Foresting of Staffing Needs

April, 2019

Contributors: Luo (Iris) Yang, Marcelle Chiriboga, Patrick Tung, Weifeng (Davy) Guo

Mentor: Rodolfo Lourenzutti

Executive Summary

The purpose of the project is to help the People Analytics and Innovation Team from Providence Health Care (PHC) to predict the short-term staff needs in order to prepare for potential cost rising and staff shortage. The predictions will be made based on the known history of schedule exceptions, i.e. staff absences due to unexpected or previously arranged reasons such as sick time or vacation, respectively.

Introduction

For most positions in the healthcare business, any staff absences must always be filled in by another person. More than 70% of the operation costs in health care are tied to staffing, and the costs of substituting absences on a short notice are usually significantly higher than regular staffing costs. Hence, predicting the short-term staffing needs in order to prepare for potential shortages can significantly improve the operational efficiency of healthcare institutions.

Providence Health Care is a government agency that operates 16 healthcare facilities in British Columbia, with almost 7,000 staff, including 1,000 medical staff. At their scale, under or over staffing can have significant impact both in terms of cost to the organization as in quality of care provided to patients. For these reasons, it makes sense to predict future staffing needs to the best of their ability, which enables intelligent hiring decisions both for permanent and temporary staff.

In this project, we are partnering with PHC to predict staff needs based in their historical scheduling data. As suggested by our partner, we will focus our predictions on the operational level. i.e. short term needs, specifically on a time horizon of less than a month.

Based on the data provided by Providence Health, the question this project will aim to answer is “How many backup staff we need on a weekly basis to have a full staff for the next four weeks?”. We will start by exploring the data to identify potential features to be used for the predictions. Then, using a subset of the data provided, we will implement and train a set of different candidate models, which we will evaluate by comparing their predictions with actual known values. In the end, we will select the best model based on a combination of accuracy and interpretability.

The final product will consist of three components:

- a dashboard (developed in R Shiny or Tableau);
- the scripts containing the code used to proceed with the analysis
- a report outlining the methodologies and findings.

Data Science Techniques

The dataset consists of more than 2 millions records of exceptions since 2012, and we will split the original data based on years. This way, not only will we have a smaller dataset to generate some insight from, but we will also be able to tell the difference caused by time (facility opening, system development, increases in staff size, etc).

We will start from more general questions and then answer more specific questions as we obtain a better understanding of the data and process and mature our models. For example:

Step 1: How many exceptions will happen each week for the next four weeks?

Step 2: How many exceptions will happen each week for the next four weeks for each exception group (e.g. vacation, maternity leave, sick time, etc)?

Step 3: How many exceptions will happen each week for the next four weeks for each job family (physician, nurse, physiotherapist, etc)?

We are considering the following three approaches for the problem:

- Time Series: For this model, we assume that for every year, there is a pattern for exception occurrences. We will explore the trend throughout the years, both for all exceptions in aggregate, as for separate exception groups, in order to make predictions for each of the next four weeks.

- Linear Regression: We will fit a linear regression model to predict the number of exceptions for each one of the next four weeks based on the history of past exceptions and the known scheduled future exceptions.
- Neural Network: We will train a LSTM model in order to learn the history of exceptions, and use that to make predictions of the number of exceptions for each of the next four weeks.

Initially, we will implement the first two (simpler) approaches, and move on to the Neural Network solution if we evaluate that the more complex model has the potential to yield better results.

Timeline and Evaluation

Below is our proposed timeline for the project, as a starting point. The actual dates may be updated depending on whether particular activities turn out to be more or less time intensive than anticipated.

Time Period	Milestone
Week 1	Review documentation, and finalize the proposal reports to mentor and partner
Week 2	Data wrangling, feature selection, EDA and implement baseline model
Weeks 3 - 4	Explore different approaches to fit the models
Week 5	Build algorithms, testing, adjusting
Week 6	Improve the dashboard, wrapping up
Week 7	Presentations and reports