情報処理学会論文誌　数理モデル化と応用　Vol.8　No.1　87–95 (Mar. 2015)

ウィンドウ遷移ログを用いた従業員のユーザ行動モデルの時系列変化分析

齋藤 良平[1,a)]

受付日 2014年8月28日，再受付日 2014年10月15日，
採録日 2014年12月24日

概要：ユーザ行動モデルは単位時間内に使用されたアプリケーションの組合せとその遷移で PC 操作を表現する確率的モデルである．このモデルを企業内 PC から取得されたウィンドウ遷移ログを用いて構築することで業務上の操作の特徴を抽出することが可能である．本稿では，ユーザ行動モデルがどのように変化していくかを可視化する手法を述べ，実際の企業から取得した約 4 年間分のウィンドウ遷移ログを用い，ユーザ行動モデルの変化を分析することで，従業員の業務と PC 操作の関連を検証した．

キーワード：ユーザ行動モデル，隠れマルコフモデル，イベント要約，グラフカーネル，カーネル PCA，ログ分析

User Behavior Model Analysis Using Long Term Window Transition Logs

Ryohei Saito[1,a)]

Received: August 28, 2014, Revised: October 15, 2014,
Accepted: December 24, 2014

Abstract: This paper describes about a method of inferring long term changes of working styles/behaviors of employees, through analysis of User Behavior Model generated from long term window transition logs in the user PCs. User Behavior Model is a probabilistic model to represent PC operations by combinations of applications and their transitions. User Behavior Model created from window transition logs recorded in individual computers for employees in a company represents working styles/behaviors. In this paper, by analyzing User Behavior Model using window transition logs recorded for about 4 years, relations between PC operations and working styles/behavior are clarified.

Keywords: User Behavior Model, hidden Markov model, event summarization, graph kernel, kernel PCA, log analysis

1. はじめに

近年の IT の発展により，企業における PC の使用はごくありふれた光景となっている．PC の使用はメールや資料作成に始まり，ブラウザを用いたイントラネットシステムの利用や業務アプリケーションの利用等多岐にわたり，業務の多くを PC を用いて行う従業員も少なくない．しかしながら，PC を用いた業務の評価は成果物をもって判断されることが多く，その過程が評価されることは少ないといえる．これは，PC 上の操作は他人が把握することが難しく，個人作業であることが多いためである．同じ業務であっても，それに関わる PC 上の操作は人により異なり，また，同一人物であっても習熟度や環境によって大きく異なることが考えられ，この違いを客観的に判断できるならば業務の生産性や効率を推定する手がかりとなる．

一方で，これら企業内の PC に記録されている重要情報を狙う外部からの不正アクセスや従業員による内部情報漏えい対策のために，多くの企業でセキュリティ対策ソフトウェアの導入が行われている．これらのソフトウェアではインシデント発生時の状況のトレースや従業員の不正な行動の抑止のために，PC 上の操作ログを記録する機能を

[1] ハミングヘッズ株式会社
Humming Heads, Inc., Chuo, Tokyo 104–0052, Japan
[a)] ryouhei-s@hummingheads.co.jp

持っている．この操作ログの 1 つとして操作したウィンドウをトレースするためのウィンドウ遷移ログがある．

一般的な企業では従業員は担当する業務が決まっており，その担当業務において使用するアプリケーションは限定的である．このため家庭における PC の使用に比べて PC の使用方法が限定的であるといえる．

本研究では，企業内で取得したウィンドウ遷移ログに着目し，筆者らが提案したユーザ行動モデル [1] を約 4 年間分のログを用いて構築，分析する手法を示し，ユーザ行動モデルの変化と業務の関連を検証した．ユーザ行動モデルは単位時間内に使用されたアプリケーションとその遷移を用いて業務の作業を表現することから，同一の従業員が同一の業務を行っていても異なるモデルが構築される．このような違いが発生する要因を分析できるならば，生産性や作業効率の推定に有益な手がかりとなる．これを行うためには，ログと業務の間に強い関連性が必要であり，提案する手法でこの関連性が導けることを示す．

また，ウィンドウ遷移ログのように低レベルなログは短期間においても膨大に記録され，その傾向やパターンを直感的に人が判断するのは困難である．ユーザ行動モデルを用いてログの情報を抽象化，可視化することで長期間の大量のログに対する分析を容易にすることが可能であることを示す．

本稿では次節で関連研究について述べ，2 章で分析対象とするウィンドウ遷移ログとユーザ行動モデルの概要を述べる．3 章ではユーザ行動モデルの時系列変化を可視化する手法と実際に企業から取得したログを用いた分析結果について述べる．

1.1　関連研究

ログとして代表される Web のアクセスログやモバイルデバイスの加速度センサ等のライフログからユーザの行動を抽出や分類する研究は数多く行われている [2], [3]．

PC の操作ログから人の行動や状態を分析する試みは古くから研究されており，2000 年に Hilbert らによってまとめられているように，ユーザインタフェースのイベントからソフトウェアのユーザビリティを推定することを目的に行われることが多い [4]．

近年では，PC 操作ログから人の状態や行動パターンを抽出する研究が多くなされている．Singh らは Web サーバや Firewall の通信ログから利用者が攻撃者かを判断する手法を提案している [5]．Beauvisage はアプリケーションの利用ログを用いて PC の利用状況および目的を分析している [6]．ウィンドウ遷移に着目した研究としては，Tak らによってウィンドウの切替えを可視化するツールの作成が行われている [7]．また，Suzuki らによってウィンドウの切替え頻度から従業員が休憩中かどうかを推定する手法が提案されている [8]．

企業内の操作ログに着目した研究としては鳥羽らによってキーボードのログから得られる特徴量と従業員のストレス量の相関関係が報告されている [9]．また，平山らは PC 操作ログの時系列データから特徴的なパターンを検出する手法の提案を行っている [10]．

ログ分析を行う際の，膨大なログデータを可視化，要約する研究は非常にさかんに行われている．本研究で使用するユーザ行動モデルも強く依存する要約手法として Wang らの提案した手法があげられる [11]．

これらのように，ログの分析に関する研究は非常に注目されているが，PC の操作ログとユーザ行動を関連付ける研究はまだまだ少ない．それゆえ，本研究で扱うような数年に及ぶ実社会での PC 操作ログを対象とした行動分析手法の提案や分析結果は筆者の知る限りでは存在しない．

2.　ウィンドウ遷移ログとユーザ行動モデル

2.1　ウィンドウ遷移ログ

本節では分析対象となるウィンドウ遷移ログについて説明する．ここでは Microsoft 社製 OS Windows®を対象として説明する．

ウィンドウ遷移ログはウィンドウシステムを搭載した OS における，フォアグラウンドなウィンドウの遷移を記録したログである．

フォアグラウンドなウィンドウとは通常キーボード入力を受付けるウィンドウであり，多くの場合でディスプレイ上の最前面に表示される．

フォアグラウンドウィンドウはマウスやキーボード操作等で切り替えることができる．**図 1** の例では，左図は Mail がフォアグラウンドであり，マウスで Web Browser のウィンドウをクリックすると右図のように Web Browser がフォアグラウンドのウィンドウとなる．この右図の状態でキーボードから ‘4’ の入力を行うと，Web Browser に ‘4’ という入力が送られる．

本研究ではフォアグラウンドウィンドウをユーザが実際に作業しているウィンドウと見なして，その遷移に関する分析を行う．

ウィンドウ遷移ログはフォアグラウンドであったウィンドウを保持するアプリケーション名とタイムスタンプおよびフォアグラウンドであった時間（秒）が時系列的に記録される．**表 1** はウィンドウ遷移ログの例である．この例で

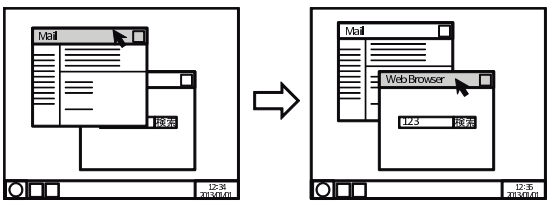

図 1　フォアグラウンドウィンドウの切り替わり
Fig. 1　The switching of the foreground window.

表 1　ウィンドウ遷移ログ
Table 1　Window trainsition logs.

Start Time	Application	Duration (sec)
2010-04-01 8:00:30	Web	403
2010-04-01 8:07:13	Mail	165
2010-04-01 8:10:15	Editor	503
2010-04-01 8:18:38	Web	386
2010-04-01 8:25:04	Editor	303
2010-04-01 8:30:20	Mail	440
⋮	⋮	⋮

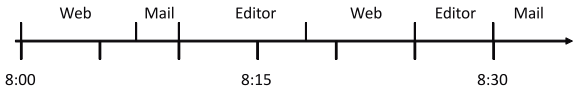

図 2　ウィンドウ遷移
Fig. 2　Window trainsition.

は図 **2** のように朝 8:00:30 から 403 秒間 Web ブラウザを使用し，その後メーラを 165 秒間使用していたということが推定できる．また，マウスやキーボードの活動を記録したログを合わせて用いることで，離席時のようなフォアグラウンドウィンドウであっても使用していない時間を除外することが可能である．

ウィンドウ遷移ログは市販のログ取得ソフトウェアで取得可能である．本研究ではログ取得ソフトウェアを導入している企業から取得したログを用いて分析を行う．

2.2　ユーザ行動モデル

本節ではウィンドウ遷移ログから作成されるユーザ行動モデルの概要とその構築手法および分析手法について記述する．

筆者らが提案したユーザ行動モデルは隠れマルコフモデル（Hidden Markov Model: HMM）を基とした確率的モデルであり，ウィンドウ遷移ログから得られる単位時間内に使用されたアプリケーションの組合せとその遷移を表現する [1]．

従業員が業務を行う際には，各業務に応じた PC 操作の作業手順がある．たとえば，メールで依頼された資料を作成し，送り返す場合にはメーラと文章作成ソフトウェアが使用されることが予想できる．これら 2 つのアプリケーションが近傍の時間内で利用される場合は，上記業務の作業を行っているとして，1 つのタスクとして考える．ユーザ行動モデルではこのタスクを抽象化し，単純にメーラと文章作成ソフトウェアを使用する作業として扱う．このため，1 つの抽象化したタスクで複数の実作業を含む可能性もあるが，そのタスク間の遷移も合わせて用いることで抽象化されたユーザの行動と業務を表現することができる．

このように，作成されたユーザ行動モデルは膨大な量のウィンドウ遷移ログを抽象化したモデルであり，このモデルに対して分析を行うことで，ログから特徴やパターンを検出する．

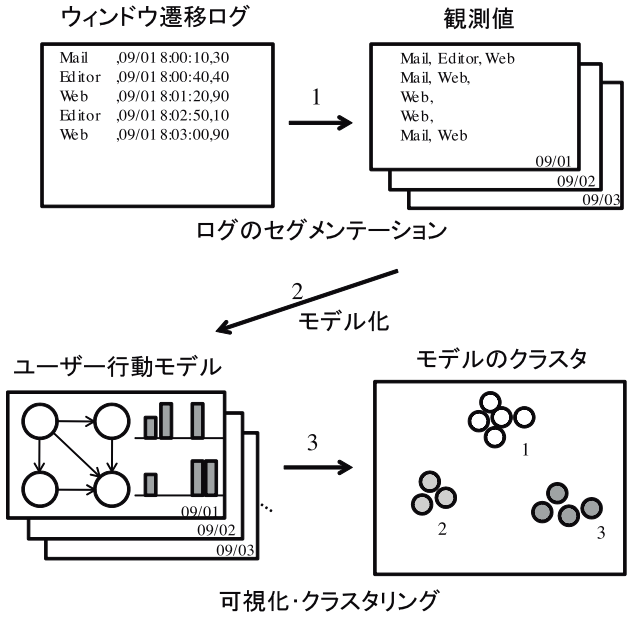

図 3　分析の流れ
Fig. 3　Approach.

ユーザ行動モデルは図 **3** のようにウィンドウ遷移ログから HMM のパラメータ推定を用いて抽象度の高いユーザ行動モデルを作成することで，可視化，比較を行うことが可能である．

2.3　ログデータからの観測値

本節ではウィンドウ遷移ログからユーザ行動モデルの観測値の作成方法について説明する．

前述のとおり，ユーザ行動モデルでは単位時間内に使用されたアプリケーションの組合せを表現するため，観測値としてアプリケーションの組合せを用いる．

$\mathcal{E}$ をログ中 Mail や Web のように記録されているアプリケーション $\{e_1, \ldots, e_m\}$ の集合とする．ある時刻におけるフォアグラウンドウィンドウに対応するアプリケーションをその時刻に観測されたイベントとして $\mathcal{E}$ を用いて表す．

等間隔に測定された離散的なインターバル $t \in \{0, \ldots, T\}$ に対して，インターバル $[t, t+1)$ 中に観測されたイベントの集合を観測値とする．ここで，$[t, t+1)$ は $t \leq t' < t+1$ を満たす期間 t' とする．例として $t \in \{8\,00, 8\,10 \ldots, 21\,00\}$ とすると $[8\,00, 8\,10)$ は 8:00 から 8:09:59 までの期間である．インターバル $[t, t+1)$ に観測されたイベントの集合を o_t とし，$o_t = \{e \in \mathcal{E} \mid e \text{ occurs in } [t, t+1)\}$ と定義する．これにより，ログから観測値としてイベント集合の系列を得る．

$$\boldsymbol{O} = (o_1, \ldots, o_T), \text{ where } o_t \subseteq \mathcal{E} \text{ for } t \in \{1, \ldots, T\}.$$

ウィンドウ遷移をタイムテーブルで表した図とその時刻におけるイベント集合を図 **4** に示す．

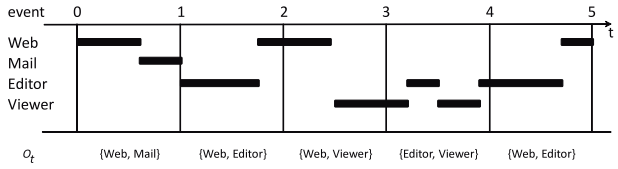

図 4　ログから作成される観測値
Fig. 4　Observations obtained from log data.

2.3.1　タスク要約

業務において従業員は K 個のタスク状態を持つとして，それぞれに固有の状態 $s_i \in \{s_1, \ldots, s_K\}$ を与えることとする．したがって，タスク状態は状態の集合として $S = \{s_1, \ldots, s_K\}$ と表記される．

ユーザ行動モデルでは，タスク状態は各タスク状態において観測されたアプリケーションの出現確率によって表現される．これは直感的な観点から自然なモデル化といえる．たとえば，ウェブページを作成するという業務を行う場合，ウェブブラウザと HTML エディタを主に使用し，それ以外のアプリケーションの出現頻度に比べ圧倒的に高くなることが容易に予想できる．それゆえ，タスク $s \in S$ でユーザが使用するアプリケーションの集合 $o \subseteq \mathcal{E}$ は出力確率 $b_s(o)$ に従うと想定する．ここで，すべての状態 $s \in S$ に対する出力確率 $b_s(o)$ の集合を B とし，$o \subseteq \mathcal{E}$ とする．

与えられた集合 B に対して状態 s でアプリケーション e が使用される確率は $P_s(e)$ で与えられ，以下のように推定される．

$$P_s(e) = \sum_{o \subseteq \mathcal{E} \text{ such that } e \in o} b_s(o)$$

したがって，アプリケーションの数を m としたとき，タスク s においてそれぞれのアプリケーション $e_1, \ldots, e_m$ は確率 $P_s(e_1), \ldots, P_s(e_m)$ で出現する．タスク s に対して，これら出現確率を m 次元のベクトルとして

$$M_s = (P_s(e_1), \ldots, P_s(e_m))$$

と定義し，s の**タスク要約**と呼ぶことにする．

2.4　ユーザ行動モデルの推定

HMM はそのパラメータ $\lambda = (\Pi, A, B)$ で定義される．HMM の各隠れ状態はそれぞれタスク状態と対応し，その状態がアプリケーションの組合せを生成する確率は B によって決定され，初期確率 Π，遷移確率 A によって系列が生成される．

ユーザ行動モデルはウィンドウ遷移ログから得られた観測値の系列を用いて HMM のパラメータ λ の推定を行うことで得ることができる．パラメータ推定は非常によく知られた Baum-Welch アルゴリズムを用いて行う [12]．

2.5　ユーザ行動モデルの比較

前節で述べた方法で得られるユーザ行動モデルはエルゴード的マルコフモデルである．

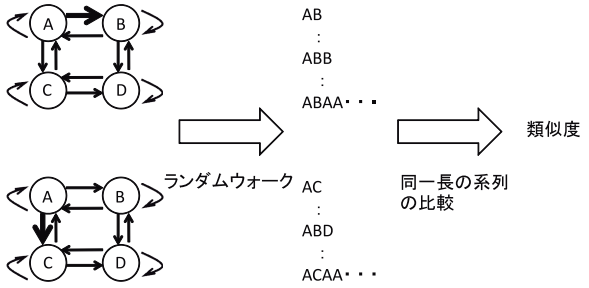

図 5　ランダムウォーカを用いたグラフ間の類似度計算
Fig. 5　The similarity measure between two graphs by random walker.

ユーザ行動モデルを各頂点がタスク要約でラベル付けされ，各辺がタスク間の遷移確率でラベル付けされた有向グラフであるとし，このグラフをタスク遷移グラフと呼ぶことにする．2 つのタスク遷移グラフの類似性測度として，Kashima らが提案したラベル付きグラフ間の周辺化カーネル [13] を用いる．

直感的には**図 5** のように，このカーネルは 2 つのグラフ上を 2 つのランダムウォーカが移動した軌跡を比較することに相当する．ランダムウォーカはある頂点を出発し，辺の遷移確率に従って次の頂点へと移動する．τ ステップ後にはランダムウォーカは τ タスクを得て移動を終了する．タスクの各系列は系列内のタスク間遷移確率のすべての積の形で重みづけされる．カーネルは 2 つのグラフ上の長さが同じすべての重みづけされた系列どうしに対して類似度を計算し，その長さが 1 から無限大に対してその類似度を足し合わせることでグラフ間の類似度を求める．

このように HMM より得られたユーザ行動モデルの類似度を計算しカーネル PCA [14] を行い，各モデルの主成分得点をプロットすることで可視化およびクラスタリングを行う．

3.　実データを用いた分析

本章では長期間にわたるウィンドウ遷移ログからユーザ行動モデルを構築し，このモデルの変化を分析する手法と実際に日本の IT 企業から取得したウィンドウ遷移ログを用いて分析した結果について述べる．

3.1　対象ログ

ログを収集した企業はソフトウェアの設計，開発，製造，販売を行う日本のソフトウェアメーカである．各従業員は各自専用のデスクトップ PC とユーザアカウントを使用している．

分析においては**表 2** のように平日の 8:00 から 21:00 までのログを対象に定常的な業務を行う同一部署に所属する 4 人の約 4 年間にわたる長期間のログを用いて行った．

表 2　ログの情報
Table 2　Information about the logs.

期間	2010/09～2014/06
対象時刻	8:00～21:00
対象人数	4 人
ログ総数	3,062,420 行

表 3　ログ取得対象者
Table 3　Target log data.

ユーザ名	入社月	有効日数	ログ数	平均ログ数
USER A	2010/09	824 日	727,426 行	882 行/日
USER B	2010/09	684 日	784,328 行	1,146 行/日
USER C	2011/04	616 日	828,426 行	1,344 行/日
USER D	2011/04	717 日	722,240 行	1,007 行/日

表 3 は取得したログにおける各ユーザの内訳である．有効日数は取得対象時刻内に 2 時間以上の PC を使用していた日数であり，2 時間未満の日のログはノイズとなるため分析対象から除外した．この要因としては休暇，外出，自席以外での作業，PC の故障によるログ欠損等がある．平均ログ数は 1 日あたりのウィンドウ遷移ログの記録数を表し，これは 1 日あたりのウィンドウ切替え回数に相当する．この値は USER A と USER C では 1.5 倍以上の差があり，切替え頻度に個人差があることが分かる．

本研究では分析作業の効率化のため，ログ全体のアプリケーション使用率を計算することにより上位 8 種類のアプリケーションを抜粋し，これら以外のアプリケーションは Etc として下記 9 つのカテゴリを用いてアプリケーション名を置換し分析に用いた．この 8 種類でアプリケーション使用時間の約 90%を占める．

{Mail, Web, Document, Spreadsheet, Presentation,
Editor, Viewer, Explore, Etc}

これらのうち Mail，Web，Editor，Viewer はアプリケーショングループであり，それぞれ業務上果たす役割が同一であると考えられるアプリケーションをまとめたものである．Web はウェブブラウザの集まりであり，Mail はメールクライアントソフトウェア，Editor は各種エディタソフトウェア，Viewer は Adobe Reader®および画像表示ソフトウェアを含む．

Explore は Windows®の標準のシェルソフトウェアであり，デスクトップや標準のスクリーンセーバ，ファイラとしての機能を持つ．Document，Spreadseet，Presentation は，それぞれ文章作成，表計算，プレゼンテーション用ソフトウェアである．

3.2　分析手法

本節では，前章で述べたユーザ行動モデルを用いて同一の従業員の時系列的なモデルの変化を分析する手法について説明する．

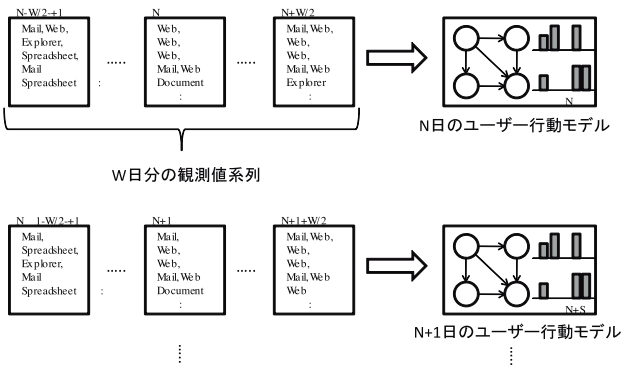

図 6　モデル構築に使用する観測値
Fig. 6　Observations for the modeling.

ユーザ行動モデルは HMM を基としているため，パラメータ推定に使用する観測値に大きく依存する．このため日別や週別のような少量の観測値を用いた場合，大きな偏りが発生する場合がある．特に日々同じ業務にあたる従業員の場合はこれらの偏りが分析の際に大きなノイズとして発生する．

この問題を解決するため，**図 6** のようにウィンドウ幅 W を用いて前後の観測値を合わせてパラメータ推定を行った．この方法を用いることでモデルを時系列的にゆっくりと変化させることが可能となり，時系列変化が視覚的に判断しやすくなる．

ウィンドウ幅を導入する際に，注意すべき点としてログの欠損があげられる．単純にカレンダーどおりウィンドウ幅を適用すると前後 W 日以内にログの欠損がある場合に観測値が減ってしまい前述の効果が期待できなくなる．このため本研究では次のような方法を用いた．

まず，ログから日別に観測値系列 I_n を作成し日付順に並べると，ログの有効日数を T とすると系列 $\boldsymbol{I} = (I_1, \ldots, I_T)$ を得ることができる．n 日目のユーザ行動モデルを構築する際には $(I_{n-W/2+1}, \ldots, I_n, \ldots, I_{n+W/2})$ の観測値系列を用いてパラメータ推定を行い，I_n に対応する日付をユーザ行動モデルにラベル付けする．このようにすることで，たとえ長期間のログ欠損があった場合でも，その前後で大きく傾向が変わっているかを判断すればよく，可視化の際に日付からログの欠損を判断可能となる．

分析としては，まず，対象とするユーザのウィンドウ遷移ログから日別の観測値系列を作成する．次にスライド幅を 1 としてそれぞれ W 日分の観測値系列を用いて，ユーザ行動モデルを作成する．このとき，パラメータ推定の初期モデルとしてすべての観測値から作成したユーザ行動モデルを用いる．これはエルゴード性マルコフモデルに対して Baum-Welch アルゴリズムを適用する際には初期モデルが大きく影響するためであり，共通のモデルを用いることでモデルのばらつきを小さくする効果がある．また，パラメータ推定の際に出力確率 B を変化させないで行うことが可能であり，この場合においてはすべてのモデルでタスク

要約が共有される．このため，モデル間の違いは遷移確率のみとなるため比較が容易となる．

最後に作成された $T-W$ 個のユーザ行動モデルに対して，それぞれの類似度を計算しカーネル PCA を行い，可視化およびクラスタリングを行う．

本分析では，ユーザ行動モデルを構築する際のパラメータである従業員のタスク状態数はタスク要約間の類似度を考慮した実験の結果，$K=12$ として行った．

3.3　分析結果

まず，4 人分のウィンドウ遷移ログから月別の観測値系列を作成し，各ユーザの月別のユーザ行動モデルを作成した．**図 7** はカーネル PCA から得られる第 1，第 2 主成分得点のユーザごとのプロットである．この図から USER A の各点は図中左側および上部付近に分布していることが分かる．同様に他のユーザもある程度の固まった分布が確認でき，重なりあっている箇所が少ないことから，それぞれ固有のパターンを保持していることが予想できる．

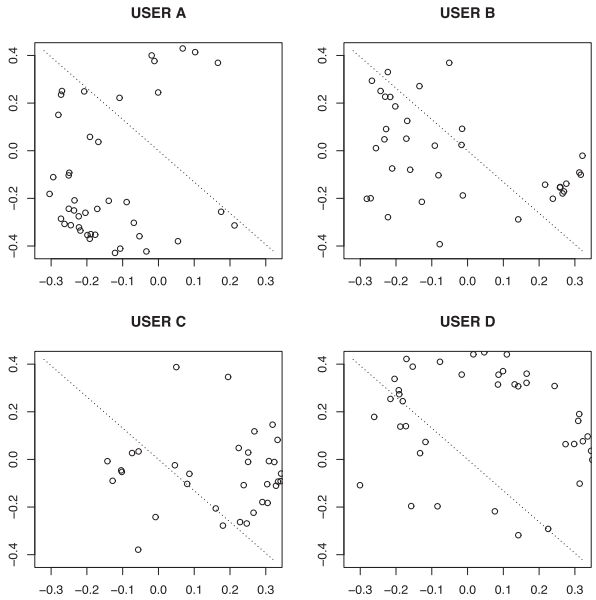

図 7　4 人のカーネル PCA 結果
Fig. 7　Results of four users by kernel PCA.

また，ヒアリング調査の結果 USER A と USER B, USER C と USER D はそれぞれ同一の業務を行うことが多いということが分かった．図からも左上から右下に対角線を引くことで，多くの点がペアに従って分割可能なことが分かる．

次項では，各ペアのうち USER A と USER C の時系列変化についての分析結果を示す．

3.3.1　時系列変化分析

前述した手法を用いて，USER A と USER C それぞれのウィンドウ遷移ログを用いてユーザ行動モデルを作成し，カーネル PCA およびクラスタリングを行った結果を**図 8** に示す．ウィンドウ幅 W は実験的に約 1 カ月の営業日数にあたる 20 を使用した．左図が USER A，右図が USER C のプロットである．各点の色は日付を示しており，図中右下のカラーバーのように日付と色が対応している．各点の形状はクラスタリングの結果を示しており，どちらの図でも 5 つのクラスタに分割している．

USER A は綺麗に色ごとに固まっており，この塊ごとに別々のクラスタが割り当てられている．一方，USER C はある程度色ごとに固まってはいるが離れた日付の点が混在している．

これは，USER A に関してはカーネル PCA で得られる主成分が時系列と関連性があることを意味する．すなわち，モデルの分散は時系列的に大きく時系列に沿ってモデルが変化していることを示している．一方 USER C は時間の経過とモデルの違いにあまり関連性がない．USER A のモデル間の平均類似度は 0.4596，標準偏差 0.2381 であるのに

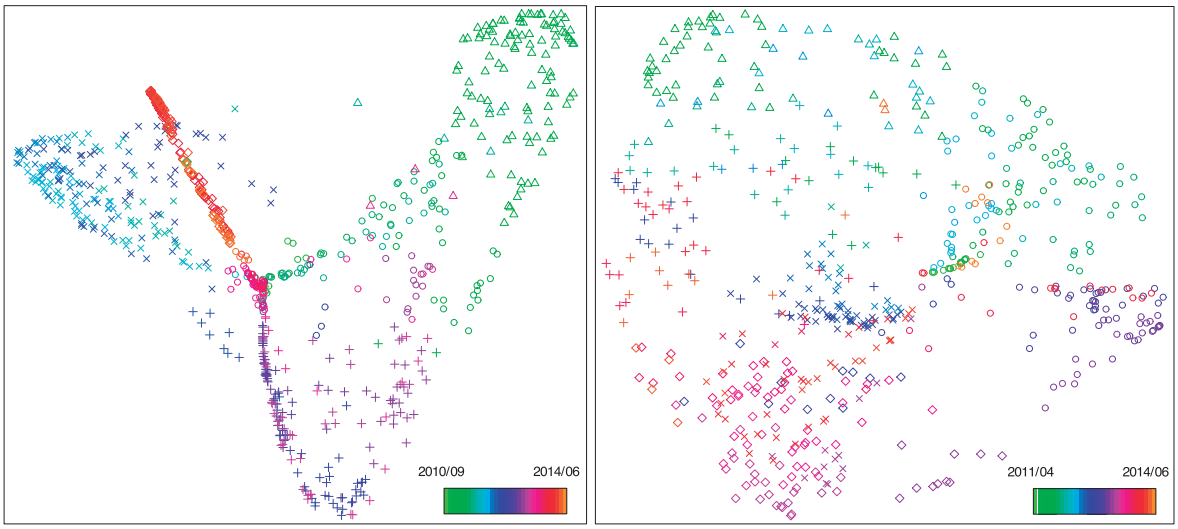

図 8　カーネル PCA の主成分得点のプロット（左：USER A，右：USER C）
Fig. 8　Scatter plot of two users by kernel PCA (Left: USERA, Right: USER C).

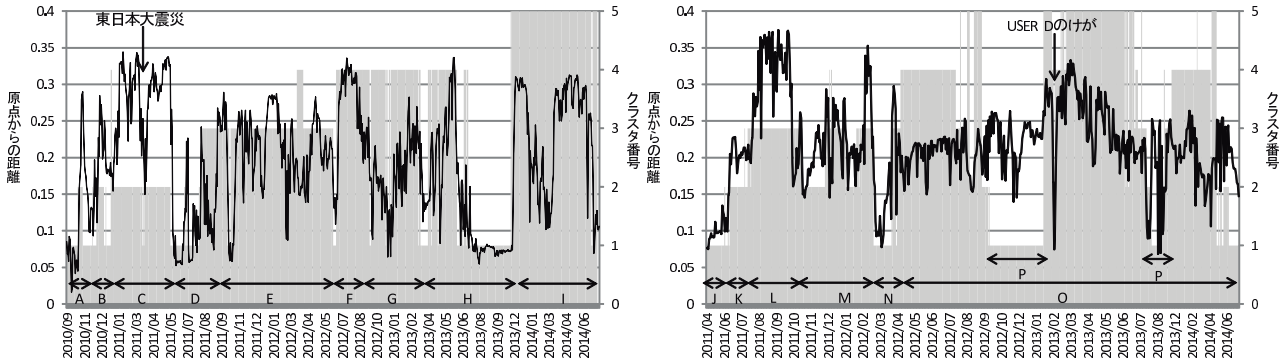

図 9　原点からの距離とクラスタ（左：USER A，右：USER C）
Fig. 9　Distance from the origin and cluster number (Left: USERA, Right: USER C).

対して，USER C では平均類似度 0.6668，標準偏差 0.1794 と全体的に類似性が高いことからも分かるように USER A に比べてモデルの変化が小さいということを示している．

次に各モデルの第 1，第 2，第 3 主成分からなる部分空間上の原点からの距離とクラスタの関連を**図 9** に示す．x 軸には日付をとり，実線が原点からの距離（左の軸），灰色の網がけはクラスタのインデックス（右の軸）を示す．

図中の実線は原点からの距離であるため，同一の値であってもモデル間の距離が近いということを意味しない．しかし，近傍でこの値が大きく異なるということは部分空間内での位置が大きくことなることを意味し，ユーザ行動モデルが変化していることを示す．クラスタリングは主成分得点に対して階層化クラスタリングを用いて行っているため，時系列的に見てクラスタの切替えが起こる点の近傍では原点からの距離も大きく変動している．また USER A の 2011/03 付近では近傍に比べ大きく減少している．これは東日本大震災の影響で，企業内での業務体制が変わったことが関連している．USER C では 2013/02 付近にペアとして勤務している USER D のけがの影響で業務パターンが変わった形跡が見られる．

図中 A から O で示されている区間はヒアリングの結果から得られた実際の担当業務を行っていた期間である．各期間に該当する業務を**表 4** に示す．USER A に関しては各期間とクラスタの切り替わる地点がおおむね一致していることが分かる．ヒアリング結果からも該当業務内容が大きく変わっており，PC の使用方法も変化していることが予想できる．一方 USER C は入社以来，主に事務業務を行っており，特に期間 O では事務業務のみを行っている．これは USER C の全体的なモデル間の類似性が高いことと合致する．

同一の業務を行っている中でのパターンの違いを分析するため，次項では期間 O 内でもクラスタが異なる期間 P に着目し他の部分との違いを分析する．

3.3.2　モデル間比較

モデル間の比較を行うにあたり，タスク要約が全モデル

表 4　実業務のヒアリング調査結果
Table 4　Results of hearing investigation.

USER A	該当業務	USER C	該当業務
A	研修	J	研修
B	引き継ぎ	K	引き継ぎ
C	事務	L	事務
D	引き継ぎ	M	事務+自席外作業
E	経理業務	N	資料整理
F	調査業務 1	O	事務
G	調査業務 2		
H	資料作成 1		
I	資料作成 2		

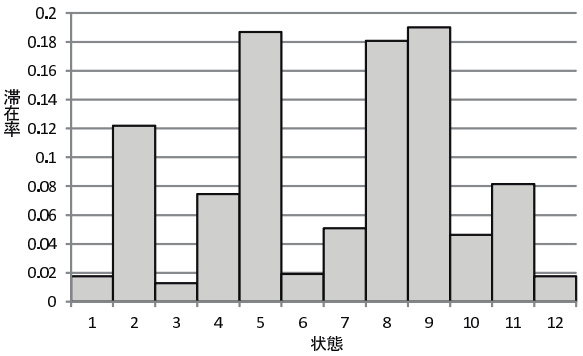

図 10　USER C の定常状態分布
Fig. 10　The steady-state distribution of the task transition of USER C.

間で共有されるとモデル間の違いは遷移確率のみとなり比較が容易になることから，パラメータ推定の際に出力確率 B を更新せずに同一の観測値系列からモデルを再度作成した．

まず，USER C の全ログから作成した全体的なモデルの定常状態分布を**図 10** に示す．この図から USER C の滞在率が高い状態が状態 2，状態 5，状態 8，状態 9 であることが分かる．次に，作成した各モデルから定常状態分布を求め上記 4 つの状態に対する滞在率の変化と各状態のタスク要約を**図 11** に示す．タスク要約の各グラフ上の値は

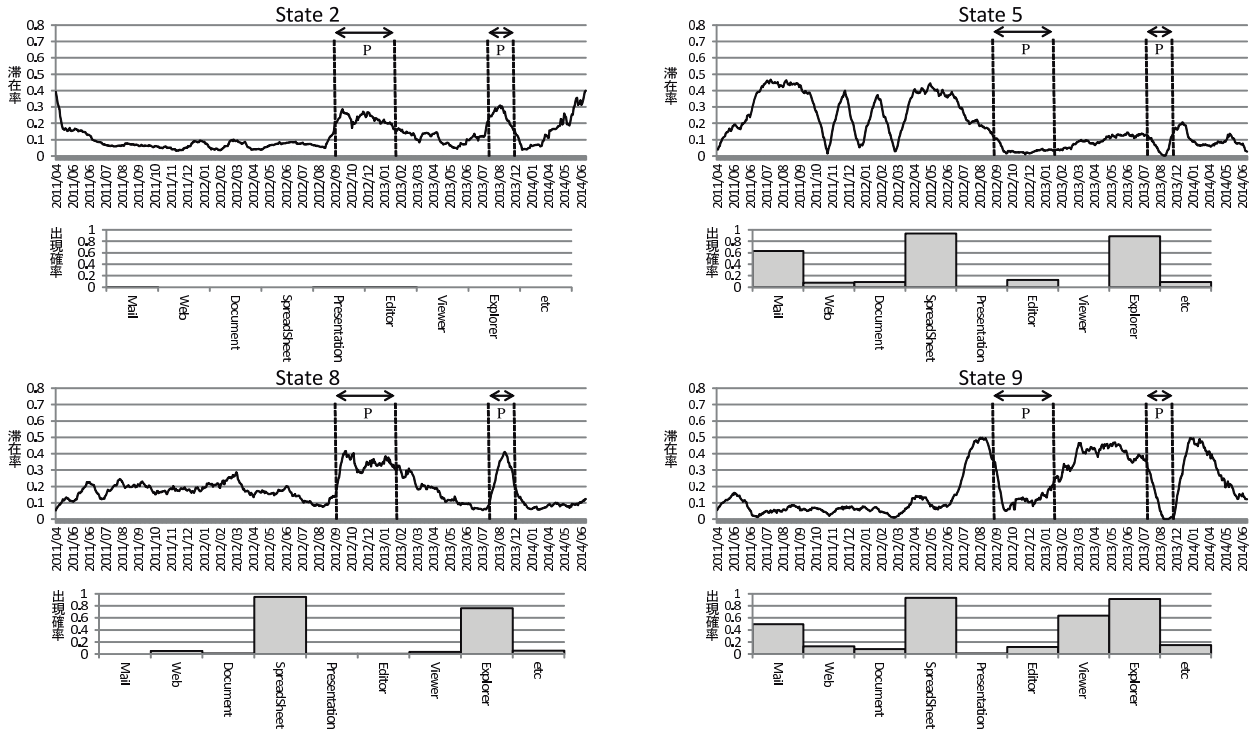

図 11　滞在率の変化とタスク要約
Fig. 11　Task summary and changes of stay rate.

そのタスク状態でのアプリケーションの使用頻度を示しており，値が 1 の場合には必ずそのアプリケーションが使用されることを示す．状態 2 はすべてのアプリケーションでほぼ 0 であり，離席や PC を使用しない業務を意味する．図中の期間 P に該当する箇所で状態 5 と状態 9 の値が下がり，状態 2 と状態 8 の値が上がっていることが分かる．これは期間 P においては Mail と Viewer の使用傾向が変わり，離席または PC を使用しない作業が増えたことを意味する．これは明らかに他の期間とは異なるパターンである．ヒアリング調査時にこのような話はいっさい出てこなかったことから，環境や忙しさ等で働き方が変わる等，無自覚な変化であるといえる．

4.　おわりに

本稿では，長期間にわたるウィンドウ遷移ログから作成したユーザ行動モデルの時系列変化分析を行うことで，ユーザ行動モデルと実際の業務との関連性を検証した．ユーザ行動モデルを用いて，4 年間–約 300 万行に及ぶ実際のログを抽象化した情報は，ヒアリングの結果と合致する点が多く，ログと業務の間の関連性は強いことを示した．また，同一業務期間内における異なった作業パターンを見つけることができた．

最後に，本手法の長所と短所についての考察と今後の応用分野について述べる．

4.1　考察

前章の分析結果より，ユーザ行動モデルの時系列変化は実際の業務と関連性が強いといえる．しかしながら，複雑なユーザ行動モデルを作成しなくても，もっと簡単な手法でも可能ではなかったのかという疑問が残る．

単純な方法としては，ウィンドウ遷移ログからアプリケーションの使用時間を計算しその比率に対して PCA を用いることが考えられる．実際に分析を行ってみたところ，ある程度のクラスタに分割は可能であった．しかし，前章の USER C における期間 P に対する違いは検出できなかった．これは USER C は期間 O においてはほとんどアプリケーションの使用頻度が変わらないためである．使用時間による手法では短時間の使用を高頻度で行う場合と長時間低頻度で使用する場合を区別できないため，上記の違いを検出することができない．一方，ユーザ行動モデルを用いる場合には使用頻度と遷移の情報を保持するためこのような違いが検出可能である．

このような同一業務における PC 利用傾向の違いは，業務の効率性を推定する手がかりとなる可能性がある．仮に期間 P において仕事量が多く，非常に忙しい状況であった場合，成果物の質が期間 P 以外と同一であるならば，期間 P では作業効率が良いと考えられる．しかしながら，この違いの要因を特定するためには，ヒアリングだけではなく，業務における記録や環境の変化，業務の評価等の付加情報を用いる必要がある．本分析では蓄積されていた実際の業務のログを使用しており，これらの情報が乏しいため，こ

れ以上の言及は難しい．今後は付加情報を取得可能な実験的環境を用いての分析を行う予定である．

また，今回用いたウィンドウ幅を用いた分析手法は比較的長い期間同じ業務を担当する従業員にのみ適用可能である．前述のとおりウィンドウ幅が小さいとサンプル数が少ないため多くのノイズが混入してしまう．一方，業務が1週間や数日で変わるような従業員に対しては，1つのユーザ行動モデルの中に複数の業務が含まれてしまうため時系列的に近傍のモデルに対してばらつきが生まれてしまう．このような業務を行う従業員に対しては短期間でモデルを作成し，業務別にクラスタリングした後，同一クラスタ内の時系列変化をみる等の別の手法を考える必要がある．

4.2　今後の展望

今後の展望としては，ユーザ行動モデルを用いた分析を実習や研修中のログを対象として行うことを考えている．企業で行われる実習や研修は，固定されたPC上で操作を行い，様々なバックグラウンドの人が同一の成果物を作成することを目標とすることが多い．たとえばプログラミングの研修において，Webページにプログラミングの解説や課題が用意されており，用意された開発環境を使用してプログラミングを学び，課題のプログラムの作成を行うことが想定できる．このような環境の場合ではWebページ上の資料や書籍の閲覧頻度やデバッグの頻度，コンパイルの頻度，テスト実行の頻度等が開始時と終了時では大きく異なることが予想できる．

また，ウィンドウ遷移と合わせてマウスとキーボードのログを取得することでキー入力の速度やショートカットキーの使用頻度等の作業効率に影響する変化も観測することができる．さらには，実際の成果物の評価や試験を行うことで学習の効果もランク付け可能であるため，これらの情報を基に優秀な人やそうでない人の特徴パターン抽出を行うことが可能である．

謝辞　本稿の作成にあたり，多大なご指摘とアドバイスを賜りました東京電機大学の安田浩教授と望月氏，学習院大学の久保山哲二教授に感謝いたします．また，ご多忙の中査読していただいた査読者および編集委員の方々に感謝いたします．

参考文献

[1] Saito, R., Kuboyama, T., Yamakawa, Y. and Yasuda, H.: Understanding User Behavior through Summarization of Window Transition Logs, *Databases in Networked Information Systems*, Lecture Notes in Computer Science, Vol.7108, pp.162–178, Springer Berlin Heidelberg (2011).

[2] Rana, C.: A Study of Web Usage Mining Research Tools, *Information Retrieval*, Vol.1429, No.06, pp.1422–1429 (2012).

[3] Chu, H.L., Raman, V., Shen, J., Choudhury, R., Kansal, A. and Bahl, V.: In-vehicle driver detection using mobile phone sensors, *ACM MobiSys* (2011).

[4] Hilbert, D.M. and Redmiles, D.F.: Extracting usability information from user interface events, *ACM Computing Surveys* (*CSUR*), Vol.32, No.4, pp.384–421 (2000).

[5] Singh, N., Tomar, D. and Roy, B.: An Approach to Understand the End User Behavior through Log Analysis, *International Journal of Computer Applications IJCA*, Vol.5, No.11, pp.9–13 (2010).

[6] Beauvisage, T.: Computer usage in daily life, *Proc. 27th international conference on Human factors in computing systems*, pp.575–584, ACM (2009).

[7] Tak, S. and Cockburn, A.: Window Watcher : A Visualisation Tool for Understanding Windowing Activities, *Time*, pp.241–248 (2009).

[8] Suzuki, K., Yasuda, H., Shin, K. and Kuboyama, T.: Discriminating User Behavior through PC Operation Logs by PageRank Convergence Patterns, *International Journal*, Vol.3, No.1, pp.37–40 (2014).

[9] 鳥羽美奈子，櫻井隆雄，森　靖英：PC操作ログの特徴量とオフィスワーカーのストレス量の相関分析（データマイニング，＜特集＞ライフログ処理技術とその活用システム論文），電子情報通信学会論文誌D，情報・システム，Vol.95, No.4, pp.747–757 (2012).

[10] 平山明彦，原　直，阿部匡伸：非負値行列因子分解によるPC操作ログからの勤務パタン抽出（ライフインテリジェンスとオフィス情報システム），電子情報通信学会技術研究報告 = IEICE technical report: 信学技報，Vol.114, No.32, pp.33–38 (2014).

[11] Wang, P., Wang, H., Liu, M. and Wang, W.: An algorithmic approach to event summarization, *Proc. 2010 international conference on Management of data*, pp.183–194, ACM (2010).

[12] Rabiner, L.: A tutorial on hidden Markov models and selected applications in speech recognition, *Proc. IEEE*, Vol.77, No.2, pp.257–286 (1989).

[13] Kashima, H., Tsuda, K. and Inokuchi, A.: Marginalized kernels between labeled graphs, *Proc. 20th International Conference on Machine Learning* (*ICML*), pp.321–328 (2003).

[14] Schölkopf, B. and Smola, A.: *Learning with kernels*, MIT Press (2002).

齋藤 良平

1981年生．2004年電気通信大学電気通信学部情報通信工学科卒業．2006年同大学大学院修士課程修了．同年ハミングヘッズ株式会社入社．セキュリティ関連ソフトウェアの研究開発に従事．2010年4月から2014年3月まで東京電機大学研究員を兼務．PC操作ログの分析に関する研究に従事．電子情報通信学会会員．