

# 区役所の窓口を対象とした来庁者の滞在時間分析モデルに関する研究

1X12C062-8 数藤 光太郎  
指導教員 後藤 正幸

## 1 研究背景・目的

近年、区役所等の行政機関においても情報システムの導入が進み、来庁者の行動ログデータを蓄積し、活用する動きが見られる。しかし、現状では、区役所において膨大な量の行動ログデータが活用されている例は少なく、これらのデータをサービス改善に活用することが求められている。そこで、本研究では、某区役所における行動ログデータを用いて、経営改善のための一助となるモデルの提案を目指す。

本研究で対象とする区役所では Web ページを通じ、現時点での窓口ごとの予測待ち時間を情報提供している。しかし、来庁者にとっては待ち時間のみならず、各窓口での待ち時間に各窓口で行われる業務の所要時間（以下、業務所要時間）を加えた滞在時間も知ることが有益であると考えられる。このため、本研究では滞在時間の予測を考えるものとする。

滞在時間の予測モデルを構築する際、考慮すべき点が2点ある。1点目は、来庁者が複数窓口を訪れる（以下、これを巡回と呼び、このパターンを巡回パターンとして定義する）場合は単一窓口のみ訪れる場合と業務所要時間が異なるケースが存在するという点である。2点目は、行動ログデータには業務や巡回パターンによってはデータ数が少ないものがあり、これも含めて正確な予測を行えるモデルを考える必要があるという点がある。

1点目に対して、現行のシステムを用いて来庁者の滞在時間を予測することを考えた場合、待ち時間は現時点での待ち人数と窓口ごとの平均処理時間を用いて算出している。しかし、現行のシステムでは単一窓口のみ訪問するケースと区別せず平均処理時間を算出するため、予測精度が低くなると考えられる。本研究では巡回の有無により業務所要時間に差異が生じる（以下、この差異を巡回効果と呼ぶ）という仮定に基づき、滞在時間予測モデルを構築することを考える。2点目に対しては、Aspect Model(以下、AM)[1]を導入することを考える。AMは、主に推薦システムにおいてユーザとアイテムの共起関係を潜在クラスを用いて表すモデルとして用いられており、本研究では業務と業務所要時間、巡回パターンと巡回効果の間の共起関係を表すモデルとして応用する。このモデルでは類似した傾向を持つ業務所要時間、巡回効果を持つ業務、巡回パターンが存在し、グルーピング可能であるという仮定に基づき、業務、巡回パターンについて潜在クラスへの所属確率をもとに予測値を算出する。これにより、データ数の少ない業務、巡回パターンについても正確な予測が行えると期待できる。

以上より、本研究では巡回効果の存在を仮定し、AMを用いることで巡回効果を考慮可能な滞在時間予測手法を提案する。また、提案手法の有効性を某区役所の行動ログデータを用いた実験により評価する。

## 2 基本分析

本研究では、経営科学系研究部会連合協議会主催、平成27年度データ解析コンペティションにおいて提供された某区役所の2014年11月～2015年7月までの行動ログデータを分析対象とする。本分析で対象とする区役所では、窓口番号が付与された複数の窓口があり、各窓口で1つ以上の業務が処理されている。区役所の来庁者は、それぞれの用件に応じて1つまたは複数の窓口を巡回する。ここで、これらの来庁者の行動ログデータには1つの窓口を訪問して用件を終える来庁者と複数の窓口を巡回する来庁者が混在しているが、本研究では巡回するデータのうち9割を占める2つの窓口を巡回する場合の来庁者の行動ログに着目して予備分析を行った。その結果、来庁者が複数の窓口を巡回する場合、処理された個々の業務の業務所要時間と、巡回しない場合で各々の

業務のみを処理する場合の業務所要時間の間に差異が生じることが分かった。また、各業務ごとのデータ数にばらつきが大きいことも分かった。以上のことから、滞在時間を予測する場合巡回の有無と各業務のデータ数のばらつきを考慮することが予測モデルを構築する際に必要であると示唆される。

## 3 提案手法

### 3.1 概要

予備分析の結果を踏まえ、ある巡回パターンについて、巡回した各業務の業務所要時間の合計と巡回しない場合の同業務の業務所要時間の合計との差異を残差、残差の予測値を巡回効果と定義し、この巡回効果を考慮した滞在時間の予測モデルを構築する。ここで、業務または巡回パターンの中にはデータ数が十分に確保されていないものが存在しており、単純に業務または巡回パターンごとのデータで予測値を算出した場合、予測精度に差が出る可能性があるため、このデータ数の差を補うための予測モデルが必要になる。このため、類似した業務所要時間、残差を持つ業務、巡回パターンが存在し、グルーピング可能であるという仮定を置き、AMを用いて業務とその業務所要時間、もしくは巡回パターンとその残差を潜在クラスに対して確率的にクラスタリングすることで、業務、巡回パターンの各クラスに対する所属確率を考慮して予測を行える。このため、データが少ない業務もしくはパターンについても、業務所要時間もしくは巡回効果に対する予測の精度向上が期待できる。

### 3.2 Aspect Model を用いた滞在時間予測モデル

提案モデルでは、まず各業務について巡回しない場合の業務所要時間の予測値を算出し、それをもとに巡回する場合での業務所要時間の残差を計算する。次に、巡回パターンと算出された残差を用いて、業務所要時間の巡回効果を算出し、これを用いて滞在時間の予測値を算出するという4ステップからなる。以下、各ステップについて説明する。なお、本モデルにおいて  $\mathcal{A} = \{a_i : 1 \leq i \leq I\}$  を業務集合、 $\mathcal{B} = \{b_\ell : 1 \leq \ell \leq L\}$  を2業務を巡回する場合の巡回パターンの集合とする。ただし、 $b_\ell = (b_{\ell_1}, b_{\ell_2})$ 、 $b_{\ell_1}, b_{\ell_2} \in \mathcal{A}$  であり、 $b_{\ell_1}$  を1番目に行われる業務、 $b_{\ell_2}$  を2番目に行われる業務とする。また、 $\mathcal{B}$  には実際に行動ログデータに出現したもののみが含まれるものとする。

#### Step1 業務所要時間の予測値の算出

$j$  番目の単一窓口を利用する来庁者が訪れた窓口で処理した業務  $x_j \in \mathcal{A}$  と、業務所要時間  $u_j \in \mathcal{R}$  の組を1レコードとし、これら  $n_1$  個からなる学習データの集合  $\{(x_j, u_j)\}_{j=1}^{n_1}$  を用意する。本研究では  $\mathcal{Z}^\alpha = \{z_{g_1}^\alpha : 1 \leq g_1 \leq G_1\}$  を潜在クラス集合として、学習データ  $(x_j, u_j)$  に関する確率モデルとして、式(1)を仮定する。なお、 $P(a_{i(x_j)} | z_{g_1}^\alpha)$  には  $I$  項分布を、 $P(u_j | z_{g_1}^\alpha)$  には平均  $\mu_{g_1}^\alpha$ 、標準偏差  $\sigma_{g_1}^\alpha$  をもつ正規分布を仮定するものとする。また、 $i(x)$  は  $x = a_i$  となるような  $i$  を表すものとする。

$$P(x_j, u_j) = \sum_{g_1=1}^{G_1} P(z_{g_1}^\alpha) P(a_{i(x_j)} | z_{g_1}^\alpha) P(u_j | z_{g_1}^\alpha) \quad (1)$$

ここで、式(1)のパラメータは  $\mu_{g_1}^\alpha, (\sigma_{g_1}^\alpha)^2, P(z_{g_1}^\alpha), P(a_{i(x_j)} | z_{g_1}^\alpha)$  であり、EMアルゴリズム[2]によりこれらを推定する。

また、得られた推定値を  $\hat{\mu}_{g_1}^\alpha, \hat{P}(z_{g_1}^\alpha | a_i)$  として、業務  $a_i$  に対する業務所要時間の予測値  $\hat{t}_i$  を式(2)で与える。

$$\hat{t}_i = \sum_{g_1=1}^{G_1} \hat{\mu}_{g_1}^\alpha \hat{P}(z_{g_1}^\alpha | a_i) \quad (2)$$

## Step2 業務所要時間の巡回効果の計算

$k$  番目の巡回する来庁者について、 $\mathbf{y}_k = (y_{k_1}, y_{k_2}) \in \mathcal{B}$  を  $y_{k_1}, y_{k_2}$  の順に巡回した場合の巡回パターンとし、この巡回に関する残差を  $\epsilon_k$  とする。このとき、 $\epsilon_k$  は式 (3) で算出される。なお、 $v_{k_1}, v_{k_2}$  をそれぞれ業務  $y_{k_1}, y_{k_2}$  に対する業務所要時間の実測値とする。

$$\epsilon_k = (v_{k_1} + v_{k_2}) - \{\hat{t}_{i(y_{k_1})} + \hat{t}_{i(y_{k_2})}\} \quad (3)$$

巡回パターンとその残差の組の集合  $\{(\mathbf{y}_k, \epsilon_k)\}_{k=1}^{n_2}$  を Step3 での学習データとする。

## Step3 巡回パターンに対する業務所要時間の巡回効果の予測値の算出

$\mathcal{Z}^\beta = \{z_{g_2}^\beta : 1 \leq g_2 \leq G_2\}$  を潜在クラス集合とすると、確率モデルは  $\mathcal{Z}^\beta$  を用いて式 (4) で表される。なお、 $P(\mathbf{b}_{\ell(\mathbf{y}_k)} | z_{g_2}^\beta)$  には  $L$  項分布を、 $P(\epsilon_k | z_{g_2}^\beta)$  は平均を  $\mu_{g_2}^\beta$ 、標準偏差を  $\sigma_{g_2}^\beta$  とする正規分布を仮定するものとする。また  $\ell(\mathbf{y})$  を  $y_1 = b_{\ell_1}, y_2 = b_{\ell_2}$  となるような  $\ell$  を表すものとする。

$$P(\mathbf{y}_k, \epsilon_k) = \sum_{g_2=1}^{G_2} P(z_{g_2}^\beta) P(\mathbf{b}_{\ell(\mathbf{y}_k)} | z_{g_2}^\beta) P(\epsilon_k | z_{g_2}^\beta) \quad (4)$$

式 (4) のパラメータは  $\mu_{g_2}^\beta, (\sigma_{g_2}^\beta)^2, P(z_{g_2}^\beta), P(\mathbf{b}_{\ell(\mathbf{y}_k)} | z_{g_2}^\beta)$  であり、EM アルゴリズムによりこれらを推定する。

また、得られた推定値を  $\hat{\mu}_{g_2}^\beta, \hat{P}(z_{g_2}^\beta | \mathbf{b}_\ell)$  とし、巡回パターン  $\mathbf{b}_\ell$  に対する巡回効果  $\hat{r}_\ell$  を式 (5) で与える。

$$\hat{r}_\ell = \sum_{g_2=1}^{G_2} \hat{\mu}_{g_2}^\beta \hat{P}(z_{g_2}^\beta | \mathbf{b}_\ell) \quad (5)$$

## Step4 滞在時間の予測値の算出

来庁者  $f$  が  $\mathbf{y}_f = (y_{f_1}, y_{f_2})$  と巡回する時の滞在時間の予測値  $\hat{T}_f$  を式 (6) で算出する。

$$\begin{aligned} \hat{T}_f &= \hat{r}_{\ell(\mathbf{y}_f)} + \hat{t}_{i(y_{f_1})} + \hat{t}_{i(y_{f_2})} \\ &+ \frac{m_{i(y_{f_1})} \hat{t}_{i(y_{f_1})}}{H_{i(y_{f_1})}} + \frac{m_{i(y_{f_2})} \hat{t}_{i(y_{f_2})}}{H_{i(y_{f_2})}} \end{aligned} \quad (6)$$

ただし、 $H_{i(y_{f_1})}, H_{i(y_{f_2})}$  はそれぞれ業務  $y_{f_1}, y_{f_2}$  における窓口数、 $m_{i(y_{f_1})}, m_{i(y_{f_2})}$  はそれぞれ業務  $y_{f_1}, y_{f_2}$  の窓口における、来庁者  $f$  の来庁時点での待ち人数とする。

## 4 実験と考察

提案手法の有効性を検証するため、2 節で用いた行動ログデータを利用し、実験を行った。

### 4.1 実験条件

2014 年 11 月 25 日から 2015 年 6 月 30 日を学習に用いる期間とし、2015 年 7 月 1 日から 31 日をテストに用いる期間とする。業務数は  $I = 28$ 、巡回パターン数は  $L = 359$ 、Step1, Step3 での学習データ数はそれぞれ  $n_1 = 846, 392$ 、 $n_2 = 13, 684$ 、Step4 でテスト対象とするデータ数は  $F = 2, 306$  である。また、評価値は平均絶対誤差 (MAE) とする。 $f$  番目のデータに関する実測値を  $T_f$  とすると、MAE は式 (7) により与えられる。

$$MAE = \frac{1}{F} \sum_{f=1}^F |T_f - \hat{T}_f| \quad (7)$$

比較手法は次に示す 2 手法を用いた。

**比較手法 1** : 提案手法において巡回効果を導入せず、滞在時間の予測値を算出する手法

**比較手法 2** : 学習データから業務ごとの業務所要時間の平均値を  $\hat{t}_i$  とし、 $\hat{t}_i$  の値を元に残差  $\epsilon_k$  を算出する。巡回パターンごとに残差の平均値をとり  $\hat{r}_\ell$  とし、滞在時間の予測値を算出する。

## 4.2 実験結果と考察

実験結果を図 1 に示す。

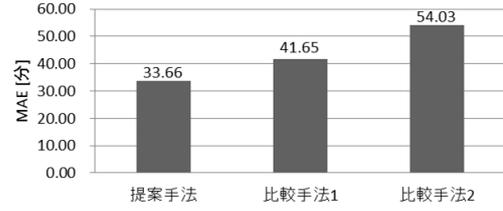


図 1. 実験結果

図 1 より、提案手法は業務所要時間の予測に平均値を用いる比較手法 2 より 20 分ほど小さい MAE を示した。これは、データ数が少ない業務または巡回パターンについて、AM を用いてクラスターへの所属確率をもとに予測値を算出し、推定すべきパラメータ数を減らしたことで、より正確な業務所要時間の予測を行えたためであると考えられる。また、巡回を行う場合と行わない場合の業務所要時間の誤差を考慮しない場合の滞在時間予測モデル (比較手法 1) と比較して、提案手法が 8 分ほど小さい MAE を示した。このことから、巡回効果を考慮する意義が示せ、これにより提案手法の有効性が示された。図 2 に、Step3 において計算した巡回効果のうち、絶対値の大きい巡回パターンについて上位 20 件をまとめ、表 1 にその内訳の中で、同じ業務が複数回出現しているものについて示す。

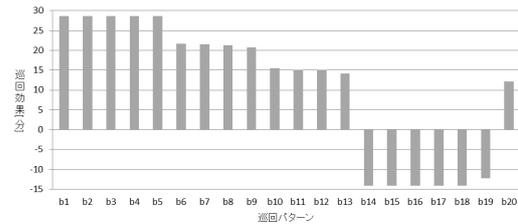


図 2. 巡回効果の絶対値の大きい上位 20 パターン

表 1. 上位 20 パターンの内訳 (抜粋)

	$\mathbf{b}_3$	$\mathbf{b}_5$	$\mathbf{b}_{10}$	$\mathbf{b}_{12}$	$\mathbf{b}_{13}$	$\mathbf{b}_{14}$	$\mathbf{b}_{16}$	$\mathbf{b}_{17}$	$\mathbf{b}_{19}$
1 番目	$a_{15}$	$a_{21}$	$a_{10}$	$a_{15}$	$a_{25}$	$a_4$	$a_{23}$	$a_{26}$	$a_{14}$
2 番目	$a_{23}$	$a_{10}$	$a_{20}$	$a_6$	$a_{10}$	$a_{17}$	$a_{17}$	$a_{14}$	$a_1$

図 2 において、 $\hat{r}_\ell$  が正の値を取る巡回パターンには、 $a_{10}, a_{15}$  などの特定の業務が多く含まれていることがわかる。同様に、 $\hat{r}_\ell$  が負となる巡回パターンには、 $a_{14}, a_{17}$  などの業務が多く見られる。これらのことから、巡回効果  $\hat{r}_\ell$  の値は、特定の業務に影響を受けていると考えられ、巡回効果  $\hat{r}_\ell$  が正となる業務を扱う窓口は他の窓口との連携を工夫することで、来庁者の滞在時間を削減でき、顧客満足度の向上を図ることができると考えられる。

## 5 まとめと今後の課題

本研究では区役所で 2 業務について巡回する場合の滞在時間予測のためのモデルを提案した。また、実際のデータを用い、従来の待ち時間の計算法に対する優位性と、巡回効果を考慮することの意義を示せた。しかし、本研究では窓口ごとの待ち人数の時間変化を直接的に評価しておらず、また 3 業務以上の巡回は対象外であった。このため、今後の課題としては窓口ごとの待ち人数の変化を考慮した予測モデルを構築すること、また 3 業務以上の巡回についても考慮することなどが挙げられる。

### 参考文献

- [1] T. Hofmann, "Probabilistic Latent Semantic Analysis," *Proc. of UAI'99*, pp.289-296, 1999.
- [2] 宮川雅巳, "EM アルゴリズムとその周辺," 応用統計学, Vol.16, No.1, pp.1-21, 1987.