A Bayesian Approach to Predicting Occupational Transitions

Lilia Chang
AtheyLab
Stanford University
lilia.chang@stanford.edu

Lisa Simon
AtheyLab
Stanford University
lksimon@stanford.edu

Karthik Rajkumar
AtheyLab
Stanford University
krajkumar@stanford.edu

1 Introduction

The future of the labor market is uncertain for many, particularly for low-wage workers in sectors that technology has already begun to disrupt. Workers displaced by AI technologies will be forced to transition to new jobs. For this reason we are interested in understanding what factors make job transitions more likely and estimating the "lifetime values" for a person starting out in a particular "state" defined by education level and current occupation.

Adapting Athey, et. al.'s Bayesian, Poisson choice model for person-restaurant choice, we model a person's one-year transition between states. This model estimates latent variables for time, individual characteristics, start and end state attributes, which allows for faster computation time and personalization of predictions compared to other discrete choice models.

4 Models

We model this problem as a persona state (occup x education level) choosing their next persona state. The attributes of personas are taken as the mean of individuals in that persona.

We limit the number of potential choices available at each starting persona state to only those destination personas that we observe actual transitions to in the data.

In the course of this project we have tried the following models.

(a) Travel-Time Factorization Model (Athey, et. al. 2017)

TTFM estimates a persona $i$'s utility of transitioning to persona $j$ at time $t$ as

$$ U_{ijt} = \theta_{ij} \delta_{ij} + \frac{W}{2} \rho^T \rho + \sum_{t \neq j} \frac{W}{2} \rho^T \delta_{it} X_t + \sum_{t < j} \frac{W}{2} \rho^T \delta_{it} Z_t + \text{noise} $$

such that the probability of persona $i$ choosing $j$ at time $t$ is

$$ p(y_{it} = j) \propto \exp(U_{ijt}) $$

As with most Bayesian models, the exact posterior over the latent variables is not available in closed form. As Athey, et. al. do, we approximate using variational inference, namely by minimizing Kullback-Leibler (KL) divergence or equivalently, maximizing the approximate using variational inference, namely by minimizing evidence lower bound (ELBO):

$$ \mathcal{L} = \mathbb{E}_{q(\mathcal{H})} \left[ \log p(y, \mathcal{H}) - \log q(\mathcal{H}) \right] $$

(b) Conditional Logit Model

The utility for conditional logit does not have time-varying or latent-latent terms. Additionally, due to the size of the attribute space the number of estimated variables by taking the Euclidean distance with the mean attributes of the end-state

$$ U_{ijt} = \alpha_{i} + \beta_{j} \cdot d_{ijt} $$

Then the probability of the choice is computed using softmax.

4 (cont.)

For these models, we classified individual persons to occupations a a baseline metric.

(c) "Interaction" Logit Model

Instead of the classical multinomial model we have features representing present occupation x attribute terms.

(d) Classical multinomial logit model

5 Data

Current Population Survey (CPS) dataset: a national U.S. labor force survey, from the years 1991-2018. Each example in the data consists of a person and their characteristics (note: we only observe a person's occupation for two years). Features include age, sex, income, race, education level.

Autor and Dorn Occupation indices: We have supplemented the CPS data with the dataset produced by Autor and Dorn in their 2013 analysis of U.S. occupations. We specifically include their indices on an occupation's level of "abstract," "routine," and "manual" work.

We split the data into 70% train, 20% test, 10% validation.

- Years: 1991 - 2018
- Number of individuals: 1,670,526
- Possible years of education: 11
- Number of occupations: ~400

5 Results

It should be noted that this is essentially a classification problem for an average, a possibility set of size 236 end states per start state (with standard deviation 169). The best TTFM model was trained with 3000 iterations, decaying learning rate, dimension 10 for the latent vectors on time, and dimension 20 for the latent vectors per persona.

<table>
<thead>
<tr>
<th>Accuracy</th>
<th>Log-likelihood</th>
<th>F1</th>
<th>Precision</th>
<th>Recall</th>
<th>Train size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train</td>
<td>0.366</td>
<td>-116.9</td>
<td>0.22</td>
<td>0.386</td>
<td>0.16</td>
</tr>
<tr>
<td>Test</td>
<td>0.364</td>
<td>-118.1</td>
<td>0.24</td>
<td>0.367</td>
<td>0.18</td>
</tr>
</tbody>
</table>

Due to the limitations of the baseline models we could only train on a 10% sub-sample of the data.

<table>
<thead>
<tr>
<th>Baseline Model</th>
<th>Train, test accuracy</th>
<th>Train, test size</th>
<th>Num features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interaction Logit</td>
<td>3.3%, 3.0%</td>
<td>150000, 150000</td>
<td>25 * 20</td>
</tr>
<tr>
<td>Classical Multinom</td>
<td>2.5%, 2.4%</td>
<td>150000, 150000</td>
<td>20</td>
</tr>
</tbody>
</table>

At the time of reporting, our conditional logit model, using Stata's off-the-shelf program, was unable to converge.

6 Future Work

The fact that the TTFM can train on the full set of data is already significant improvement to the other logit models. Our accuracy and F1/precision/recall scores are good considering the complexity of the decision space. We still wish to have a comparable baseline and so going forth we will implement the logit models instead of using off-the-shelf packages. We also plan on including a distance factor for the TTFM utility, similar to our currently conditional logit utility, but with a Mahalonobis distance, to incorporate individual-level characteristics (instead of mean characteristics).

7 References


(Author, et. al. 2017)