Forecasting of Staffing Needs

April, 2019

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

Mentor: Rodolfo Lourenzutti

Executive Summary

For most positions in the healthcare business, any staff absences must always be filled in by another staff and the costs of substituting absences with short notice are usually significantly higher than regular staffing. Hence, preparing for potential shortages by predicting the short-term staffing needs can significantly improve the operational efficiency of healthcare institutions.

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 unexpected potential costs and staff shortages. The predictions will be made based on the historical records of scheduled exceptions, i.e. staff absences due to unexpected or previously arranged reasons such as sick time or vacation, etc.

Introduction

An increase in patients' waiting times at hospitals or the postponement of important procedures, such as surgeries, are known to be extremely critical, which is why hospitals try to make sure that their clinical health care positions have backups at all times. On the other hand, more than 70% of the operation costs in health care are tied to staffing and, overstaffing can result in a significant increase in operational costs. For this reason, accurately forecasting staffing needs can have a very positive impact both in the quality of care provided to patients, and in the cost efficiency of running a health care organization.

PHC is a government agency that operates more than 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 impacts both in terms of cost to the organization and in quality of care provided to patients. Hence, it makes sense to predict future staffing needs to reach their best utilities, 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 week. Based on the data provided by PHC, the question of this project is "How many backup staff do 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). Since the raw dataset contains more than 50 columns (e.g., job family, exception group, site, etc), we will perform data wrangling in order to process feature selection. We have found several features which might potentially impact the exception occurrences, for example, notice period. Furthermore, there is an obvious difference between weekday exceptions and weekend exceptions which could indicate that weekday/weekend could be a feature. We also noticed that different types of staffing positions could have an impact on exceptions as well as their location/facility.

After wrangling the dataset, we will start with slightly general questions and then answer more specific questions. We will obtain a better understanding of the data and process and mature our models as the following steps:

Step 1: How many total exceptions considering PHC as a whole will happen each week for the next four weeks?

Step 2: How many exceptions will happen each week for the next four week with 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)?

Step 4: For each predicted exception, will PHC be able to find a relief?

We are considering the following approaches for the problem:

- Time Series: We observed that the number of exceptions show some regular changing patterns every year. We will decompose the trend and seasonalities using time series for both the entire data set and separate exception groups, to make predictions of the number of exceptions in future weeks. We will also implement classifications to label whether these exceptions can find a relief or not.

- Linear Regression: For the residuals in the decomposition of time series, we will use linear regression models to bring in new variables to explain the pattern. The prediction of errors by regression models will be combined with the forecasting of time series to produce the final result.

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.

- Neural Network: We will train an 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.

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