機械学習による食堂食事数予測問題への個人行動モデルの適用と

その効果分析

Application of individual behavior model in machine learning based cafeteria visitor prediction model and its effectiveness analysis

松江 清高*1	野田 英樹 <sup>*1</sup>	近藤 浩一*1
Kiyotaka Matsue	Hideki Noda	Koichi Kondo

\*1 東芝インフラシステムズ株式会社 Toshiba Infrastructure Systems & Solutions Corporation

Various data about individual behavior are becoming more easily available based on IoT related technologies. In this study, a machine learning based cafeteria visitor prediction model using individual behavior models has been developed. Each individual behavior model is built up from personal information in our office building. The effectiveness analysis of the model has been studied by comparing to another prediction model using conventional Random Forest algorithm. Authors clarified that RMSE of the conventional model and the developed model were 47.0 and 44.3, respectively. These results were better than the error in the cafeteria's prediction.

### 1. はじめに

近年 IoT (Internet of Things) の普及に伴い, さまざまなデー タが取得できるようになってきている.たとえば外食産業への適 用事例では,飲食店の売上予測を行うために,飲食店舗の隣 接催事場でのイベントの予想平均入場者数, 天候, 曜日などの 情報を用いて状態空間モデルから売上を予測する手法[山口 04]などが発表されている. 我々が入居するラゾーナ川崎東芝ビ ルは近隣の飲食店街に加え同ビル内での弁当販売やコンビニ 利用が可能なことから、日々5000名程度で変動している同ビル の在館者のうち同ビルの食堂利用者は半数程度,人間の経験 に基づく食堂食事数の予測誤差は 100 食程度であり、食堂事 業者より予測精度向上のニーズがあった.一方,同ビルは IoT 基盤を導入しており個人データを活用できる環境にある. 個人 データを食堂食事数予測に適用した事例は無いことから,筆者 らは個人の食堂利用有無を推定するモデル(以降,個人行動 モデルという)を適用した食堂食事数予測モデル(以降,検討モ デルという)を開発した.本稿は検討モデルと従来の食堂食事 数予測モデル(以降,ベースモデルという)との比較分析結果を まとめたものである.

# 2. 使用データと食事数の定義

ラゾーナ川崎東芝ビルの利用者はセキュリティゲートを通らな いと入館,退館できない仕組みになっている.セキュリティゲート 通過時に読み取る個人 ID データを利用することで入館,退館 を追跡でき,任意の時刻の在館者数を把握できる.また,同ビ ル内の食堂には POS (Point of sale)端末が設置され,入館者 は現金に加えて個人 ID カードでの支払いも可能である.その ため POS データを利用して個人の食堂利用回数や日々の食 堂食事数を取得できる.これらのデータに加えて天気情報も利 用してモデルを作成する.これは個人行動が天気や気温などに 依存する可能性を考慮するためである. また、食事数とは「昼間(11時00分~14時00分)の食堂での支払い回数」と定義する.支払い回数は支払い金額や食事メニューに依存しないものとする.

### 3. 予測モデル

### 3.1 ベースモデル

個人行動を加味せずマクロな視点で予測するモデル(以降, ベースモデルという)を作成する.ベースモデルの作成にはラン ダムフォレスト(以降, RF という)[Breiman 01]を用いて,変数選 択とモデル作成を繰り返すアルゴリズム[Diaz-Uriarte 06]を採 用する.全説明変数を用いて RFモデルを作成した後に説明変 数の重要度を算出し,その重要度の上位 80%の変数を用いて 再度 RF モデルを作成する,という処理を繰り返し実施し,作成 した複数のモデルから最適なモデルを選択するアルゴリズムで ある.これにより,直近 90 日間の食堂食事数やビル内の在館 者数などの実績値から食堂利用率を算出するモデルを作成す る.食堂食事数は当日の在館者数に食堂利用率を乗算して算 出する.

#### 3.2 個人行動モデルを用いた検討モデル

個人行動モデルを用いた検討モデルでは各個人を表1に示 す4つのいずれかに分類して食堂食事数を予測する.支払い 方法が現金の個人はベースモデルと同一手法で作成した RF 回帰モデルを作成する(分類1).支払い方法が個人 ID カード の個人は各個人の食堂利用状況から食堂利用有無を推定する 個人行動モデル(分類2-1~2-3)を作成する.これら分類1の 予測値と分類2-1~2-3の食堂利用有りと推定した個人の人数 を足し合わせることで食堂食事数を予測するモデル(検討モデ ル)を作成する.

分類 2-1 は食堂利用回数の少ない個人を対象にしている. データ数が少ないため,個人ごとにモデルを作成せず,個人 IDカードを利用した利用者全員のデータを使用して RF 分類モ デルを作成する.同様に,分類 2-2 は食堂利用率に偏りがある 個人を対象にしており,食堂利用回数または未利用回数のどち らか一方のデータが極端に少ない又は存在しない個人を対象

連絡先:松江 清高,東芝インフラシステムズ株式会社 技術 企画部,川崎市幸区堀川町 72 番地 34,044-331-1365, kiyotaka.matsue@toshiba.co.jp

にしている. 分類 2-1 同様, データ数が少なく特徴を捉えたモ デル作成が難しいため食堂利用率から食堂利用有無を推定す るモデルを個人ごとに作成する. 分類 2-3 は上記以外の個人を 対象にしたモデルであり個人ごとに RF 分類モデルを作成する. モデル 2~4を分類している表 1 の数値について, 食堂利用 回数 10 回とは 10 営業日に相当し, 個人行動が曜日に依存す ると想定した場合に祝日等を考慮するとモデルを作成する上で の必要最低限のデータ数と考えている. また, 食堂利用率 0.2 や 0.8 は, 一ヶ月間(営業日換算で 20 日間)を想定した場合,

食堂利用率 0.8 超は食堂未利用回数が 3 日間以下を意味し, 週 5 日間, 食堂を利用するケースが必ず存在することを意味す る. 食堂未利用データが少ない上に行動に偏りのある個人は特 徴抽出が難しい個人と考えられる. ま1 検討モデルの詳細

X 1 换时 C / / C / 开柏					
分類	支払い	訓練	食堂利用	食堂	
	方法	データ	回数 N	利用率 $\alpha$	
1.現金払い (RF回帰モデル)	現金	現金 全員	_	_	
2-1. 訓練データ少 (RF分類モデル)	個人 ID	個人 ID 全員	N<10	_	
2-2. 訓練データ偏り (食堂利用率モデル)	個人 ID	個人 ID	10≦N	α<0.2, 0.8<α	
2-3. 上記以外 (RF分類モデル)	個人 ID	個人 ID	10≦N	$0.2 \leq \alpha \leq 0.8$	

### 4. 評価結果と効果分析

ベースモデルおよび検討モデルを R 言語の汎用ライブラリ ranger[Wright 15]を用いて作成し食堂食事数の予測結果を評 価した.両モデルの RMSE (Root Mean Square Error) と MAPE (Mean Absolute Percentage Error)の値を表 2 に示す.ベースモ デルは RMSE で 47.0 となり,人間の経験に基づく予測値の 100 食と比較して良好な結果が得られた.

表2 各モデルの RMSEと MAPE

	RMSE	MAPE
ベースモデル	47.0	1.5
検討モデル	67.2	2.2

しかし検討モデルはベースモデルと比較して RMSE で 20.2, MAPE で 0.7 悪化する結果となった. そこで精度悪化の理由を 探るため個人の動きを調べた. 図 1 に 2016 年 12 月 1 日の個 人行動モデル適用者の食堂利用率のヒストグラムと個人行動モ デルの予測値,実績値,利用率 0.5 以上の人数の推移を示す.



(a)食堂利用率ヒストグラム (b)利用人数推移(予測と実績) 図1 個人行動モデル適用者の分析結果

図 1(a)の食堂利用率ヒストグラムより食堂利用率 0.5 以上の 人数のほうが 0.5 未満の人数より多い. また図 1(b)を見ると全日 程において予測値が実績値を上回っており,さらに利用率 0.5 以上の人数が予測値を上回っている. つまり食堂利用率の高 い個人が多いために,食堂利用有りと推定する個人が増えてし まい,結果として全体の予測食事数を押し上げていると考えら れる. そこで,この推定値と実績値の差分を学習し予測モデル を補正することで予測精度の向上を試みた. 一ヶ月間の平均残 差を個人行動モデルに加算して補正したモデルの評価結果を 図 2 と表 3 に示す. その結果, RMSE で 22.92, MAPE で 0.8 改善し,ベースモデルの予測精度を若干上回った.



表3 補正を適用した場合の RMSE と MAPE

	RMSE	MAPE
検討モデル(補正有り)	44.3	1.4

## 5. おわりに

食堂食事数予測に個人行動モデルを適用し,以下の 3 点を 明らかにした.

- ・マクロな視点で予測したベースモデルによる予測値は人間 の経験に基づく予測結果より良好な結果であること
- ・食堂利用率ヒストグラムが 0.5 を中心に正規分布していない 場合には個人行動モデルに補正項が必要なこと
- ・個人行動モデルに補正項を付加することでベースモデルの 予測精度を若干上回る結果が得られること

更なる精度向上を目指し、個人行動モデル作成時の数値最 適化(食堂利用回数および食堂利用率によるモデル分類時の 最適化)や食堂利用率ヒストグラム形状に基づいた個人行動モ デルの補正方法の検討などの課題を解決していく.

### 参考文献

- [山口 04] 山口類, 土屋映子, 樋口和之: 状態空間モデルを 用いた飲食店売上の要因分解, 日本オペレーションズ・リサ ーチ学会, 2004.
- [Breiman 01] Breiman, L.,: Random Forest, Machine Learning 45, 2001.
- [Diaz-Uriarte 06] Diaz-Uriarte, R., and Andres, S. A., Gene selection and classification of microarray data using random forest, BMC Bioinformatics, January 6, 2006.
- [Wright 15] Wright, M. N. and Ziegler, A., ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R, arXiv: 1508.04409v1, 18 Aug 2015.