Optimizing Refugee Integration

Jason Bud Ginsberg
Department of Computer Science
Stanford University
jasong2@stanford.edu

Max Darling
Department of Computer Science
Stanford University
darlingm@stanford.edu

1 Problem Statement and Motivation

Our project aims to improve the process by which refugees are assigned to resettlement locations within a given country. Our hope is that accurate predictions in placement outcomes will improve policy recommendations and planning for future refugee crises.

Refugee resettlement is a complex issue comprised of multiple factors. Today, major sources of mass migrations are myriad, whether they stem from conflicts (e.g. the Syrian Civil War), demographic changes (e.g. the youth boom in Africa), or climate change [3]. One issue found in immigration research is that two refugees of similar backgrounds can face very different outcomes based solely on where they placed. One interpretation is that the complex interplay of geographic context and personal characteristics affect employment rates. Current policies in the United States, however, are agnostic to these complex factors, deciding settlement locations on proximity, capacity, and even by random [1].

Though our project will necessarily have to suspend disbelief with regard to political realities, our aim is to aid in solving this issue of skill allocation. Additionally, for feasibility's sake, we will have to simplify some significant constraints (e.g those seeking temporary resettlement, connections among family members in the anonymous dataset, etc). The implementation of our approach may raise some issues: left unexplained, it may be viewed as ignorant toward complex regional histories and or for allocating humans as "resources." Care will need to be taken in adhering to ethical best practices with regard to the above.

With that, our approach is as follows. Using the Office of Refugee Resettlement's Annual Survey of Refugees (ASR) from 2016, we will train an algorithm to assign unseen refugees to an optimal region within the United States for resettlement. The data set contains roughly 4,800 refugees along with surveys results for country of origin, family size, level of education, known languages, gender, age, assignment location, date of arrival, employment history, and current level of employment within the US.

In further detail, we will split the dataset across geographic regions, and train separate neural networks for each region. The output of the networks will be a probability representing how likely a refugee is to be employed in that region. For each new refugee not in the training set, we will then compute an employment success score across all regions. In theory, we could simply assign each refugee to the region of maximum success. But in practice, geographic distribution and regional capacity limits set strict constraints. As such, we will feed the predicted scores to a matching algorithm that will find the optimal assignments given the regional restrictions.

Finally, we hope to apply our model as a toolkit for policymaking. In particular, we will generate new test datasets based on realistic future scenarios. For example, we might generate a dataset of

CS 221 Artificial Intelligence (Fall 2019), Stanford, CA.

10,000 refugees fleeing Nicaragua, sampled from available population statistics. We will then use our model to see how well these refugees could be accommodated by the US in the best case scenario.

In terms of related work, we identify two similar projects: *Placement Optimization in Refugee* Resettlement (Trapp et al., 2018) and *Improving Refugee Integration through Data-Driven Algorithmic Assignment* (Bansak et al., 2018). The latter paper follows a similar approach of ML paired with matching using different non-public data sets. This work, however, does not specify the model architecture or how it was trained. Our work will focus first on replicating this research and second on contributing a new extension by making predictions for future scenarios.

The difficulty of this approach lies mainly in the available dataset. 5,000 refugees is a small number especially when we will have to split the data not only across train, test, and validation sets, but also per region. We may also face difficulties when trying to sample from population data. If our model is trained on inputs of different dimensions and features, we will not be able to apply our model to the new data. This will require careful curation to coalesce the datasets while preserving expressive features. Finally, the matching algorithm is not within the scope of this course and will require some further research into the best implementation given our data and needs. Fortunately, the aforementioned paper by Bansak et. al was conducted at the Stanford Immigration Policy Lab, who we hope to reach out for support in this project and with finding more data.

As a preliminary step, we have implemented a baseline and oracle. The baseline consists of a linear regression model that predicts emplyoment outcome based on two hand-picked paramters: english ability and level of education. This model achieved 56% accuracy. The oracle performed k-nearest neighbors on the data set split 80-20 between a training and test set. The data set formatted into a CSV from questionnaire data. Questions related directly to employment outcomes were removed from the inputs (with current employment representing the labeled output). This oracle was implemented in PyTorch as vectorized code. With k = 25, we found that the model could classify the employment of a refugee with 70.44% accuracy across all regions combined.

References

[1] Bansak, K, Ferwerda, J., Hainmueller, J., Dillon, A., Hangartner, D., Lawrence, D., & Weinstein, J. (2018) Improving refugee integration through data-driven algorithmic assignment. *Science*. **359**: 325-329.

[2] Trapp, A., Teytelboym, A., Martinello, A., Andersson, T., Ahani, N. (2018). Placement Optimization in Refugee Resettlement. Working Paper 2018:23, Lund University, Department of Economics.

[3] "Modern Immigration to the United Kingdom." Wikipedia, Wikimedia Foundation, 1 Oct. 2019, en.wikipedia.org/wiki/Modern_immigration_to_the_United_Kingdom.