Intro: Goal and Data Preprocessing, Sherry

Methodology: MC, PageRank, Transition Matrix, Leo

Migration Map: Division and probability of movement between divisions, Shawn

Results: Tech and Food, heatmaps and histograms, Jessica

Summary: Evaluation, contribution, limitation, Lawrence

Insight

1. Goal: Predict geographical employment trend for different occupations across US.
2. Methodology: We use the Markov Chain model by treating each division as a node and movement between two divisions as a link. Following PageRank algorithm, we model the movement between two states
3. as a transition matrix and arrive at a stable equilibrium state that indicates the ranking of each division for one occupation.

Visualization

1. Map with initial states with arrows (transition matrix) . Formula
2. Map with final states with histogram. Compare with the external data

Present steps:

1. Talk about our goal and what we think about our data(how to deal with missing values, what variables are important and what industries to analyze)
2. Form a formula indicating the attractiveness of each state based on important indicators and get out transition matrix(methodology, graph on first page)
3. Talk about our final equilibrium state for each industry and results from comparing with historical data, also use other government data to back up(graph on second page)
4. Summarize the meaning and application of our model and some limitations

Sherry

Hi everyone, we are team StochasticFisher. My name is … I am.. I am .. I am.. I am.. We are happy to present here. With the given data, we want to study the employment attractiveness in different areas. We focused on locations, estimated salaries, and the number of clicks as these variables can reflect the supply and demand in labor market and they have less missing values. With these variables, we want to predict geographical employment trend for different occupations across the US. We use the markov chain model to simulate the movement of labor. We reference the employment data from U.S. Bureau of Labor Statistics. The initial page rank vector is obtained by normalizing employment in 2016. We trace out the final equilibrium state and compare it with the employment in 2017 to evaluate our model.

Leo

1. Markov Chain on graphs
2. Treat each division as a node and the movement as a link
3. Random Walk and PageRank: likelihood of a person in this occupation finally working in a certain division
4. “Attractiveness score” for each division
5. Transition Matrix entry: difference of the scores

[To predict the employment trend for a certain occupation, we have to find a Mathematical model that can show the stochastic movement between different locations. So we thought of Markov Chain Model and applied it to a graph or network setting. According to United States Census Bureau, we considered nine divisions of the US and treated each division as a node and the movement between two divisions as a link in the graph. Basically, it is a random walk problem on a graph. And finally, we applied the PageRank Algorithm that gives us the final equilibrium probability distribution across areas and used output to measure the likelihood of a person in this occupation finally working in a certain division.

To use PageRank Algorithm, we have to find the state transition matrix of the nine divisions. The transition matrix is a nine by nine matrix where the ij-th entry represents the probability of a worker moving from division i to division j for each state in the Markov Chain. According PageRank, if we continuously multiply the initial state by the transition matrix, and add some damping factor, we can get a final equilibrium state for each division.

Now we want to find a way to measure the “attractiveness” of a division, and we considered the supply of jobs, demand of workers, and the salary. Finally, the entry-ij for the transition matrix is calculated by the difference between the score for division j minus score for entry i.

And Shawn will talk about the visualization for our transition matrix]

Shawn

To better illustrate the effectiveness of our model, we divided the United States into 9 regions according to Bureau Census Divisions, labeled by color here.

Here is the **circular migration plot that describes the migration trend between the nine divisions for technology software employee in 2016.** The color of the ribbon represents where the labors comes from; the color of the bottom of the ribbon represents where the labors go; and the width of the ribbon represents how likely people are gonna move from one region to another.

Let’s take Pacific region for example. The incoming ribbons are much more wider than the outgoing ones, which means that technology employee are much more likely to move in rather than move out.

With this plot, the process and the trend of labor migration can be more intuitively understood.

Now Jessica will talk about the final results of our predictions.

Jessica

Now we try to measure the predictive power of our model using 2 years of employment rate for each region from the bureau of labor statistics for the 2 industries tech software and food. First the graph on the upper left corner shows the equilibrium state of our page rank for the tech sector.

From the graph, we expect the proportion of labor force employed in the Pacific Region, which has the deepest color, to be the highest, and there is an increase in employment rates across the nation from left to right. Compared with short term data from 2017, we noticed that for the historically strong tech regions with small employment rate variances like the Pacific region, our predictions aligned with the rates of 2017. Besides, since our model generally does a long term forecasts, deviations from the short term data are acceptable, and from the graph we expect a huge influx of labor force into the New England, where educational resources are rich, and significant outflow from the South Atlantic.

Then we applied the same techniques onto the food sector and from our final equilibrium state graph on the lower left corner, the Pacific region still takes the lead, and the middle regions historically known for food services, has deeper color as compared to the tech graph above, implying more opportunities there. Therefore, our model have good prediction power for different industries

Lawrence

As my teammates have mentioned, we used the Markov Chain Model and the PageRank theory to model the geographical employment trend for different occupations across US. We believe our model is also powerful for identifying future opportunities in different areas Our model can also be applied to other industries as well. We note that our model is still simplistic as it is based on only one year of data and we only consider the supply, demand, and salary as our metric for the attractiveness of each division. Needless to say, our model will be more robust if we are given more data.

Thanks for listening to our presentation.