Georgia Tech | Master of Science in Analytics

Homework Problem: Week 8

Overview

This document discusses the possible modelling approaches the may be taken for the Power Company Case presented in lectures. The modelling is to aid decision makers to determine which sites/residences should have power turned off given they have not paid. At a high level this requires us to determine three things:

1. Whether the customer is unlikely to resolve the issue with unpaid bills – that is to say is this customer a candidate to have their power shut off
2. The likely cost of the unpaid customer power consumption over the next period and the cost of shutting the power off
3. The sites that should have their power disconnected – this is subject to the findings of points 1 and 2 above.

The remainder of this document will step through each of the above sub-problems, finally proposing a model to determine which sites should be switched off.

Candidate Identification

It is understood in this business problem not all unpaid accounts should be considered candidates to be paid off. Some accounts are highly likely to be paid in the next few periods and some accounts are linked to social factors or extenuating circumstances which should be allowed for. It is proposed that two models are built to filter accounts which can reasonably be disconnected. The first model is a simple classification model to determine if it is conscionable to turn the power off. This will eliminate candidates that should be supported on socio-economic grounds. Features for each data point could include income, demographic data or otherwise. On appeal or application accounts may have the power company explicitly classify them supported and therefore not candidates for disconnection. Accounts that are classified as not supported will then proceed to the logistic regression model. The features for the logistic regression can include payment history, credit score, demographic information, location, and power consumption data. The reason a logistic

model is preferred is because the probability prediction can be used as a weight in the valuing of the account (step 2).

Given the socio-economic, payment history and demographic information we would use a classification model to predict if power shutdown was conscionable.

For conscionable cases, given data on payment history, credit history, power consumption data and demographics we would use logistic regression model to predict the likelihood of account payment in the next few periods.

Cost Estimation

A cost estimate will be produced by a regression model. The cost if power is left on by building a model which is a product of (potentially) payment history, credit score and demographic factors. This will be multiplied by the likelihood estimate from the previous stage to produce the expected value of taking no action on this customer.

Given data on payment history credit score and demographic factor we would produce a regression model to predict cost if this account remained open. This would be combined with the previous model to produce an expected cost.

Site Selection

The approach taken for this stage is optimisation. We will optimise to the objective function that is the minimum sum of all accounts lost. The constraints are limited to resource hours. Resource hours are consumed by jobs and the transfer between jobs. From the literature, we know that the map optimisation problem is a nondeterministic polynomial time (NP) complete problem. As such, it may require experimentation in the optimisation technique. We would not expect a optimal result but rather hope from a good result using a heuristic approach. This is a shortest path problem.
We could consider building a model treating resources as cost in the objective function rather than fixed. If this were the cause, we would increase resource allocations to the point where the next driver would save less money disconnecting accounts than the resource costs to deploy.

Given the resources allocated we would build an optimisation model to minimise cost while selecting routes for drivers. This could be expanded further by treating the number of drivers deployed as a variable.