BC Stats Proposal | Quantifying the Responses to Open-Ended Survey Questions

Executive Summary

The BC Government conducts a Work Environment Survey with the goals of understanding their employees experience, celebrating their successes and to identify areas for improvement. A great deal of work has been performed by BC stats to analyze the open ended survey responses and to model the multiple choice questions. The insights that BC stats obtain currently achieve the survey goals, however we propose to leverage current data science techniques to help automate part of the analysis and enhance the overall insight. To do this we will apply natural language processing and machine learning techniques to automate the labeling of the text responses and uncover useful sentiment. We also propose to build a dashboard to visualize and better communicate the results of the survey. We are confident that by using cutting edge data science tools and techniques we can further understand your employees experience, celebrate their successes, and identify areas for improvement.

Introduction

The BC Public Service is committed to understanding the challenges within the workplace so they can make continuous improvements for their employees. One of the ways this is quantified is through the work environment survey (WES) which aims to measure key drivers like job satisfaction and engagement. The results of the survey are summarized into two main reports explaining the qualitative and quantitative data at the organization, ministry and department level.

This project has two main objectives which is to automate qualitative labeling and to gain new insights about the qualitative and quantitative data. Currently before the qualitative report can be generated the responses have to be manually coded into more than 60 sub-themes, we would like to automate this process to reduce the time and cost of generating the qualitative report. There are two new insights we will provide, the first is relating the qualitative and quantitative responses which will help the BC Public Service to understand how the qualitative and quantitative reports support each other. The second analysis will look at how the key drivers have changed over time and what factors may have contributed, this analysis will help the BC Public Service to see historic trends within their organization.

The objectives have been broken into three research questions:

1. Coding Themes - What is the best method to code the themes and sub-themes to the qualitative responses?

2. Linking Quantitative to Qualitative - How well does the sentiment of the qualitative responses agree with the quantitative responses?

3. Trends Across Ministries and Overtime - What trends in key engagement drivers exist over time and across departments from the 2008 to 2018 quantitative survey data?

Data Science Techniques

Each of the three research questions require different data science techniques. Generally, the focus of question one deals with predictive statistics while questions two and three are more descriptive. For all questions our approach will utilize the quantitative or qualitative data provided by the WES survey. The survey has over 15,000 respondents across 26 ministries and has been conducted over 8 survey cycles starting in 2007. There are approximately 80 multiple choice questions and one open ended response question. Our approaches to the proposed questions are discussed below.

Coding Themes

The labels to the open ended survey responses have been provided, therefore coding the themes can be described as a supervised learning problem. We will train a model to automate this task by processing the text data to be used as features in our model with the theme as our prediction target. In our initial approach we will use a bag of words analysis with a linear classifier. Building on this approach we will investigate the optimal model and pre-processing technique to increase our classification accuracy. The deliverable for this work will be a data pipeline and model that can be used to label the open ended survey questions for future cycles of the WES survey.

Linking Quantitative to Qualtitative

The free form nature of the open ended survey responses offer many insights that are ripe for natural language processing. Using sentiment analysis we plan to quantify these responses and tie them to the quantitative data. Making the connection between the open ended question and the multiple choice questions is important to add support to the current survey design. Consistency across both types of questions reinforces the current analysis (engagement model) that is done with the quantitative data.

To best answer this question we are going to investigate inter-rater reliability metrics such as percent agreement, Cohen's Kappa, and Krippendorff's Alpha. The final deliverable for this question will be a discussion in our report.

Trends Across Ministries and Overtime

To discover trends both through time and across departments, we will build a dash board to visualize all the survey cycles to date. The particular challenges for this problem is the state of the data and inconsistency in questions across the surveys. This will require extensive data cleaning. The final deliverable will be an interactive dashboard highlighting the trends and allowing for future survey data to be incorporated into the dashboard.

Timeline

To stay on target complete the specified tasks we have scheduled the following meetings:

- weekly touch base meetings with the mentor(Varada Kolhatkar) on Friday morning

- weekly meetings on Tuesday morning with the partner (BC Stats)

A brief outline of out milestones and deadlines are tabulated below:

Table 1. Project Timeline

Task	Expected number of weeks to take
Coding Themes	4
Linking Quantitative to Qualitative	4
rends Across Ministries and Overtime	5

Table 2. Milestone Timeline

Milestone	Submission Due date
Proposal report	Friday May 3, 2019 *To partner*
Final presentation	June 17 or 18, 2019 *To partner*
Final report	Wednesday June 26, 2019 *To partner*
Data product	Wednesday June 26, 2019 18:00 *To partner*