

AND 合流ゲートウェイと連結するアクティビティの 平均潜在待ち時間とサービス時間の推定

野ヶ山 尊秀^{1,a)} 高橋 治久^{2,b)}

概要：開始時刻または完了時刻のみが記録されたイベントログからは、アクティビティの待ち時間とサービス時間は計算できない。筆者らはこのようなログから統計的に平均待ち時間と平均サービス時間を推定する方法を提案したが、AND 合流ゲートウェイが含まれる場合には仮定にない並列実行の同期待ち時間が含まれるため適用できない。この論文では、指数分布を仮定して AND 合流ゲートウェイの確率モデルを構築し、平均待ち時間とサービス時間を EM アルゴリズムを用いて統計的に推定する方法を提案する。人工的に生成させたログを用いた数値実験により、提案手法が正しく推定できることを示した。

1. はじめに

ビジネスプロセスのモデル化と応用の研究は、欧米において活発に研究されており^{*1}、プロセスマイニングと呼ばれている。ビジネスプロセスは、処理の基本単位であるアクティビティ（例えば保険の査定受付、審査、保険料見積もり、稟議プロセスの承認など）と AND, OR, XOR の分岐や合流のゲートウェイをもつグラフとしてモデル化される。モデルを記述する表記法として、これまでの多くの研究や標準化の試みの結果（詳しくは [3]）として、Business Process Modeling Notation (BPMN) [15] が主流となっている。BPMN を入力としてビジネスプロセス環境を自動生成する、ワークフローマネジメントシステム (WFMS) やビジネスプロセスマネジメントシステム (BPMS) などは、プロセス志向型情報システム (PAIS) [4] と呼ばれている。PAIS では、プロセスの実行の履歴をイベントログとして詳細に記録する。プロセスマイニングは、主にこのイベントログから知識を獲得するための技術である [9], [23]。

品質の高い PAIS は、アクティビティの実行履歴だけでなく、その状態遷移 (transaction types of life-cycle と呼ばれる) も記録する。例えば、schedule (アクティビティの実

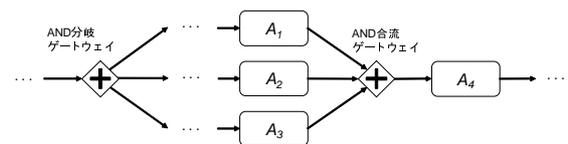


図 1 プロセスモデル内の And 分岐と And 合流ゲートウェイ
Fig. 1 And-split and And-join gateways in process model

表 1 開始イベントしか記録されないイベントログの一部
Table 1 An event log that has only start events of activities

案件 ID	時刻	アクティビティ	...
1	2015-08-01 09:00:00	A ₂	...
1	2015-08-01 09:01:19	A ₃	...
1	2015-08-01 09:02:10	A ₁	...
1	2015-08-01 09:11:58	A ₄	...

行が確定), assign (担当者が決定), start (開始), complete (完了) などの代表的な遷移が書籍 [19] やログフォーマット [6] で標準化されている。これらの状態遷移の時間間隔からは、アクティビティの待ち時間やサービス時間といった性能指標を得ることができる。こうした指標を用いた性能分析は人件費や設備費に直接関係するため、ビジネスプロセス改善のうち特に重要な工程である。

しかし、多くのプロセスシステムは不完全なイベントログを生成するため、こうした性能指標は信頼できるものにはならないか、計算不可能であることが多い。性能指標に関わるイベントログの不完全さには主に以下の 2 つが考えられる：

- (1) アクティビティに対して一つの時刻しか記録されずアクティビティ所要時間が計算できない場合。多くのプロセスシステムでは、ログの記録の目的が監査や問

¹ 日本アイ・ビー・エム株式会社 東京基礎研究所
19-21 Nihonbashi, Hakozaiki-cho, Chuo-ku, Tokyo, 103-8510, Japan

² 電気通信大学
1-5-1 Chofugaoka, Chofu, Tokyo 182-8585, Japan

a) nogayama@jp.ibm.com

b) takaharuroka@uec.ac.jp

^{*1} 主な国際会議として Business Process Management (BPM), Conference on Advanced Information Systems Engineering (CAiSE) などがある。

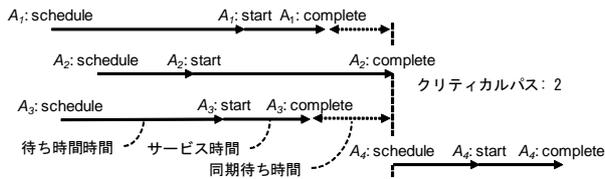


図 2 アクティビティの状態遷移が記録されたイベントログ
Fig. 2 An event log that has life-cycle transitions of activities

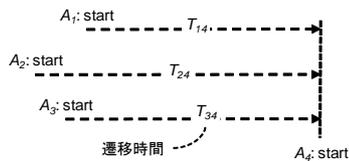


図 3 開始イベントしか記録されないイベントログ
Fig. 3 An event log that has only start events of activities

題判別であるため、開始または終了時刻のみが記録される。その一つの時刻が開始時刻なのか終了時刻なのか分からない場合もある。典型的なイベントログは、表 1 のように、アクティビティに対して一つの時刻が記録される。例えば、Kuo ら [10] は開始時刻のみが記録される医療システムに直面した。他にも、プロセスマイニングの普及を意図したマニフェスト [9], [23] でも同様のイベントログを典型的な例としてあげている。

(2) 制御がアクティビティに移った正確な時刻や、制御を他のアクティビティに移した正確な時刻が記録されない場合。例えば、処理依頼を受け取ったにも関わらずすぐに受領処理を行わなかった場合や、アクティビティが終了したのちにすぐ次のアクティビティに処理依頼を出さずに溜めてしまった場合は、ログの上では遷移時間が長くなる。

BPMS ベンダーにとってこのような不完全イベントログの分析は、他社との差別化を図るため必須である。例えば、ある顧客が古いプロセスシステムを更改するとき、同時に現行システムの問題を改善した新システムの設計を望む事が多い。現行システムの不完全イベントログを分析し、移行案に加えて改善案を提案できれば、BPMS ベンダーはシステム移行プロジェクトを獲得できる見込みが高まる。

我々は、このような不完全なイベントログであっても統計的に分析する事でサービス時間と待ち時間の平均値を推定する方法を提案した [14], [24]。この方法の本質は、遷移時間が遷移元アクティビティの潜在サービス時間と遷移先アクティビティの潜在待ち時間によって構成されると仮定することと、分岐または合流ゲートウェイを通る遷移時間は同じ確率分布に従う潜在待ち時間と潜在サービス時間を共有するという特徴を利用することにある。この仮定は BPMN で定義されているほとんどの要素間の遷移で成

り立つが、AND 合流ゲートウェイ^{*2}を経由するアクティビティ遷移では成り立たない。なぜならば、先に到着したすべてのサブプロセスが最後に到着したサブプロセスを待つためにそれより前に到着したサブプロセスには同期待ち時間が加わるため、確率モデルに不整合がでてくるためである。

例えば、図 1 のような AND 合流ゲートウェイとその前後にアクティビティがあったとき、アクティビティの状態遷移が記録されれば、図 2 を描くことができる。そして、ゲートウェイ通過後のアクティビティ A_4 の開始時刻と A_1, A_2, A_3 の完了時刻を比べることで、クリティカルパスとなったアクティビティが A_2 であることもわかる。ところがイベントログに表 1 のように開始時刻のみが記録された場合、図 3 のような関係しか復元できないため、開始時刻間の遷移時間以外は、待ち時間、サービス時間、同期待ち時間、クリティカルパスがわからない。Leemans ら [12] も、アクティビティの状態遷移を利用したプロセスディスカバリーと性能分析の研究において、状態遷移が記録されない不完全イベントログが多い事に触れ、状態遷移を推定する方法も提案している。彼らはまた、そうしたイベントログの上で AND 合流ゲートウェイが存在したときの同期待ち時間の観測の難しさについても触れている。

AND 合流ゲートウェイは主にプロセスの並列実行による高速化のために用いられ、依存関係の無いアクティビティの直列実行が性能上ボトルネックになった場合などに採用される。最も簡便に適用可能で効果も大きい性能改善であるため、ビジネスプロセス改善の現場では頻りに利用される。一方、並列実行に設計されたビジネスプロセスの性能改善はさらに難しくなる。待ち時間、サービス時間に加えて、同期待ち時間やクリティカルパスによる影響を考慮しなければならないからである。それにも関わらず不完全イベントログの影響でこれらの性能指標が利用できないため、並列実行の性能改善は極めて困難な課題となり、解決が望まれている。

本研究では、こうした現実的な問題を解決するため、[24] を拡張し潜在時間として指数分布を採用した、AND 合流ゲートウェイを含むプロセスの不完全イベントログから、EM アルゴリズムにより平均潜在待ち時間とサービス時間を算出する方法を提案する。提案手法の検証は、AND 合流ゲートウェイを通過するプロセスのログを生成し、提案手法が正しく推定できることにより示す。

次の節でビジネスプロセスの性能分析に関する研究を紹介し、3 節では AND 合流ゲートウェイを含むプロセスの確率的なモデルを与える。4 節ではそのモデルのパラメタ

*2 OR 合流ゲートウェイは AND 合流と XOR 合流のいずれかとして動作するため、AND 合流として動作したときの OR 合流ゲートウェイも同様であるが、この論文ではそれも含めて AND 合流ゲートウェイと呼ぶことにする。

を推定方法を提案し、5節でその推定法の検証実験の結果を示す。

2. 関連研究

イベントログから性能を要約する研究は多く行われている。例えば、Ferreiraらは医療プロセス内の各アクティビティにおける所要時間の最大値、最小値、平均値を算出してプロセスモデル上に可視化した[5]。所要時間に加え、イベントログから抽出できる多くの性能指標について、Lanzら[11]がよく整理し分類している。また、多くのビジネスプロセスの性能指標やその上での分析法について、Hornix[8]がよくまとめている。

性能の要約データは、プロセスの再設計時のほかにも実行中のプロセスの予測や制御にも用いられる事がある。van der Aalstら[20], [21]は、要約データを元にして、実行中のプロセスインスタンス全体の所要時間の予測方法を提案している。また、Rogge-soltiら[16]はアクティビティの平均サービス時間を用いて実行中のプロセスの異常を予測する方法を提案している。予測だけでなく、動的にプロセスに介入する事でより良いプロセスに変えることもできる。例えば、Sindhgatta[18]はアクティビティの担当者ごとの平均時間を用いて、実行中のアクティビティの適切な担当者を割り当てる手法を提案している。

待ち行列理論によるビジネスプロセスの性能分析も研究されている。トランザクションタイプが記録されたイベントログの上では、アクティビティの処理依頼の到着や処理の開始や完了の時刻が観測されるため、待ち時間やサービス時間を算出できる。これらの情報を用いて平均行列長などをシミュレーションによって算出できる[17], [22]。シミュレーションでは到着過程、サービス(担当者)の数、アクティビティ内の行列の構造などはあらかじめ仮定する必要がある。しかし、ビジネスプロセスでは担当者の数が時間帯や日によって変わるため、単純な待ち行列理論で扱うことは難しい。また、これらの仮定を補強する情報は通常イベントログに記録されないため、それ以外の情報源、例えば詳細を知る者へのインタビューや設計文書、が必要となる。

3. 確率モデル

本節では、アクティビティの遷移時間を平均潜在待ち時間とサービス時間に分解し、AND合流ゲートウェイの定式化を行う。

潜在待ち時間と潜在サービス時間の確率密度関数として、処理時間の分布として最もシンプルで多く用いられている指数分布を用いる。パラメタ $\lambda > 0$ の指数分布に従う確率変数 $X > 0$ の確率密度関数は $p(X; \lambda) = \frac{1}{\lambda} e^{-\frac{X}{\lambda}}$ である。これを $X \sim \text{Exp}(\lambda)$ と書く。 X の期待値 $E[X]$ は λ と等しい。またある時点よりも前に処理が完了した確率は分

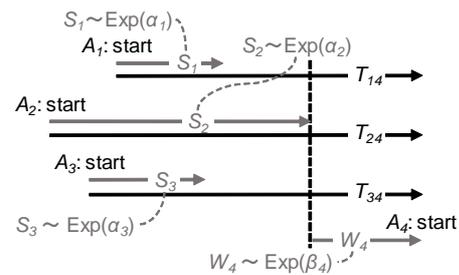


図4 AND合流ゲートウェイの確率モデル

Fig. 4 The probabilistic model on AND-join gateway

布関数 $p(x \leq X; \lambda) = 1 - e^{-\frac{x}{\lambda}}$ により得られる。

いま、ビジネスプロセスに N 個のアクティビティ A_1, \dots, A_N があり、各アクティビティの潜在待ち時間 W_i は $\text{Exp}(\beta_i)$ に、潜在サービス時間 S_i は $\text{Exp}(\alpha_i)$ に従うとする。状態遷移が複数観測される場合は、連続する最初の状態遷移から最後の状態遷移までの時間を処理時間 D_i とする。

AND合流ゲートウェイを通らない A_i から A_j への遷移は、文献[24]と同様に $T_{ij} = S_i + W_j$ と仮定すれば同時確率は

$$p(T_{ij}, W_j) = p(S_i = T_{ij} - W_j)p(W_j) \quad (1)$$

となる。 S_i と W_j は観測されないため、 W_j に関して周辺化を行い

$$p(T_{ij}) = \int_0^{T_{ij}} p(T_{ij}, W_j) dW_j \quad (2)$$

を得る。

AND合流ゲートウェイを通る場合は複数の遷移時間を一つの事象として扱う。他のゲートウェイと組み合わせたときAND合流ゲートウェイの遷移元アクティビティは実際のプロセスに依存して決まり、事前には決まらない。そのため同じAND合流ゲートウェイでも、観測された遷移元アクティビティ集合と遷移先アクティビティの組み合わせごとに異なる確率モデルを構成する。

あるとき観測された遷移元アクティビティの添え字集合を I 、遷移先アクティビティを A_j 、そのときクリティカルパスとなった遷移元アクティビティの添え字を $C \in I$ 、観測された遷移時間を集合 $T_{Ij} = \{T_{ij} | i \in I\}$ とする。例えば、表1で観測されたイベントログでは、遷移元添え字集合は $I = \{1, 2, 3\}$ 、遷移先は A_4 、遷移時間集合は $T_{Ij} = \{T_{14}, T_{24}, T_{34}\}$ 、クリティカルパスを $C = 2$ と仮定した場合の関係を図4に図示した。

A_c の遷移時間 T_{cj} と潜在時間 S_c, W_j の関係は $T_{cj} = S_c + W_j$ であるため同時確率は式(1)と同じである。クリティカルパスとならない $i \in I$ の遷移時間 T_{ij} と潜在時間 S_i, W_j の関係は $S_i \leq T_{ij} - W_j$ である。これは A_i の潜在サービスが $T_{ij} - W_j$ よりも前に終了したため、クリティカルパスにならなかったことを意味し、同時確率は

$p(T_{ij}, W_j) = p(S_i \leq T_{ij} - W_j)p(W_j)$ となる。関係する全ての事象 T_{I_j}, S_i, W_j, C の同時確率は、全ての $p(T_{ij}, W_j) \ i \in I$ の積により

$$p(T_{I_j}, C, W_j) = \prod_{i \in I} \begin{cases} p(S_i = T_{ij} - W_j)p(W_j) & i = C \\ p(S_i \leq T_{ij} - W_j)p(W_j) & i \neq C \end{cases}$$

となる。クリティカルパス C と潜在時間 S_i, W_j は観測されないため、周辺化することで

$$p(T_{I_j}) = \sum_{C \in I} \int_0^{\min(t_{I_j})} p(T_{I_j}, C, W_j) dW_j \quad (3)$$

を得る。ここで、 W_j のとりうる値は 0 から遷移時間 T_{I_j} の最小値 $\min(T_{I_j})$ である。なぜならばそれよりも大きい場合、最後に開始した遷移元アクティビティの開始時刻よりも前に遷移先アクティビティが開始することになってしまうからである。

4. パラメタ推定

本論文では、最尤原理に基づき観測された遷移時間を最も尤もらしく説明するパラメタ $\hat{\alpha}_i, \hat{\beta}_i \ (i = 1, \dots, N)$ を求め、そのパラメタを用いてアクティビティ A_i の平均潜在待ち時間 $E[W_i] = \hat{\alpha}_i$ とサービス時間 $E[S_i] = \hat{\beta}_i$ を推定する。アクティビティの状態遷移が一つしか観測されなければ、アクティビティ A_i の平均潜在処理時間を $E[W_i] + E[S_i]$ として推定し、状態遷移が複数観測され、処理時間を算出できるのであれば、平均処理時間 \bar{D}_i を加え平均潜在処理時間を $E[W_i] + \bar{D}_i + E[S_i]$ として推定すればよい。

4.1 最尤推定

AND 合流ゲートウェイを通らない A_i から A_j への遷移を (i, j) を用いて表し、あるイベントログ内で観測された全ての AND 合流ゲートウェイを通らない遷移の集合を \mathcal{T}_1 とする。そのイベントログにおいて、遷移 $(i, j) \in \mathcal{T}_1$ の遷移時間 $t_{ij}^{(1)}, \dots, t_{ij}^{(n_{ij})}$ がお互いに独立に n_{ij} 個観測されたとする。これらの遷移時間の尤度を L_1 とする。

AND 合流を通るアクティビティ $A_i \ i \in I$ から A_j への遷移を (I, j) を用いて表し、あるイベントログ内で観測された全ての AND 合流ゲートウェイを通る遷移の集合を \mathcal{T}_2 とする。そのイベントログにおいて、遷移 $(I, j) \in \mathcal{T}_2$ の遷移時間集合 $t_{Ij}^{(1)}, \dots, t_{Ij}^{(n_{Ij})}$ がお互いに独立に n_{Ij} 個観測されたとする。これらの遷移時間の尤度を L_2 とする。

観測された遷移時間の尤もらしさを計る対数尤度 $\log L = \log L_1 + \log L_2$ は、前節の式 (2) と式 (3) を用いて

$$\log L_1 = \sum_{(i,j) \in \mathcal{T}_1} \sum_{k=1}^{n_{ij}} \log p(t_{ij}^{(k)}; \theta) \quad (4)$$

$$\log L_2 = \sum_{(I,j) \in \mathcal{T}_2} \sum_{k=1}^{n_{Ij}} \log p(t_{Ij}^{(k)}; \theta) \quad (5)$$

となる。ここで、 $\theta = \{\alpha_1, \dots, \alpha_N, \beta_1, \dots, \beta_N\}$ とした。この尤度関数を最大にする $\hat{\theta}$ が求めたいパラメタである。しかし、この最大化問題は解析的に解くことができない。そのため、本研究では expectation maximization (EM) アルゴリズムを用いてこの問題を解く。

4.2 EM アルゴリズム

EM アルゴリズム [2] は観測されない確率変数が含まれる最尤推定問題を逐次的に解く手法である適当な初期値からスタートし、E ステップと M ステップを繰り返し適用してパラメタを更新することで、より尤度の高いパラメタに更新できることが証明されている [1], [7], [13]。この節では、観測される確率変数を T_{ij} 、観測されない確率変数を S_i, W_j, C として、EM アルゴリズムに当てはめる。アルゴリズムは、まずパラメタ θ に適当な値を初期値として与え、M ステップにて新たな θ' に逐次更新し、尤度関数の極大点に到達したら停止する。

E ステップでは、現在のパラメタ θ を基にした確率を用いて、次のパラメタ θ' の完全対数尤度関数の期待値 Q を導出し、不完全対数尤度関数 (4)(5) を近似する。すなわち、式 (4) の期待値 $Q_1(\theta, \theta')$ は

$$\sum_{(i,j) \in \mathcal{T}_1} \sum_{k=1}^{n_{ij}} \int_0^{t_{ij}^{(k)}} p(W_j | t_{ij}^{(k)}; \theta) \log p(t_{ij}^{(k)}, W_j; \theta') dW_j,$$

式 (5) の期待値 $Q_2(\theta, \theta')$ は

$$\sum_{(I,j) \in \mathcal{T}_2} \sum_{k=1}^{n_{Ij}} \sum_{c \in I} \int_0^{\min(t_{Ij}^{(k)})} p(C, W_j | t_{Ij}^{(k)}; \theta) \log p(t_{Ij}^{(k)}, C, W_j; \theta') dW_j$$

となり、あわせて $Q(\theta, \theta') = Q_1(\theta, \theta') + Q_2(\theta, \theta')$ となる。ここで $p(W_j | t_{Ij}) = p(t_{Ij}, W_j) / p(t_{Ij})$ である。

M ステップでは、期待値 $Q(\theta, \theta')$ を最大化する θ' を求める。Q は上に凸な関数であるため、極大値条件 $\frac{\partial Q}{\partial \theta'} = 0$ を満たすパラメタを求めればよい。Q₁(θ, θ') の偏微分を式 (6)、式 (7) に、Q₂(θ, θ') の偏微分を式 (8)、式 (9) に示した。

この条件式は解析的に解くことはできない。そこで一般化 EM アルゴリズムと呼ばれる緩和を採用する。すなわち M ステップでの極大化をあきらめ、Q が増加する θ' に更新すればよいことにする。この論文では最も単純な手法である勾配法を用いて $\theta' = \theta + \eta \frac{\partial Q}{\partial \theta}$ と更新する。ここで $\eta > 0$ は数値安定性を確保するため、移動量を調整する小さな正の値の定数である。

5. 実験

この節では、真の値が分かっているデータを最もシンプルな構成の AND 合流ゲートウェイで生成し、提案手法が潜在平均待ち時間とサービス時間を推定できることを実験により示す。提案手法は GNU Octave を用いて実装した。

$$\frac{\partial Q_1}{\partial \alpha'_h} = \sum_{\substack{(i,j) \in \mathcal{T}_1 \\ i=h}} \sum_{k=1}^{n_{ij}} \left(-\frac{1}{\alpha'_h} + \frac{1}{\alpha_h'^2} \left(t_{hj}^{(k)} - \frac{1}{p(t_{hj}^{(k)}; \theta)} \int_0^{t_{hj}^{(k)}} p(t_{hj}^{(k)}, W_j; \theta) W_j dW_j \right) \right) \quad (6)$$

$$\frac{\partial Q_1}{\partial \beta'_h} = \sum_{\substack{(i,j) \in \mathcal{T}_1 \\ j=h}} \sum_{k=1}^{n_{ij}} \left(-\frac{1}{\beta'_h} + \frac{1}{\beta_h'^2 p(t_{ih}^{(k)}; \theta)} \int_0^{t_{ih}^{(k)}} p(t_{ih}^{(k)}, W_h; \theta) W_h dW_h \right) \quad (7)$$

$$\frac{\partial Q_2}{\partial \alpha'_h} = \sum_{\substack{(I,j) \in \mathcal{T}_2 \\ I \ni h}} \sum_{k=1}^{n_{Ij}} \left(\frac{1}{p(\mathbf{t}_{Ij}^{(k)}; \theta)} \sum_{C \in I} \int_0^{\min(t_{Ij}^{(k)})} p(\mathbf{t}_{Ij}, C, W_j; \theta) \left\{ \begin{array}{l} \left(-\frac{1}{\alpha'_h} + \frac{t_{hj}^{(k)} - W_j}{\alpha_h'^2} \right) \quad C = h \\ \left(-\frac{t_{hj}^{(k)} - W_j}{\alpha_h'^2 (e^{(t_{hj}^{(k)} - W_j)/\alpha_h'} - 1)} \right) \quad C \neq h \end{array} \right\} dW_j \right) \quad (8)$$

$$\frac{\partial Q_2}{\partial \beta'_h} = \sum_{\substack{(I,j) \in \mathcal{T}_2 \\ j=h}} \sum_{k=1}^{n_{Ij}} |I| \left(-\frac{1}{\beta'_h} + \frac{1}{\beta_h'^2 p(\mathbf{t}_{Ih}^{(k)}; \theta)} \sum_{C \in I} \int_0^{\min(t_{Ih}^{(k)})} p(\mathbf{t}_{Ih}, C, W_h; \theta) W_h dW_h \right) \quad (9)$$

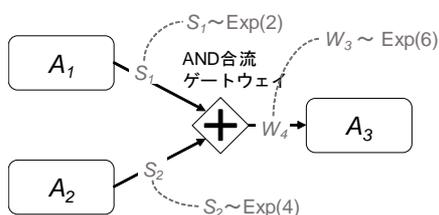


図 5 実験に使用した AND 合流ゲートウェイと確率分布

Fig. 5 An AND-join gateway and probabilistic distributions used in experiment

図 5 に示すような AND 合流ゲートウェイを構成し、 A_1 の潜在サービス時間は $S_1 \sim \text{Exp}(2)$ 、 A_2 の潜在サービス時間は $S_2 \sim \text{Exp}(4)$ 、 A_3 の潜在待ち時間は $W_3 \sim \text{Exp}(6)$ とした。提案手法が推定する値の真の値は、 A_1 の平均潜在サービス時間は $E[S_1] = 2$ 、 A_2 の平均潜在サービス時間は $E[S_2] = 4$ 、 A_3 の平均潜在待ち時間は $E[W_3] = 6$ であり、これらの数値に近ければ近いほど良い推定値であると言える。AND 合流ゲートウェイを通過する遷移時間のセットを乱数を用いて 1000 セット生成し、そのデータを元に推定を行った。

図 6 に、M ステップで繰り返し更新される推定値の履歴を示した。実線が推定値、破線が真の値を示す。繰り返しごとに推定値が真の値に近づいており、提案手法が平均潜在待ち時間とサービス時間を推定できる事がわかった。また、本論文で提案した尤度関数が正しくモデル化できている事を示している。

一方で、真の値に近づけば近づくほど、推定値の変化が少なくなり、収束が遅くなっている。これは、尤度関数が極大値付近で平らになっている事を示している。一般に隠れ変数が増え、周辺化を行えば行うほど尤度関数が平らになり、推定が困難になる事が知られている。本研究では現実の問題からの要請として、待ち時間とサービス時間、クリティカルパス、同期待ち時間の観測されない 3 種類の事象を導入しているため、推定問題自体が難しい問題となっ

たとえられる。

6. おわりに

本研究では、AND 合流ゲートウェイを含むビジネスプロセスからの不完全なイベントログであっても、平均潜在待ち時間とサービス時間を推定する方法を提案した。これまで現実問題として典型的に観測されるイベントログが不完全であるためにできなかった性能分析が、提案手法を利用する事で可能になった。このような性能指標はビジネスプロセスの性能改善だけでなく、性能に関する分析の特徴量としても用いることができるため、多くの応用が期待できる。

また、人工的に生成させたログを用いた数値実験により、提案手法が正しく推定できることを示した。この研究で取り組んだ課題は、待ち時間、サービス時間、クリティカルパス、同期待ち時間が観測されないという厳しい条件であるにも関わらず、正しい推定値を得ることができる事がわかった。

一方で、このような未観測事象を周辺化して取り扱ったため、推定問題自体が解きにくい構造を持ち、収束が遅くなるといったことが分かった。標本数を増やすことで推定問題の構造の難しさは緩和されると予想されるが、一方で尤度や勾配の計算は標本数に比例して増大するため、収束を早める解決方法は自明ではない。

AND 合流ゲートウェイは他のゲートウェイと共に使用することが可能であり、この論文で実験した構成よりも複雑なアクティビティ遷移が起こりうる。本論文での定式化でそういった複雑な場合についても扱えるため、検証を行う予定である。

本論文では時間間隔の確率分布として指数分布を仮定した。この仮定は多くの場合で現実的だが、より複雑な場合は整合性を保てなくなる。例えば複数の小さなタスクがアクティビティ内に内包されているが外からそれを観測できない場合、処理時間は指数分布の和になる。こうした一般

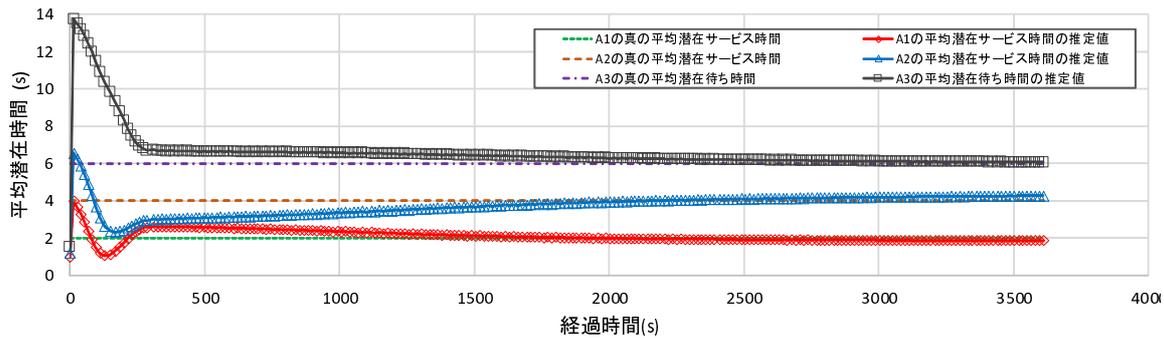


図 6 繰り返し更新する推定値の推移
Fig. 6 Estimators of each iteration

の処理時間の分布としてはガンマ分布が知られており，提案手法をガンマ分布を基にした手法に拡張すればより多くの場面で整合性を失うことなく推定が可能になるため，今後の研究課題である．

参考文献

- [1] Csiszár, I. and Tusnady, G.: Information geometry and alternating minimization procedures, *Statistics and Decisions*, Vol. Supplement, pp. 205–237 (1984).
- [2] Dempster, A. P., Laird, N. M. and Rubin, D. B.: Maximum likelihood from incomplete data via the EM algorithm, *Journal of the royal statistical society. Series B*, Vol. 39, No. 1, pp. 1–38 (1977).
- [3] Dumas, M.: From models to data and back: the journey of the BPM discipline and the tangled road to BPM 2020, *BPM 2015. LNCS*, Vol. 9253, Springer, Heidelberg (2015).
- [4] Dumas, M., van der Aalst, W. M. P. and ter Hofstede, A. H. M.: *Process-Aware Information Systems: Bridging People and Software through Process Technology*, Wiley (2005).
- [5] Ferreira, D. R.: Performance analysis of healthcare processes through process mining, *ERCIM News 89*, pp. 18–19 (2012).
- [6] Günther, C. W. and Verbeek, H.: Extensible event stream (XES) standard, <http://www.xes-standard.org/>.
- [7] Hathaway, R. J.: Another interpretation of the EM algorithm for mixture distributions, *Statistics & Probability Letters*, Vol. 4, No. 2, pp. 53–56 (1986).
- [8] Hornix, P. T.: Performance Analysis of Business Processes through Process Mining, *Master's Thesis, Eindhoven University of Technology* (2007).
- [9] IEEE Task Force on Process Mining: Process mining manifesto, *BPM 2011 Workshops, LNBIP* (Daniel, F., Barkaoui, K. and Dustdar, S., eds.), Vol. 99, Springer, Heidelberg, pp. 169–194 (2012).
- [10] Kuo, Y.-H., Leung, J. M. Y. and Graham, C. A.: Simulation with data scarcity: Developing a simulation model of a hospital emergency department, *Proceedings of the 2012 Winter Simulation Conference*, IEEE, pp. 1–12 (2012).
- [11] Lanz, A., Weber, B. and Reichert, M.: Time patterns for process-aware information systems, *Requirements Engineering*, Vol. 19, No. 2, pp. 113–141 (2014).
- [12] Leemans, S. J., Fahland, D. and van der Aalst, W. M.: Using Life Cycle Information in Process Discovery, *BPM 2015 Workshops, LNBIP*, Springer, Heidelberg (2015).
- [13] Neal, R. M. and Hinton, G. E.: A new view of the EM algorithm that justifies incremental and other variants, *Learning in Graphical Models*, Kluwer Academic Publishers, pp. 355–368 (1993).
- [14] Nogayama, T. and Takahashi, H.: Estimation of average latent waiting and service times of activities from event logs, *BPM 2015. LNCS*, Vol. 9253, Springer, Heidelberg, pp. 172–179 (2015).
- [15] OMG: Business Process Model and Notation (BPMN) (2010).
- [16] Rogge-solti, A. and Kasneci, G.: Temporal anomaly detection in business processes, *BPM2014. LNCS* (Sadiq, S., Soffer, P. and Hagen, V., eds.), Vol. 8659, Springer, Heidelberg, pp. 234–249 (2014).
- [17] Senderovich, A., Weidlich, M., Gal, A. and Mandelbaum, A.: Queue mining - predicting delays in service processes, *CAiSE 2014. LNCS* (Jarke, M., Mylopoulos, J., Quix, C., Rolland, C., Manolopoulos, Y., Mouratidis, H. and Horkoff, J., eds.), Vol. 8484, Springer, Heidelberg, pp. 42–57 (2014).
- [18] Sindhgatta, R., Dasgupta, G. B. and Ghose, A.: Analysis of operational data for expertise aware staffing, *BPM 2014. LNCS* (Sadiq, S., Soffer, P. and Hagen, V., eds.), Vol. 8659, Springer, Heidelberg, pp. 317–332 (2014).
- [19] van der Aalst, W. M. P.: *Process Mining: Discovery, Conformance and Enhancement of Business Processes*, Vol. 136, Springer (2011).
- [20] van der Aalst, W. M. P., Pesic, M. and Song, M.: Beyond process mining: from the past to present and future, *CAiSE 2010. LNCS* (Pernici, B., ed.), Vol. 6051, Springer, Heidelberg, pp. 38–52 (2010).
- [21] van der Aalst, W. M. P., Schonenberg, M. H. and Song, M.: Time prediction based on process mining, *Information Systems*, Vol. 36, No. 2, pp. 450–475 (2011).
- [22] Zerguini, L.: On the estimation of the response time of the business process, *17th UK Performance Engineering Workshop, University of Leeds* (2001).
- [23] 加藤光幾 (訳), IEEE Task Force on Process Mining: プロセスマイニングマニフェスト, <http://www.win.tue.nl/ieeetfpm/lib/exe/fetch.php?media=shared:pmm-japanese-v1.pdf>.
- [24] 野ヶ山尊秀: アクセスログの滞在時間を処理時間と次の遷移ページの前処理時間とに分解する方法, 電子情報通信学会技術研究報告. LOIS, ライフインテリジェンスとオフィス情報システム, Vol. 114, No. 32, pp. 39–43 (2014).