Heuristics

Xiaoyi Yuan

November 7, 2017

Contents

1 Employee distribution in workplaces at a specific tract

On tract level, we have data on number of workplaces and number of employees. We don't have data on the distribution of employees in the workplaces in that tract. So based on a research article (`http://www.haas.berkeley.`

edu/faculty/pdf/wallace_dynamics.pdf), we come up with a lognormal distribution to assign workplaces. [This is done in preprocessing, establishment countes and sizes]

2 Limits of school age and the usage of "school district"

When assigning students to school, we don't have data on which school that children (under 18) in a specific household attend. We came up with the age range for each level of education so that we find the NEAREST (note: another assumption) school that fits the child's age. The age range for daycare is 0-5 years old and the age range for school is as follows: {'01': 6, '02': 7, '03': 8, '04': 9, '05': 10, '06': 11, '07': 12, '08': 13, '09': 14, '10': 15, '11': 16, '12': 17, 'KG': 5, 'PK': 3, 'T1': 6, 'TK': 5, 'UG': 5} ("01" - "12" = Grade 1-12; "KG" = Kindergarten; "PK" = Prekindergarten; "T1" = Transitional first grade; "TK" = Transitional kindergarten; "UG" = ?) [This is done in preprocessing, school] When assigning students to a specific school (note that we have specific locations and enrollment data for each school, unlike workplaces we only have tract id), we choose nearest one that is not full. If it's full, then choose a secondary nearest. If no school is available, then choose one that is the least crowded. [This is done in main method, assigning students to schools]

3 Age distribution within each age range

On track level, we have data on number of people for each age range (range size is 4 years) but we don't know the number of people for each age. The distribution of people within each age range follows uniform distribution. [This is done in main method, create individuals]

4 Age constraints between wife, husband, and their childre in households

This step happens after we have synthetic population with two basic features, age and sex. When assigning everyone to households, age and sex is two basic constraints. However, we added more constrains other than these to make the synthetic households to be more realistic. One of them is the age limits between wife, husband, and their children. Note that these constraints only

apply to household type of family with children under 18. The specifics of the constraints are:

1. both husband and wife in a household has to be above 18
2. the difference between husband and wife is based on the US national husband wife age disparity in heterosexual relationships from 2013.

-9 < husband.age - wife.age < 10 if cannot find enough pairs, search: -14 < husband.age - wife.age < 15 if still cannot find enough pairs, search: -24 < husband.age - wife.age < 25 if still cannot find enough pairs, search the unpaired female (above 18) left and pair them. father.age - child.age <= 50 mother.age - child.age <= 40 [This is done in main method, $\text{create}_{\text{household}}$,$\text{gen}_{\text{households}}$]

5 Number of children in the households

Although from census data, we know the number of households for each household type, for those with children, we don't know the number of children. When generating these households, we use uniform distribution. More specifically, household type 1: husband & wife with kids ($\text{num}_{\text{ofchild}}$ = max(1,abs(int(np.random.normal(2)))) household type 3: male with kids and wife is not present ($\text{num}_{\text{ofchild}}$ = max(1,abs(int(np.random.normal(1.6)))) household type 5: female with kids and husband is not present $\text{num}_{\text{ofchild}}$ = max(1,abs(int(np.random.normal(1.8)))) reference: `https://www.census.gov/hhes/families/files/graphics/FM-3.pdf` [This is done in main method, $\text{create}_{\text{household}}$, $\text{gen}_{\text{households}}$]

6 Institutional group quarters

group quarters are correctional facilities for adults, juvenile facilities, nursing facilities/skilled-nursing facilities, and other institutional facilities or noninstitutional group quarters (college/university student housing,military quarters, and other noninstitutional facilities). We assume that for each tract, there's only one group quarters and those who belongs to living in group quarters all live in this location. [This is done in main method, $\text{create}_{\text{household}}$, $\text{populate}_{\text{households}}$]

7 Treating "leftover" population as relatives in households and those living in group quarters

After assigning people to all kinds of households, there are a number of people left (suppose it's n). We first randomly sample one household with members smaller than 7 n times with replacement = True, meaning households can have more than one relatives. Then add these relatives to correponding households. If after this step, there are still leftover people, add all of them to the group quarter for that tract. [This is done in main method, populate households]

8 Employment population estimation and employees selection:

At the tract level, we don't know how many people and who are employeed. One dataset that gives indirect information is the origin destination profiles where it has the number of employeed population who commute from/to certain tracts. Therefore, for people living in the tract we are generating, we have the proportion who commute. We took this commute population as employeed population. Those who commute outside of our One way to validate the effectiveness of this method is that by using this heuritics, we usually end up having half of the population (>18) that are employeed, which fit the general statistics of 58.5% employeed population (16+) in the US. See the reference here: `https://data.bls.gov/timeseries/LNS12300000`. When selecting who are employeed, we sample randomly from population with age 18+. [This is done in main method, assign workplaces]

9 Household and workplace location creation in physical space

After creating and populating households, the physical locations for these households are not created. The locations are created on residential roads (the road shapefile has categories of roads). Potential locations are created by choosing a location every 50 meters from residential roads and 20 meters from secondary (non-residential) roads. Then we sample from the potential locations based on number of households and workplaces we created earlier. Houses can be on top of each other. [This is done in main method, space creation]

10 Social network typologies

When a household has less than 5 people, then create a fully connected network for every one. If more than 5, we create a small world network. Also, for school and workplace where there are more than 5 members, we create same small world network among them. All the small world network has k=4 and p=0.3. See reference: `http://www.sciencedirect.com/science/article/pii/S0375960199007574` [This is done in main methods, create networks]