

家庭用スマートメータへの適用を目的とする 世帯の在宅推定手法の開発

大島 悠^{1,a)} 石曾根 毅^{2,b)} 樋口 知之^{1,c)}

受付日 2021年12月2日, 再受付日 2022年2月14日,

採録日 2022年2月25日

概要：近年、スマートメータからの電力需要値を用いた在宅推定の需要が高まっている。時刻ごとの世帯の在宅状況を把握することは、災害時の避難計画の策定、配送業と営業活動の経路最適化などといった応用可能性がある。先行研究として、15分や60分間隔の電力需要値の平均値や標準偏差といった説明変数を用いて、在不在を判別する教師あり学習の2値判別モデルを構築することが行われてきた。そこで本稿では新たな教師あり学習の判別モデルを構築した。先行研究との違いとしては2つある。1つ目は、国内のスマートメータの規格に従い、測定時間頻度が30分間隔であり3相交流の和として定義される電力需要値を用いたことである。これは服部・篠原による研究を除く先行研究との違いである。2つ目は電力比という1変数を主たる説明変数としたことである。短い時間間隔における電力需要値の絶対量とその付近の変動だけでなく、在宅時と比較した生活パターンの逸脱度が在宅状況の判別に重要な情報であることを示した。結果として新たに構築した2つの判別モデルの精度が、先行研究と同程度以上のものであることが示された。

キーワード：スマートメータ, 非侵入型在宅推定, 機械学習

Development of an Occupancy Detection Method Applying to Smart Meters for General Households

HISASHI OSHIMA^{1,a)} TSUYOSHI ISHIZONE^{2,b)} TOMOYUKI HIGUCHI^{1,c)}

Received: December 2, 2021, Revised: February 14, 2022,

Accepted: February 25, 2022

Abstract: Recently, a demand for occupancy detection is increased, since it is beneficial to society in many ways. Knowing whether a household is occupied by its residents in real time contributes to removing absent delivery, route optimization for work activities. The existing supervised classification models can perform at from about 70 percent accuracy to 90. On the other hand, There are two problems. First, The previous classification models always require high-resolution electricity data such as 1 second interval measurement and each data from three phases of smart meter. The requirement is too difficult to get in Japan, due to the installation methods of the sensor. Second, the previous models only use electricity data's absolute and near-absolute variation for short time intervals as explanatory variables. It is not enough to explain occupancy, since it will lead to a misclassification in the situation where a household is occupied regularly without using electricity.

In this paper, we propose new classification models (method (a) and method (b)) which has two differences to overcome the previous problems. First, our models only require electricity data measured at 30 minutes intervals as main trunk electricity data, which has been measured from smart meter in general household. Second, we propose a new variable called an electricity data ratio. The electricity data ratio at given time can be regarded as the degree of deviation from the lifestyle pattern seen from the time when the person is at home. It can reduce misclassification when a household is occupied regularly without appliances.

The method (a) is to set a threshold on electricity data ratio. The numerator of the ratio at given time is electricity data that is observed at 30 minutes intervals, and the denominator is electricity data at the same time when a household may be occupied. Definition of when a household may be occupied is when electricity data is maximum value for 2 weeks before and after and also 30 minutes before and after. Given that i is number of days and j is the observation point which is measured at 30 minutes intervals in the i , electricity data ratio is defined by $R_{i,j} = \frac{y_{i,j}}{y_{i_*,j_*}}$, $i_*, j_* = \operatorname{argmax}_{k=i-14, i-13, \dots, i+14, l=j-1, j, j+1} (y_{k,l})$. The method (b) is based on Random Forest with eight explanatory variables. The ratio is added to the representative seven explanatory variables of the previous studies. As a result, method (a) and method (b)

detected occupancy status with equally or more accurately compared with previous studies.

Keywords: smartmeter, non-intrusive occupancy monitoring, machine learning

1. 序論

近年、スマートメータを代表とする電子式電力量計が多く世帯に導入されている。導入の期待の1つとして世帯内人物の在宅状況を推定する在宅推定 [1] がある。時刻ごとの各世帯の在宅状況を把握することが出来れば、災害時の避難計画の策定や配送業と営業活動の経路最適化、電力の需要と供給の管理自動化、高齢者見守りシステム [2], [3] につながる。これらは市民の人命救助、企業や世帯のコスト大幅削減、高齢者の健康状態・生活状況の把握といった観点から有益である。日本の配送業者全体では不在配達によって年間数千億円単位で損失が生まれている [4]。

在宅推定の研究は使用する説明変数の観点により2つに大別される。1つ目は、人感センサーの物体認識情報やGPSからのリアルタイム位置情報により、直接的に在宅状況を示す説明変数から推定を行う研究である。2つ目は、電力需要値や二酸化炭素排出量などにより、間接的に在宅状況に影響を与える説明変数から推定を行う研究である。二酸化炭素排出量の測定方法の例は非分散型赤外線 (NDIR) による方法である。[5] では、二酸化炭素の濃度を説明変数に機械学習の手法で、在宅している人数を推定している。電力需要値を中心とした在宅推定の研究については本稿と密接に関係するため、次の段落以降で詳細に述べる。

電力需要値を中心とした在宅推定の典型例は、電力需要値を含む説明変数により目的変数の在宅状況を判別する教師あり学習の2値判別モデルである [4], [6]–[10]。ここで在宅状況とは、世帯構成員のうち1人でも在宅ならば1、そうでなければ0という2値変数である。教師なし学習や半教師あり学習による在宅推定の手法も研究が進められている [7], [11]–[14]。多くの先行研究 [4], [6]–[9] では、本稿と同様の ECO data set [15] を使用している。これはスイスの5世帯に対して電力需要値と在宅状況を数カ月記録したものである。電力需要値は測定時間頻度が1秒間隔、測定値の単位が1 w 未満単位である。在宅状況も同様の測定時間

頻度で測定されている。データについての詳細は3章で述べる。

電力需要値を中心とした教師あり学習の先行研究について推定の手法を詳しく述べる。説明変数は、1分または15分間隔でありなおかつ3相交流の各相から得られる、電力需要値の平均値や標準偏差などである。[4], [7]–[10]。判別モデルは、閾値 [7]、サポートベクトルマシン (SVM) [4], [7], [9]、K 近傍法 (KNN) [4], [7], [10]、主成分分析 (PCA) により次元削減を加えて SVM を構築する手法 (SVM-PCA) [8]、Random Forest (RF) [4], [10] などである。[6] においては、測定時間頻度が60分間隔でありさらに3相交流の主幹の電力需要値を説明変数として、SVM と RF を適用している。なお主幹の電力需要値とは3相の相ごとの電力需要値ではなくその和としての電力需要値を示す。

電力需要値を中心とした教師あり学習の先行研究についての問題点を2つ提示する。1つ目は、日本国内での在宅推定の場面を考えると説明変数のデータ収集のコストが高いことである。データ収集のコストが高い説明変数は、1秒間隔などの高測定時間頻度の電力需要値や3相交流の相ごとの電力需要値である。これは [6] を除く先行研究で共通に採用されたものである。国内の在宅推定を行う場面においてデータ収集のコストが高い説明変数を得ようとすると、追加のセンサーが必要となり入手が難しい。国内の一般世帯向けのスマートメータで測定される電力需要値は、30分間隔といった低測定時間頻度であり、さらに3相交流の主幹の値が採用されている。これはAルート計測値と呼ばれている。

2つ目に説明変数について述べると、15分や60分間隔などの短い時間間隔における電力需要値の絶対量とその付近の変動しか捉えることができていないことである。[4], [6]–[10] これは在宅状況の説明に不十分である。なぜなら、家電製品未使用で活動しているとき/就寝/昼寝などが安定した生活パターンとして存在する場合を説明できないと推察されるからである [16]。これらの場合では、短い時間間隔の中で電力需要値の絶対量が小さくなりモデルは不在と推定しうる。しかし実際には在宅であり誤推定となることが多いであろう。

そこで本稿では、一般家庭の生活パターンからの考察により「電力比」を用いた新しい教師あり学習の判別モデルを提案し、ECO data set を用いてその有効性を評価する。これは日本国内での大規模な在宅推定を本稿の先の最終目的として見据え、その途中実験を行うという位置づけであ

¹ 中央大学理工学研究科
Graduate School of Science and Engineering Chuo University,
Bunkyo, Tokyo 112-8551, Japan

² 明治大学先端数理科学研究科
Graduate School of Advanced Mathematical Sciences Meiji University, Nakano, Tokyo 164-8525, Japan

a) a16.kk5e@g.chuo-u.ac.jp

b) tsuyoshi.ishizone@gmail.com

c) higuchi@kc.chuo-u.ac.jp

る。先行研究との違いとして主に2つある。1つ目に、測定時間頻度が30分間隔であり、なおかつ主幹の電力需要値を用いたことを挙げる。これは国内スマートメータの規格に従ったものである。データ収集にコストのかかる説明変数を用いなかったため、ビジネス展開を視野に入れた汎用性の高い手法となっている。

2つ目に、電力比という1変数を主たる説明変数としたことである。電力比とは、ある推定日推定時刻の電力需要値について、およそ同時刻の在宅と思われるときの電力需要値をもとにした割合である。およそ同時刻の在宅と思われるときの定義として前後2週間前後30分の電力需要値の最大値をとるときとした。またこの指標は在宅時と比較した生活パターンの逸脱度と捉えることができる。在宅時と比較した生活パターンの逸脱度を説明変数にすることにより、家電製品未使用で活動しているとき/就寝/昼寝などが安定した生活パターンとして存在する場合、をうまく判別できるようになる。これらの場合では、短い時間間隔で電力需要値の絶対量は小さくなるが、在宅時と比べると小さくないことが多くモデルは正しく在宅を推定するであろう。電力比の構築には測定時間頻度が30分間隔でありなおかつ主幹の電力需要値を用いた。

測定時間頻度が30分間隔でありなおかつ主幹の電力需要値からなる、電力比を説明変数とした判別モデルが実用的な精度を持つか否かを2つの判別手法により検証する。1つ目の手法(a)は電力比の1変数で閾値により判別を行う手法、2つ目の手法(b)は[6]の説明変数に電力比を加えRFを構築する手法である。手法(a)において、電力比の1変数のみで判別モデルが十分な精度を持つという仮説の可否を明らかにする。手法(b)においては、先行研究と電力比を組み合わせることで、判別モデルの精度が先行研究と比較して向上するという仮説の可否を明らかにする。さらにRFの特徴量重要度において電力比が相対的に重要か否かを確認する。

2. 先行研究と本稿の位置づけ

類似度の高い先行研究と本稿で提案する2つの手法—手法(a)と手法(b)—について、電力需要値の測定時間頻度、在宅状況の測定時間頻度、電力需要値が主幹のものか

そうでないか、に焦点を当て、表1にて整理した。手法(a)、(b)とは4.2.1項および4.2.2項に対応する本稿が提案する手法である。類似度の高い先行研究とは、電力需要値を中心とした在宅推定のうち教師あり学習による手法とする。

表の項目「電力需要値の測定時間頻度」とは、国内スマートメータからの電力需要値に相当するものと、それ以外の説明変数構築に必要な電力需要値の測定時間頻度の中で最も高い測定時間頻度を示す。国内スマートメータからの電力需要値に相当するものは、30分間隔/15分間隔/1分間隔の低測定時間頻度な電力需要値の平均値または和である。これは国内スマートメータから直接入手可能な変数であるため、測定時間頻度としてはそのまま、30分間隔ならば30分と表記する。一方たとえば15分(=900秒)間隔の電力需要値の最大値を1秒単位で構築するときは、国内スマートメータからの電力需要値に相当するものとは別に1秒間隔の電力需要値を必要とするため、その他と区別をして「1秒」と表記する。その他、「在宅状況の測定時間頻度」は目的変数となる在宅状況の測定時間頻度を示し、「主幹(True)か相ごと(False)」は電力需要値が3相交流の主幹の値かそれとも相ごとの値かをTrueとFalseの2値で示したものである。また、両方使用する場合はFalseとし、主幹か否かの判断が出来ない場合は「?」と表記した。

電力需要値の測定時間頻度、主幹か相ごとか、という2つの観点で表1について説明を行う。国内スマートメータから得られる電力需要値は30分間隔の主幹の値である。SVM, RF [6]がその条件を満たしており、本稿と類似度が最も高いと言える。その他は類似度が低い。改めてここで本稿が想定する日本国内の在宅推定についての条件を下記に整理する。

条件1 電力需要値の測定時間頻度は30分とする

条件2 3相交流の主幹の電力需要値とする

3. ECO data set

本稿では、[7], [8], [17]の筆者であるWilhelm Kleimingerが在籍していた、スイス連邦工科大学チューリッヒ校が公開しているデータセットのEco data set [15]

表1 類似度の高い先行研究
Table 1 Previous studies with high degree of similarity to this paper.

手法名	電力需要値の 測定時間頻度	在宅状況の 測定時間頻度	主幹 (True) か相ごと (False)
RF, DesicionTree (DT), KNN, Naive Bayes (NB) など [10]	1 秒	1 分	?
SVM, KNN, 閾値 [7]	”	15 分	False
KNN, SVM, RF, Multiple Layered Perceptron (MLP) [4]	”	”	False
SVM-PCA, KNN-PCA, Gaussian Mixture Models (GMMs) -PCA など [8]	”	”	False
SVM [9]	”	”	False
SVM, RF [6]	30 分	60 分	True
手法 (a)	”	30 分	True
手法 (b)	”	60 分	True

を使用する。Eco data set の中で本稿に関わる要素としては 2 つある。1 つ目に powerallphases である。これは、スイス国内のある 5 世帯に対して、3 相交流の主幹の電力需要値を 1 w 未満単位 1 秒間隔で数カ月記録したものである。欠損値は、使用したデータのうち、月単位の割合から調和平均値を算出した値で 0.0001 ほど存在する。これに対して 1 日単位で、時刻方向の線形補完を施す処理を行った。測定器はスマートメータである model E750 from Landis+Gyr を使用しており、無線で接続されるサーバーでそのデータを集計している。

2 つ目に occupancy である。これは、在宅ならば 1 不在ならば 0 と実在宅状況をラベル化したものである。少なくとも 1 人でも在宅していれば在宅、それ以外が不在である。測定時間頻度は電力需要値と同様である。在宅状況の測定は、赤外線センサーの Roving RN-134 およびアンケートにより行われた。アンケートは Galaxy Tab P7510 というタブレットを戸口付近に設置し、専用のユーザインタフェースに対して、世帯内人物が在宅状況を入力する形で測定された。

また本稿の中では、電力需要値の測定時間頻度を 1 秒間隔から 30 分間隔へと情報を落として実験を行う。電力需要値は 30 分ごとの和をとって再集計した。在宅状況は 1800 秒間すべて在宅ならば在宅、それ以外は不在とし再集計した。

次に、5 世帯の家族構成と物件形態については表 2 に、在宅推定を行う日数については表 3 に、それぞれまとめた。ここでは夏を 7 月～9 月、冬を 11 月～1 月としている。在宅推定を行う日数は、在宅状況の測定が行われ、さ

表 2 5 世帯の家族構成と物件形態

Table 2 Family structure and property type of the sample households.

世帯 ID	家族構成	物件形態
1	大人 2 人子供 2 人	戸建て
2	大人 2 人	集合住宅
3	大人 2 人子供 2 人	戸建て
4	大人 2 人	戸建て
5	大人 2 人	戸建て

表 3 在宅推定を行う期間

Table 3 Time period for which occupancy detection is to be applied.

世帯 ID	在宅推定日数 (冬)	在宅推定日数 (夏)
1	46	39
2	30	83
3	18	47
4	48	38
5	31	43

らに各日数の前後 2 週間の電力需要値が測定されている期間とする。

4. 在宅推定アルゴリズム

4.1 電力比

推定のアイディアは、電力需要値から成る電力比が在宅状況の目安になり得るのではないかとという自然なものを元に行っている。分子は推定日推定時刻の電力需要値自身であり、分母はおよそ同時刻の「在宅と思われる」ときの電力需要値である。電力比は「在宅と思われる」ときの電力需要値に対して、ある推定時刻において電力需要値をどれくらいの割合で使っているかという、在宅時と比べた生活パターンの逸脱度を示している。

およそ同時刻の「在宅と思われる」ときの定義として、ある推定日推定時刻において、その推定日の window 幅前後 2 週間およびその推定時刻の window 幅前後 30 分で最大の電力需要値となるときとする。直感的に想像すると、「在宅と思われる」ときは電力需要値が高くなる傾向を持つ日曜日から多く選出される。実際、同サンプル内ではそうになっている。しかし職種によっては、日曜日に不在で電力需要値が高くない場合もあるためこのような柔軟性を持たせた。推定日の前後 2 週間とした理由は、季節に依存する値の変化に上手く対応するためである。日本やスイスの 9 月は前半と後半で平均気温が異なることが多く、電力需要値のパターンや値に変化が生まれる。window 幅を同月内と設定し、9 月後半のある日付に対して、9 月前半の電力需要値を分母に選択する状況を考える。季節の変わり目は電力需要値パターンが変化するため、これは不自然である。一方 window 幅を狭くしすぎると window 幅内に在宅しているときがなくなり、結果として比の値に意味がなくなる。そんなトレードオフの関係があるため、本稿では前後 2 週間とした。

また推定時刻の前後 30 分とした理由は、世帯内人物の行動に 30 分程の前後の変動があることを考慮し、その変動に上手く対応するためである。起床や就寝、帰宅等の行動は 30 分早くなったり遅くなったりし得る。下記で電力比アイディアを数式で表現し、図 1 および図 2 で電力比を説明変数とすることの合理性について述べる。

以上説明した電力比を数式で表現する。 y は電力需要値である。添え字 i は推定期間における日付の初日を起点とした日数を表し、初日を $i=1$ とした。以降 $i=1,2,3,\dots,N$ となり、 N は推定期間における日付の最終日となる日数を示す。また $i-14$ は日数 i から 2 週間前の日数を示し、 $i+14$ は日数 i から 2 週間後の日数を示す。添え字 j は、それぞれの日数 i に対応し、1 日 24 時間の時刻を 30 分間隔で刻んだ測定点であり、 $j=1,2,3,\dots,48$ となる。 $j+1$ は時刻 j から 30 分後の測定点を示し、 $j-1$ は時刻 j から 30 分前の測定点を示す。本アイディアにおける電力比は下記のよ

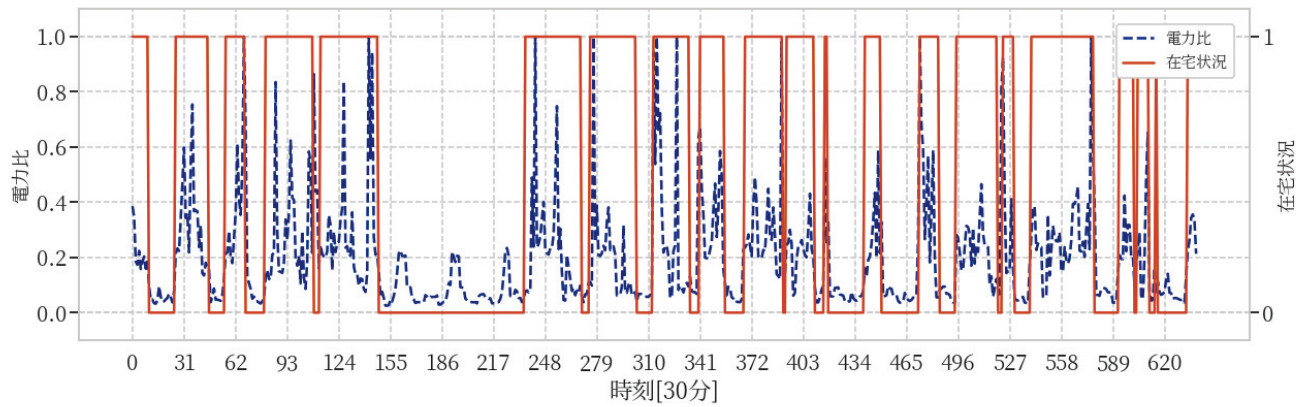


図 1 世帯 ID:2 についてある 20 日間の電力比と在宅状況の対応

Fig. 1 Correspondence between electricity data ratio and occupancy for 20 days in the household with ID 2.

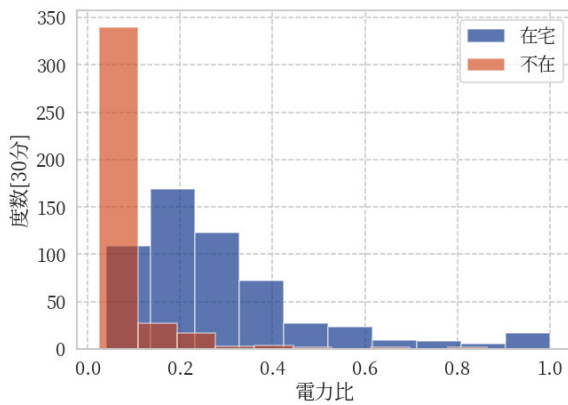


図 2 電力比の分布

Fig. 2 Histogram of electricity data ratio.

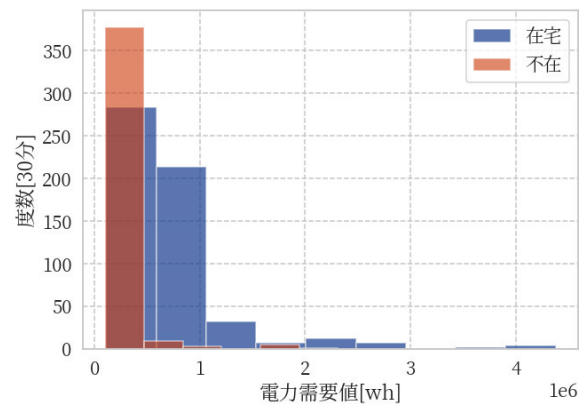


図 3 電力需要値の分布

Fig. 3 Histogram of electricity data.

うに数式で表される.

$$R_{i,j} = \frac{y_{i,j}}{y_{i_*,j_*}}$$

$$i_*, j_* = \operatorname{argmax}_{k=i-14, i-13, \dots, i+14, l=j-1, j, j+1} (y_{k,l})$$

図 1 は世帯 ID:2 について、冬の在宅推定を行うすべての期間のうちある 20 日間の、さらに 6~22 時の、電力比と在宅状況をプロットしたものである。横軸時刻 [30 分] が 30 分間隔の測定点を示し、縦軸電力比 (在宅状況) が電力比と在宅状況 (在宅:1, 不在:0) を示す。図 1 より、電力比がおおよそ 0.1 未満まで減少したとき、在宅状況は不在を示すことが多い。

次に図 2 は世帯 ID:2 について、冬の在宅推定を行うすべての期間の、さらに 6~22 時の、電力比をヒストグラムとして表示したものである。横軸が電力比、縦軸がその度数 [30 分] を示す。図中のラベル: 在宅が在宅時の電力比、ラベル: 不在が不在時の電力比を示している。不在時においては、電力比のとり値が 0.0~おおよそ 0.1 のときに単峰のピークをもち、それ以外の範囲では密度が低い。一

方在宅時には、おおよそ 0.2 のときにピークをもち右に裾の長い形状を持つことが見て取れる。両者の分布の違いは明らかである。

また同期間において、図 2 と同様の形式で、電力需要値をヒストグラムとして表示したものが図 3 である。図 3 によると、電力需要値において、不在時の分布は電力需要値が 0 に近いときにピークをもち、しかし在宅時についても同様に 0 に近いときにピークをもち、この点が電力比と電力需要値の違いである。電力需要値が下がっても在宅の場合が多いことは直感的に明らかであり納得のいく結果であると言える。その例として、家電製品未使用で活動しているとき/就寝/昼寝などが安定した生活パターンとして存在する場合は挙げられる。電力比は、日付方向の時系列の情報を組み込むことにより、在宅時の分布と不在時の分布の違いをより顕著に表現するような特徴量であると言える。

4.1.1 電力比の window 幅についての分析

4.1 節の定義では、電力比の分母について window 幅前

後2週間の最大値というように設定した。本項では window 幅をいくつか変更しその影響について分析を行う。window 幅を前後1, 2, 3週間に変更し、それぞれの場合の電力比と在宅状況の対応を図示した。図4は window 幅を前後1・2週間にしたときの、夏の世帯ID:2について14日間分の電力比と在宅状況の対応を示した図である。図の形式は図1と同様である。図4より、window 幅を1週間にしたときの不在時の電力比の値は、window 幅2週間の系列より大きくなる傾向をもつ。これは前後1週間に在宅と思われるときが存在せず電力比の分母が過小評価されている影響であろう。また、その他の世帯・期間でも同様の傾向がみられた。

次に図5は window 幅を前後2・3週間にしたときの、図4と同じ世帯および期間の電力比と在宅状況の対応を示した図である。図5より、window 幅を3週間にしたときの在宅時の電力比の値は、window 幅2週間の系列より小さくなる傾向をもつ。これは前後3週間と window 幅を広げたとき、季節の変わり目などが存在し、電力比の分母が過大評価されてしまっている影響であろう。また、その他の世帯・期間でも同様の傾向がみられた。

また5.1節の評価方法に従い本稿が提案する2手法に対して、説明変数である電力比の window 幅を前後1, 2, 3

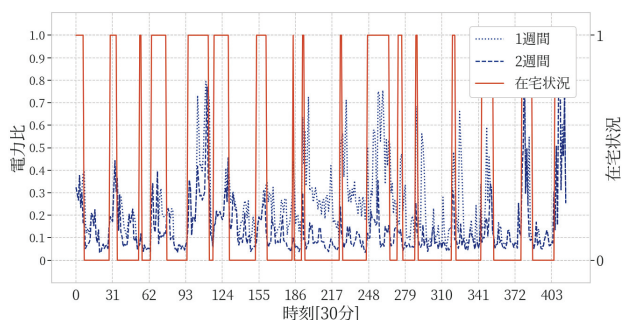


図4 window 幅前後1週間の電力比

Fig. 4 Electricity data ratio when window width is set to 1 week before and after.

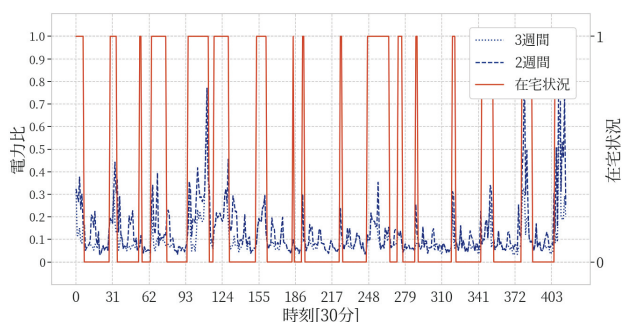


図5 window 幅前後3週間の電力比

Fig. 5 Electricity data ratio when window width is set to 3 weeks before and after.

週間に変更し、それぞれの場合の正答率の平均値を算出した。手法(a)においては window 幅前後2週間が正答率が最大であり、前後1週間と比べて0.04、前後3週間と比べて0.02ほどの差が生まれた。手法(b)においても window 幅前後2週間が正答率が最大であり、前後1週間と比べて0.04、前後3週間と比べて0.02ほどの差が生まれた。

不在時の電力比が小さく在宅時の電力比が大きいという傾向をもつこと、5.1節の評価により正答率が最も高くなることより本稿では前後2週間が妥当であると考えることとする。一般的には適用するデータに対してこれらの分析を行って、window 幅を決定する工程が求められる。

4.2 電力比から在宅状況への判別方法

4.2.1 手法(a): 電力比のみの閾値

世帯ごとに、また在宅推定を行う期間における冬と夏ごとに、電力比のみを説明変数に、目的変数の在宅状況を判別する。手法(a)の検証の狙いは、電力比の1変数のみで判別モデルが十分な精度を持つという仮説の可否を明らかにすることである。判別は閾値により行う。在宅推定の目的の1つである訪問が求められる営業活動を見据えると、在宅であることを正しく推定することが一般的に有益であると考えられている。閾値の決定方法はいくつかあるが、ここではその場面を想定し、推定が在宅のうち実際のラベルも在宅であること(適合率)と実際のラベルが在宅のうち推定も在宅であること(再現率)の両者を最大化することを考える。そこで、Precision-Recall Curve (PR-Curve, PR 曲線)による手法を採用する。ビジネス場面で適合率を解釈すると、モデルが在宅を推定したため従業員が訪問を行ったとき、実際に世帯が在宅で営業活動が正しく行える確率と言える。

2値判別問題における閾値の決定方法の1つである PR 曲線による方法を述べる。PR 曲線は横軸に再現率 (Recall)、縦軸に適合率 (Precision) を取り、閾値を変化させてそれぞれの場合の点をプロットしたものである。閾値の決定については、点 (1.0, 1.0) と各点とのユークリッド距離を求め、距離が最小となる点のときの閾値を採用する。図6はそれを模式図として表したものであり、(1.0, 1.0) との最小の距離 (灰色の矢印) と、そのとき選択されるべき点 (赤色の三角の打点) を示している。横軸 Recall と縦軸 Precision について、両者のスケールが異なっている点に注意されたい。

4.2.2 手法(b): 8変数のRF

世帯ごとに、また在宅推定を行う期間における冬と夏ごとに、合計8変数でRFにより在宅状況の判別を行う。電力比に加えて、[6]で採用された、主幹電力需要値の平均値/最大値/最小値/標準偏差/範囲および気温と時刻(6~22)を説明変数とする。気温については、測定対象の地域

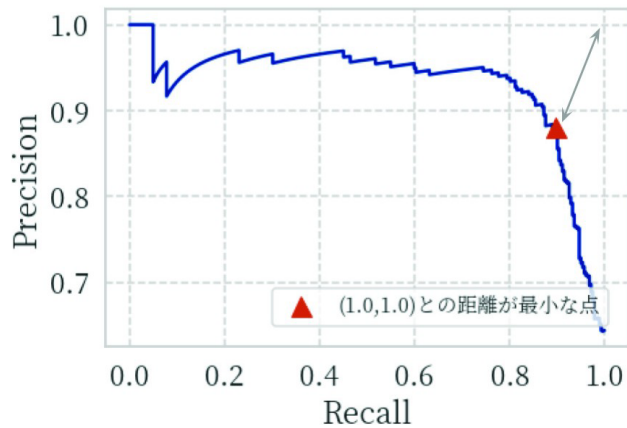


図6 PR 曲線による閾値決定の様子

Fig. 6 How we determine the threshold by Precision-Recall Curve.

であるスイスのチューリッヒの 30 分間隔の気温を [18] から入手した。日本国内でも 30 分間隔や 1 時間間隔の気温は容易に入手できる。手法 (b) による検証の狙いは、先行研究 [6] と電力比を組み合わせることで、判別モデルの精度が先行研究と比較して向上するという仮説の可否を明らかにすることである。さらに RF の特徴量重要度において電力比が相対的に重要か否かを確認する。

各変数の構築方法について述べる。ここでは各変数と在宅状況は 60 分間隔に変換する。まず電力比については、30 分間隔の電力需要値の和をとって 60 分間隔へ変換し、その後 4.1 節と同様の方法で電力比を構築した。ここで、時刻を表す j は $j=1,2,3,\dots,24$ のように変化する。その他は同様である。電力需要値の平均値/最大値/最小値/標準偏差/範囲については、30 分間隔の電力需要値からそれぞれ構築した。気温は 30 分間隔の気温データから、平均をとることで 60 分間隔へと変換した。30 分間隔の気温データに欠損値がある場合は、時刻方向で線形補間を行った。時刻は、6~22 時のみを用いてモデルの学習とテストを行うため、各測定点が従う時刻を 6~22 時として構築した。

RF の計算機上のプログラムは python の scikit-learn [19] が持つクラスの 1 つである RandomForestClassifier を使用する。これはオープンソースとして公開されているプログラムである。

5. 評価

5.1 評価方法

判別の精度についての評価方法を述べる。手法 (a) と (b) とともに stratified k-fold cross validation を採用し、正答率/適合率/再現率を k 個のテストデータへの平均値として計算しこれを評価に用いる。手法 (a) は $k=2$ 、手法 (b) は $k=5$ とした。手法 (a) の閾値については、4.2.1 項で述べた PR 曲線による閾値決定手法に従い、それぞれのトレーニングデータに対して学習を行い決定する。手法

表 4 手法 (a) の正答率/適合率/再現率
Table 4 Method(a)'s Accuracy/Precision/Recall.

世帯 ID	冬			夏		
	正答率	適合率	再現率	正答率	適合率	再現率
1	0.80	0.88	0.85	0.74	0.89	0.78
2	0.86	0.87	0.88	0.82	0.88	0.83
3	0.73	0.94	0.74	0.72	0.83	0.75
4	0.89	0.98	0.90	0.85	0.97	0.86
5	0.79	0.96	0.80	0.87	0.99	0.87

(b) のハイパーパラメータとして決定木の数や深さなどがあるが、これらは固定してモデルの学習を行い評価することとした。

また本稿では学習・テストに使用する時間帯を 6~22 時に限定した。これは先行研究との比較をすること、さらに営業活動が行われる一般的な時間帯を考慮したことによるものである。

5.2 結果

5.2.1 手法 (a)

5.1 節に従って手法 (a) の判別の精度を評価する。各世帯の冬と夏ごとに、さらに 6~22 時について、正答率/適合率/再現率の算出を行った。冬は 11 月~1 月であり、夏は 7 月~9 月と定義している。各値は小数第 3 位を四捨五入する形で表記した。

表 4 より、正答率の範囲は、冬において 0.73~0.89、夏において 0.72~0.87 となった。平均値はそれぞれ 0.81 と 0.80 となった。世帯間の差は冬において 0.03~0.16 であり、夏において 0.02~0.15 である。

また世帯 ID:2 の冬の 20 日間の電力比と在宅状況の対応を示した図 1 について、実際の判別の様子を図 1 と同様の形式で図 7 にて可視化する。ここでは 2-fold cross validation の各 iteration にて得られる閾値の平均値を、仮の閾値として図示している。4.1 節では、電力比がおおよそ 0.1 未満まで減少したとき、在宅状況が不在を示すことが多いと述べた。図 7 より、その推測のとおり、おおよそ 0.1 に閾値 (図中のドット付き実線) が設けられたことが見て取れる。閾値以上の値の大きさをもつ電力比 (図中の点線) に対しては在宅を推定し、閾値未満の値の大きさをもつ電力比に対しては不在の推定を行っている。この判別の様子は、実在宅状況 (図中の実線) とおおよそ一致している。

5.2.2 手法 (b)

次に、5.1 節に従って手法 (b) の判別の精度を評価する。各世帯の冬と夏ごとに、さらに 6~22 時について、正答率/適合率/再現率の算出を行った。表 5 より、正答率の範囲は、冬において 0.84~0.91、夏において 0.81~0.87 となった。平均値はそれぞれ 0.87 と 0.85 となった。世帯間の差は冬において 0.01~0.07 であり、夏において 0.01~0.06 である。

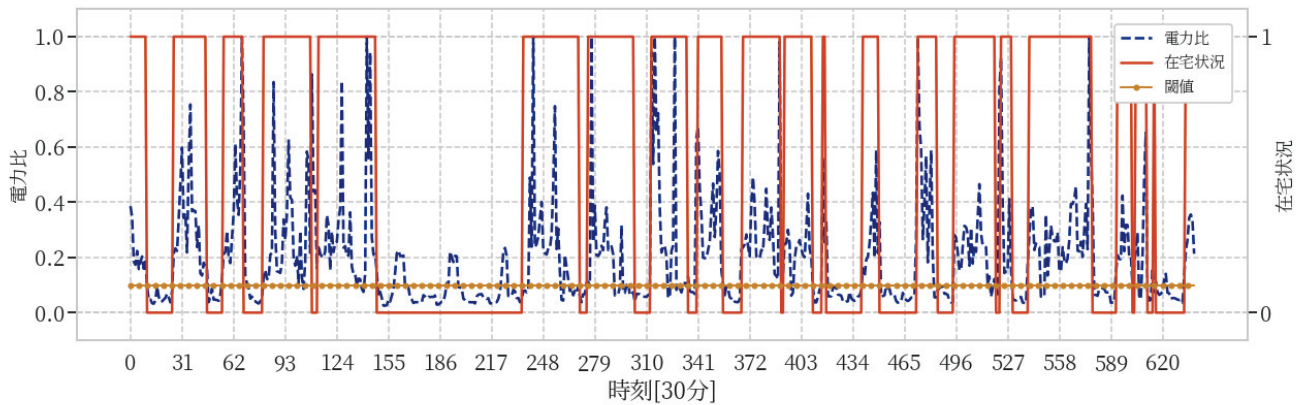


図7 図1に対応する実際の判別の様子

Fig. 7 How method(a) actually detects occupancy in Fig. 1.

表5 手法 (b) の正答率/適合率/再現率

Table 5 Method(b)'s Accuracy/Precision/Recall.

世帯 ID	冬			夏		
	正答率	適合率	再現率	正答率	適合率	再現率
1	0.88	0.89	0.93	0.81	0.84	0.87
2	0.91	0.91	0.93	0.87	0.86	0.93
3	0.85	0.86	0.92	0.86	0.86	0.91
4	0.89	0.89	0.99	0.85	0.87	0.96
5	0.84	0.86	0.94	0.87	0.87	0.99

また、世帯 ID:2 の冬の推定において、構築された RF のモデルの内 1 つの木を図 8 に掲載する。各矢印が現在のノードから 1 つ深いノードへの分割/判別を示している。また図中の Ratio が電力比、Range が範囲、Max が最大値を示す。ヒストグラムは判別対象のデータを示し、ラベル Occupancy の Occupied が在宅時、Unoccupied が不在時のデータを示す。ここから RF の決定木の 1 つについて、判別の様子を木の深さ 2 までに限定し解釈する。まず木の深さ 1 において (1 つめの矢印を示す)、電力比が 0.15 より大きいかそれ以下かで判別している。次に木の深さ 2 において、電力比が 0.15 より大きい集合に対して、さらに範囲がある値より大きいかそれ以下かで判別している。同時に電力比が 0.15 以下である集合に対して、最大値がある値より大きいかそれ以下かで判別をしている。

次に特徴量重要度について図 9 で示す。特徴量重要度とは、各説明変数により判別を行ったときのジニ不純度の減少量を木の数で平均し、さらに正規化した値である。ジニ不純度は、データが各ノードに到達する確率で重み付けされた値を使っている。特徴量重要度の意義としては、各説明変数が在不在の分割においてどれほど貢献したのかを明らかにすることにある。また図 9 では、世帯数/夏と冬について平均値を算出したものを図示している。図中の棒線はプラスマイナス標準偏差を示したエラーバーである。図 9 より、電力比は平均値とともに重要度が 0.16 となり、

8 変数の中で最も重要度が高く在不在の分割において最も貢献している。また、最大値・最小値との重要度の差は 0.03 以内に収まり在不在の分割に対する貢献度の差は小さく、標準偏差・範囲・気温・時刻との差は 0.05 以上であり貢献度の差が大きい。

5.2.3 手法 (a), (b) と先行研究との比較

手法 (a), (b) と先行研究のモデルとの正答率の比較を行う。先行研究のモデルとして、参考文献 [6] から RF を比較対象として選定した。前述のとおり、この先行研究は本稿との類似度が特に高い。また RF は [6] だけでなく、条件 1 と 2 には従わないが類似の研究である [10], [4] でも同様に採用されており、先行研究らの中において精度が高いことが示されている。[6] の RF の推定手法について述べる。主幹電力需要値の 60 分間隔の平均値/最大値/最小値/標準偏差/範囲/気温/時刻 (7~22)/世帯 ID (1~5)/季節 (夏か冬) を説明変数に、同時間間隔の在宅状況を判別する、RF を構築する手法である。ハイパーパラメータは本稿と同様に固定して扱っている。評価方法としては、本稿と同様に $k=5$ の k -fold cross validation を採用している。

また補足として、RF の正答率は筆者による再現実装の結果を掲載した。この理由として、[6] に世帯ごとの正答率が記載されておらず比較ができないこと、手法 (b) と RF でハイパーパラメータなどの条件を統一した上で比較することが挙げられる。また [6] においては世帯 ID と季節も説明変数に入れまとめて 1 つのモデルを構築しているが、本稿では世帯ごとまた季節ごとに別々のモデルを構築することとした。これは決定木のある深さにおいて、ある世帯 ID およびある季節のデータすべてに在宅または不在を判別してしまうことを避けるためである。図 10 と図 11 に比較結果を掲載する。

図 10 より、冬の RF の正答率は 0.80~0.88 の範囲であり、平均値は 0.84 である。平均値で比較すると手法 (a) は 0.03 値が小さく、手法 (b) は 0.03 値が大きい。同様

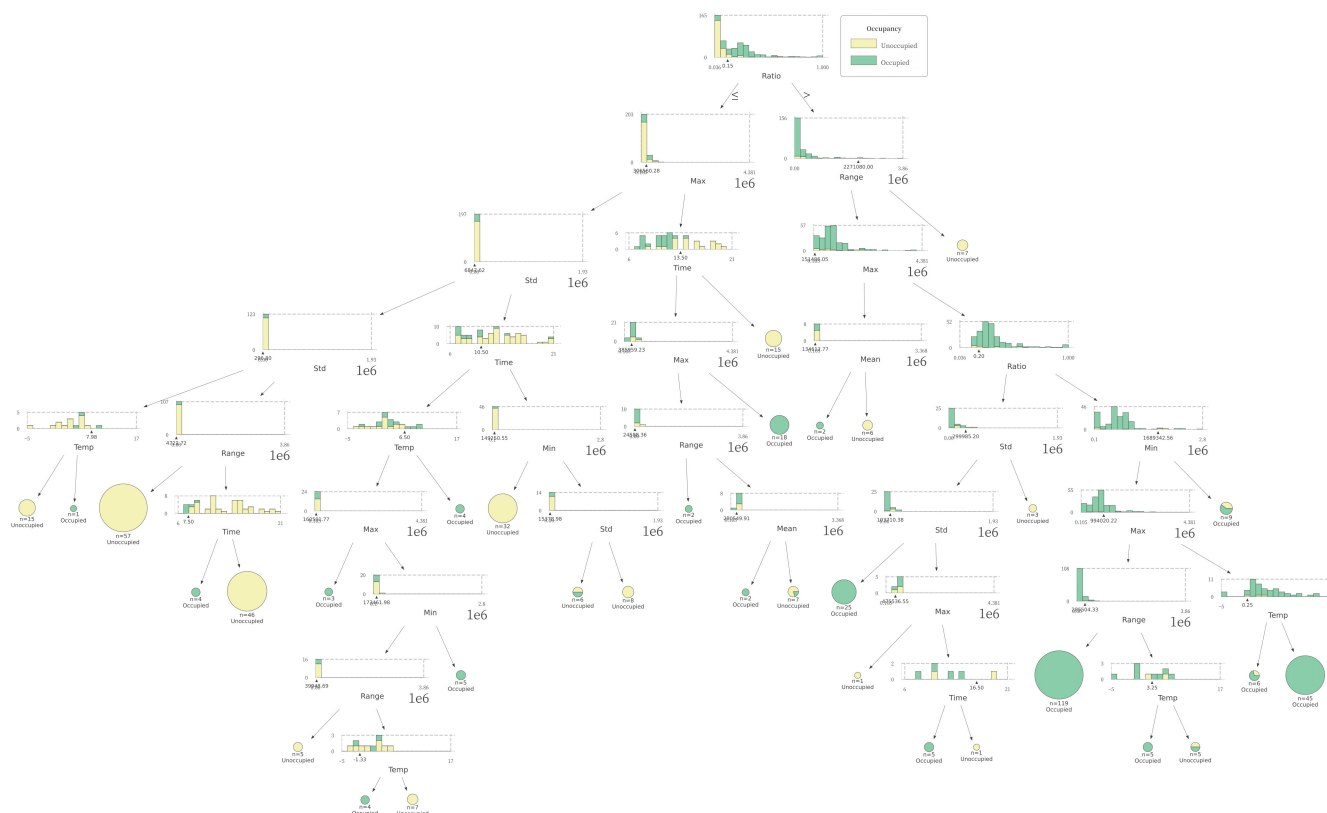


図 8 決定木の 1 つの可視化
Fig. 8 One of the decision trees in the method(b).

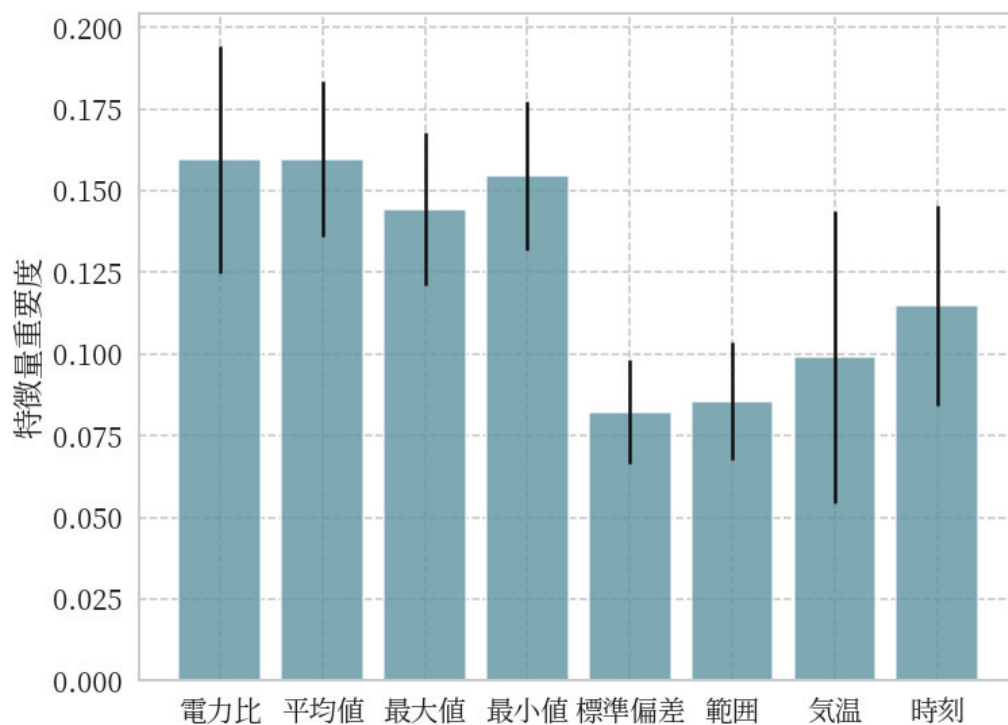


図 9 特徴量重要度
Fig. 9 Random forest's feature importance calculated by gini impurity.

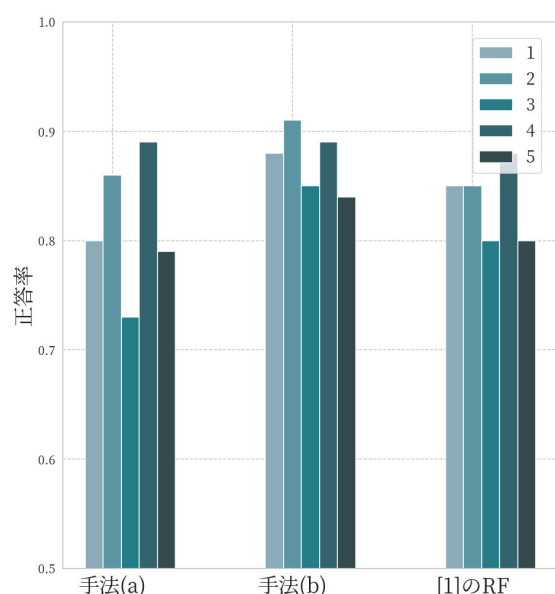


図 10 冬の正答率比較

Fig. 10 Comparison of the accuracy between method(a), method(b), and [1]'s RF in winter.

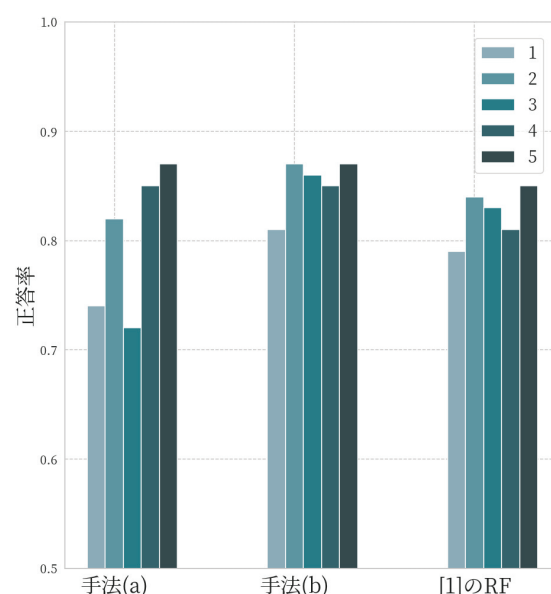


図 11 夏の正答率比較

Fig. 11 Comparison of the accuracy between method(a), method(b), and [1]'s RF in summer.

に図 11 より、夏の RF の正答率は 0.79~0.85 の範囲であり、平均値は 0.82 である。平均値で比較すると手法 (a) は 0.02 値が小さく、手法 (b) は 0.03 値が大きい。

5.3 考察

電力比を元にする判別モデルが示す正答率の結果より、手法 (a) と手法 (b) において、それぞれ正答率の平均値で 0.81 と 0.86 の精度をもつ判別モデルの構築に成功した。以下手法 (a)、手法 (b) それぞれの結果に対して考察を行う。

5.3.1 手法 (a)

表 4 より、手法 (a) の夏冬および世帯間の正答率の範囲は 0.72~0.89 となった。平均値は 0.81 である。また適合率は 0.87~0.99 となった。手法 (a) の検証の狙いは電力比の 1 変数のみで判別モデルが十分な精度を持つという仮説の可否を明らかにすることであった。結果として、[6]と比較して冬において 0.03、夏において 0.02 ほどだけ精度が下がるものの、8 割以上の精度をもつ判別モデルを構築することができた。このことより、電力比の 1 変数のみで十分に判別が行えると考えられる。適合率の範囲は 0.87~0.99 であったがこれが意味することは、モデルが在宅と推定し実際に従業員が世帯に訪問した際、0.87~0.99 の割合で世帯が在宅しており営業活動が実際に行えることを示している。

一方およそ 1~3 割の不正解のサンプルが存在すること

と、世帯に依存し正答率が 0.02~0.16 程変動してしまうことが判明した。不正解とは、推定結果は在宅だが実際は不在の場合またはその逆の場合を示す。

不正解の理由としては、2 つある。1 つ目は、電力比の値が大きくモデルが在宅を推定したときに、実は不在であったケースが存在するためである。家電製品の電源の消し忘れ、不在時に作動する家電製品の存在などがあつたのであろう。2 つ目は逆に、電力比の値が小さくモデルが不在を推定したときに、実は在宅であったケースがあるためである。在宅しているが家電製品を使用していないという状況が不定期に存在した可能性がある。

世帯に依存し正答率が変動してしまう理由としては、在宅時と不在時の在宅状況ごとの電力比の分布の形状が、世帯によって変動するからである。3.1 節で述べた、電力比が電力需要値より在不在の分布差をはっきりと表現するという特徴をとりわけ示したのは、冬の世帯 ID が 1/2/4/5 と夏の世帯 ID:2 である。これらの世帯は比較的正答率が高くなっている。手法 (a) をそのまま適用する際は、世帯間や季節間の差を意識する必要がある。不在時は電力比が小さく、在宅時は電力比が大きくなるという特徴を持つ場合は有用だが、そうでない場合は不適切である。データに在宅状況のラベリングが成されてる場合はそれを確認し、そうでない場合は別の方法でその特徴を確認すべきである。たとえばラベルによらない電力需要値・電力比の時系列上のパターンが安定している際は同様の特徴を持つ

可能性が高いと言える。生活パターンが安定しているならば、在不在のパターンも安定している可能性が高いと考えられるためである。

また世帯 ID:4 の夏冬と世帯 ID:5 の夏においては、不在時の割合がおよそ 1 割以下でその他の世帯に比べて小さい。その他はおおよそ 1 割 7 分～3 割である。当枠組みでは PR 曲線による閾値決定手法を採用したため、これらの世帯に対して不在より在宅を優先して推定するよう設定した。実際、閾値決定手法はいくつもあり、ビジネスでの在宅推定の場面においては、求められる目的に応じて柔軟に変化させるということが必要である。最後に世帯 ID:3 に対して、参考文献 [17] が電力需要値の測定単位が低解像度であると指摘している。その影響が電力比の分布の形状および正答率に及んでいると推察する。

5.3.2 手法 (b)

表 5 より、手法 (b) の夏冬および世帯間の正答率の範囲は 0.81～0.91 となった。平均値は 0.86 である。手法 (b) の検証の狙いは、先行研究と電力比を組み合わせることで、判別モデルの精度が先行研究と比較して向上するという仮説の可否を明らかにすることであった。結果として [6] と比べて夏と冬ごとの正答率の平均値がそれぞれ 0.03 ほど向上し、正答率の向上を確認できた。[6] の説明変数に電力比を加えることで、8 つの説明変数と在宅状況の間の複雑な関係を捉えることが出来るようになったからではないかと推察される。さらに RF の特徴量重要度において電力比が平均値とともに最も重要であることが示された。

また手法 (a) と手法 (b) を比較すると、世帯間の正答率の差が手法 (a) は 0.02～0.16 だが、手法 (b) は 0.01～0.07 である。手法 (b) は手法 (a) より世帯によらず高い正答率をもつことを示唆している。

6. 結論と今後の展望

6.1 結論

本稿は先行研究との 2 点の違いを持つ新しい教師あり学習の判別モデルを構築した。1 点目は、国内の一般家庭への適用を見据え、測定時間頻度が 30 分間隔でありなおかつ主幹の電力需要値を使用したことである。これは [6] を除く先行研究との違いである。2 点目は、電力比の提案により、短い時間間隔での電力需要値の絶対量とその付近の変動しか捉えられないという問題を解決したことである。電力比の 1 変数に閾値を設ける手法 (a) と、合計 8 変数で RF を構築する手法 (b) において、それぞれ正答率の平均値が 0.81 と 0.86 となり先行研究と同程度以上の精度となった。

6.2 今後の展望

今後の展望としては主に 5 つ見据えている。1 つ目は、

列挙された現状の課題についてより深く知見を得ることである。たとえば、不正解のサンプルについて、電力需要値以外の収集可能な変数から共通の特徴を見つけることが出来れば有用である。

2 つ目はオンライン処理への対応についてである。本手法は推定する日時その後 2 週間の電力需要値が推定に必要である。リアルタイムのオンライン処理を行う場面では、ある推定日でそれは手に入らない。そこで代替案として 2 つ提示する。1 つ目は電力需要値の予測値を使用するというものである。これは、各世帯の電力需要値のデータが数カ月や数年単位で入手可能な場面を想定している。予測モデルは多数想定されるが、過去のデータから推定日時その後 2 週間の予測値を得ておけば、後 2 週間の電力需要値の代替案として使用することができる。2 つ目は過去の電力需要値のみで推定を行うことである。たとえば、前 2 週間の電力需要値のみを使用することを挙げる。この場合は長期不在の影響を受けやすくなるため、前 3 週間ほどに window 幅を大きくしてもよい。

3 つ目は本稿の最終的な目的であるが、国内世帯の電力需要値と在宅状況の実データに対して本手法を適用することである。その際国内の電力需要値の規格である A ルート計測値には、条件 1 と 2 の特徴の他に見せかけの需要変動が生じるという特徴がある。30 分間隔で測定値が 100 wh に満たないとき 0 wh と測定し、次の時刻にその値分を繰り越すことにより、実際の需要変動とは異なる変動が生まれてしまうというものである。これに対しては、[20] などによりデータの内挿を行うことが想定される。また、手法 (a) の結果より世帯に依存し正答率変動してしまうことが判明した。さらに考察より、それが在宅状況ごとの電力比分布の形状によるものと判明した。世帯によりどのような在宅状況ごとの電力比分布の形状が存在するのかといった所も併せて調査する必要がある。

4 つ目は本アイデアを軸に、正答率向上に重きを置き、より複雑な手法に挑戦することも興味深い。本稿では、国内スマートメータの規格に従い測定時間頻度が 30 分間隔でありなおかつ主幹の電力需要値を使用し、さらに短い時間間隔での電力需要値の絶対量とその付近の変動しか捉えられないという先行研究の問題点を解決することを目的とした。具体案は検討中であるが、説明変数の数を増やすこと、より複雑な判別モデルを採用することなども有望である。

5 つ目は、教師なし/半教師あり学習の枠組みに挑戦することである。本稿を含めた教師あり学習の手法においては実在宅状況の測定が不可欠であり、その部分のデータ収集のコストは依然高い。教師なし/半教師あり学習の枠組みで在宅推定を行うことができればデータ収集のコストの観点から有益である。またこの場合も、説明変数の 1 つとして電力比を採用することを視野に入れている。現在まで

に [7], [11]-[13], [15], [18] により教師なし学習, 半教師あり学習の研究が行われており, これらを参考文献として取り扱い進めていくことを視野に入れている。

謝辞 東京電力ホールディングス株式会社および東京電力エナジーパートナー株式会社の担当者の皆様には, 在宅推定が持つ課題やその社会的な意義について, 多くの示唆をいただいたことに感謝いたします。また, 中央大学理工学研究科の2回生の四ツ谷さんには, 評価方法についての有益なアドバイスをいただきました。ありがとうございました。

参考文献

- [1] Nguyen, T. A. and MarcoAiello: Energy intelligent buildings based on user activity: A survey, *Energy and Buildings*, Vol.56, pp.244-257 (2013).
- [2] 佐野芳樹, 松方直樹, 酒井貴洋, 増田 陸, 濱本望絵, 杉村 博, 一色正男: スマートメーターとHEMSを利用した実住宅での生活行動推定, マルチメディア, 分散協調とモバイルシンポジウム 2019 論文集, Vol.2019, pp.863-868 (2019).
- [3] 中野幸夫, 上野 剛: 電気の使い方から独居高齢者を見守るシステム, 電気学会論文誌 C, Vol.135, No.5, pp.471-480 (2015).
- [4] Ohsugi, S. and Koshizuka, N.: Delivery route optimization through occupancy prediction from electricity usage, 2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC) (2018).
- [5] Jin, M., Bekiaris-Liberis, N., Weekly, K., Spanos, C. and Bayen, A.: Sensing by Proxy: Occupancy Detection Based on Indoor CO2 Concentration, The 9th International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies (UBICOMM'15) (2015).
- [6] 服部俊一, 篠原靖志: 研究成果 研究報告書(電力中央研究所報告) 詳細情報 電力中央研究所, 技術報告, 電力中央研究所 (2016).
- [7] Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings: *Occupancy Detection from Electricity Consumption Data* (2013).
- [8] UbiComp '15: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing: *Household occupancy monitoring using electricity meters* (2015).
- [9] Gao, Y., Schay, A. and Hou, D.: Occupancy Detection in Smart Housing Using Both Aggregated and Appliance-Specific Power Consumption Data, 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA) (2018).
- [10] Yang, L., Ting, K. and Srivastava, M. B.: Inferring Occupancy from Opportunistically Available Sensor Data, 2014 IEEE International Conference on Pervasive Computing and Communications (PerCom) (2014).
- [11] Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings: *Non-Intrusive Occupancy Monitoring using Smart Meters* (2013).
- [12] Jin, M., Jia, R. and Spanos, C. J.: Virtual Occupancy Sensing: Using Smart Meters to Indicate Your Presence, *IEEE Transactions on Mobile Computing*, No.16, p.11 (2017).
- [13] Tang, G., Wu, K., Lei, J. and Xiao, W.: The Meter Tells You Are at Home! Non-Intrusive Occupancy Detection via Load Curve Data, 2015 IEEE International Conference on Smart Grid Communications (Smart-Grid-Comm): Data Management, Grid Analytics, and Dynamic Pricing (2014).
- [14] 石津紘太郎, 山口弘純, 東野輝夫: 主幹電力時系列データからの家庭内行動推定手法, マルチメディア, 分散協調とモバイルシンポジウム 2019 論文集, Vol.2019, pp.1011-1020 (2019).
- [15] DSG - Research Project: The ECO data set, ETHzurich (online), available from <<https://www.vs.inf.ethz.ch/res/show.html?what=eco-data>> (accessed 2021-02-28).
- [16] Occupancy Detection Methods - Resources-SoftServe, SoftServe (online), available from <<https://www.softserveinc.com/en-us/resources/occupancy-detection-methods>> (accessed 2021-10-31).
- [17] Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings: *The ECO Data Set and the Performance of Non-Intrusive Load Monitoring Algorithms* (2014).
- [18] EuroMETEO - Previsioni del tempo, EuroMETEO (online), available from <<http://www.eurometeo.com/italian/home>> (accessed 2021-02-28).
- [19] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and douard Duchesnay: Scikit-learn: Machine Learning in Python, *Journal of Machine Learning Research*, Vol.12, No.85, pp.2825-2830 (2011).
- [20] 篠原靖志, 上野 剛, 近藤修平: 住宅におけるガス流量計測に基づく需要推定 (その2)一用途分解手法の高精度化一, 【C】平成27年電気学会電子・情報・システム部門大会講演論文集, pp.1224-1229 (2015).



大島 悠

2020年中央大学理工学部経営システム工学科卒業。同年同大学大学院博士前期課程入学。在宅推定の研究に従事。統計学会会員。



石曾根 毅 (学生会員)

2019年明治大学総合数理学部現象数理学科卒業。2021年同大学大学院博士前期課程修了。同年同大学大学院博士後期課程入学。時系列解析の研究に従事。JSIAM,

JSS 各会員。



樋口 知之

1989 年東京大学大学院理学系研究科地球物理学専攻博士課程修了。同年統計数理研究所助手に採用，同研究所所長を務め，現在中央大学理工学部教授および同大学 AI・データサイエンスセンター所長。主としてベイジアンモデリングの研究に従事。電子情報通信学会などの会員。理学博士。