I cleaned the data set using Python, removed the content included in parentheses (attached: day\_in\_the\_life\_drop\_suffix.csv), and merged all the locations a person stays in one day to a single sequence, and only remained the column Age and Gender (attached: day\_in\_the\_life\_after\_processing.csv). There are only 83 pieces data after processing. Without considering about when a person is arriving to a location and how long it stays, I put the data set into Weka and do supervised machine learning: Naïve Bayes, decision tree and random forest with cross-validation. Probably due to the small amount of data, all the algorithms don’t give good results.

A picture containing text

Description automatically generated

Figure1: Result for Naïve Bayes

A screenshot of a cell phone

Description automatically generated

Figure2: Result for decision tree

A screenshot of a cell phone

Description automatically generated

Figure3: Result for random forest

After doing supervised machine learning, I also try two unsupervised machine learning algorithms: cluster (result is in attached file: result\_for\_cluster) and perceptron. As imagined, the result is also not satisfied. Because these kinds of algorithms need huge amount of data to train.

A screenshot of a cell phone

Description automatically generated

Figure4: Result for Multilayer perceptron

Then, I change every age to age interval (day\_in\_the\_life\_drop\_suffix\_age\_processed.csv). And I was trying to extract some information about trajectory patterns with different age and gender by putting the data set into R and draw a graphic (Figure 1 and Figure 2). X axis represents different trajectory patterns while Y axis is the count of how many persons have a particular pattern with different age interval. As you can see, there are so many different movement patterns. For female, pattern ”Home-Private Transport- Private Transport-other- Private Transport-other” have the most frequency. There are two others in this pattern. We don’t know information about this “other”, because people are unwilling to tell, maybe this is the reason why this pattern is with the highest frequency.

A screenshot of a cell phone

Description automatically generated

Figure 1: Trajectory pattern for female

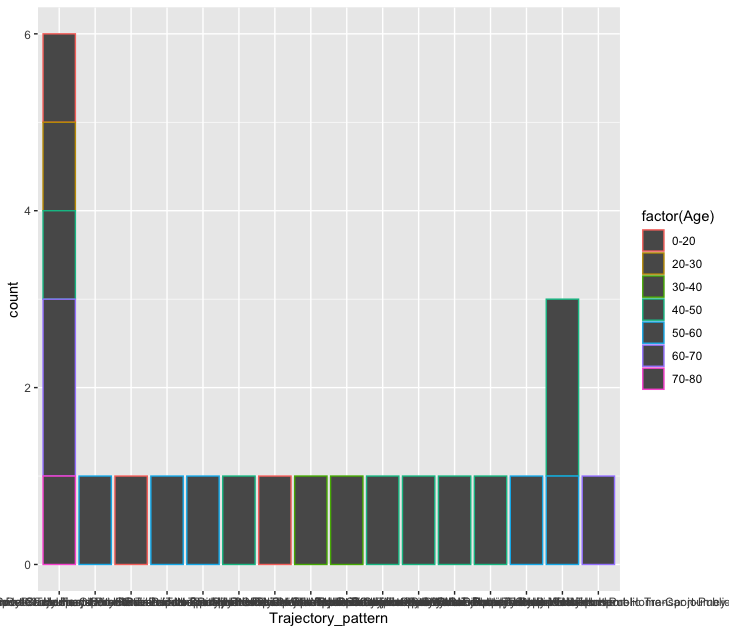


Figure 2: Trajectory pattern for male

All of the code and documents and result file attached here will also be uploaded to Github.

<https://github.com/liyi19950329/Research-Project.git>