longitude from the data set. About 91.77% of the requests have recorded zip codes.

Population in an area is one predictive feature to the

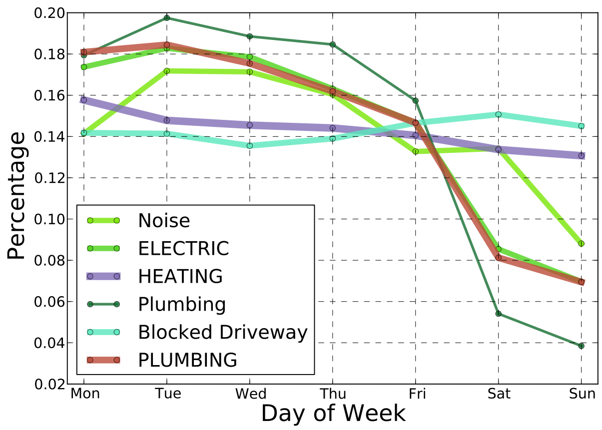

Figure 3: The request volumes of different complaint types distributed over days of the week. The thickness of the lines reflects the total number of requests in the associated complaint type.

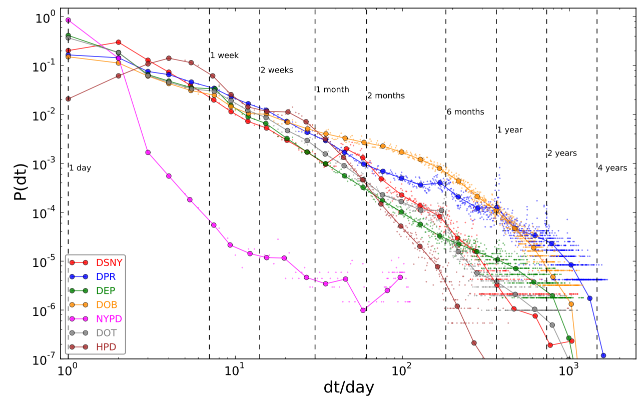

Figure 4: The distribution of request processing time in 7 agencies with largest number of requests. The plot is in a log-log scale. Each color is related to an agency. The dots represents the percentage of given dt, while the line-circles are the binned percentages.

number of 311 requests in that area. In order to see how strong the correlation is between population and different types of event, we calculate the geo-correlation coefficient according to Equation 1:

$$C = \frac{\sum_{z \in Z}(N_z - \bar{N})(P_z - \bar{P})}{\sqrt{\sum_{z \in Z}(N_z - \bar{N})^2 \sum_{z \in Z}(P_z - \bar{P})^2}}, \quad (1)$$

where Z is the set of all zip codes in NYC, N_z is the number of requests at zip code z, $\bar{N}$ is the mean of N_z, P_z is the population at zip code z and $\bar{P}$ is the mean of P_z.

The result of the geo-correlation between the population and different types of requests is given in Figure 5. According to this figure, requests in most of the complaint types have a positive geo-correlation with the population, the *Dirty Conditions* has the highest one, while *taxi complaint* is least correlated, with a negative value of about -0.04.

Prediction

As with most data problems, we would like to use existing data to make predictions about the unknown data. We need to decide on the features of the data to use. Before digging into which features would be useful, we first identify what target to be predicted. In this 311 data set, there are many different types of targets that can be predicted. One could predict the number of calls at different levels of granularity, namely in terms of different time intervals (e.g., weekly, daily, or hourly volumes), or in terms of geo-features (e.g., Boroughs, zip codes, or census tract), or in terms of volume for specific complaint types or for specific agencies. We continue to work on predictions along these different aspects, and here we focus on presenting the prediction of daily number of calls for the top most frequent 50 complaint types. We explore the correlations of the predictions with a complete and ablated input feature sets.

Feature Preparation

To prepare for the features for prediction, external data sources will be used along with the 311 data set. They are U.S. calendar and NYC historical weather data from National Weather Service Forecast Office.

These two data sources have long been used in modeling and predicting the public services such as urban forest recreation (Dwyer 1988) and daily visits to a walk-in clinic (Holleman, Bowling, and Gathy 1996). In these works, linear regression models are used to evaluate the feature importance and predict the future event volumes.

From our data sources, 12 features are created, including *day of the week* (*DOW*), *public holiday* (*Holiday*), *mean temperature*, *temperature range*, *snow* and the request volumes of the last 7 days (*Last Week*). These features will be used to predict the request volume of each day. Meanwhile, we can calculate the real daily request volumes since the date of each request is given in the *Created Date*.

By associating a date with the calendar, we determine what *day of the week* and *public holiday* are related to the date. Both these two temporal characteristics are used as categorical features, where *day of week* have 7 levels and the *public holiday* has 13 levels (12 for holidays and 1 for normal days).

The weather information is recorded daily. We use the *average temperature*, *temperature range* and *snow* of each given day as our input features. Both *average temperature* and *temperature range* are numerical values, while *snow* is categorical with two levels indicating whether there is snow or not in given day.

At last, the request volumes in previous 7 days are also used, which will be the most predictive feature according to our experiments.

Combining these features, we generate the inputs with 12 columns, which will be used for predicting the 311 request volume of a given day.

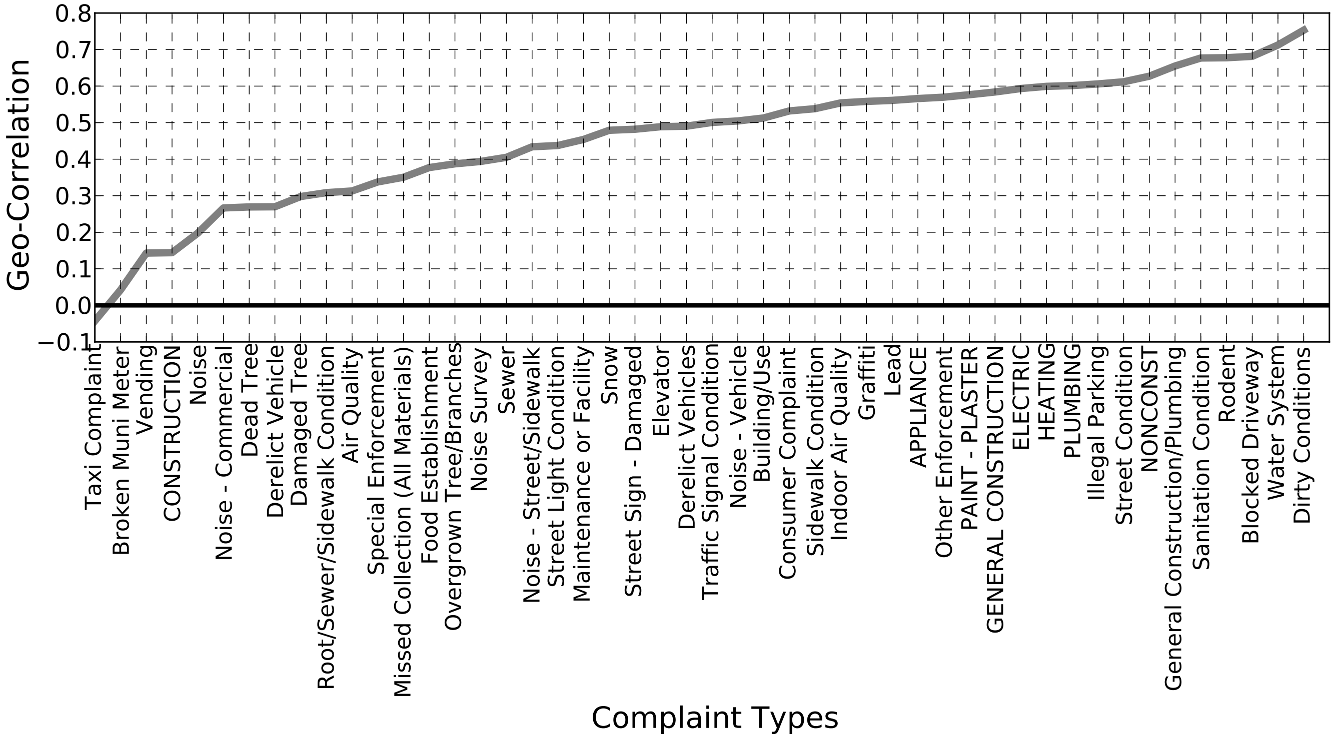

Figure 5: Geo-correlation between features. The y-axis is the geo-correlation between the distribution of request volumes of a complaint type and that of the population in zip codes. The x-axis is the complaint type ordered by the value of the geo-correlation in an increasing manner.

Type Separation

At this point, our prediction could be further improved by taking complaint types into consideration. The idea is inspired by different temporal patterns observed in different complaint types in the previous section.

Since each daily request volume is the sum of individual request volumes in all different complaint types, it will be easier to learn to predict request volumes in each types and sum all predicted volumes as the total daily volume, than to learn to predict the total volumes in one step. However, the latter approach will gradually learn the similar relationship when we have more years of data for training and the patterns in each complaint types are unchanged in different years.

In practice, the types of complaints are ranked against their request volumes, and the top 50 types with largest request volumes are chosen. Then, the volumes of these types on given days are used as labels, meanwhile the complaint types are extended to the original features as a categorical feature with 50 levels. In this way, the Random Forest will predict 50 volumes in each day, and the total volumes of that day can be predicted as the sum of the 50 outcomes.

Experiment

A Random Forest (Breiman 2001) is trained with data of the first 3 years and used to predict the daily request volumes of the fourth year.

This method is an ensemble of classification or regression trees created by using bootstrap samples of the training data and random feature selection in tree induction. Prediction is made by taking the mode or the average value of the votes from each trees. It is widely used in different fields, like chemical informatics (Svetnik et al. 2003) and bioinformatics (Díaz-Uriarte and De Andres 2006; Saeys, Inza, and Larrañaga 2007). This is mainly due to the measurement of relative importance of features provided by this method (Archer and Kimes 2008).

However, it is pointed out that there is bias in the traditional Random Forest variable importance measures (Strobl et al. 2007). An alternative method, namely the permutation accuracy importance measuring the difference in the prediction accuracy before and after randomly permuting the variables can be used as a more reliable alternative measure (Strobl et al. 2008). In this work, a similar approach will be applied to measure the feature importance in our prediction.

Our experiment is conducted using a Random Forest with 500 trees. The results show that by considering the complaint types, the mean square error (MSE) of the prediction will reduce approximately 9% percent, from 351331.90 to 326372.63. The result of the best prediction is given in Figure 6.

Table 1: MSE of Type Prediction By Features

Feature Name	Complete	Holiday	Weather	DoW	Last Week
Total	326372.63	**368826.91**	355800.49	**365654.76**	**597134.07**
HEATING	114757.76	112703.67	**138414.12**	109863.79	**137031.83**
GENERAL CONSTRUCTION	3833.5	**4937.24**	3947.77	**4275.48**	**5879.91**
PLUMBING	3030.07	**3457.41**	3115.16	**3608.06**	**7399.4**
Street Light Condition	6270.38	**8572.13**	6262.65	**6942.06**	**8624.92**
PAINT - PLASTER	2742.89	**3375.4**	2734.93	3001.64	**4902.6**
Street Condition	4862.66	4785.72	4589.44	4801.77	**20490.69**
NONCONST	966.56	**1085.88**	1021.18	999.99	**2505.34**
Water System	4345.57	4329.23	**6484.08**	4616.33	**13448.81**
Blocked Driveway	551.27	553.14	528.56	558.12	**1344.29**
Traffic Signal Condition	2083.9	2063.55	2137.51	**2309.35**	**2862.53**
Sewer	4410.98	4344.71	4700.96	4545.11	**6840.85**
ELECTRIC	567.2	**640.22**	608.48	**641.65**	618.26
Dirty Conditions	2892.08	2762.26	2813.96	2901.16	*2376.7*
Damaged Tree	1957.28	*1622.43*	*1075.1*	*1065.17*	**25382.71**
Noise	964.27	955.0	952.39	**1072.79**	**1430.26**
Illegal Parking	450.06	458.16	449.75	450.82	**2498.43**
General Construction/Plumbing	441.89	484.27	449.11	**497.05**	**1966.57**
Building/Use	525.32	549.18	531.69	**602.79**	**1063.91**
Sanitation Condition	259.37	269.29	281.25	**292.89**	257.9
Rodent	202.84	220.62	212.93	**230.95**	**234.38**
Noise - Commercial	499.44	511.0	480.76	**554.4**	**881.95**
Broken Muni Meter	5294.09	5288.12	5722.42	5210.1	**6575.39**
Noise - Street/Sidewalk	1398.1	1429.81	1499.79	1316.71	1360.53
Taxi Complaint	177.7	182.73	190.63	169.66	**270.59**
Consumer Complaint	103.6	**118.34**	98.86	**121.02**	**128.66**
DOF Literature Request	3920.98	4217.27	3898.16	**4697.41**	**20457.28**
Missed Collection (All Materials)	371.36	378.08	385.74	**430.23**	**461.41**
Graffiti	710.76	716.39	699.16	762.1	**1664.42**
Overgrown Tree/Branches	154.43	**170.96**	163.39	**184.34**	**386.71**
Noise - Vehicle	183.22	180.0	199.78	195.25	*164.17*
Derelict Vehicle	66.46	69.98	66.78	**77.31**	**94.17**
Derelict Vehicles	103.73	**115.38**	105.61	**127.43**	**131.51**
APPLIANCE	39.14	40.84	37.0	**44.31**	**461.47**
Root/Sewer/Sidewalk Condition	65.1	**74.89**	63.4	**88.1**	**77.3**
Dead Tree	62.74	**69.26**	63.8	**75.2**	**120.2**
Maintenance or Facility	116.41	115.11	123.22	118.34	**140.83**
Elevator	88.16	95.07	**106.95**	**98.38**	**102.48**
Sidewalk Condition	28.22	30.09	25.51	**35.33**	**216.18**
Snow	926.88	910.77	951.77	977.03	**9421.77**
Street Sign - Damaged	154.85	152.71	153.45	165.58	**320.84**
Air Quality	47.68	46.45	46.2	51.09	**57.64**
Food Establishment	34.11	35.08	33.78	**40.4**	37.26
Special Enforcement	732.71	718.75	671.73	702.41	*329.31*
Lead	12.14	11.84	**13.89**	12.66	**1445.55**
CONSTRUCTION	4.95	5.04	5.09	5.38	**5.57**
Other Enforcement	44.32	44.68	45.9	47.04	**50.04**
Indoor Air Quality	19.09	19.8	18.87	**23.19**	**24.16**
DCA / DOH New License Application Request	34.47	34.65	*19.74*	**48.19**	**58.31**
Noise Survey	0.1	*0.08*	*0.0*	*0.08*	**453.0**
Vending	22.13	21.87	**24.55**	22.85	**31.78**

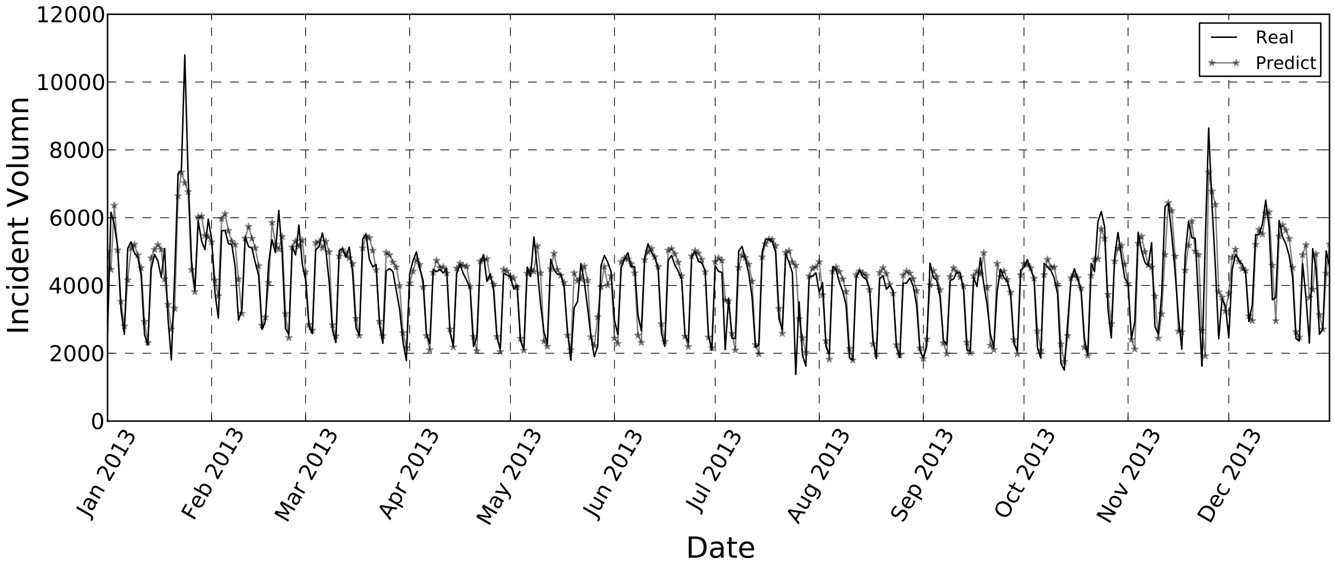

Figure 6: Comparison between the prediction of the daily 311 request volumes through the whole year 2014 and the reality. The black real curve shows the request volumes in the real data, and the grey curve with stars represents the one from the prediction.

Feature Evaluation

In order to measure the feature importance, several predictions are conducted with the absence of different features. The accuracy of these predictions is measured using the MSE. Then, the relative importance of the features can be evaluated by comparing the MSEs of the predictions with the absence of features and the prediction with complete features.

Table 1 shows the MSEs of predictions in both the total request volumes and that in each complaint types. The column *Complete* refers to the prediction with a complete feature set, and the rest of the columns refers to the prediction with the absence of certain features. The names of these columns indicate the features that are absent. For easier comparison, any of the MSEs that are 10% higher than that of the predictions using the complete features are bonded.

According to this table, the most important feature is undoubtedly the request volumes in the past 7-days (column *Last Week*). One reason may be that the 311 requests have an obvious 7 day cycle. In other words, one could achieve a relatively good prediction by simply using the request volume of 7 days ago as the volume tomorrow.

The public holiday is the second most important feature. It can be observed in the 311 data that there is a significant decrease of requests during holidays. Also, the holidays are relatively independent of the weather, day of the week, and the request volumes of previous days.

The feature *Day of Week* (column *DoW*) is less important in this comparison. This is possibly because the information contained in *Last Week* is highly correlated to that of *DoW*.

Finally, the weather is the least important feature according to Table 1. One reason is that the request volumes in the last 7 days will be able to predict those of the next week if the weather doesn't change much.

However, the weather matters in some of the complaint types, including the *HEATING* and the *Water system*. The request volumes of these types are sensitive to slight changes of weather.

Conclusion

We demonstrate that the separate-and-combine approach actually extracts more effective features when there are variant patterns in subsets of the data. This approach is based on the observation and understanding of the individual patterns in subsets of data. It would work best when there is big variation in patterns over each subsets and the subsets are independent of each other.

For the 311 data set, the predictive model we proposed is a prototype. The model can be extended to predict other useful quantities, such as the number of requests by zip code. Another implication of the model could be a better understanding to the city, such as to what extent events happen over and over again in the city and how to detect uncommon events automatically.

References

Archer, K. J., and Kimes, R. V. 2008. Empirical characterization of random forest variable importance measures. *Computational Statistics & Data Analysis* 52(4):2249–2260.

Breiman, L. 2001. Random forests. *Machine learning* 45(1):5–32.

Clauset, A.; Shalizi, C. R.; and Newman, M. E. 2009. Power-law distributions in empirical data. *SIAM review* 51(4):661–703.

Díaz-Uriarte, R., and De Andres, S. A. 2006. Gene selection and classification of microarray data using random forest. *BMC bioinformatics* 7(1):3.

Dwyer, J. F. 1988. Predicting daily use of urban forest recreation sites. *Landscape and Urban Planning* 15(1):127–138.

Holleman, D. R.; Bowling, R. L.; and Gathy, C. 1996. Predicting daily visits to a waik-in clinic and emergency department using calendar and weather data. *Journal of General Internal Medicine* 11(4):237–239.

Oliveira, J. G., and Barabási, A.-L. 2005. Human dynamics: Darwin and einstein correspondence patterns. *Nature* 437(7063):1251–1251.

Saeys, Y.; Inza, I.; and Larrañaga, P. 2007. A review of feature selection techniques in bioinformatics. *bioinformatics* 23(19):2507–2517.

Strobl, C.; Boulesteix, A.-L.; Zeileis, A.; and Hothorn, T. 2007. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC bioinformatics* 8(1):25.

Strobl, C.; Boulesteix, A.-L.; Kneib, T.; Augustin, T.; and Zeileis, A. 2008. Conditional variable importance for random forests. *BMC bioinformatics* 9(1):307.

Svetnik, V.; Liaw, A.; Tong, C.; Culberson, J. C.; Sheridan, R. P.; and Feuston, B. P. 2003. Random forest: a classification and regression tool for compound classification and qsar modeling. *Journal of chemical information and computer sciences* 43(6):1947–1958.