Exceeding Indeed's Needs

GroupH[x]H: Hang Su[1], Rui Wang[2], Xu Lu[3], Yunran Chen[2], Junhong Chen[3]

[1]Department of Economics, Duke University

[2]Department of Statistical Science, Duke University

[3]Department of Electrical and Computer Engineering, Duke University

Abstract

In this project, we focus on how Indeed can better serve its existing customers and possibly expand the boundary of data applications in the real market. In the first part, we give some suggestions to job seekers using the mismatch between demand and supply in different cities. In the second part, we aim to provide some advice for Indeed to improve the number of clicks of posted jobs for different companies. Finally, we try to find out how critical data given by Indeed can be incorporated into existing models to predict macro-level indices.

Introduction

Demand and Supply: Dynamic Visualization

We formed a interaction platform to present the dynamic visualization on how the total the number of clicks, local clicks and posted jobs vary among different locations in the United States, Canada and Germany (cities level and states level) and different time. The size, the color of each bubbles, and the color of each states can be adjusted as you want.

Customized Strategies: Bayesian Hierarchical Model

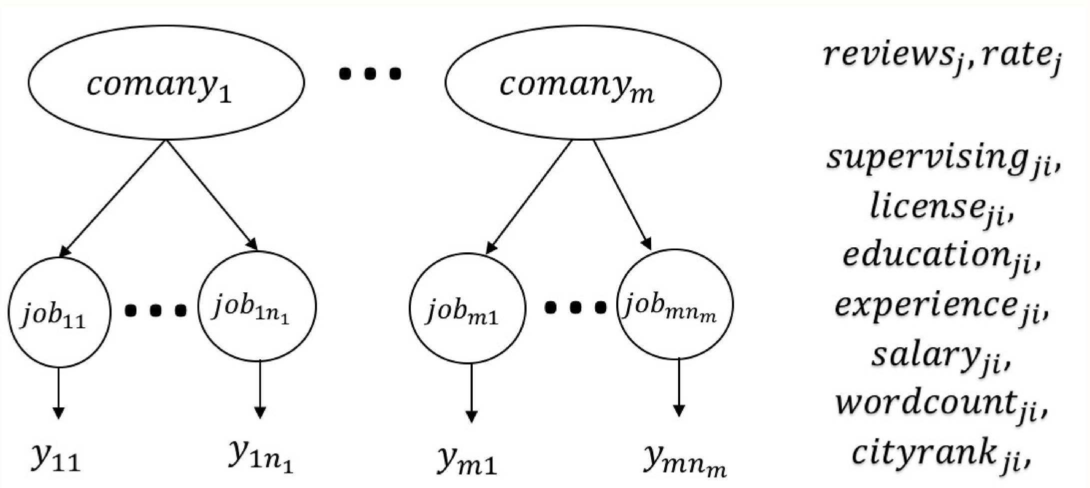

Figure 1: Hierarchical Structure

We build a Bayesian Hierarchical Model to capture the differences and similarities among different companies. Especially how the number of reviews, average ratings, requirements (licence, education, supervised, experience), salaries, description (word count), and locations would influence the number of clicks per day. And how this influence differs from each companies. Based on this we could give some suggestions to Indeed to make Customized strategies to help improve the clicks of each company.

We classified different jobs into ten categories according to Dow Jones Sector Indices. We classified the different cities into six groups according to population to incorporate the location influence into our model. We picked the fifteen largest companies in Financial industry in the United States served as an example. Our response variable is the number of clicks per-day.

Macro-level Insights: Fama-French Three Factor Model

We incorporated data from outside sources to explore whether Indeed data can be connected to variation of Macro indices. We group the data into ten sectors according to Dow Jones Sector Indices. Then we use daily data of total clicks and average salaries for different industries to explain the corresponding next-day's returns. We applied Fama-French three factor model in empirical asset pricing as our baseline, after adding clicks and salary as the independent variables, the R-squares of the models are all improved. We also use the data from both inside America and outside America to test the robustness of the regression.

Result and discussions

Demand and Supply: Dynamic Visualization

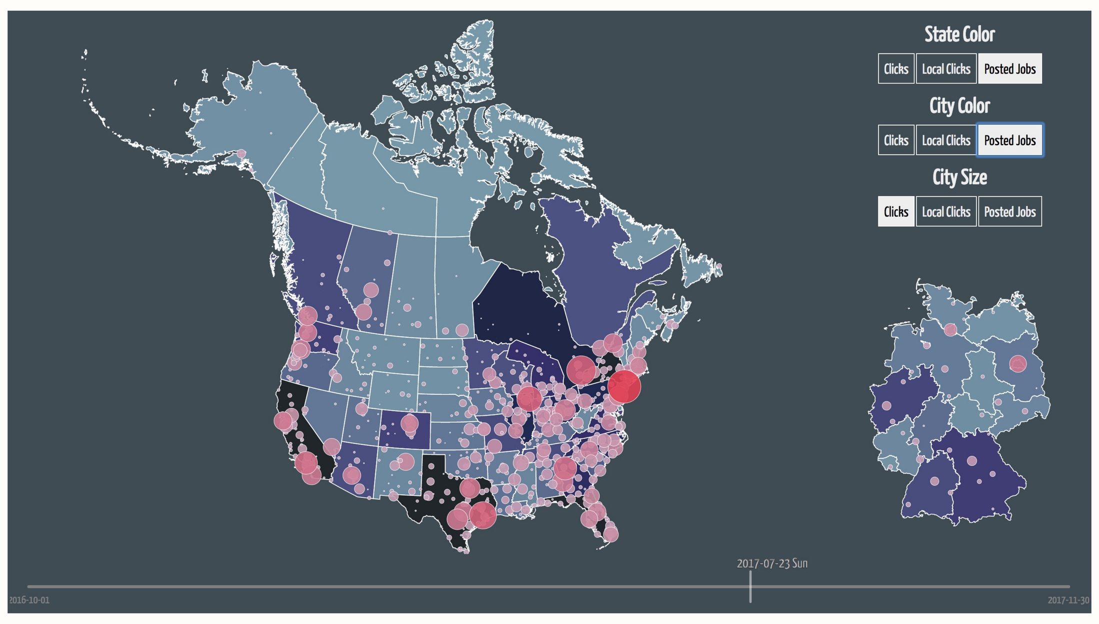

Figure 2: Dynamic Visualization

The first thing to notice, even though a bit trivial, is a pulse-like pattern in the employment market, where both the job searching and job posting undergo a decrease during the weekend. Thought the explorative visualization tool we have created, here are some of the findings which we believe would be helpful for job seekers:
1.Alberta, Quebec, New Mexico, Oregon, Kentucky, Berlin has more job requests than jobs available, which suggested that job seeker should better look for opportunities in other areas. 2. Ideal places where there are oversupply of jobs are: Bayern, Nordrhein-Westfalen, Quebec, Illinois. Employment is optimistic in these areas. 3. It is obvious that during graduation period(April-June), the number of job requests reaches the peak of the year. 4. Common hot places for job seekers and employers are: California(LA, SF, SD), Texas(Houston), New York, Ontario(Toronto), Illinois(Chicago), Bayern(Germany).

Customized Strategies: Bayesian Hierarchical Model

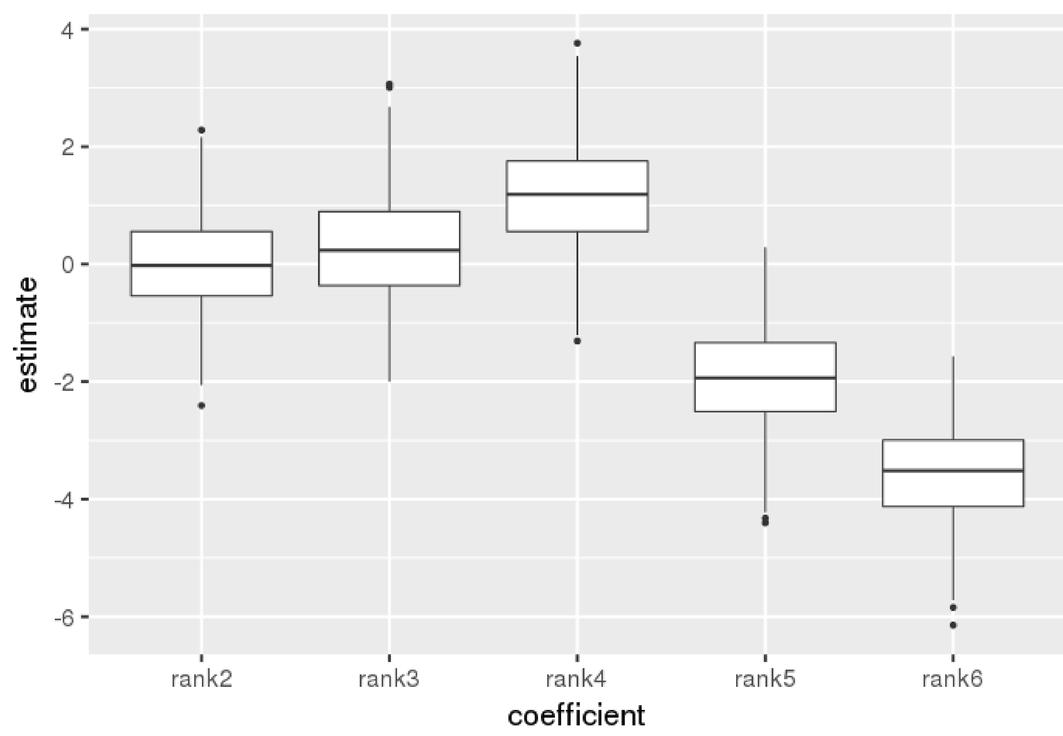

Figure 3: Coefficients for Different cities

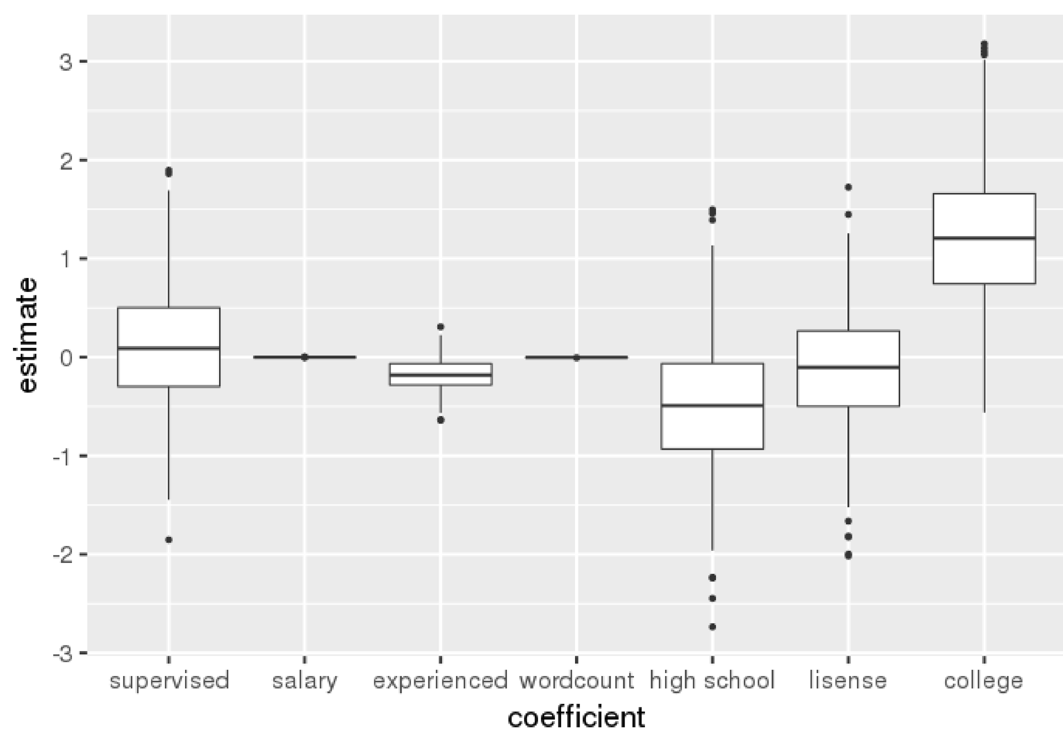

Figure 4: Coefficients for Other Covariate

The above two plots shows posterior distributions of job level covariates. As we can see, some covariates such as "city rank" and "high level education requirement" significantly deviate from 0 while predictors like "word count of job description" have little influence on clicks. This can help us do variable selection if we want to find out important predictors in estimating clicks for future job post.

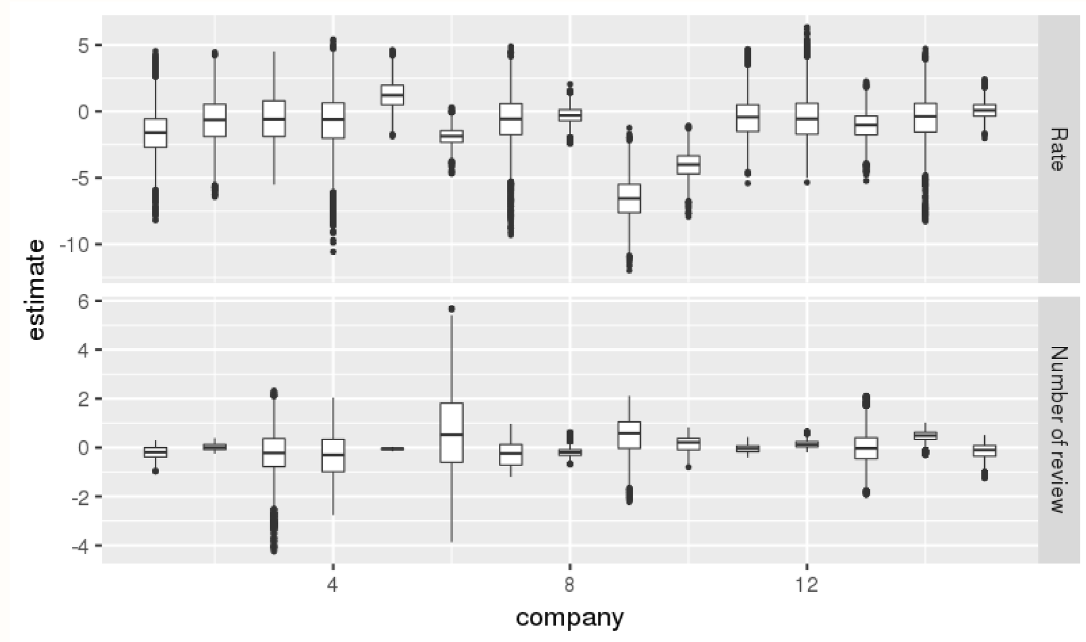

Figure 5: Coefficient for Different Companies

The plot shows that the posterior distributions of the coefficients of company level covariates are different among different companies. This indicates that company rating and review amounts affect number of clicks per-day in different level. For example,if we look at covariates "Rate", for some companies, the 75 percent confidence interval contain 0, meaning that the effect of company rating on clicks is not significant, while for some other company, the effect of company rating on clicks are significantly different from 0.

Thus, apart from taking job level characteristics into consideration, Indeed should make customized strategies to account for company characteristics. More company information can be incorporated into our model so that we can find out what company characteristics have similar effect on all companies' jobs and what characteristics have their unique influence on clicks for each company. In the meantime, companies who want to promote their job clicking can use this model as instructions to make changes accordingly.

Macro-level Insights: Fama-French Three Factor Model

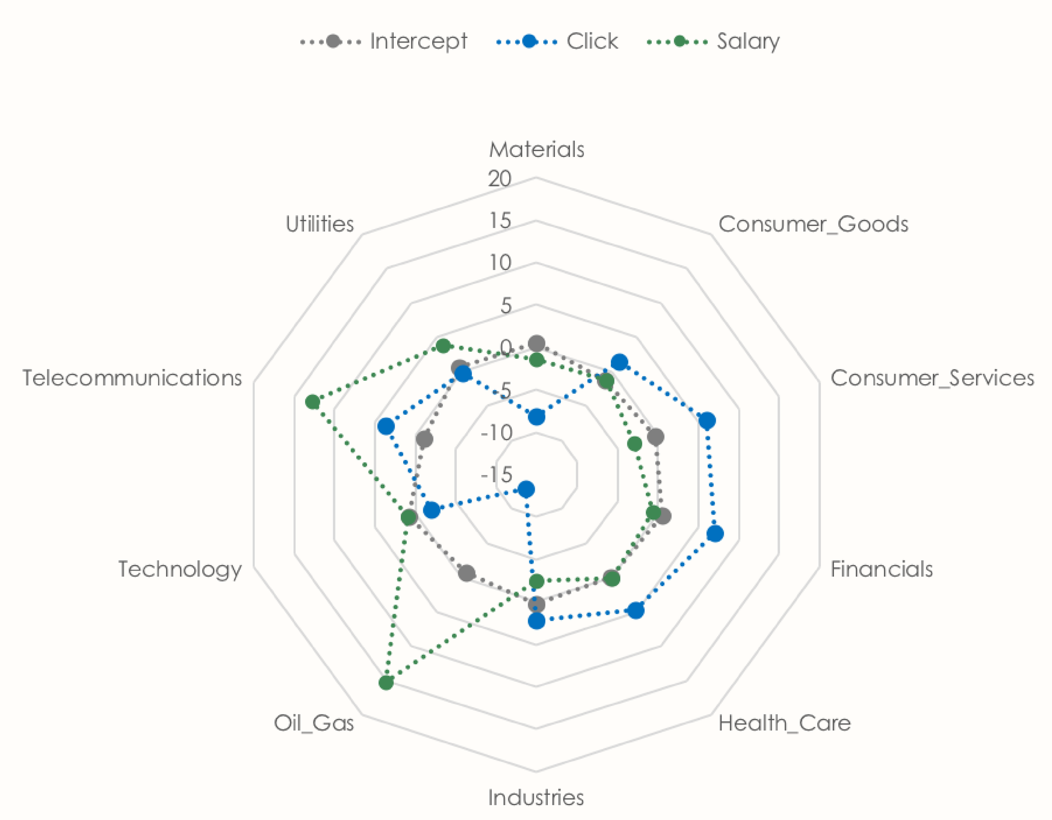

Figure 6: T-stats For Different Sectors - U.S. Data

We found out that Indeed data does well in predicting future return trends in different industries in the macro-level generally. According to the result, this model gives really good prediction of returns except utilities and industries. We can also see that clicks have the largest positive influence on the returns of Financial sector and Telecommunications, while salary has the largest positive influence on the returns of Telecommunications.