DATA-DRIVEN HUMAN ACTIVITY PREDICTION

DHIVYA GOVINDARAJAN
Project Advisor: Dr. Leonid Reznik

Introduction

• With advancements in information and communication technologies (ICT), massive volume of human activity data is generated everyday.
• This data provides an opportunity to understand and predict human activities.
• Predicting human activity is gaining importance in several fields such as robotics, city planning, transportation, surveillance, targeted advertisements, wearable devices, etc.
• Goal: To build a Human Activity Prediction (HAP) model given the history of activities done and other contextual information.

Data

• American Time Use Survey (ATUS) [1] is conducted by Bureau of Labor Statistics (BLS).
• ATUS surveys people on how they spend their time doing various activities in a given day.
• ATUS contains activities information such as where, when, duration, start and stop time, etc. and respondents demographics information such as age, income, occupation, etc.
• Minimum 11,000 respondents participate each year chosen based on three-stage stratified sampling from Current Population Survey (CPS).
• ATUS consists of 466 activities, with 17 major activities.
• Activities are six digit codes that are classified into 3 tiers: Tier1 (major), Tier2 and Tier3 are additional levels of detail of Tier1.
• For example: Activity code 070101
  • 07 – Consumer Purchases
  • 01 – Shopping
  • 01 – Grocery Shopping

Data Analysis

6:00am: Home, Unemployed
7:30am: Breakfast
8:30am: Work
10:00am: Lunch
12:30pm: Lunch
3:00pm: Lunch
6:00pm: Dinners
7:00pm: Entertainment
10:00pm: Beds

Fig 1 With whom did an average American spent in a day
Fig 2 Where did an average American spent in a day
Fig 3. Average time spent in a day grouped by Gender

Methodology

• Data cleaning and statistical analysis are performed using R programming. It involves detecting and removing anomalies and outliers, normalizing values, consistency check, correct misspelled data and eliminating irrelevant attributes.
• Associated files: Activity, Respondent, Roster, Roster Eldercare, Summary, Who and CPS file are merged together.
• Dataset is split into training (70%) and testing (30%).
• Three modelling algorithms are used to build the prediction model: Bayesian Networks, Decision Tree and Naive Bayes.

Test Results: Illustrative Example of HAP

Input: Socio-economic & contextual information

Output: Activity

Results

• Prediction model accuracy obtained are:
  - Bayesian Networks (90%), J48(93%) and Naïve Bayes (88%).
  - HAP model is successfully generated with 90% - 93% accuracy for major activities.
  - This model uses 11 features to predict. In future, training with more contextual information is needed.
  - Due to the complexity involved, predicting sub-categories is a big challenge. However, in real world, sub-activities prediction is necessary.

Discussions & Conclusion

• Prediction model accuracy obtained are:
  - Bayesian Networks (90%), J48(93%) and Naïve Bayes (88%).
  - HAP model is successfully generated with 90% - 93% accuracy for major activities.
  - This model uses 11 features to predict. In future, training with more contextual information is needed.
  - Due to the complexity involved, predicting sub-categories is a big challenge. However, in real world, sub-activities prediction is necessary.

References


Contact details
Dhivya Govindarajan: dng2378@rit.edu
Leonid Reznik: lrvcs@rit.edu