

# 家庭内行動センシングにおける 機械学習データの家庭間転移について

井上創造<sup>1</sup> 潘新程<sup>1</sup> 花沢明俊<sup>1</sup>

概要：本稿では、家庭内に簡易センサを設置して家庭内の行動をセンシングする技術において、既存の家庭で取得した訓練データセットを新たな家庭に適用する際に、家庭間の差異を考慮して認識精度を向上する手法を提案する。異なる家庭の間では、得られるセンサデータの特徴にも違いがあると考えられるため、センサと行動ラベルの訓練データ間、またはセンサのみの入力データとの間において、異なる家庭間でも認識精度を向上するためのデータの転移の考え方を導入する。提案手法では、推定対象家庭の行動ラベルが与えられない場合と与えられる場合の2つを考え、行動ラベルありのデータどうしで交差検証により精度が良くなるサンプルのみを採用する教師あり転移の手法と、行動ラベルありのデータの分布を補正することで行動ラベルなしのデータに分布を近似させる教師なし転移を組み合わせる。あらかじめ、家庭に設置したタブレット端末のセンサデータとスマートメーターの電力データを同時に収集し、そのデータから生活行動を推定し、住人に提示して記録/修正してもらうシステムを開発し、35軒の家庭から約4ヶ月実験を行った。このデータのうち行動ラベルの入力が一定以上の家庭と日について提案手法を評価したところ、両方の場合において精度向上が確認された。

## 1. はじめに

近年重要視されている、家庭内の電力消費の効率化や、高齢者の見守り、健康分野の生活習慣改善のために、家庭内での居住者の行動を認識する需要が高まっている。

我々は、家庭内に、照度センサやスマートメータのような簡易センサを設置して、家庭内の居住者の行動を把握する研究に取り組んでいる。これまでに、簡易センサから行動を推定するアルゴリズムを機械学習により生成するため、生活行動センシング・可視化システムを開発し、実験を行った。

システム開発においては、タブレット端末に搭載されているセンサデータをインターネットを介して収集するタブレット端末アプリと、行動を推定し利用者に提示しながら利用者が編集することで行動ラベルを得ることが出来るWebシステムを開発した。同時に、家庭の分電盤から1分ごとに電力データをサーバに送信する機器を用いた。

開発システムを用いて、35軒の被験者にタブレット端末を貸与し、端末を家庭のよく使用する場所に置いてもらい、そのセンサデータおよび電力データを収集した。行動ラベルについては1日に数分程度行動の入力を依頼した。結果として、約4ヶ月間分、約11,745件の行動入力と約7.14GBのセンサデータと約72,554時間分の消費電力デー

タを得た。

これらのデータをもとに行動推定アルゴリズムを機械学習できるが、その行動推定アルゴリズムを新しい家庭に用いようとした場合、家庭ごとにセンサデータや行動の減少は異なることが考えられるため、良い精度を得られにくい。また、新しい家庭においても上記のようなシステムを用いれば行動ラベルを得ることが出来るが、当初は訓練データサンプルが少ないため、やはり他の家庭のデータを機械学習に用いる必要がある。

そこで本研究では、センサと行動ラベルの訓練データ、またはセンサのみの入力データにおいて、異なる家庭間でも認識精度を向上するためのデータの適応、つまり転移の手法をいくつか提案し、評価を行う。具体的には、推定対象家庭 $h$ の行動ラベルが与えられない場合と与えられる場合に分け、次の2つの問題を考える。

「自家庭ラベルなし転移」 他の家庭の訓練データ  $D_{-h}$  のみが存在し、自分の家庭  $h$  はセンサデータ  $X_h$  のみでラベルデータは存在しない。この状況で、 $X_h$  についての推定精度を最大化する。

「自家庭ラベルあり転移」 システムを運用する中で、 $D_{-h}$  に加えて自分の家庭の訓練データ  $D_h$  も得られ始める。この状況で、新しい入力  $X'_h$  についての推定精度を最大化する。

これらの問題に対して、行動ラベルありのデータどうしで交差検証により精度が良くなるサンプルのみを採用する

<sup>1</sup> 九州工業大学  
〒804-8580, 北九州市戸畑区仙水町 1-1

教師あり転移の手法と、行動ラベルありのデータの分布を補正することで行動ラベルなしのデータに分布を近似させる教師なし転移を提案し、これらを組み合わせることで精度向上できるかを、上記データセットのうち行動ラベルの入力が一定以上の家庭と日について検証した。

その結果、前者において最大 3.13%、後者において最大 2.15% の精度向上が得られた。

## 2. 家庭内行動センシング実験

生活行動を実データから機械学習により見いだすため、我々は生活行動センシングシステム EneAct を開発し、一般家庭において約 4 ヶ月の実験を行った [7], [14]。本節では、本システムの設計および実証実験の内容を述べる。

### 2.1 生活行動センシングシステム EneAct

行動・電力センシングシステム EneAct は、加速度センサ、角速度センサと照度センサデータを継続的に収集するタブレット端末アプリおよび、そのデータから推測された行動情報を遠隔に置かれたサーバに保存し、利用者の Web ブラウザに可視化し、編集することで行動ラベルを記録できる Web システムからなる。

家庭から電力データと行動データおよび行動に関するセンサデータを負担無く収集するためには、設置時および運用時のどちらにおいても利用者の負担を極力少なくすることが求められる。このために、タブレット端末から照度センサデータを継続的に収集するタブレット端末アプリおよび、様々なデータから推測された行動情報を遠隔に置かれたサーバに保存し、利用者の Web ブラウザに可視化・編集できる Web システムを開発した。

#### 2.1.1 Web システム

利用者がシステムを初めて使う場合は、Web サービス上で利用者登録を行う。後述のタブレット端末アプリのホーム画面に配置されたアイコンをクリックするか、毎日利用者に届くメールの本文に記載してある URL から、EneAct の Web サービスにアクセスできる。Web サービスにおいては、利用者の推定された一日の生活行動が表示され、利用者はそれを確認、または修正することで行動ラベルを記録することが出来る。

- **EneAct** へのログイン: 実験参加登録時に作成したアカウントを入力して、ログインすることができる。
- 行動推定・可視化機能: 過去の行動入力とセンサデータを訓練データとし機械学習を一日に一度行い、一日の利用者の行動が推測されて、図 1 の下部のようやすい色で表示される。
- 行動入力機能: 図 1 から行動追加ボタンを押すか、または推定された行動を選んで確定または修正することにより、正しい行動を入力することが出来る。この入力結果は後日の行動推定のための機械学習における教

師行動ラベルとして用いられる。なお、複数人の家庭の場合、ある行動を同居する複数人が行うことも考えられるため、行動を同時に行った人数も記入できる。行動を別々に行った場合には、それらの行動を個人を特定することなく個別に入力してもらった。

- **メール配信機能:** 一日に一度、その日の推定行動と次の日の天気予報をメールで送信し、利用者に行動ラベルの入力を促す。
- **質問機能:** 運用にあたって質問がある場合は、Web システム上の掲示板から実験実施者に質問などのやり取りをすることができる。実験実施者は全利用者に一斉送信できるため、トラブル対応など細かいやり取りを行うことで双方がスムーズに実験を実施できるようにする。

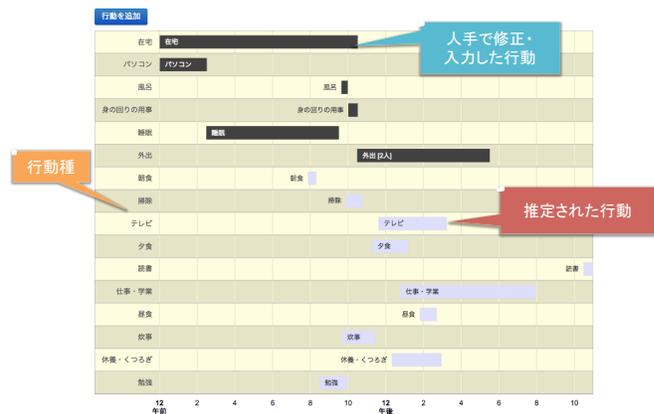


図 1 EneAct アクティビティ画面

#### 2.1.2 タブレット端末アプリ

タブレット端末上のアプリにおいては、上記の要件を満たすために、次のような設計を行った。

- (1) 利用者の認証の手間を出来るだけ減らす。  
利用者認証は後述の Web システムに HTTPS プロトコルによって行い、一度ログインした後はアプリのクッキーおよびファイルにセッション情報を保存しておくことで、次回起動時にも自動的にログインできるようにする。
- (2) 端末がどのような状態でもデータ収集を続ける。  
利用者が別のアプリを使っている場合、画面を消している場合でも、データ収集を続けるため、センシング中はバックグラウンド実行状態に入り、動作を続けるようにする。また、不意にアプリが停止したり、端末を再起動しても、自動的にアプリが起動するように、OS の提供する API を通じて設定した。
- (3) ネットワーク状態がどのような状態でもデータを紛失しない。

センサデータは毎分ごとに HTTP プロトコルでサーバにアップロードされるが、ネットワーク接続が中断してもアプリが不意に停止しても、データを紛失してはならない。また、データが何度も重複してアップロードされることも避けなければならない。このため、実行スレッドを、センシングするスレッドとサーバに送信するスレッドの2つに分け、前者は端末内のファイルにデータを追記し、後者はそのファイルから FIFO (First In First Out) 方式で、既に、先に取得されたデータを先に処理してサーバに送信し、後から取得されたデータは先に取得されたデータより後に処理して送信する。データを読み出しサーバに接続し、データ送信に成功した時のみファイルの既読部分を削除する動作を、並行して実行させることにした。この動作を、図2に示した計測画面のように設計した。



図3 ユビキタス社 NaviEne Master の設置の様子



図2 タブレット端末アプリの計測画面

## 2.2 実験

生活行動と消費電力の関係を調べるために実際に 35 軒の家庭で、システムを利用してもらいデータの取得を行った。2014 年 12 月 5 日から約 4 ヶ月の期間、一般家庭から被験者を募って、被験者にはタブレット端末を貸与し、家庭のよく使用する場所に図4に示したように置いてもらい、1日に数分程度行動の入力をするよう依頼して、行動ラベルデータとしてサーバに保存した。なお、電力データの収集については、家庭の分電盤に設置し一分に一度消費電力を無線 LAN 経由でサーバに送ることの出来るユビキタス社の Navi-Ene Master を用いた。図3に装置を示す。

## 2.3 結果

2015 年 3 月 26 日の時点で約 4 ヶ月間分、35 軒の家庭から約 11,745 件の行動入力と約 7.14GB のセンサデータと約 72,554 時間分の消費電力データを得ることができた。本論文ではこれらのうち、2015 年 3 月 26 日の時点で約 4 ヶ



図4 タブレット端末の置き方 (画面ははめ込み)

月間分、23 軒の家庭から 10,250 件の行動入力と約 39,480 時間分の消費電力データを用いて、実験分析を行った。入力された行動ラベルは 52 種類であったが、2 家庭以上が入力したものは 40 種類、3 家庭以上が入力したものは 35 種類であった。

## 2.4 課題

この実験で得られた照度および消費電力データは行動ラベルを同時に持つため、このデータを用いて教師あり機械学習を行って推定モデルを構築することで、新しい家庭の行動推定に用いることが出来る。

しかし、現実には、家庭ごとに電力の消費や照度のパターンは異なり、ある家庭集合で訓練した推定モデルを新しい家庭にそのまま適用するのでは、良い精度を得られないことが考えられる。

また、新しい家庭においては場合によって行動ラベルを得られることもある、しかしその場合でも、当初は訓練データサンプルが少ないため、他の家庭のデータと組み合

わせて機械学習を行う必要がある。

### 3. 提案手法

本節では、複数の家庭の訓練データが与えられたときに、家庭間でそれらのデータを教師ありまたはなしで転移する手法を述べる。

#### 3.1 アプローチ

提案手法では、自分の家庭の訓練データが得られる前と、得られた後で手法を切り替える。自分の家庭の訓練データは、転移をせずにそのまま自分の家庭の訓練データとして用いることが出来ると考えられるため、自分の家庭の訓練データが得られた後は、自分の家庭の訓練データを相対的に信頼できるためである。

それぞれ、以下のようなアプローチをとる。

「自家庭ラベルなし転移」 他の家庭の訓練データ  $D_{-h}$  のみが存在し、自分の家庭  $h$  は評価用入力データ  $X_h^{new}$  のみでラベルデータは存在しない場合は、まず他の家庭  $D_{-h}$  と自分の家庭  $X_h^{new}$  の間で教師なし転移 **UT** を行って訓練データを更新する。次にその訓練データ内の家庭どうして相互に教師あり転移 **ST** を行い、訓練データを更新する。

「自家庭ラベルあり転移」 システムを運用する中で、 $D_{-h}$  に加えて自分の家庭のラベルありデータ  $D_h$  も得られ始めた後は、まず他の家庭  $D_{-h}$  と自分の家庭の評価用データ  $X_h^{new}$  の間で教師なし転移 **UT** を行って訓練データを更新する。次にその訓練データと自分の家庭のラベルありデータ  $D_h$  との間で教師あり転移 **ST** を行って再度訓練データを更新する。それに自分の家庭  $D_h$  を加えて訓練データとする。

提案手法の概要を、図5に示す。

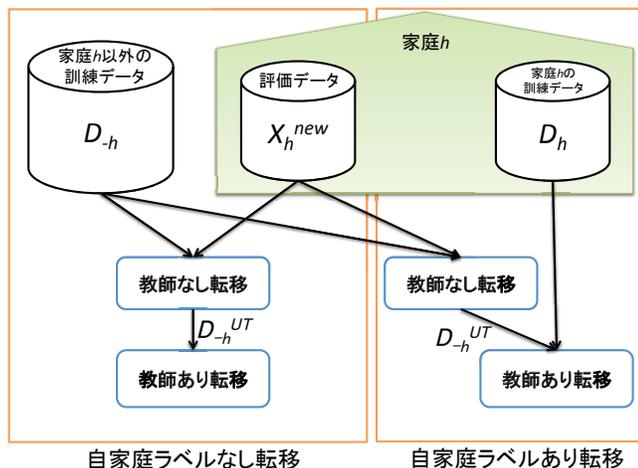


図5 提案手法の概観

表1 表記のまとめ

Symbol	Summary
$X \subseteq \mathcal{X}$	入力サンプルの集合.
$Y \subseteq \mathcal{Y}$	行動ラベルサンプルの集合.
$\ X\ $	集合 $X$ のサンプル数.
$D = (X, Y) \subseteq \mathcal{D}$	訓練データ, ただし $\ X\  = \ Y\ $ .
$H \subseteq \mathcal{H}$	家庭の集合.
$D_h = (X_h, Y_h) \subseteq \mathcal{D}$	家庭 $h \in H$ の訓練データ
$D_{-h} = D - D_h$	家庭 $h \in H$ 以外の訓練データ
$\psi(D_1, D_2)$	データ $D_1$ でモデルを訓練し, データ $D_2$ で評価したときの精度.
$\mathbf{ST}(D_1, D_2) \subseteq \mathcal{D}$	データ $D_1$ をデータ $D_2$ に教師あり転移した結果のデータセット
$\mathbf{UT}(D, X) \subseteq \mathcal{D}$	データ $D$ をデータ $X$ に教師なし転移した結果のデータセット

#### 3.2 準備

準備として、本稿全体で扱う数学表記を述べる。表1にまとめを記載する。

入力データサンプルを  $X$  とする。この各要素  $x \in X$  は、特定の家庭の特定の時刻前後のセンサデータから計算された特徴量ベクトルである。各入力データサンプルには、行動ラベル  $y \in Y$  が対応する。特に断らない場合は、 $X$  と  $Y$  のサンプル数  $\|X\|$  と  $\|Y\|$  は等しいとする。行動推定のための訓練データは、入力とそれに対応する正解行動のペアなので、 $D = (X, Y)$  と表される。

取り扱う家庭の集合を  $H$  とする。データ  $D$  のうち、ある家庭  $h \in H$  についての訓練データを  $D_h = (X_h, Y_h)$  とする。また家庭  $h$  以外の訓練データを  $D_{-h}$  と表記する。

通常の行動推定においては、訓練データ  $D_1$  で推定モデルを構築し、その推定モデルに評価用データ  $D_2$  の入力  $X_2$  を入力し推定した行動を、 $Y_2$  と比較することで精度を計算する。このように交差検証をした結果の精度を、 $\psi(D_1, D_2)$  と表す。

#### 3.3 転移のための基本アルゴリズム

以下では、自家庭ラベルなし転移または自家庭ラベルあり転移の両方で用いる、基本アルゴリズムを提案する。

##### 3.3.1 教師あり転移アルゴリズム **ST**

訓練データ  $D$  が与えられたときに、このうち全てのサンプルを利用することも考えられるが、各家庭に特化されたサンプルが混ざっていけば、それは他の家庭に用いた際には精度を低下させてしまう。そこで、訓練データ  $D$  内の家庭間で交差検証を行い、そこで精度が良いサンプルのみを使うことにする。

そのための基本アルゴリズム **ST** を、Algorithm 1 に示す。ここでは、訓練データ  $D$  内から  $k$  回サンプリングを行い、それを訓練後、転移先データ  $D_T$  で評価し、精度が良い場合のサンプルだけを採用する。

このアルゴリズムをある家庭  $h \in H$  とそれ以外の家庭

---

**Algorithm 1** 教師あり転移  $\mathbf{ST}(D, D_T)$

---

**Input:** 転移元データ  $D \in \mathcal{D}$ , 転移先データ  $D_T \in \mathcal{D}$ , パラメータ  
 $k \geq l \geq 1, m > 1$

**Output:** 転移後のデータセット  $D^* \in \mathcal{D}$

- 1: **for**  $i = 1$  **to**  $k$  **do**
  - 2:  $D$  から  $m$  点サンプリングする. これを  $D^i$  とする.
  - 3:  $D^i$  を訓練データとしたときの精度  $\psi(D^i, D_T)$  を測る. これを  $\psi^i$  と呼ぶ.
  - 4: **end for**
  - 5:  $D^1, \dots, D^k$  のうち, 精度  $\psi^i$  が良い順に  $l$  個のサンプル  $D^i$  を採用し, これらの和集合を  $D^* \subseteq D$  とする.
  - 6: **return**  $D^*$
- 

---

**Algorithm 2** 教師なし転移  $\mathbf{UT}(D_H, X)$

---

**Input:** 家庭集合  $H$  に対応する転移元データ  $D_H$ , 転移先入力データ  $X$

**Output:** 転移後のデータセット  $D_H^*$

- 1: サンプル  $x \in X$  についての尤度  $P(x)$  の統計パラメータ  $\theta_x$  を求める.
  - 2: **for all** 家庭  $h \in H$  **do**
  - 3:  $D$  のうち家庭  $h$  に対応した入力  $x_h \in X_h$  の尤度  $P(x_h)$  の統計パラメータ  $\theta_{x_h}$  を,  $\theta_x$  に近づける. 具体的には,  $P(x)$  と  $P(x_h)$  の各変数を正規分布とみなし, 平均値と標準偏差が同一になるように標準化する. この結果の入力データを  $X_h^*$  とする.
  - 4: **end for**
  - 5: すべての  $X_h^*$  の和集合を  $X_H^*$  とする.
  - 6:  $X_H^*$  にラベルを付加したデータを,  $D_H^* \subseteq D_H$  とする.
  - 7: **return**  $D_H^*$
- 

の間で適用し, その和集合, つまり

$$\bigcup_{h \in H} \mathbf{ST}(D_h, D_{-h})$$

を新たな訓練データとすることが出来る.

ここでは交差検証の考え方をういたが, 4節の評価の際はもちろん, これらの精度を評価するために別途, 評価用の家庭を用意し交差検証を行うこととする.

### 3.3.2 教師なし転移アルゴリズム $\mathbf{UT}$

上記の方法だと, 教師ありデータ間の転移はできるが, 評価用の入力データが与えられたときにそれへの転移は出来ない. 評価用データはもちろん教師なしデータであるが, このようなラベルなしデータとの間でデータを適用するアルゴリズム  $\mathbf{UT}$  を, Algorithm 2 に示す.

$\mathbf{UT}$  では, 転移元入力データの家庭ごとの分布を, 転移先入力データの分布に近づけることを試みる. このために, 転移先データの統計パラメータ  $\theta_x$  を求めて, 転移元データを  $\theta_x$  に標準化することで, 転移元の入力を転移先の分布に近似させる. 今回は簡単のために, 各変数を単変量正規分布と見なして, 平均値と標準偏差について標準化する.

このアルゴリズムを用いて, 家庭  $h$  以外の家庭についての訓練データ  $D_{-h}$  および家庭  $h$  についての入力データ  $X_h$  が与えられたときに,

$$\mathbf{UT}(D_{-h}, X_h)$$

を新たな訓練データとすることが出来る.

### 3.4 家庭間転移

ここでは, 上記の転移手法を組み合わせると, システム運用当初の対象家庭の訓練データが無い場合に自家庭ラベルなし転移をする方法と, そのうちに対象家庭から教師ラベルが得られて訓練データがある場合に自家庭ラベルあり転移を行う方法を述べる.

#### 3.4.1 自家庭ラベルなし転移

家庭  $h$  以外の家庭  $H'$  についての訓練データ  $D_{-h}$  および家庭  $h$  についての評価用入力データ  $X_h^{new}$  が与えられたときに, 以下の処理を行う.

- (1) まず訓練データ  $D_{-h}$  を,

$$D^U \leftarrow \mathbf{UT}(D_{-h}, X_h^{new})$$

により  $X_h^{new}$  に転移させる.

- (2) 次に  $D^U$  の訓練データの家庭間の汎化性能を考慮した訓練データ

$$D^S \leftarrow \bigcup_{h' \in H'} \mathbf{ST}(D_{h'}^U, D_{-h'}^U)$$

を得る. これを新たな訓練データとして, 行動推定モデルを機械学習する.

#### 3.4.2 自家庭ラベルあり転移

対象家庭の教師ありデータが得られる際には, 下記の手法に切り替える.

家庭  $h$  以外の家庭  $H'$  についての訓練データ  $D_{-h}$  および家庭  $h$  についての入力データ  $X_h^{new}$  に加えて, 家庭  $h$  についての教師ラベルありデータ  $D_h$  が与えられたときに, 以下の処理を行う.

- (1) まず訓練データ  $D_{-h}$  を,

$$D^U \leftarrow \mathbf{UT}(D_{-h}, X_h^{new})$$

により  $X_h^{new}$  に転移させる.

- (2) 次に  $D^U$  を家庭  $h$  に転移したデータおよび  $D_h$  の両方

$$D^S \leftarrow \mathbf{ST}(D^U, D_h) \cup D_h$$

を得る. これを新たな訓練データとして, モデルを作成する.

## 4. 評価

本節では, 2節で得られたデータに3節の手法を適用し, その精度を評価する.

具体的には, 以下の点を評価する.

- (1) 他の家庭の訓練データ  $D_{-h}$  のみか使える自家庭ラベルなし転移において, アルゴリズム  $\mathbf{UT}$  および  $\mathbf{ST}$  の適用が, 推定精度を向上させるか?
- (2)  $D_{-h}$  に加えて自分の家庭の訓練データ  $D_h$  も得られ

ている自家庭ラベルあり転移において、アルゴリズム **UT** および **ST** の適用が、推定精度を向上させるか？ またそれらが、1 の  $D_h$  のみを用いた訓練に比べて精度を向上させるか？

#### 4.1 前処理

データセットとしては、2 節で得られた照度、電力消費データおよび、ユーザがつけた行動ラベルを用いた。

限られた家庭でしか行われないう行動クラスについては、特に上記の **ST** アルゴリズムは適用出来ない。このため、あらかじめ 3 家庭以上が行っていない行動クラスは除外した。また、各ユーザに対して、行動クラスが 8 種類以上記録されている日のみを用いた。その結果、7 家庭ののべ 41 日、15 行動クラスのデータセットとなった。

これらのデータセットから特徴量を計算した。特徴量は、1 分ごとに次の値を計算した。

- 一日の中の時刻 (0-24 時 × 60 分)
- 照度の平均値、標準偏差、最大値、最小値
- 一分ごとに得られる消費電力値

この結果、50,098 サンプルの多次元データとなった。

#### 4.2 提案手法の適用

各アルゴリズムの中で用いる基本的な機械学習としては、決定木を用いた。

**ST** アルゴリズムにおいては、パラメータを  $k = 50$ ,  $l = 10$ ,  $m = 10$  とした。ただし、自家庭ラベルなし転移においては簡単のために、 $D_{-h}$  中の家庭ごとに **ST** を適用するのではなく、家庭  $h'$  とその中のサンプルを  $k$  回サンプリングし、それらをまとめて精度の良い  $l$  回を採用した。

評価にあたっては、交差検証を厳密に行うために、

- 自家庭ラベルなし転移においては 1-family-left-out 交差検証、つまり「1 家庭を評価データとし、それ以外の家庭の全データを教師なし転移および訓練に用いる」ことを全家庭について繰り返し、
- 自家庭ラベルあり転移においては、「1 家庭の前半のデータとそれ以外の家庭の全データを教師あり転移および訓練に用いる」ことを同様に全家庭について繰り返す

ことを行った。

なお、各検証において、正例と負例が同数になるように、各例からブートストラップ法によりサンプリングを行った。また、自家庭ラベルあり転移においては、自分の家庭の訓練データ  $D_h$  を一日だけ使うことにしたが、この日にある行動を行っていないと、**ST** による転移が難しいため、どの日を訓練データに使うかは、行動によってその行動数が最も多い日を  $D_h$  とみなした。

#### 4.3 評価尺度

通常の行動推定の評価においては、適合率や再現率、その調和平均である F-measure が用いられることが多い。しかし今回のデータにおいてはこの尺度は以下の理由で用いにくい。

- 「在宅」と「テレビ」のように、一時刻に対して複数の行動ラベルが付加されることを許している。
- 適合率と F-measure は、均一でない行動ラベルサンプルに影響されてしまう。これらの尺度はもともと数が少ない正例に対して敏感になるため、「睡眠」のようなもともと長時間の行動は見た目の精度が高く、逆に短時間の行動に対しては低くなる傾向にある。

このため、行動ごとに精度を計算でき、かつ不均一な正例と負例のサンプル数にも影響されない指標である BCR (Balanced Classification Rate) を用いることにする。この指標は以下のように定義される。

$$BCR = \frac{TP\text{-rate} + TN\text{-rate}}{2}$$

ただし、 $TP$ ,  $FP$ ,  $TN$ ,  $FN$  をそれぞれ True Positive, False Positive, True Negative, False Negative のサンプル数とした時に、 $TP\text{-rate}$  は  $TP/(FN + TP)$ ,  $TN\text{-rate}$  は  $TN/(TN + FP)$  である。 $TP\text{-rate}$  も  $TN\text{-rate}$  も分母はそれぞれ正例と負例のサンプル数となるため、どちらも正例と負例の不均一に影響されない。その平均である BCR も同様である。

#### 4.4 結果

上記の方針に基づいた評価結果を、図 6 と図 7 に示す。

図 6 は、自家庭ラベルなし転移の比較結果である。比較として、ランダムフォレストを単純に適用した方法 (RandomForest), 決定木を単純に適用した方法 (決定木), 提案手法のうち **UT** アルゴリズムのみを適用した方法 (**UT**), 提案手法のうち **ST** アルゴリズムのみを適用した方法 (**ST**), 提案手法 **UT** と **ST** を両方適用した方法 (**UT + ST**) を対象とした。

結果の精度は行動によって大分異なるが、右端の平均の精度を見ると、RandomForest 54.48%, 決定木 64.46%, **UT** 65.59%, **ST** 66.15%, **UT + ST** 67.59% となり、徐々に精度が上がっている。

図 6 は、自家庭ラベルあり転移の比較結果である。比較として、自家庭ラベルなし転移と同様に、自分の家庭の訓練データ  $D_h$  のみ (自家庭), 他の家庭の訓練データ  $D_{-h}$  を加えた場合 (自家庭+他家庭), 他の家庭に **UT** を適用した場合 (自家庭+他家庭 **UT**), 他の家庭に **ST** を適用した場合 (自家庭+他家庭 **ST**), 他の家庭に **UT** と **ST** の両方を適用した場合 (自家庭+他家庭 **UT + ST**) を対象とした。

こちらも結果の精度は行動によって異なるが、右端の平均の精度を見ると、自家庭 66.79%, 自家庭+他家庭 68.05%,

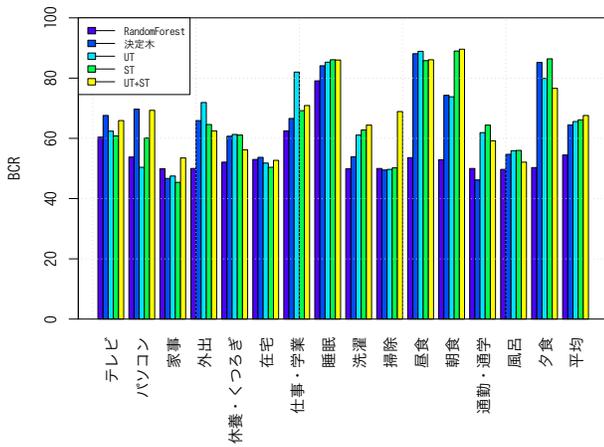


図 6 自家庭ラベルなし転移の精度比較

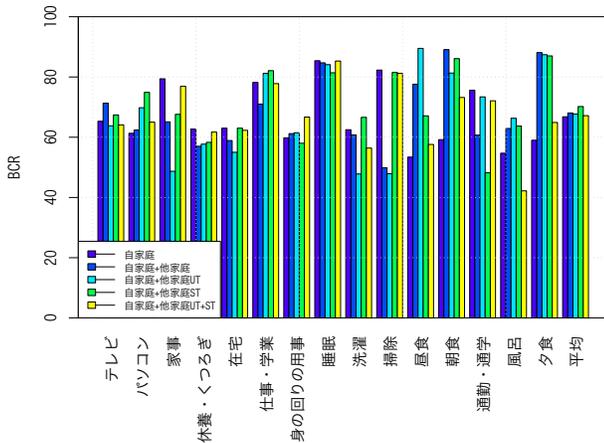


図 7 自家庭ラベルあり転移の精度比較

自家庭+他家庭 UT 67.69%, 自家庭+他家庭 ST 70.19%, 自家庭+他家庭 UT + ST 67.16%となった。

#### 4.5 考察

図 6 においても図 7 においても、全般に精度が高い行動は、「睡眠」「夕食」である。照度センサがリビングやダイニングなどのよく使う場所に置かれたこと、夜は暗くなることからこれらは普遍的に推定可能であることが分かる。また「昼食」「朝食」は自家庭ラベルなし転移においては精度が高い傾向にあるが、自家庭ラベルあり転移においては必ずしもそうではない。後者についても前者と同じ処理は出来るはずなので、改善の余地がある。

また、図 6 において精度が低い行動は、「家事」「在宅」「通勤・通学」「風呂」であるが、これらは図 7 においても精度は低いものの、改善の傾向が見られている。このような、照度センサや電力使用量と容易に関連づけにくい行動について、精度が向上していることが分かる。

図 6 の自家庭ラベルなし転移においては、UT+ST, ST, UT, 決定木, RandomForest の順に精度が良く、UT+ST と決定木の差は 3.13%であった。BCR の指標を用いると微小ではあるが精度が向上していることが分かる。なお、

F-measure を用いると決定木は 48.27%に対して UT + ST は 62.21%となっており、13%の向上になっている。このように見える主な原因は、BCR の指標に用いられる TN-rate が 10.51% 減少しており、提案手法では本当は行われていない行動を誤って肯定してしまう (False Positive) が増えることが分かる。

RandomForest はアンサンブル学習を行っており、通常は決定木学習より良い精度が出ることが多いが、今回は逆の結果となっている。この理由に付いては今後検証が必要だが、今回は特徴量ベクトルが 6 次元と高次元ではなかったため、RandomForest におけるランダムな変数選択が効かなかったこと、転移を考慮しないランダムサンプリングにおいても過学習を起こしてしまっている可能性がある。

図 7 の自家庭ラベルあり転移においては、全体として自家庭ラベルなし転移より精度が上がる傾向にあるものの、ST を用いた場合は精度が自家庭+他家庭に比べて 2.05%上がっているが、UT を使った 2 つの手法に置いては、自家庭+他家庭に比べて精度が下がっている。これについては、自分の家庭の訓練データが一日程度では、教師なし転移 UT が逆効果になっている可能性が考えられる。自家庭の訓練データがどの程度あれば最適になるのか、今後検証が必要である。

## 5. 関連研究

家庭内での行動センシングは文献 [9] をはじめ多く提案されている。特にライフラインの使用状況をセンシングして行動を認識する研究は、ガス [1], 水道 [10], [3] のように種々の入力を用いて試みられている。我々の研究はこれらのように、専用のセンサを用いずに家庭内の行動を認識するという研究の延長線上にあり、絶対的な精度は未だ高くないかもしれないが、複数の家庭の訓練データを用いて精度を向上させることを目指している。

我々の研究 [7], [14] では、本論文と同じ実験で得られたデータについて、行動ラベルと電力消費の相関を解析している。また文献 [11] では、このデータをもとに行動が電力需要予測にどれだけ影響するかを調べている。本論文ではこのデータを行動推定における家庭間の転移に用いた。

転移学習については、文献 [12] や [6] が網羅的に解説している。これに関連して、提案する教師あり転移アルゴリズム ST に類似する既存手法としては、TrBagging[4] や文献 [5] がある。TrBagging[4] においては、転移元データから複数回ブートストラップサンプリングを行い弱分類器を構成し、転移先データにおいて交差検証を行い精度が良い弱分類器のみを残し多数決をとる。通常のバギングに対して、転移元と先で交差検証を行い弱分類器をふるいにかける点が新しい。文献 [5] においては、TrBagging のような転移の過程を、弱分類器の重要度を定める選択のフェーズと、生き残った分類器に使われたサンプルについて分類器を生

成する訓練フェーズに分けることで、精度を最大6%程度向上している。我々のST手法は、この手法と同様の考え方を、自家庭ラベルあり転移と自家庭ラベルなし転移という異なる設定において利用し、特に自家庭ラベルなし転移においては訓練データの複数の家庭間で適用して精度を検証している。

一方、我々の教師なし転移アルゴリズムUTに類似する既存手法としては、共変量シフト下での教師あり学習[8][13]があげられる。この手法は、入力データの分布が訓練時と評価時で異なる共変量シフト状況において、転移先の入力データの分布を重点サンプリングにより転移元に近似させている。ただし、サンプル数によっては精度が向上しにくいことも報告されている。我々の提案手法は、入力データの分布が単変量の正規分布と見なし、その平均と標準偏差を転移元から転移先に近似するという単純な手法である。我々は重点サンプリングによる手法も試したが、精度はあがらなかった。今後さらなる検証が必要である。

## 6. まとめ

本稿では、家庭内に簡易センサを設置して家庭内の行動をセンシングする技術において、既存の家庭で取得した教師ありデータセットを新たな家庭に適用する際に、家庭間の差異を考慮して認識精度を向上する手法を提案した。センサと行動ラベルの訓練データ間、またはセンサのみの入力データとの間において、異なる家庭間でも認識精度を向上するためのデータの転移の考え方を導入する。推定対象家庭の行動ラベルが与えられない場合と与えられる場合の2つを考え、それぞれにおいて教師あり転移の手法と教師なし転移を組み合わせた。実際に我々が取得した約5万サンプルの照度と消費電力データについて適用し、評価したところ、両方の場合において精度向上が確認された。

今後の課題としては、よりよい特徴量の探索や、他に多く提案されている転移アルゴリズムの探索、そして家庭によって行動ラベルの種類にも違いがあるため、この違いを吸収、利用する手法を検討することが考えられる。

## 謝辞

本研究の一部は、CREST分散協調型エネルギー管理システム構築のための理論及び基盤技術の創出と融合展開「需要家の行動変容に影響を与える要因に関する基礎的研究(研究代表者:日高一義)」および基盤研究(B)「物理層と意味層の2階層からなるセンサコンテキスト推定技術(研究代表者:井上創造)」による。

## 参考文献

[1] Cohn, G., Gupta, S., Froehlich, J., Larson, E. and Patel, S. N.: GasSense: Appliance-level, single-point sensing of gas activity in the home, *Lecture Notes in*

*Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 6030 LNCS, pp. 265–282 (online), DOI: 10.1007/978-3-642-12654-3\_16 (2010).

- [2] Fawcett, T.: An introduction to ROC analysis, *Pattern Recognition Letters*, Vol. 27, pp. 861–874 (online), DOI: 10.1016/j.patrec.2005.10.010 (2006).
- [3] Gupta, S., Reynolds, M. and Patel, S.: ElectriSense: single-point sensing using EMI for electrical event detection and classification in the home, *Proceedings of the 12th ACM international conference on Ubiquitous computing*, pp. 139–148 (online), DOI: 10.1145/1864349.1864375 (2010).
- [4] Kamishima, T., Hamasaki, M. and Akaho, S.: TrBagg: A simple transfer learning method and its application to personalization in collaborative tagging, *Proceedings - IEEE International Conference on Data Mining, ICDM*, pp. 219–228 (online), DOI: 10.1109/ICDM.2009.9 (2009).
- [5] Lin, D., An, X. and Zhang, J.: Double-bootstrapping source data selection for instance-based transfer learning, *Pattern Recognition Letters*, Vol. 34, No. 11, pp. 1279–1285 (online), DOI: 10.1016/j.patrec.2013.04.012 (2013).
- [6] Pan, S. J. and Yang, Q.: A Survey on Transfer Learning, *IEEE Transactions on Knowledge and Data Engineering*, Vol. 22, No. 10, pp. 1345–1359 (online), DOI: 10.1109/TKDE.2009.191 (2010).
- [7] Pan, X., Minezaki, T., Isoda, T., Tanaka, S., Uchino, Y. and Inoue, S.: Analyzing Daily-life Activities and Power Consumptions Using Tablet Sensors and Activity Annotation Web System, *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers, UbiComp/ISWC'15 Adjunct*, New York, NY, USA, ACM, pp. 1443–1452 (online), DOI: 10.1145/2800835.2801615 (2015).
- [8] Shimodaira, H.: Improving predictive inference under covariate shift by weighting the log-likelihood function, *Journal of Statistical Planning and Inference*, Vol. 90, No. 2, pp. 227–244 (online), DOI: 10.1016/S0378-3758(00)00115-4 (2000).
- [9] Tapia, E. M., Intille, S. S. and Larson, K.: Activity Recognition in the Home Using Simple and Ubiquitous Sensors, *Pervasive Computing*, Vol. 3001, pp. 158–175 (online), DOI: 10.1007/b96922 (2004).
- [10] Vu, T. T., Sokan, A., Nakajo, H., Fujinami, K., Suutala, J., Siirtola, P., Alasalmi, T., Pitkänen, A. and Roning, J.: Detecting water waste activities for water-efficient living, *13th International Conference on Ubiquitous Computing, {UbiComp'11} and the Co-located Workshops*, pp. 579–580 (2011).
- [11] Yong Ding, Sozo Inoue, Martin Alexander Neumann, Erwin Stamm, Xincheng Pan, M. B.: A Personalized Load Forecasting Enhanced by Activity Information, *IEEE International Smart Cities Conference (ISC2-2015)*, Guadalajara, Mexico (2015).
- [12] 神島敏弘: 転移学習, 人工知能学会誌, Vol. 25, No. 4, pp. 572–580 (2010).
- [13] 杉山将: 共変量シフト下での教師付き学習, 日本神経回路学会誌, Vol. 13, No. 3, pp. 111–118 (2006).
- [14] 潘新程, 峯崎智裕, 磯田達也, 田中翔太, 内野百里, 井上創造, 九州工業大学, 九州工業大学, 九州工業大学, 九州工業大学, Others: タブレット端末センサと行動入力 Web システムを用いた生活行動と消費電力の分析 (2015).