ISYE 6501 Homework 13

Justin Lewis

April 7, 2019

Question 18.1

Q:

Describe analytics models and data that could be used to make good recommendations to the power company.

A:

When considering which power hook ups to shut off there are a few parts to be considered. This problem can be broken down into three components:

1. How much is an account costing the company?
2. Which shutoffs should actually be done each month or rather, who is not going to pay, who cannot pay, and who will pay?
3. How do we optimize the limited manpower to perform the shutoffs?

and I will address each of these below.

1. How much is an account costing the company?

This is probably the most straight forward part of the problem. Given:

- past customer power usage data
- time of year
- usage data from the client being considered
- amount of current customer debt

Use linear regression to predict the usage for a customer in the coming month or months. By knowing the current amount of debt to the company and predicting how much they will incur in the coming month, we can predict the total amount of debt each customer will have if they do not pay in the coming month. This will provide a useful way to compare customers in the coming portion of the problem.

The time of year will be relevant since power usage follows cycles depending on the weather. Knowing which month we are considering (summer vs winter) will affect the predicted amount of power usage.

2. Which shutoffs should be done?

Before we assign resources to performing shutoffs, we need to identify which customers should be considered for a shut off. We can use past payment data from many customers, and their eventual outcome (payment/never paid/paid late) to train a model to then predict labels for new customers who are being considered.

Given:

- past data from customers (power usage, payment histories, number of missed payments, average time between billing and payment)
- rent vs own status
- marital status
- number of people living on site
- customer type (residential/business/government)

Use a one-vs-all logistic regression model to predict the probability that each customer is in one of the will pay/cannot pay/can pay but will not groups. By using historical data with known results (was power cut off? did the past customers pay eventually or not?) we can create a predictive model for new customers who are being considered for power shut offs.

It is at this point we need to include the cost to the company in our recommendations. For one thing, we need to consider the threshold for when a customer becomes a candidate for a shutoff. This threshold would need to be weighted based on how much that customer is costing the company, the cost of a false positive (shut off of a customer who would pay), and the cost of performing the shut off itself. If a customer has a large running debt or is expected to use a large amount of power in the coming month, the threshold for being considered as a shutoff candidate would be lower than for a customer who has low debt, few delinquencies, or just does not cost the company that much.

For instance a customer with multiple months of missed payments and high usage (resulting in high bills) who is predicted at 60% to be a non-payer may be considered for shutoff, while a customer with multiple months of missed payments but a small amount of debt with a 65% probability of being a non-payer may not. One possible way to calculate the threshold is below:

```
p = probability of never paying
c_debt = current outstanding balance on the account
u_p = cost of the amount of power predicted to be used in the coming month
s_c = cost to perform a shutoff or turn on power

if:

    p*(c_debt + u_p) > (1-p)*(2*s_c + u_p)

then:

Assign as a shut off candidate
```

If the predicted probability is high, or the debt and usage are high, then a customer would be flagged for potential shut offs. On the right hand side we need to consider the cost of falsely performing a shutoff, reconnecting the power, and the lost usage in that time as well (assumed to be equal to u_p for simplicity). This calculation could of course be refined and would likely be more complicated, but this is a good start for considerations.

Additionally, we would most likely want to train separate models for separate types of clients. Businesses, consumers, or government clients will likely have different payment behaviours and by considering them separately we avoid having consumer habits influence the result of a business account, and vice-versa. One likely scenario that comes to mind is budget release timing for government clients each fiscal year. If it is near the end of the fiscal year they may have drained their account for utilities and will pay out the balance with funds from the next year once the budget is released.

3. How do we optimize limited manpower resources to perform shutoffs?

Given:

- identifed customers for shut offs (from the second part of the question)
- predicted cost for the customers (from the first part of the question)
- location data for each identified customer (addresses)
- amount of time needed to perform a shut off
- approximate travel time between locations using averages of multiple routes from Google Maps/GPS applications/Waze

Use communities in graphs, modularity, clustering, and optimization to prioritize areas for shutoffs.

If we consider each customer who is a candidate for a shutoff as a node, and the calculated travel times between nodes as the arcs/edges between nodes we could use the Louvain algorithm to create communities based on modularity. We would need to set some travel time threshold to limit the length of the arcs (we do not want to consider every single node connected to every other one), and we could use a separate clustering method like K-Means to establish a realistic range for the number of communities to be considered. Once we have the communities established, then the overall predicted cost of all of the nodes within each community could be calculated, and the community with the largest expected cost could be the focus for manpower resources. This model could be updated each week or month depending on how long a shutoff actually takes, or how quickly the calculated costs in an area change.

The optimization method would then be used within supernodes to identify the most efficient way for workers to travel and work within their assigned area. Our objective function would be to minimize the amount of calculated cost within that area, knowing travel times, work time, and predicted costs per customer. The primary constraint to be considered would be total time of a worker (total travel time + shut off time * number of nodes visited - break/lunch time) in a day.

Conclusion

I believe this combination of linear regression, all-vs-one logistic classification with a sliding threshold, clustering, graph theory and communities, and optimization would provide a reasonable framework to provide good recommendations to the power company. We would need access to their own data of customer payment histories, usage records, location, and customer types, as well as easily accessible public data like GPS data and traffic data for the areas to be serviced. The network model is something that is fast to optimize, so we would not have to worry about falling into local minimums or unrealistic solutions (negative time).

There is of course possible room for improvement building on this framework. There could be queuing models incorporated as well at each node. Rather than using expected averages for work times we could model each individual job as a normal distribution (most jobs will likely take similar amounts of times, with outliers at either end), and workers as resources. If we gave each worker a method to update locations once work is finished, we could provide them with the nearest unattended shutoff within their assigned community. This would make sure workers are focused within an area, initially addressing high priority (high cost) targets, but then minimizing travel time and maximizing the number of targets addressed without the risk of assigning workers in an area to the same shutoff (over allocation of resources).

However that would involve not only data and modeling, but also a significant investment by the company in hardware and communication methods to track and provide updates in real time. As it is the suggested framework utilizes relatively simple methods of modeling, and easily available data to attempt to provide good recommendations without significant investments in other areas.