ISyE 6501 HW12

Group Member:

Jinyu, Huang | jhuang472@gatech.edu | GTID: 903522245

Chengqi, Huang | chengqihuang@gatech.edu | GTID: 903534690

Yuefan, Hu | yuefanhu@gatech.edu | GTID:903543027

Jingyu, Li | alanli@gatech.edu | GTID: 903520148

Question 18.1

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

Here are some questions to consider:

- The bottom-line question is which shutoffs should be done each month, given the capacity constraints. One consideration is that some of the capacity – the workers' time – is taken up by travel, so maybe the shutoffs can be scheduled in a way that increases the number of them that can be done.
- Not every shutoff is equal. Some shutoffs shouldn't be done at all, because if the power is left on, those people are likely to pay the bill eventually. How can you identify which shutoffs should or shouldn't be done? And among the ones to shut off, how should they be prioritized?

Think about the problem and your approach. Then talk about it with other learners, and share and combine your ideas. And then, put your approaches up on the discussion forum, and give feedback and suggestions to each other.

You can use the {given, use, to} format to guide the discussions: Given {data}, use {model} to {result}.

Have fun! Taking a real problem, and thinking through the modeling and data process to build a good solution framework, is my favorite part of analytics.

Answer:

We decomposit this problem into 3 steps: 1. Identifying if a overdue customer is never going to pay the bill. 2. Estimating the expected value of shutting off each person's power. 3. Estimating the expected cost if shutting off a person's power. For step 2 and 3, nested problems are included. The following flow chart illustrates the approach of working out the shutting off problem mentioned above..

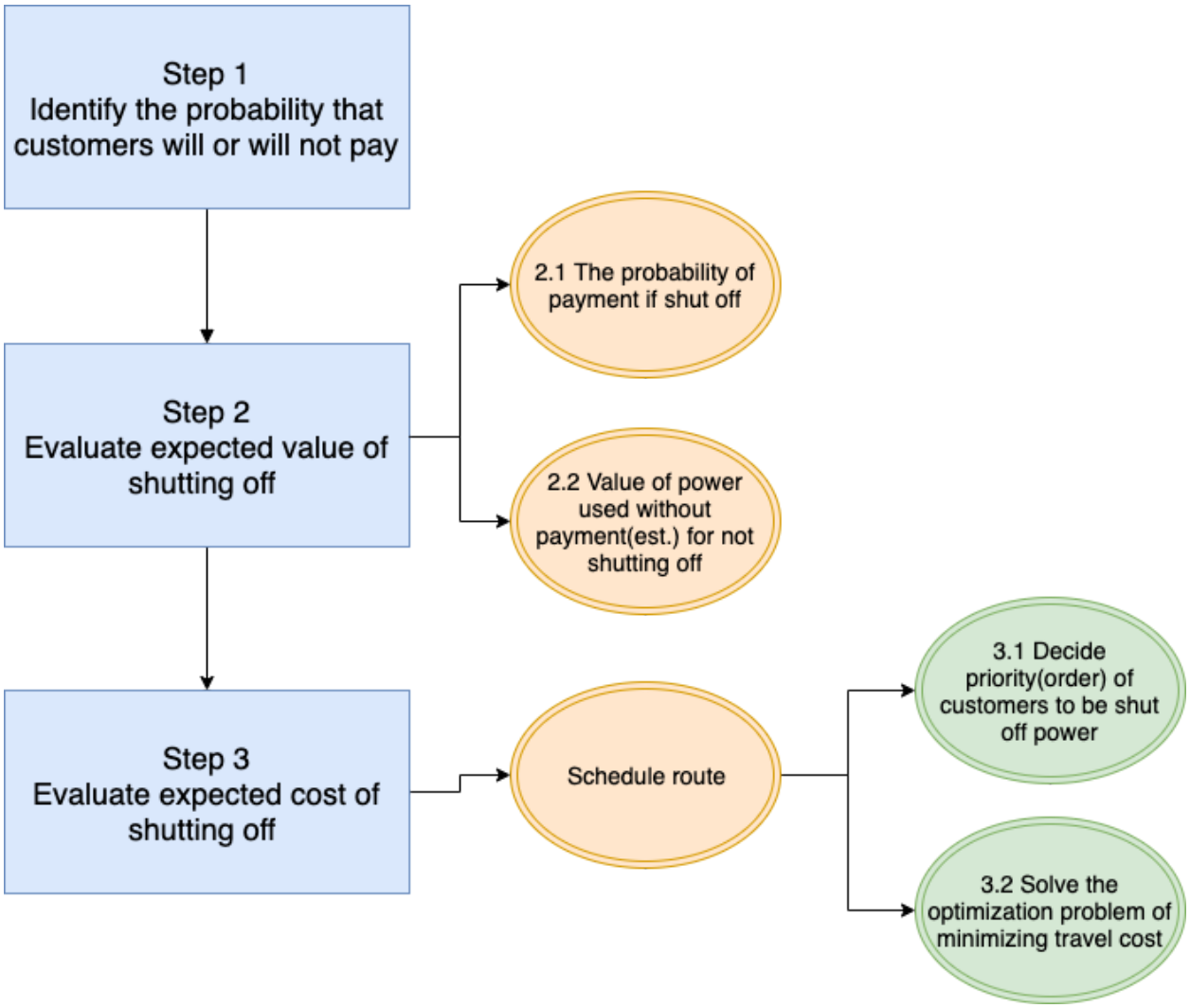

Figure 1

1. **Identify if this person is never going to pay the bill**

Data			Model	Result
Data label	**Data type**	**Data resource**		
Size of household	numerical	Community data	Logistic regression/ Distribution analysis	The possibility of payment this month
Rent or ownership	categorical	Community data		
Credit score	numerical	Credit bureau		
Salary	numerical	Registration data		
Employment status	categorical	Registration data		
Education level	categorical	Registration data		
Responsiveness of communication	categorical	Power company history data		
Payment in this month	categorical	Power company data		
Times of payment in the past 3 months	numerical	Power company history data		
Time of being customer(by month)	numerical	Power company history data		

1. Given {size of household, rent or ownership, credit score, salary, employment status, education level, responsiveness of call, Payment in this month, number of payment in the past 3 months, time of being customer(by month) }
We use {logistic regression model}
To {predict if this person is never going to pay the bill given his/her bill is overdue}.

We use logistic regression model to do the prediction and classification. We use the data listed above to train the logistic regression model. The response will be the probability that this client is going to pay the bill eventually.

In the logistic regression model, we can set up a probability P, as the classification threshold. If the probability that client i will eventually pay the bill is less than this P, we can classify this client as someone who is never going to pay.
Aside from directly building up a logistic regression model, we might also look at the distribution of payment observation and make prediction based on where the data can fit.

2. The expected value of shutting off

Data			Model	Result
Data label	**Data type**	**Data resource**		
The data used in predicting probability of none payment	multiple	multiple	Logistic regression/distribution analysis	2.1 Expected payment after shut off
The power company's share of the power market	numerical	Marketing research		
Average power used in last 3 months	numerical	Power company data	Linear regression/ Exponential smoothing	2.2 Expected usage of power without payment
Temperature	categorical	Weather Bureau		
Proportion of holidays	numerical	Calendar		

2.1 Given {the data that we have from step1 and the market share of the power company}
We use {logistic regression model or distribution analysis}
To {estimate the expected payment after shut off}.

2.2 Given {the data of average power used in last 3 months, temperature, proportion of holidays}
We use {linear regression model or exponential smoothing}
To {estimate the expected usage of power without payment}.

3. **The expected cost if shutting off a person's power**

Data			Model	Result
Data label	**Data type**	**Data resource**		
Possibility of non-payment	numerical	Logistic regression	Clustering / Louvain Algorithm	3.1 Priority(order) of customers to be shut off power
Locations of customers whose power are to be shut off	geographic coordinates	Google map/Power company data		
Household size	numerical	Registration data		
Accessibility of the community where the shut off will take place	categorical	Google map		
Locations of customers whose power is to be shut off	geographic coordinates	Google map/Power company data	Optimization	3.2 The optimal cost for shutting off to be done
Traffic pattern between each pair of locations	categorical	Google map		
Existent routes among all the locations	vector	Google map		
Number of trucks	numerical	Power company data		
Number of workers	numerical	Power company data		
Number of instruments needed for shutting off	numerical	Power company data		

power				
Priority(order) of customers to be shut off power	numerical	Previous model estimate		

3.1 Given {locations of customers whose power are to be shut off, household size, accessibility of the community where the shut off will take place},
we use {clustering or Louvain algorithm}
to {prioritize the order of customers to shut off power}.

3.2 Given {locations of customers whose power is to be shut off, traffic pattern between each pair of locations, existent routes among all the locations, the amount of resources(workers, trucks) we have,Priority(order) of customers to be shut off power},
we use {optimization model}
to {minimize the total cost of shutting off and retrieve the optimal solution}.