Student Internship Programme (SIP)
Final Project Report

at

Indeed

Reporting Period:
May 2022 to Jul 2022

by

New Jun Jie

Department of Computer Science
School of Computing
National University of Singapore

2022/2023

Project Title: Hires Needed Feature Experimentation for the Invite-to-Apply Recommendation System
Project ID: SY212131847
Project Supervisor: Ankush Girdhar

Abstract

At Indeed, the Invite-to-Apply product is a recommendation system that connects jobs with the right job seekers. In a previous offline analysis, a feature–the number of hires needed–was found to have a positive correlation with positive outcomes within a month after the job was posted. This report presents a summary of the work and contributions I have made throughout the internship.

Subject Descriptors

Recommendation System

Machine Learning

Keywords

Recommendation System, Machine Learning

Software

Apache Spark, Apache Hive, Amazon Web Services, Apache Airflow, Spring Boot, Jenkins, GitLab, Java, SQL

Acknowledgement

I would like to express my deepest appreciation for my mentors from the Match Recommendation Platform (MRP) team at Indeed. Their guidance, patience and continued support throughout my internship enabled me to progress and contribute at a faster pace than I had imagined. [todo]

I would like to thank Senior Engineering Manager Ankush Girdhar and Data Science Manager Dheeraj Toshniwal for providing me with the opportunity and platform, enabling me to work on a challenging, engaging and fun project, Staff Data Scientist Mengyuan Shi and Software Engineer Linsen Dong for guiding me towards theory and practical considerations in designing and executing the feature experiment for the Invite-to-Apply recommendation systems, and Software Engineering Senior Director Jordan Dea-Mattson for his mentorship and introducing for me to Indeed.

I would like to thank both Indeed and the school for allowing for such a meaningful internship opportunity.

Table of Contents

Introduction

Background of Indeed

Indeed is a worldwide employment website for job listings. Indeed is available in over 60 countries and 28 languages.

Position of Match Recommendation Platform within Indeed

The Match Recommendation Platform (MRP) team applies machine learning and recommendation algorithms to Internet-scale data pipelines to gain a strong understanding of jobs and job seekers and continuously run large-scale A/B experiments to improve on our recommendation systems.

Background of Project

The goal of the Match Recommendation Platform is to deliver great matches which lead to interviews and hires to all jobseekers and employers anywhere in the world. In an offline analysis, the number of hires needed for a job was found to have a positive correlation with one of our core metrics: the number of positive outcomes. Therefore, the hires needed feature is expected to be a good additional feature for the recommendation models to predict if an application will result in a positive outcome.

Internship Project - Hires Needed Feature Experimentation

About the Invite-to-Apply (I2A) Recommendation Models

The goal of the Match Recommendation Platform (MRP) is to provide great matches between jobs and jobseekers. One product of the MRP is the Invite-to-Apply (I2A) system, which generates automated invites for candidates to apply for jobs. Given a set of jobs, the recommendation system recommends the most relevant job candidates to invite them to apply for the jobs.

Project Objective

In an offline analysis, the hires needed feature was found to have a positive correlation with one of our core metrics: the number of positive outcomes. The project aims to answer the experiment question: Given the number of intended hires, how much can the match recommendation machine learning models improve?

To answer the experiment question, a series of engineering processes and scientific experiments need to be performed – collecting the new feature through backfilling historical data, understanding the data distribution through feature analysis, designing, scheduling and executing the A/B experiment, and analysing the experiment results.

About the Hires Needed Feature

When employers post jobs, they may have an intended number of hires for the job. Having a higher number of hires needed may potentially indicate a higher success rate for job seekers applying for the job, influencing job seeker decision to apply or the chance leading to a job offer. Therefore, the number of hires needed for a job should improve the employer-side models.

In the offline analysis previously, it was found that there was indeed a statistically significant positive correlation with positive outcomes, measured by the Pearson correlation coefficient.

S/N	No. of Hires Needed
1	2
2	1
3	2
4	RECURRING_HIRE
5	TEN_PLUS

Table 1: Example values of the hires needed feature.

In the feature analysis, the highest frequency for the hires needed feature was 1 hire needed, followed by 2 and 3, and decreasing in frequency as the number of hires needed value increases. A small percentage of jobs were recurring hires.

Collecting the Hires Needed Feature

The first step of the experimentation process is to first collect the feature data. To retrieve the feature to utilise it for online inference of the recommendation model and collecting it for training data, an addition was made to the job feature service to wire the new hires needed feature into the pipeline. From that day onwards, the training dataset would now contain the hires needed feature.

Over a hundred days of training data is required for training the machine learning models. To avoid waiting for the training dataset to fill up with the newly wired hires needed feature, backfilling the historical data had to be done. The hires needed feature could be found from a separate existing database, *dradisJobCreateOrUpdate*.

However, the *dradisJobCreateOrUpdate* database only contains jobs created or updated that day. Thus, to backfill for any particular day, we need to look back at a search window of days to obtain the hires needed feature for jobs that were not created or updated that day. The number of days to look back in the search window must have sufficient coverage in terms of the number of jobs in the dataset.

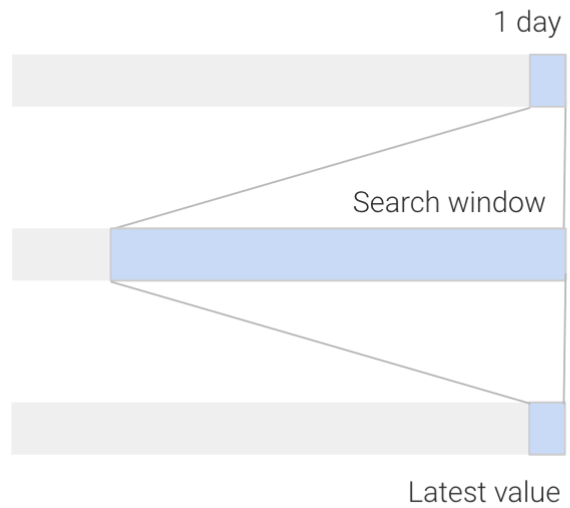

Figure 1: Backfilling the feature by looking back previous days in a search window.

Some considerations, other than the sufficiency of coverage of jobs, need to be made. For instance, since we need to load the entire search window for every day, will there be memory issues that limit the number of days to look back in the search window? Could there be days where there is missing data, thus needing a larger number of days to look back in the search window?

Model Training and Offline Evaluation

Model training is managed and executed by the internal tool at Indeed, the Butterfly machine learning platform. Butterfly is a machine learning model management platform. To ensure a fair comparison to identify a change made to the models in production, a comparison must be made between a clone of the incumbent models and the modified models.

Before model training, the hires needed feature is transformed in 3 ways. First, the feature, a *string* type, is converted into one-hot encodings, capturing information for the values "*recurring_hires*", "*ten_plus*" and "" (empty null value), to obtain the *hiresNeeded* feature. Second, the feature is also transformed, for numerical values, to the *long* type, then binned into 15 sections to obtain the *hiresNeededNumber* feature. Third, a new feature *hiresNeededRatio* is engineered, by computing the ratio of the number of hires needed to the cumulative total applies for the job, then binned into 10 sections. The 3 types of feature transformations effectively capture the information represented in the hires needed feature for machine learning.

After model training, the offline evaluation presents preliminary results to indicate whether the feature is a promising one to justify performing an A/B experiment on it. Addition of the hires needed feature results in a marginal increase in the area under the receiver operating characteristic curve (AUC) for employer-side models, but results in a marginal decrease for jobseeker-side models. The marginal change in performance is expected for adding one new feature to a model trained on tens/hundreds of thousands of features. The performance increase in only employer-side models was also as expected, since employer-side models directly model the probability of positive outcomes, which the hires needed feature was found to be positively correlated with. Therefore, the offline evaluation justifies performing a full A/B experiment on the hires needed feature.

A/B Experiment Design

The experiment question is "Given the number of intended hires, how much can the match recommendation machine learning models improve?". The goal of the experiment is thus to improve employer-side metrics, such as the probability of a positive outcome of an application.

Prior to the A/B experiment, a series of preparations need to be made. For example, I needed to schedule and de-conflict the experiment traffic from other ongoing A/B experiments. I also performed statistical power analysis to ensure that the sample size of the live traffic data is large enough for results to be statistically significant. To account for interaction effects between models, I decided on the orthogonal experiment approach, making a change to 1 model while keeping other models constant, for all models.

Experiment Results

A/B experiment for the hires needed feature required at least 3 weeks to complete. Jobseeker-side test models had a deterioration in performance as compared to the control models. Traffic for employer-side models was insufficient for statistically significant results, and more time is required for results to be conclusive.

Challenges and Takeaways

In this internship, I faced a few challenges that I eventually managed to overcome. As a result, I had also gained learning points that I would have been unable to gain otherwise.

One challenge I faced was the lack of experience with the big data technology stack. I was unfamiliar with Apache Spark, Apache Hive, Apache Airflow, and various other big data tech stack. However, this was exactly the goal of my internship – to pick up real-world, production-grade, Internet-scale hands-on experience in deploying machine learning systems. During the course of the internship, I quickly learnt to use the big data tech stack and learnt about the design considerations that went into existing code bases.

Another challenge I had faced was the running of the backfiller task to backfill historical data for the hires needed feature. The backfiller is a relatively new tool that was introduced near the start of 2022, that vastly improved velocity in conducting new experiments. However, the backfiller remains complex with many components - the dumper that dumps data from Indeed servers to Amazon Web Services (AWS) servers, the syncer that allows querying via Presto on AWS, Indeed's internal orchestration tool (Orc) that runs the dumper and syncer jobs with different parameters, and the Apache Airflow job that runs the Orc task so that the dumper and syncer jobs can be run for over hundreds of days of data. Nonetheless, through working closely with my mentors Lucas and Linsen, I finally worked through the backfiller task, ensuring that the data has been backfilled correctly.

While I had many takeaways from this internship, one key learning point was to learn to ask for help from others, persistently. It is crucial to reach out to more experienced engineers and scientists for help, once I have done my due diligence in trying to fix problems myself first. I had learnt that there is no reason why one should feel any shyness in repeatedly asking someone for help or for more details. This is a learning point that I believe can only be learnt in a company that practises a strong culture of psychological safety, which I am immensely thankful Indeed is.

Conclusion

Summary of Contributions

Over the course of my 3-month internship, I worked on the Hires Needed Feature Experimentation project, collecting the new feature through backfilling historical data, understanding the data distribution through feature analysis, designing, scheduling and executing the A/B experiment, and analysing the experiment results to determine the improvement to the recommendation models in production.

Acquired Knowledge, Skills and Experience

My internship with Indeed is my first experience with the big data technology stack, which includes Apache Spark, Apache Hive and Apache Airflow. While I initially had an introductory understanding of how to use these frameworks, I now understand various design considerations of internal tools and services, and how the different parts of the data science and machine learning pipeline come together to enable production-grade experimentation at scale. My experience with the Match Recommendation Platform team at Indeed taught me to work effectively with teams comprising data scientists and software engineers. I learnt the nuances of productionising machine learning solutions, experimenting with machine learning algorithms, and the different roles and responsibilities involved.

Concluding Remarks

As a Data Scientist, all the knowledge, skills and experience I have acquired in my internship are directly relevant and useful in my current and future studies and research. My experience at Indeed will serve advantageous in my study of data science and machine learning. Through my internship with Indeed, I have gained valuable knowledge, skills and experience and have achieved much more than I thought possible in this short timeframe. My internship at Indeed has been a very fulfilling experience and I look forward to future collaborations with Indeed.