Education Inequality Project Proposal

Christopher Tran
ctran29@uic.edu

Weixin Liu
wliu53@uic.edu

1 INTRODUCTION

Education is important in developing future generations. But, even in today's society, there is large amounts of inequality in education. This inequality is largely related to income inequality, so often times poor families do not have access to good education. We wish to study the link between school performance and income geographically, and we wish to discover properties of well performing schools that may suggest ways to decrease this education gap.

Our goal is to find patterns in school and census data that may suggest ways to decrease inequality in the education system. We will be looking at integrated school data from the National Center for Education Statistics Common Core of Data (CCD) [1] and the US Census Bureu American Community Survey (ACS) [2] data. The former data, contain administrative data such as number of students and number of teachers. The Department of Education also has an EDFacts website [3] that contains test scores and graduation rate by schools which will help in evaluating and comparing performance of schools. The ACS data are information on average income level of families in each state by geographic location. We wish to link these two datasets to discover important variables that may show ways to improve education.

To link these datasets, we propose a graphical model to capture geographic properties between schools based on average family income level. Using graduation rate and standardized test scores provided to evaluate school performance, we will study the effects of variables that increase or decrease a schools performance level. In this way, we may be able to infer some important properties which can suggest ways to improve education quality in schools that perform poorly. We suggest that average family income plays an important part, but we wish to discover other key factors that affect education.

2 PLAN OF ACTION

For our graphical model, we propose a homogenous network for measuring school performance, where a node represents a school, and links between schools are based geographic distances between schools. Node attributes will be gathered from the CCD and EDFacts dataset to get important administrative attributes to schools, such as number of teachers and students. We also will use graduation and test scores to evaluate school performance. We propose to bin schools into discrete classes based on school performance. Using the ACS data, we can assign average family income per school zone to account for income inequality between different schools. Since much of school funding comes from the state and local (district) level, we believe schools close to each other (such as in the same county) will have similar performance. Also, due to the nature of school funding, we propose to restrict links to schools that are in the same state.

To discover important features we consider the problem of education gaps as a classification problem which classifies schools into performance bin (i.e. very good, good, bad, etc.) and determine features that are important. For our baseline, we will use random forest as our non-network method for feature ranking. We will use existing methods for feature selection and collective classification techniques for graphical models.

The desired outcome would be to discover latent variables other than family income that determine education quality. We also wish to show that treating schools independently of location and other schools does not provide as good results as modeling schools as a network of other schools.

[1] Data gathered on the National Center for Education Statistics website at: https://nces.ed.gov/ccd/pubschuniv.asp.

[2] ACS Data: https://www.census.gov/programs-surveys/acs/.

[3] EDFacts Data: https://www2.ed.gov/about/inits/ed/edfacts/data-files/index.html.