行動履歴データを用いた
教師なし学習による個人属性の推定

ESTIMATION OF SOCIOECONOMIC PROPERTIES BY UNSUPERVISED LEARNING USING BEHAVIOR HISTORY DATA

Team L　山梨大学　University of Yamanashi

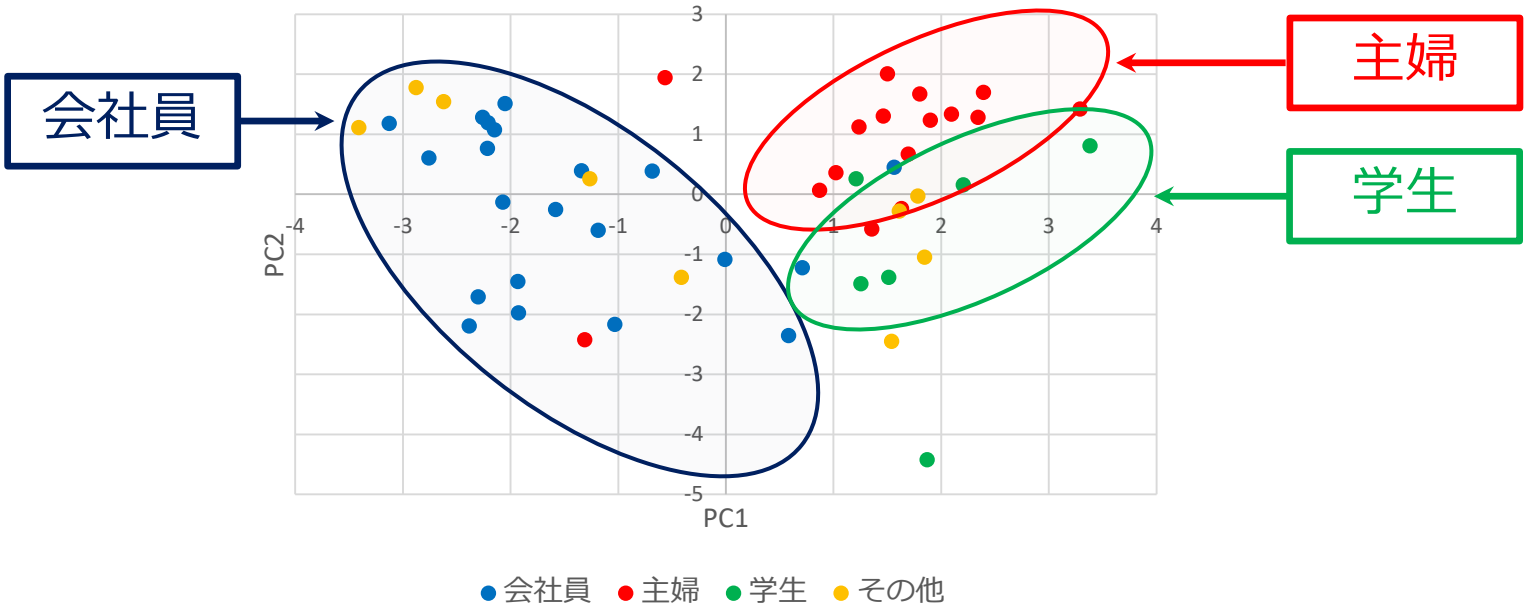
M2　Lahman
M2　末木祐多
M1　有働友哉
B4　阿保谷隆
B4　飯島翼
B4　三島隆

基礎分析　Basic analysis

行動履歴データを個人ごとに集計し，新たな個人属性を提案
Aggregate action history data for individuals and propose new personal attributes.

Research Question

行動履歴データより職業を推定できるのでは？
From the behavior history data, can I estimate occupation?



行動履歴データに主成分分析をして得られた
主成分1と主成分2を職業ごとに色分けした散布図

目的　Objective

PPデータより，個人について複数日のデータを用いることで
→　**より安定して精度の高い**

確率的潜在意味解釈（PLSI）を用いることで
→　**分類されたクラスの解釈ができる**

教師なし学習による職業の推定を行う

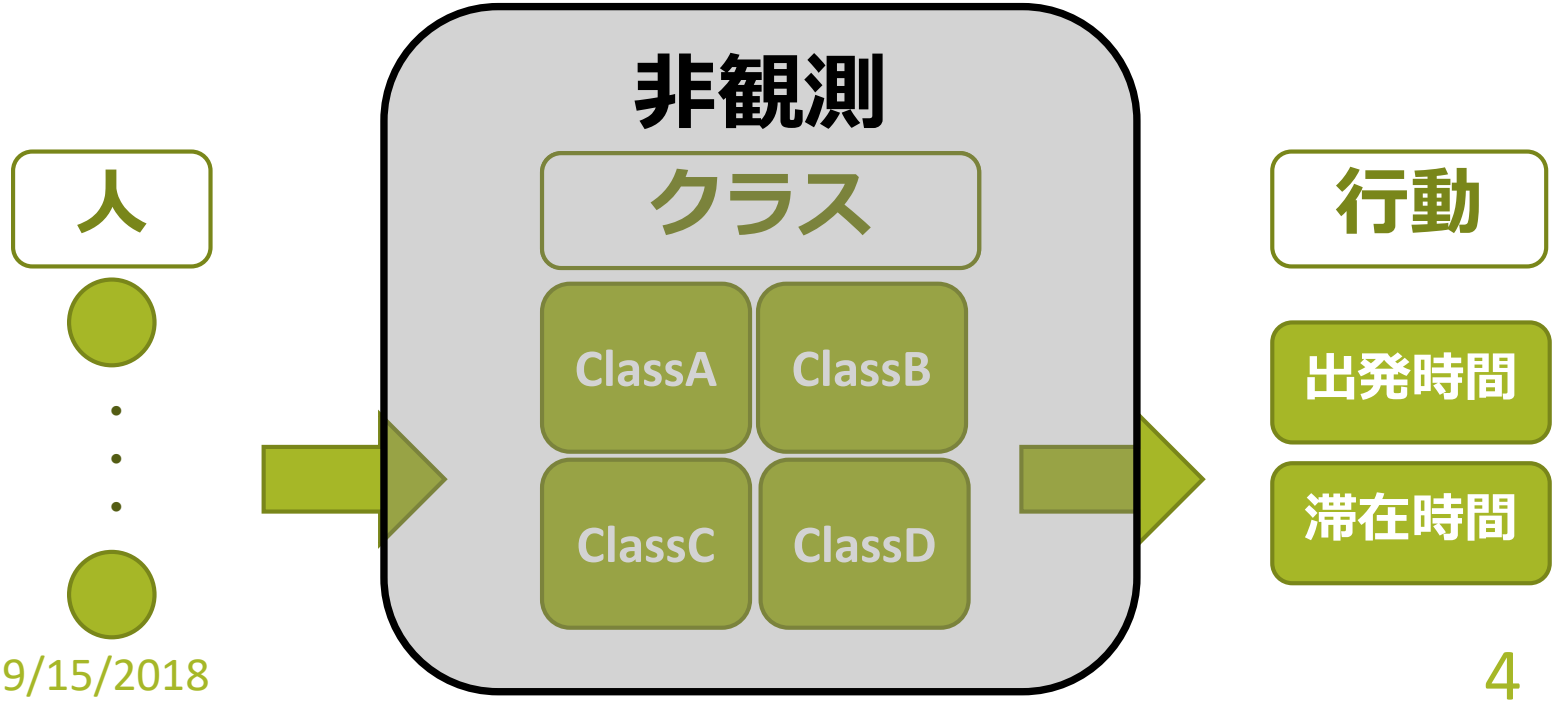
Day to Dayで個人の交通行動の変化に応じて
確率的に属性を割り当てた

分析手法　Method

使用データ　：平日の行動のみを抽出（52人・1114日）
＊7日以上平日の記録を行っている個人のみ扱う

：ツアーの出発時間，滞在時間を個人ごとに集計

クラスター分析：確率的潜在意味解釈
Probabilistic Latent Semantic Infexing



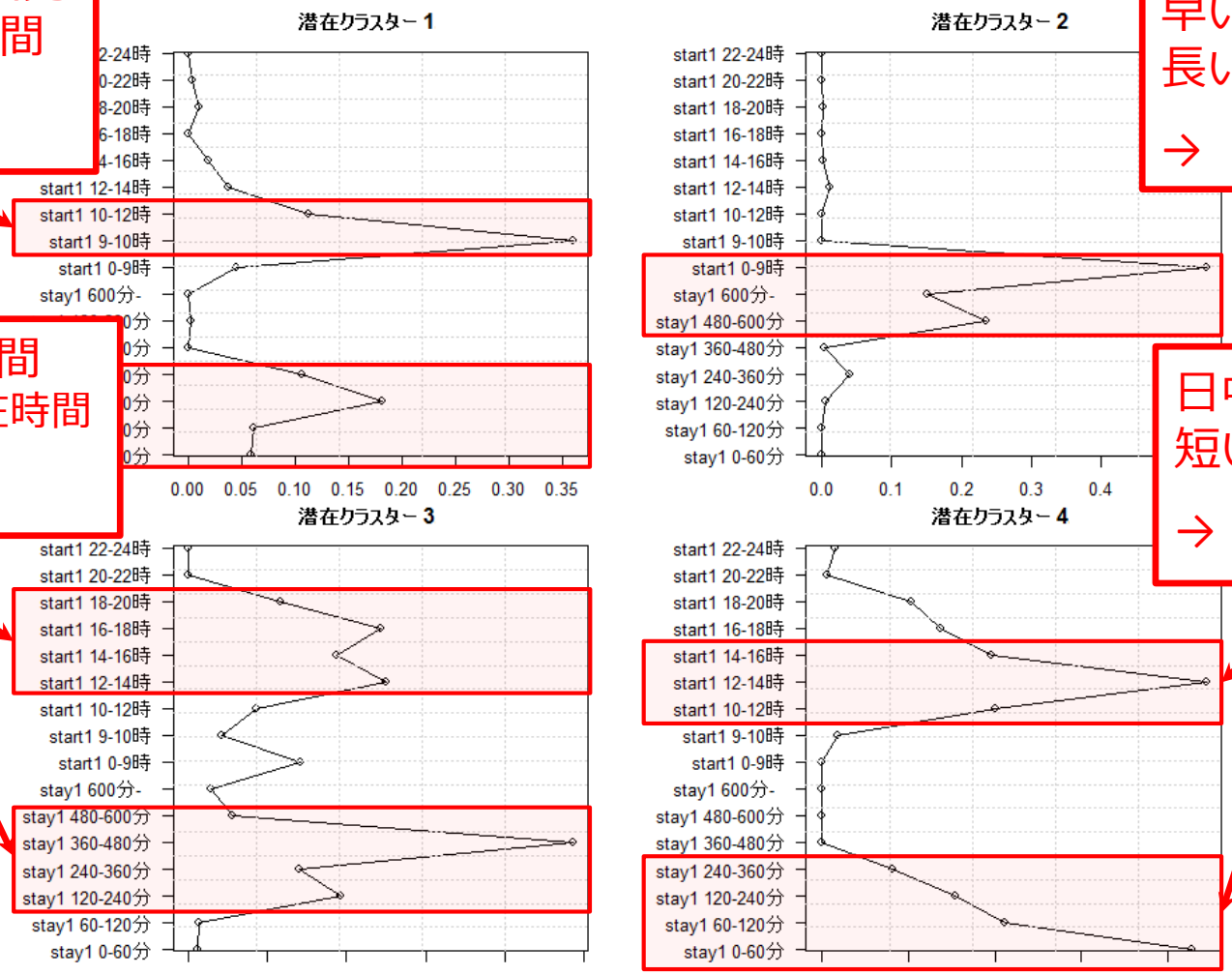
クラスターごとの行動特性
Behavioral characteristics per cluster

9時以降の出発
短い滞在時間
→ **主婦**

遅い出発時間
中程度の滞在時間
→ **学生**

早い出発時間
長い滞在時間
→ **会社員**

日中に出発
短い滞在時間
→ **主婦**



評価　Evaluation

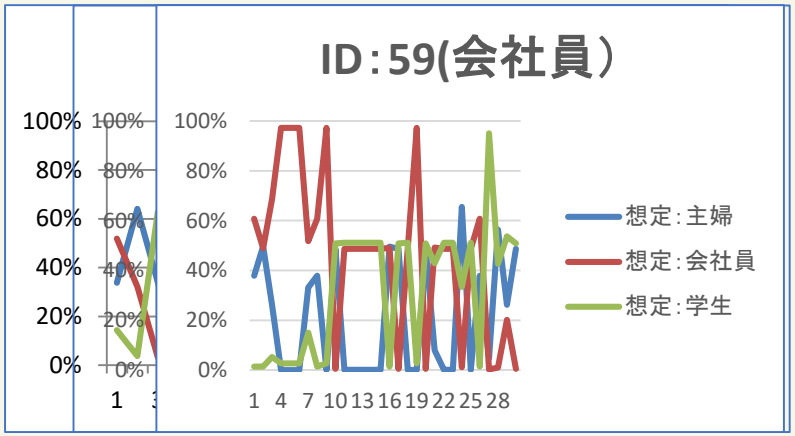
- 推定した属性は約80%の正解率
  The estimated attribute is about 80% correct answer rate.
- 行動パターンが多様な学生や主婦の分類も十分な精度
  Classification of students and housewives with diverse behavior patterns is also sufficient accuracy.

正解率 84.2%	クラスター1 主婦	クラスター2 会社員	クラスター3 学生	クラスター4 主婦
会社員	0	16	2	3
学生	0	0	5	1
主婦	10	1	0	5
パート	0	1	1	0
自営業	0	0	1	1
その他	1	2	0	2

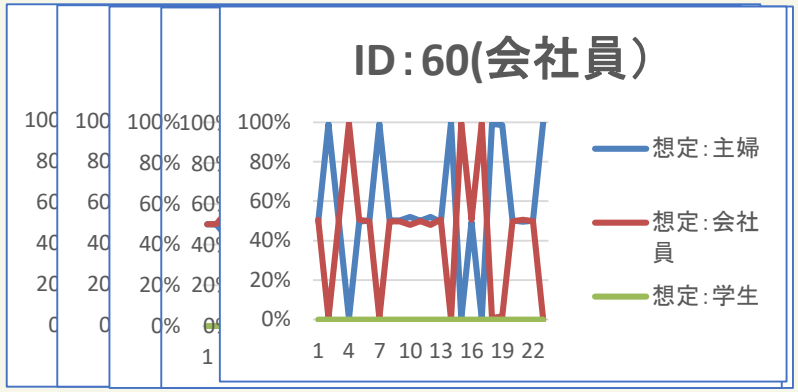
クロスバリテーションでの評価

Day-to-Dayでの属性遷移：会社員

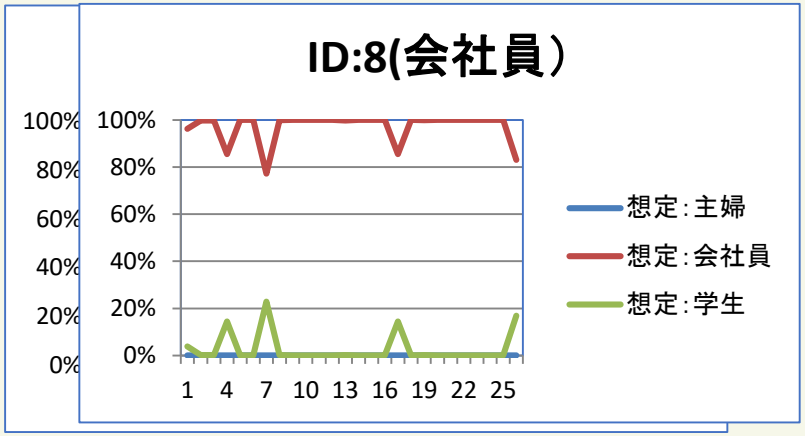
会社員で色々な行動をしている人



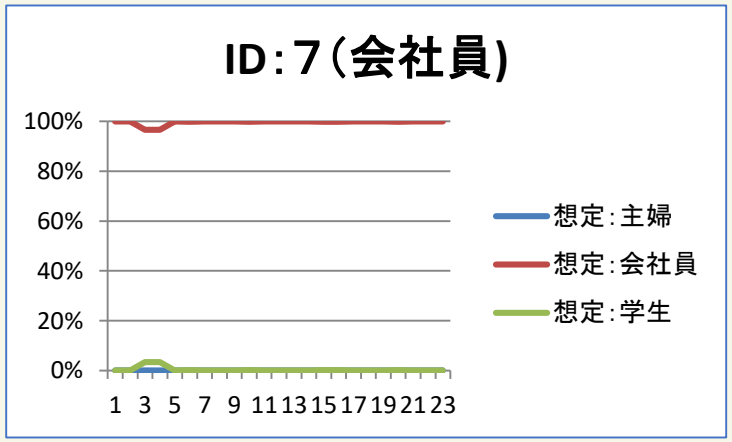
会社員でたまに主婦の行動をしている人



会社員でたまに学生の行動をしている人

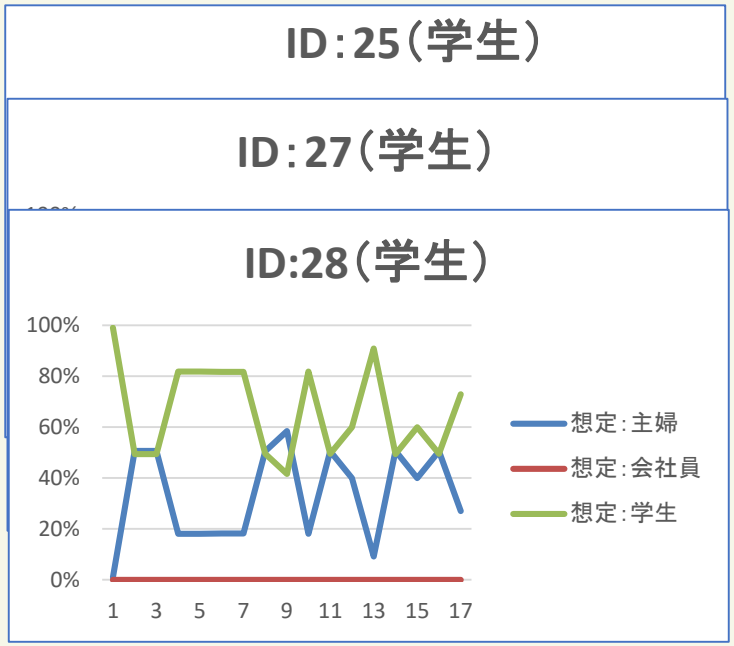


常に会社員の行動をしている人

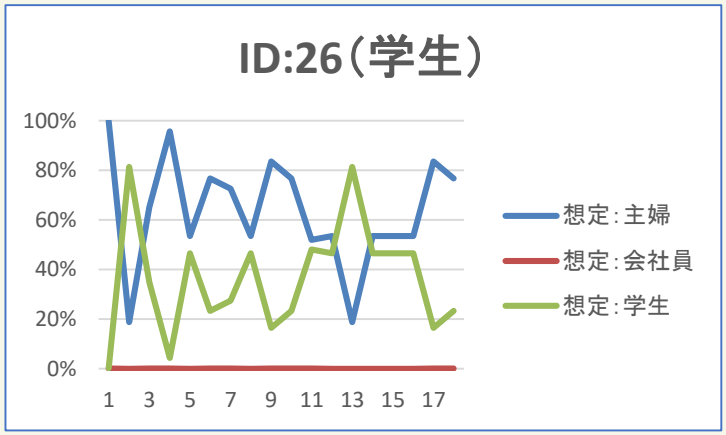


Day-to-Dayでの属性遷移：学生

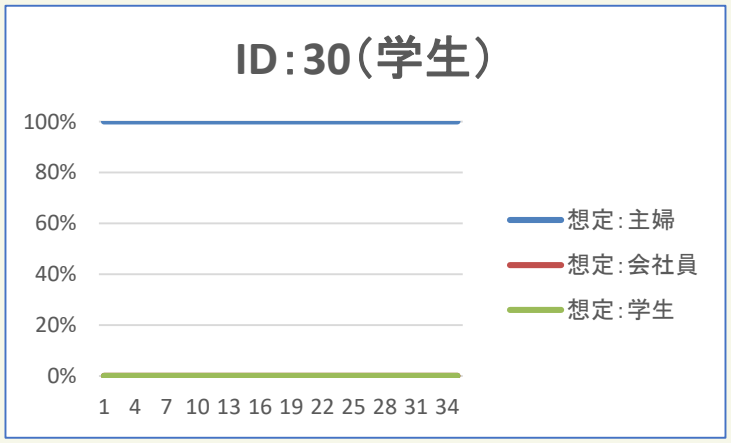
学生の中でたまに主婦の行動をしている人



学生なのにたまに主婦の行動をしている人

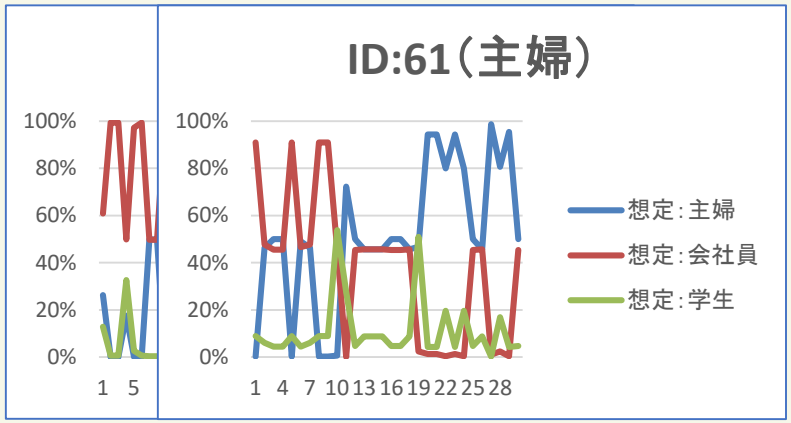


学生なのに主婦の行動を繰り返している人

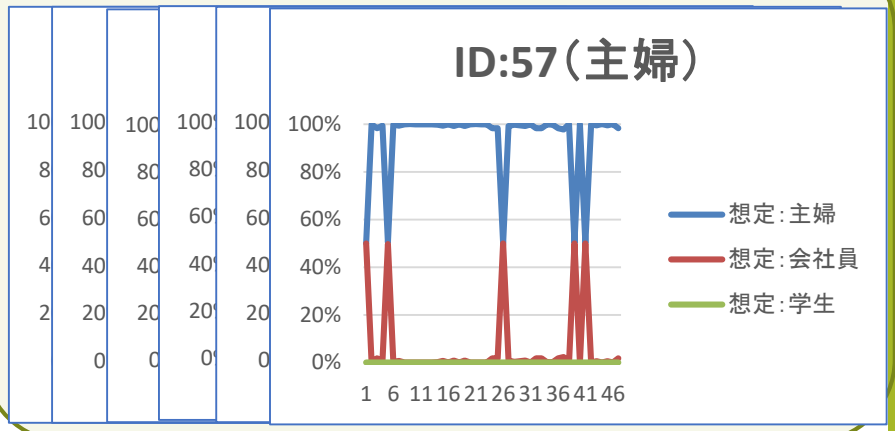


Day-to-Dayでの属性遷移：主婦

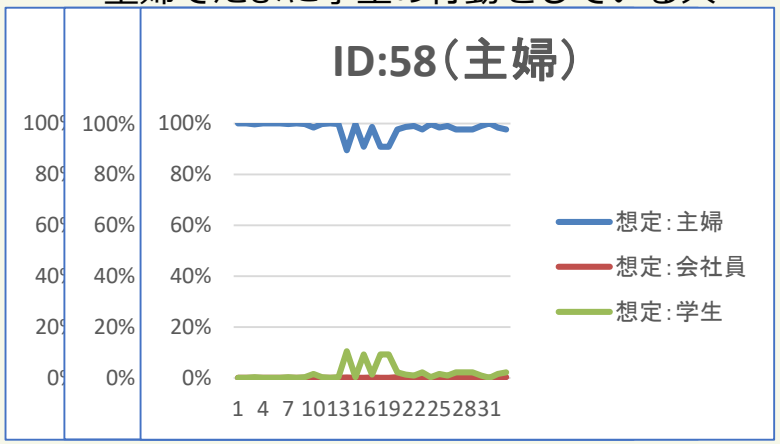
主婦で色々な行動をしている人



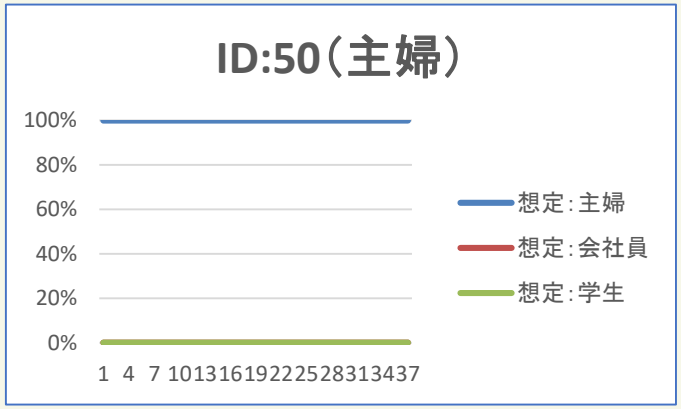
主婦でたまに会社員の行動をしている人



主婦でたまに学生の行動をしている人



主婦行動を繰り返している人



まとめと今後の課題
Summary and Future Challenges

まとめ

- PPデータにより複数日のデータを用いて，PLSIによるクラスター分析を行った.
- PLSIにより分類したクラスの行動を解釈することで，定性的，経験的にそれぞれのクラスの職業を推定した
- 推定した職業は，約80%の正解率
- Day to Dayで個人の属性を確率的に属性を割り当てた