A Regression View of Everbridge Mass Notification Service

Weiqi Weng
College of Computer and Information Science
Northeastern University

Introduction

Everbridge Mass Notification Service does critical communication during emergency. Employers may send notifications to their employees, informing danger and evacuation plan, through all kinds of paths such as cell phone, text message and office email. Then they wait for their confirmation. Currently the system sends notifications according to customers' path preference set. However, preference does not necessarily indicate their accessibility during emergency. Employee may set office email as the first option while he or she actually carries his or her cell phone all the day. Employees' safety are given top priority so this project works to figure out what factors contribute to both confirmation and quick response of the notifications and better support the service.

The project deals with three questions. How to predict whether a configured notification will get confirmed, confirmed late or not confirmed? How to predict the time for a notification to be confirmed? What is affecting the notification to be confirmed and confirmed in a shorter period of time? When we figure out the answers, notification can be configured accordingly to first make it more likely to be confirmed and second confirmed shortly.

To begin with, we merge four different data sets, carrying various information about the notification configuration, contacts' system setting and their background, into one containing all the available information of each notification. Then we clean the data set, add dummy indicators and finally split the data set into training, validation and testing data sets for further analysis.

Next, we build a Softmax Regression Model to classify the notifications into three classes: confirmed, confirmed late and not confirmed. Grid Search Cross Validation helps to locate the optimal hyper-parameters. Finally a Ridge Softmax Regression Model gets 74% accuracy on the testing data set. Finally, a Linear Regression Model is established to predict how much time does a notification take to be confirmed or confirmed late. The effect of configuration parameters are discussed in terms of their significance to make the notification more likely to be confirmed.

Pre-analysis

The data set contains notifications sent by 3 different organizations which are identified as 8928**6, 1332** and 8928**2. Number of confirmed notifications varies from organization to organization.

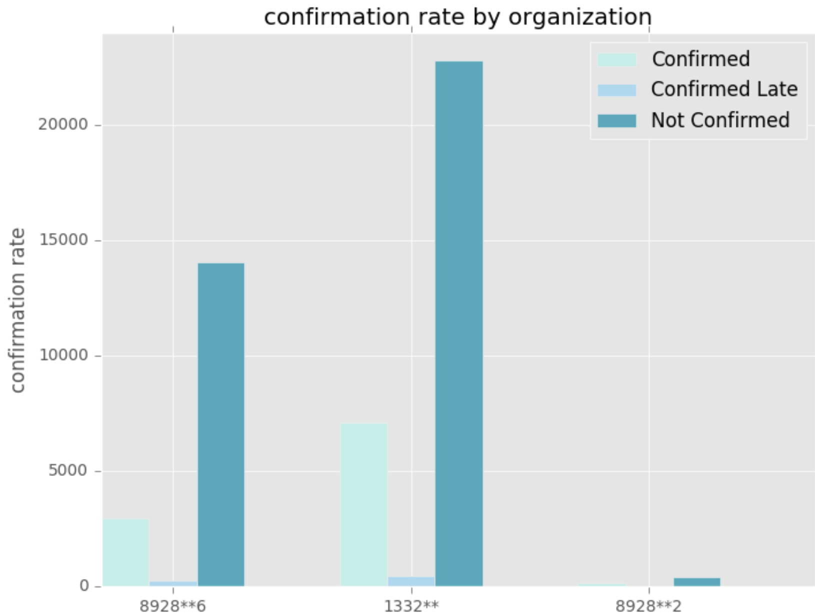

figure 1. bar chart by organization

We see that organization 1332** sent the largest number of notifications while organization 8928**2 sent the least. Besides, organization 1332** has the highest confirmation rate 24.88%. organization 8928**2 gets 24.26% and organization 8928**6 gets 18.47%. Roughly speaking, notifications sent by organization 1332* and 8928*2 are more likely to be confirmed than organization 8928**6.

More specifically, the notification configuration varies from organization to organization. For instance, organization 8928**6 has a strong preference on standard priority notifications. Also it shows little interest in Office Email, SMS, Home Phone, Cell Phone, Work Phone and Personal Cell Text which are the 5 most commonly used paths. 89.5% of its contacts are in Great Britain and United States, which probably means its major business is in these two countries. It implies that employees in GB and US probably prefer registering other means of communications such as Business Desk Phone to the system. Also, market for mass notification service is larger in these two countries for organizations with background similar to organization 8928**6.

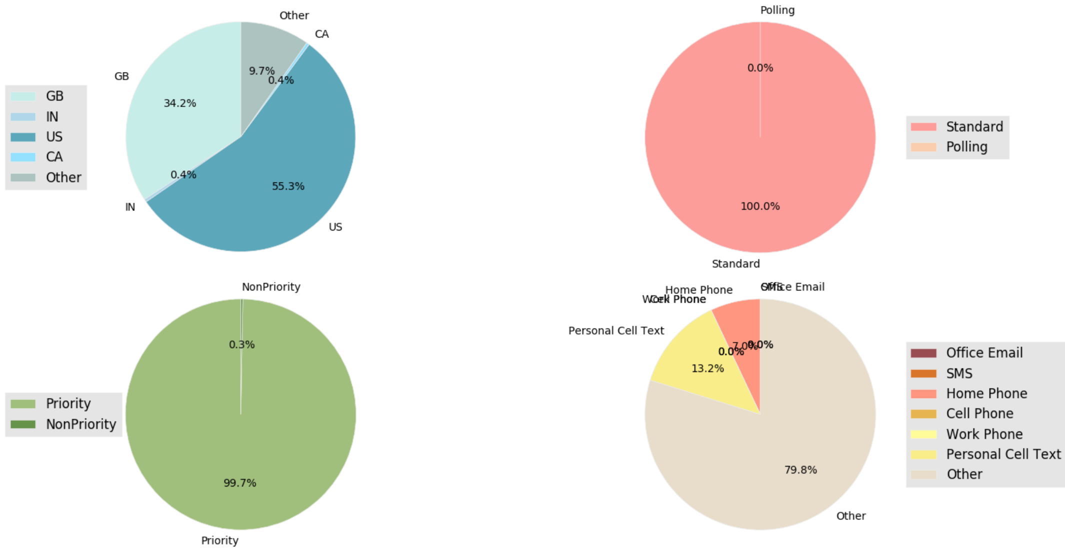

figure 2. configuration preference of organization 8928**6

The notification configuration of organization 1332** tells a different story except the preference on priority notifications. For organization 1332**, polling notification is dominant. As to paths, a variety of choices have been tried by organization 1332**. It shows a little tendency on Office Email but other paths are very close to each other. US and IN covers 76.42% of its contacts.

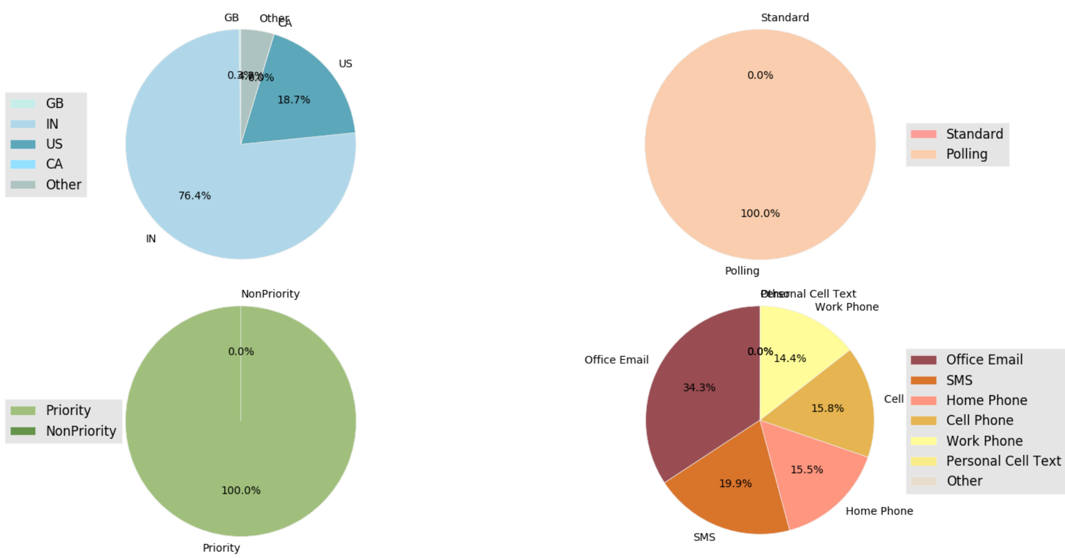

figure 3. configuration preference of organization 1332**

Distinction of organization 8928**2 is that it tries priority and non-priority notifications almost equally. In addition, Personal Cell Text is the major path and most of the Personal Cell Texts are sent with priority. Business of organization 8928**2 involves business in China, Australia and Germany. Employees in these countries are most accessible through Personal Cell Texts while Personal Cell Phone is not considered as a frequently used mean of critical communication.

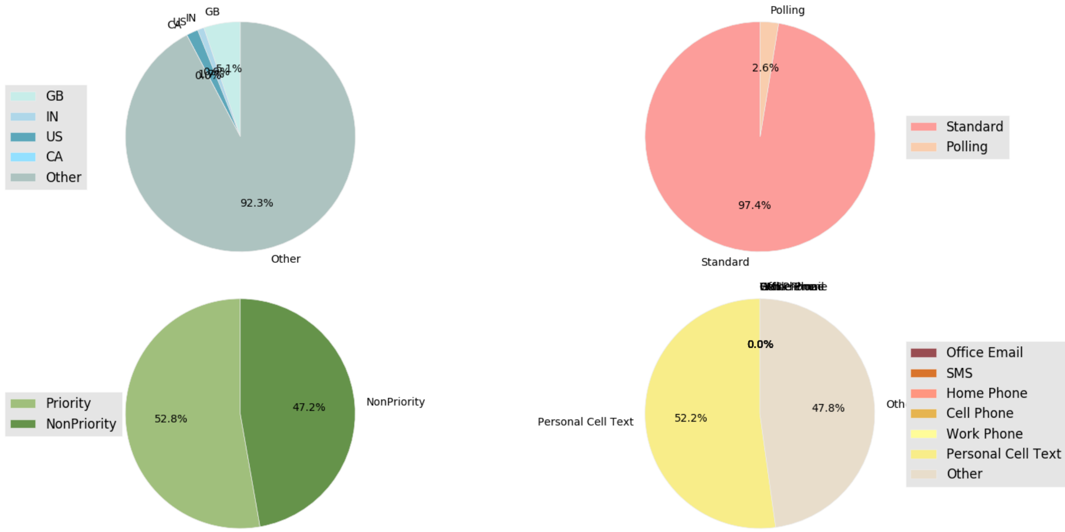

figure 4. configuration preference of organization 8928**2

From another perspective, 59.53% of the confirmed notifications are confirmed on Wednesday (6/15/16, 6/22/16 and 6/29/16), in which organization 1332** makes the most significant contribution. The records for Organization 8928**6 indicate that it issued a mass notification on 6/9/16. 90.23% of them are confirmed on the same day while 9.77% of them are confirmed on 6/10/16. From figure 1 we see that organization 1332** sends the most notifications and they are sent out on 6/23/16 and 6/29/16. The highest confirmation rate falls on Wednesday because Wednesday has the largest notification population.

The correlation between contact path, total path and confirmation of a notification will be discussed in classification model.

Data Preprocessing

Begin with Contact Attempt, we extract notification ID plus contact ID from the column named 'id' and merge with Contact Path Count on keyword contact ID which is then merged with Contact Notification on notification ID plus contact ID again. Finally the data frame is merged with Notification Result on organization ID and notification ID.

Next, duplicated features are removed.

- Created date and first attempt time are removed since they are basically the timestamp of when the event is registered to the system. Also the time can be drawn by call start date within only seconds of delay.
- Confirmed count, not confirmed count, unreachable count and confirmed late count are decomposition of notifications issued by an organization. The purpose of the model is to predict whether a notification will be confirmed. Therefore, only the total amount of notifications being sent is accessible when we are doing prediction.
- Call result is dropped since the result is revealed by attempt state.

Organization ID is remained since they can reflect some property of that organization. For example, maybe Organization 1 is a law enforcement organization while Organization 2 is a ecommerce. Probably people will pay more attention to notifications sent by Organization 1.

Numerical features (path count and total count) are severely skewed due to lack of variation in organization, which brings problem when explaining the variation of notification from different organizations. So we perform a logarithmic transformation and normalize them into z-score. In this project, for numerical features, we refer to $log_e(x)$ as x for the purpose of clean notation.

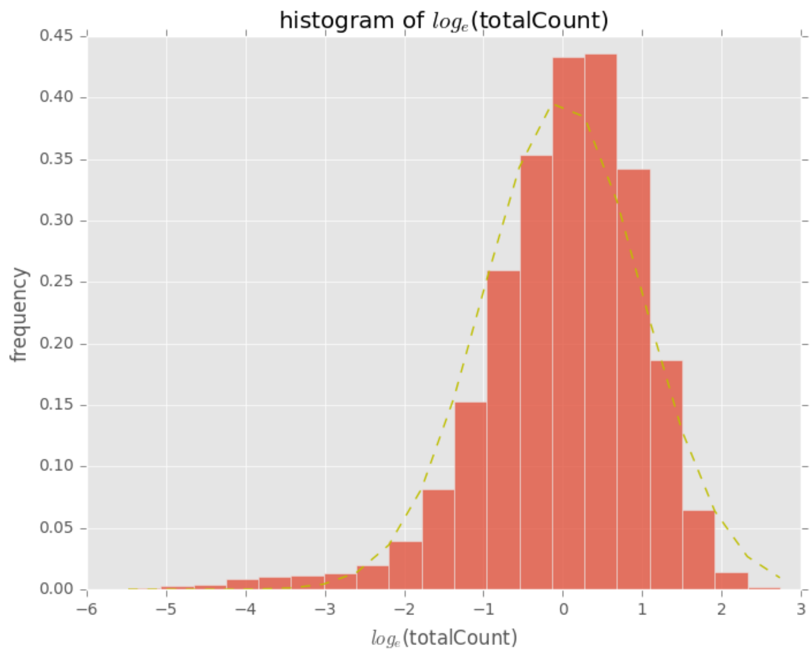

figure 5. histogram of attempt

To move on, samples with state 'Unreachable' are dropped since the corresponding call start time and path prompt are both missing. The call start time of notifications sent through SMS can be retrieved from the call result column. Dummy indicators for nominal features are added.

For each notification sent to some contact, duplicate samples (duplicate in the sense of attempt result) exist in our set. For example, the system is sending notification 2656** to contact 1345**. The process is presented in the following table.

attempt	path	result
1	Work Email	0
2	MS Cell Text	0
3	MS Cell Phone	0
4	MS Office Phone	0
5	Home Phone	0
6	Work Email	0
7	MS Cell Text	0

table 1. notification to contact

We can see that attempt 1 and 6 are duplicate. As long as two attempts through same path lead to the same result, only the last attempt of a certain path really matters. We will see that duplicates should be remained in part 2 to extract the time for the contact to respond but in this stage we remove such duplicates.

In the first stage, we have a feature set of size 59. Two of them are numerical. Others are nominal features indicating type, priority, organization issued the notification, country where the contact is and the path.

In the second part, assuming we have enough reason to believe the notification will be confirmed, we find each confirmed notification (with fixed ID), locate when it is launched and confirmed through call start time and compute the time gap. In this way, we get how much time it takes for a notification to be confirmed. Consequently we can predict the time through a predictive model.

The project is divided into two parts. If treated as a whole, the time taken for a non-confirmed notification to be confirmed will be $+\infty$, which is not appropriate for a prediction model.

Classification

Model

This section deals with the first question: How to predict whether a configured notification will get confirmed, confirmed late or not confirmed?

Based on preprocessing, it is a 3-class classification problem. Various Machine Learning techniques such as Support Vector Machine and Random Forest have been proved to be powerful classifiers. It's true that they can be very accurate. However, for these kind of classifiers, no matter in original space or mapped space, their parameters are hard to interpret, not to mention the connection between the real world and the mapped space. This is the reason for using a Softmax Regression here. The logistic function

$$f(x) = \frac{1}{1 + e^{-w^{\mathrm{T}}x}}$$

where w is the weight vector and x is sample, maps x to (0, 1) which is adding confidence to the model. It measures how likely x is classified as 1. We can also model the significance of each feature by taking the derivative.

Training

A Softmax Regression Model is trained with the following process. Denote λ_R as the Ridge weight decay.

procedure TRAIN-LOGISTIC-REGRESSION(data, range of λ_R)
Input: data frame, range of λ_R
Output: the optimal λ_R

    for all λ_R in range of λ_R **do**:
        for all $1 \leq i \leq 5, i \in \mathbb{N}$ **do**:
            hold out fold-i of data
            train a Ridge logistic regression model on the rest of data with S^*
            record the accuracy
        end for
        compute and record the mean accuracy for λ_R
    end for

    get λ_R^* corresponding to the optimal mean accuracy

    train model M_R^* with λ_R^*
return M_R^*

end procedure

The process returns $\lambda_R^* = 0.4$.

Validation and Test

This subsection runs a k-fold $k = 5, 6, 7, \cdots, 20$ cross validation on the validation set, computes mean accuracy and reports standard deviation with a error bar plot.

figure 6. error bar plot

From figure 6, we can see that the Softmax Regression Model has a validation accuracy of approximately $72.50 \pm 1.20\%$. There is no significant vibration in standard deviation, which means the model is robust and accurate enough. Then we merge training and validation set to train a final model and run it on testing set, which brings a testing accuracy of 73.49%.

Discussion

In this part, feature effect will be quantified and discussed. Now we have logistic function

$$f(x) = \frac{1}{1 + e^{-w^{j\mathrm{T}}x}}$$

for each class j, where w is the feature weight vector. Details are presented in the following table where w_i^j is the weight for feature i of class j. Due to limited space, a subset is presented. The intercept is $w_0^{1:3} = [2.08, 1.10, -3.18]$.

feature type	feature	w_i^1	w_i^2	w_i^0
numerical	Path Count	0.064	0.017	-0.081
numerical	Total Count	-0.039	-0.033	0.071
nominal	Cell Phone	0.167	0.836	-1.004
nominal	SMS	-0.228	-0.453	0.682
nominal	Office Email	-2.270	-1.311	1.538
nominal	Home Phone	0.214	0.877	-1.092
nominal	Work Phone	1.383	-0.636	-0.747
nominal	Personal Cell Text	-1.170	0.501	0.669
nominal	Credit Suisse Email	-0.533	0.607	-0.074
nominal	Personal Cell Phone	0.072	0.678	-0.750
nominal	MS Cell Text	-0.728	-0.425	1.154
nominal	Work Email	-0.752	-0.480	1.232
...	...	...	...	...
nominal	OR1332**	0.015	0.356	-0.371
nominal	OR8928**2	0.190	-0.116	-0.074
nominal	IN	-0.187	-0.122	0.309
nominal	GB	0.019	-0.133	0.114
nominal	US	0.034	0.301	-0.335
...	...	...	...	...
nominal	Priority	0.257	0.668	-0.411
nominal	Polling	-0.278	-0.656	0.378

table 2. softmax regression model

When given a new configured notification x^*, we just need to plug it in and compute

$$\underset{j}{\operatorname{argmax}} \frac{1}{1 + e^{-w^{j\mathrm{T}}x^* + 0.4\|w\|^2}}.$$

It is modeling the probability that the given notification will be confirmed, confirmed late or not confirmed then picking out the most likely one.

In general, for each class j, negative w_i^j means increase of x_i will increase the probability of the notification to be classified as class j while positive w_i means increase of x_i will decrease that probability.

Next we discuss the features by taking derivative with respect to the feature. That is

$$\frac{\partial f}{\partial x_i} = \frac{-w_i e^{-w_i x_i} e^{-(w_0+w_1x_1+\cdots+w_{i-1}x_{i-1}+w_{i+1}x_{i+1}+\cdots+w_{15}x_{15})+0.4\|w\|^2}}{(1+e^{-w^{\mathrm{T}}x+0.4\|w\|^2})^2}, \forall i \neq 0.$$

To numerical features x_1 and x_2, it measures the fluctuation in probability generated by 1-unit increase of the feature when other features are held constant. Recall that logarithmic transformation and z-score normalization have been performed. Approximately, we have

$$\frac{\partial f}{\partial x_i} \propto -w_i e^{-w_i x_i}, \forall i = 1, 2.$$

We plot $-w_i e^{-w_i x_i}$ as a function of x_i to capture how they change. Recall that class 0 is the notifications not confirmed, class 1 is the ones confirmed and class 2 is the ones confirmed late.

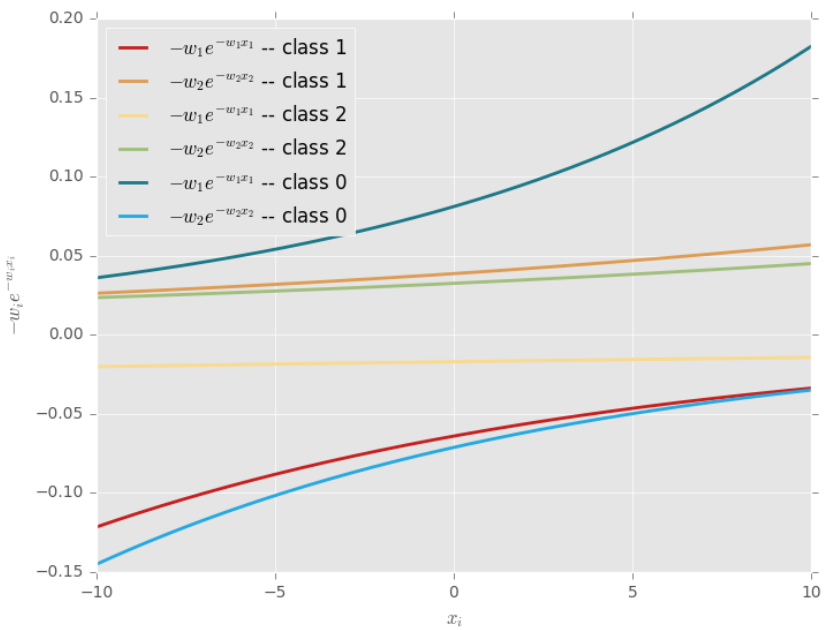

figure 7. $-w_i e^{-w_i x_i}$ vs x_i

It can be concluded that, when other configurations are fixed,

- Every time the contact adds one more path to the system, the probability of the notification to be confirmed or confirmed late will decrease. Or, equivalently, the probability of the notification to be not confirmed will increase.
- Furthermore, how much decrease in probability of the notification to be confirmed or confirmed late (when the contact adds one more path to the system) depends on the current number of paths registered in the system. The more number of paths already registered, the less probability loss of the notification to be confirmed.
- Every time the system adds one more to the notification publication, the probability of the notification to be confirmed or confirmed late will increase.

- how much increase in probability of the notification to be confirmed or confirmed late (when the system adds one more to the notification publication) depends on the current publication. The more current publication, the more probability increased of the notification to be confirmed.
- The effect of path count and publication is less significant to the notifications confirmed late than the confirmed ones.

Roughly speaking, contacts with more different paths registered in system are less likely to respond. They're encouraged to register the paths most frequently used. Also, the organization sending notifications are advised to expand the scale of notifications sent at one time. The more notifications added to each batch, the more likely each notification will be confirmed. However, if the organization has already sent a large batch of notifications, additional notification to this batch will not cause too much decrease of the notification to be confirmed.

To nominal features x_i, $\forall i = 3, \cdots, 14, 59$, $-w_i$ measures how much difference (in probability of the notification to be confirmed, confirmed late or not confirmed) the feature will make if it changes from 0 to 1, when other features are held constant. Therefore,

- In terms of path, it will increase the probability of the notification to be confirmed if the notification is sent through Cell Phone, Home Phone or Personal Cell Phone, of which Home Phone makes the most significant increase. SMS, Office Email, MS Cell Text or Work Email decrease the probability, of which Office Email makes the most significant decrease.
- As expected, endowing priority and making it standard will increase the probability of the notification to be confirmed.
- Contacts in United States are more likely to respond while the ones in India are less likely. Contacts in Great Britain is more likely to respond on time than late.
- Notification sent by organizations with similar background to Organization 1332** is probably more likely to be confirmed. Notification sent by organizations with similar background to Organization 8928**2 is more likely to be confirmed than confirmed late.

In conclusion for nominal features, organizations are encouraged to select Cell Phone, Home Phone or Personal Cell Phone as first choice if sending notification according to the preference set in system is not working. Sending standard priority notification will help. Also it has shown regional difference in people's tendency to respond. Background of organizations should be collected into the model in future works.

Prediction

Model

In this part, by duration, we mean the time for each notification to be confirmed. This part aims to correlate the features to duration through a Linear Regression Model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k.$$

On the basis of previous data preprocessing and work of classification, the time for each notification to be confirmed is extracted through the following process. Note that *duration* is in seconds and Attempt State is plugged in.

```
procedure EXTRACT-TIME(data set S)
Input: original data frame
Output: a data frame with time for each notification to be confirmed added as a new column

    get all confirmed notifications to form S_1
    for all x ∈ S_1 do:
        initialize duration = 0
        for all y ∈ S do:
            if  theny.id == x.id&&y.pathPromt == x.pathPromt:
                duration = max(y.callStartTime − x.callStartTime, duration)
            end if
        end for
        x.duration = duration
    end for
     return S_1

end procedure
```

Feature Selection and Hypothesis Testing

Exhaustive Subset Selection

Next Exhaustive Subset Selection under Bayesian information criterion (BIC) is performed. Only 4 features out of 27 are selected. They are Attempt State, Path Cell Phone (indicating whether the notification is sent through cell phone), Country US (indicating whether the contact is in US) and Country CA (indicating whether the contact is in CA). Variation Inflation Factors are also computed which suggest that there is no multicollinearity problem in the current model.

indice	variable name	VIF	coefficient	Std.	t statistic	$P(>\|t\|)$
X_1	Attempt State	1.009	-0.112	0.039	-2.845	0.005
X_2	Path Cell Phone	1.027	1.487	0.434	3.427	0.001
X_3	Country US	1.028	1.801	0.087	20.671	$<$ 2e-16
X_4	Country CA	1.007	2.772	0.607	4.569	7.4e-06

table 2. model created with the variables by Exhaustive Subset Selection

Hypothesis Testing tells that all features should be remained. The current model has an intercept -0.430 and a decision coefficient $R^2 = 0.641$, which means 64% of the variation in the time for notification to be confirmed has been explained by the introduction of the four features.

Parameter Estimation and Inference

Estimation of each parameter is presented in table 2. null hypothesis H_0:

$$\beta_k = 0,$$

X_k is not associated with the time for notification to be confirmed.
alternative hypothesis H_a:

$$\beta_k \neq 0,$$

we don't have enough evidence to show that each X_k is not associated with the time for notification to be confirmed.

test statistic:

$$t^* = \frac{b_k}{s(b_k)}$$

Require $t_{0.975}(1581 - 4) = 1.968$. In table 2, we see $|t^*|$ of each predictor is larger than 1.968, along with an extremely small $p - value$, so we conclude H_a, $\beta_k \neq 0$ for each predictor. This means each of the selected predictors is associated with the time for notification to be confirmed. Next we move on to test whether there exists a general regression relation between the predictors and duration.

F Test for Regression Relation

null hypothesis H_0:

$$\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0,$$

there is no general regression relation between the selected predictors and duration.
alternative hypothesis H_a:

$$\text{not all } \beta_k \text{ equal to } 0,$$

there is general regression relation between the selected predictors and duration.
test statistic:

$$F^* = \frac{MSR}{MSE}$$

Require $F_{0.95}(4, 275) = 2.4$. We use the test statistic $F^* = 122.7 > 2.4$ to conclude H_a, that is the existence of a regression relation.

Discussion

The model is quite straightforward.

- The coefficient of X_2 means whether the notification is sent through cell phone will make a 1.487 seconds difference.
- Similarly, the coefficient of X_3 means whether the contact is in US will make a 1.801 seconds difference.
- The coefficient of X_4 means whether the contact is in CA will make a 2.772 seconds difference.

However, the performance and interpretation of the model is limited. There are several reasons.

- We need a larger data set. First of all, duration varies too much. Most of the confirmed notifications are confirmed after 20 hours.
- The feature pool is not large enough to capture the true correlation. We still need to collect more features possibly related to duration. This also can be seen when we are doing Exhaustive Subset Selection. Only 4 out of 27 features are selected. This means most of the features are irrelevant to duration.
- The confirmed notifications lack variation. After extracting duration, 33 features vanished because they are all 0s for confirmed notifications which also means they are not contributing useful information to duration.

Diagnostic and Future Works

The classification model is achieving a 74% accuracy. The linear regression model explains above 60% of the variation in duration. Several factors are affecting the accuracy. First, the feature pool is limited. We have

- the number of notifications sent at a time,
- number of different paths the contacts registered to the system,
- priority,
- type,
- path,
- where the contact is,
- and the organization.

This feature pool is not strong enough to capture most variation in the probability of the notification to be confirmed and the time it takes. It's highly possible that other features are more related to confirmation of notification but we haven't found them yet. Secondly, curvature and interaction between these features are not studied since they are hard to interpret in a real world situation. Still, feature expansion should be tried before curvature and interaction study.

To improve the accuracy, we can try the following solutions.

- Collect more data. Especially collect more features such as organization background, reason to send notification (fire alarm, collapse, shooting incident, etc.), length of the notification and more contacts information.
- If access to more feature and data is limited, try curvature and interaction between the variables we have in hand.
- Run a ensemble method, such as AdaBoost, on the data set to increase accuracy. Ensemble methods are computational expensive so it is not yet covered in this project. Additionally, it needs to be pointed out that other model like Neural Net or even Deep Net will help to improve accuracy at the cost of interpretation.