Vol.2020-SE-205 No.9
2020/7/2

タスクマイニングに向けたタスクインスタンス識別技術の提案

應治沙織[1]　斎藤忍[1]

概要：ビジネスにおける様々な業務を実施する際，文書作成や表計算ソフトの利用やファイルの閲覧，情報システムの立ち上げ・遷移は必須の作業として業務を実施する作業者各々の PC 上で実施される．タスクマイニングとはこれらの作業を業務の遂行に必要なタスクとして捉え，可視化する技術である．しかし，タスクマイニングでは入力データとして用いる PC の操作ログから分析対象となる業務の遂行に関連する作業のまとまり（タスクインスタンス）を識別することが困難である．本研究はこの問題の解決手段として，操作ログ上の各作業に対し識別子を自動付与するタスクインスタンス識別技術を提案した．

キーワード：タスクマイニング，プロセスマイニング，業務分析，業務理解

Task instance identification method for task mining

SAORI OUJI[1]　SHINOBU SAITO[1]

1. はじめに

PC を利用した業務を実施する現場では, 業務の複雑化に伴い業務分析とそれに伴う業務改善が求められている．しかし，手作業で実施される業務分析は，分析精度が分析者のスキルに依存し，定量的な分析が困難な上，多くの稼働を要するという問題点がある．

これに対し業務に利用する情報システム（以下，業務用システムと呼ぶ）からログデータを取得し，分析することで業務プロセスを復元する技術としてプロセスマイニングが提案されている[1][2].

プロセスマイニングは業務用システムの利用履歴やトランザクションデータから作成されたイベントログと呼ばれるデータを用いる．これにより，業務用システム上で実施されたアクティビティを確認できる．

プロセスマイニングの他にも，業務を実施する作業者（以下，業務実施者と呼ぶ）各々が実施した PC を用いた情報システム（以下，システムと呼ぶ）の立ち上げ，画面遷移，ドキュメント閲覧などの一つ一つの動作（以降，作業と呼ぶ）毎に業務分析を実施する手段として新たにタスクマイニングが提案されている[3][4].

タスクマイニングは業務を実施する PC 上で行われた操作に関する情報を操作ログとして取得し，業務に関連する作業のまとまりをタスクとして抽出する．タスクマイニングは業務を実施する上で発生する作業に着目した問題点の発見に役立ち，プロセスマイニングと連携して実施することで，業務をプロセスレベル・タスクレベルで分析することができる．

しかし，タスクマイニングでは入力データとして用いる PC の操作ログから, 各作業が何を目的として実施されたかを示す情報を取得することが困難である．分析対象となる業務のために実施された作業を操作ログから抽出し，タスクを生成するためには操作ログ上の各作業がどの業務を目的として実施されたか（業務種別）を判定する情報と，同じ業務種別である業務が複数回実施された場合何回目の業務に関連する作業か（案件番号）を判定するための情報が必要となる．本研究では，分析対象業務の実施に関連する作業のまとまりを，タスクインスタンス[5]と呼ぶ．

本研究では，操作ログの各レコードに対し業務種別を識別する識別子と案件を識別する識別子を自動付与することで，操作ログに含まれるタスクインスタンスを識別する技術を提案する．提案技術を用いることで，タスクマイニングを実施する際，特定の業務・案件ごとに分析することが可能となり，業務分析の効率向上が見込まれる．

2. 背景

2.1 業務分析手法

従来業務分析では，分析目的に合わせて表 1 に示す手段が利用される[6].

表 1　業務分析手法

分析手法	概要
文書調査	業務設計時に作成された業務設計所やマニュアルなどの文書を閲覧することで業務の構造を理解する．また，現行業務と文書化された業務との乖離を発見する．
ヒアリング	分析対象業務の実施者や監督者へ業務の実施状況について聞き込みを実施することで，業務実施現場の抱える問題や改善点を発見する．

1 日本電信電話株式会社
NIPPON TELEGRAPH AND TELEPHONE CORPORATION

ⓒ2020 Information Processing Society of Japan

シャドウイング	業務の習熟者の業務実施状況を観察し，分析者が自ら業務を実施し，分析対象で行うべき作業の情報を習得し，課題を明らかにする．

これらの手法は分析者の手作業で実施されるため，分析に必要な情報を細かく収集できる．一方，分析精度は分析者のスキルに依存し，収集する情報に抜けや漏れも発生しやすくなるというデメリットがある．加えて，分析に多くの稼働を要するため，十分な分析を実施できない場合もある．

2.2 関連技術

業務遂行に必要となる作業が PC 上で実施される電子化された業務では，プロセスマイニングやタスクマイニングといった分析技術が用いられる場合がある．

一般的に電子化された業務では文書作成ソフトや表計算ソフトなどを用いてドキュメントを作成し，業務用システムへ投入することで業務実施者間での書類の共有や承認などを実施する．こういった業務では業務用システムや業務に利用した PC から取得した履歴情報をもとに，業務プロセスやタスクを自動生成する技術が提案されている．

2.2.1 プロセスマイニング

プロセスマイニングは「いつ，誰がそのシステムを利用し，どんな作業を実施したか」を示すデータを収集したイベントログと呼ばれるデータを入力情報として利用し，実施された業務を業務プロセスとして復元する技術である．

イベントログは主に業務用システム内で発生した作業や時刻情報，作業の実施者情報などが収集される．また，ERP や基幹情報システムから生成されるイベントログを用いることで，業務を実施するために業務用システム上で行われたすべての作業を業務実施者や組織に隔てられることなく横断的に業務プロセス化できる[7]. 図 1 に示す例では，イベントログに記録されている業務プロセスを可視化している．

プロセスマイニングによって生成された業務プロセスは実際に実施された業務プロセスであるため，現状の業務に含まれるボトルネックや改善箇所を発見するために有用な情報となる．

Aalst らは道路と水道に関するインフラ設備の建設と維持管理業務へプロセスマイニングを適用することで，業務中に発生する請求書の処理作業を業務プロセス・組織・ケースの観点から分析し，プロセス内の効率が望ましくない作業の特定を行った[8]．一方で，この事例では業務実施者がイベントログ上に作業を記録されないよう意図的に行動し，実際の業務プロセスから逸脱した業務プロセスを生成してしまう可能性があることも述べられている．この事例から，プロセスマイニングはイベントログ上に記録された作業の可視化，及びそれに伴う業務プロセスの分析には効果的であるものの，イベントログとして記録されない作業は分析できないといえる．しかし，イベントログとして記録されない作業は，例えば情報システムへ情報を投入するための文書作成や情報収集などの細かい作業が多く，プロセスマイニングはマクロな視点から組織・業務プロセスを分析するために有効であると考えられる．

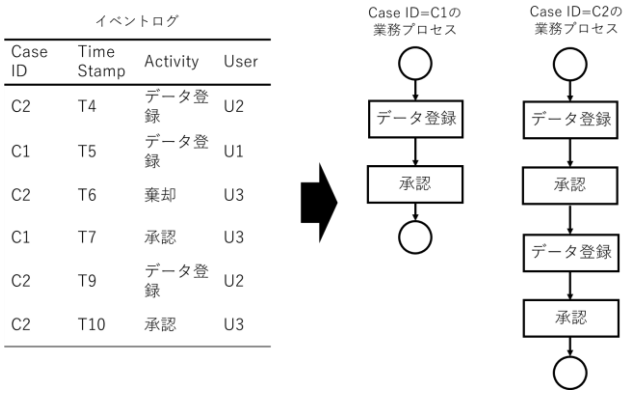


図 1　プロセスマイニング

2.2.2 タスクマイニング

プロセスマイニングに対し，より細かい粒度で業務分析を実施する新たな技術として，タスクマイニングが提唱されている．プロセスマイニングが業務用システムの記録をもとに業務プロセスを復元するのに対し，タスクマイニングでは業務実施者が作業に利用した PC から情報を収集し，タスクと呼ばれる業務を実施するために業務実施者が実施した作業のまとまりを生成する．これにより，PC を用いて実行された Web アプリケーションの立ち上げやインターネットブラウザの利用，ウィンドウの切り替え，文書の作成・閲覧などすべての作業を分析することができる．

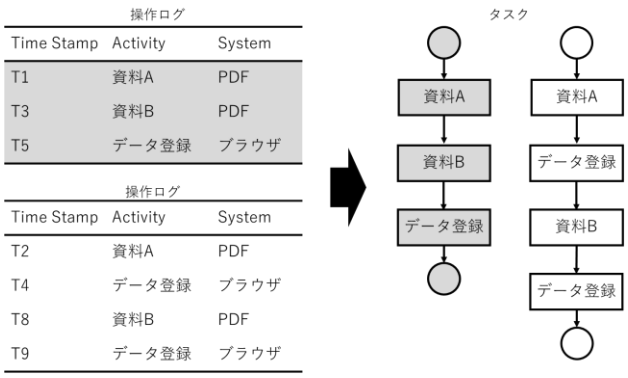


図 2　タスクマイニング

図 2 における資料 A，資料 B，データ登録などは PC を用いて実施され，操作ログへ記録された作業である．
タスクマイニングにおけるタスクとは，上記の図で示す「資料 A」「資料 B」などの最小単位の作業の集合であると言える．上記の例に示す通り，タスクマイニングを用いること

で実施された業務について，業務実施者毎のタスクレベルで分析することができる．

しかし，タスクマイニングで可視化される作業は，操作ログとして取得できるPC内で実施された作業のみであり，複数の業務実施者によって分担されて実施される業務を分析するのは困難である．さらに，タスクとして取得される作業の粒度は非常に細かいため，業務全体で実施されたタスクを可視化するため業務に携わる全業務実施者の操作ログを統合してタスクマイニングを実施すると，情報量が膨大になり業務を実施する組織におけるボトルネックや業務全体の問題点を把握するには不向きとなる．プロセスマイニングとタスクマイニングは双方を目的に応じて使分けることで効果を発揮する技術であると言える．

3. タスクマイニングの課題

2.2 節に示す通り，プロセスマイニングとタスクマイニングは双方を使い分けることで，業務プロセス上のボトルネック発見やそれに伴うリソースの調整といったトップダウンの業務改善や，業務を実施するために業務実施者が PC 上で実施した作業を補足し，改善に役立てるといったボトムアップの業務改善を実施できる．しかし，タスクマイニングには，操作ログ内の業務に関連する作業をタスクとして抽出するにあたり，操作ログに記録される作業から，分析対象となる業務のために実施された作業を識別することが困難になるという問題がある．これは，業務実施者の PC 上には業務のために実施された作業は記録されるものの，それぞれの作業が何の業務のために実施されたかという目的を示す情報が記録されないためである．図 3 に示す通り，業務実施者が操作ログの収集中に複数の業務を実施していた場合，分析対象業務のために実施された作業の特定はタスクマイニングを効果的に実施するためにも重要な作業となる．

この問題を解決するためには，操作ログ上の各作業が何の業務を目的として実施されたか（業務種別）を判断する情報と，同じ業務種別である業務が複数回実行された場合，何回目の業務に関連する作業か（案件番号）を識別する情報を取得する必要がある．

本提案ではこの問題を解決するため，操作ログの各レコードに対し業務種別を識別する識別子と案件を識別する識別子を自動付与することで，操作ログに含まれるタスクインスタンスを識別するタスクインスタンス識別技術を提案する．

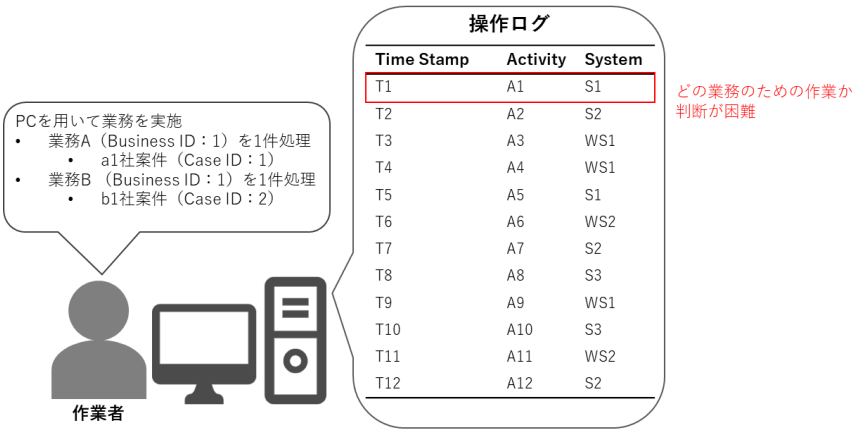
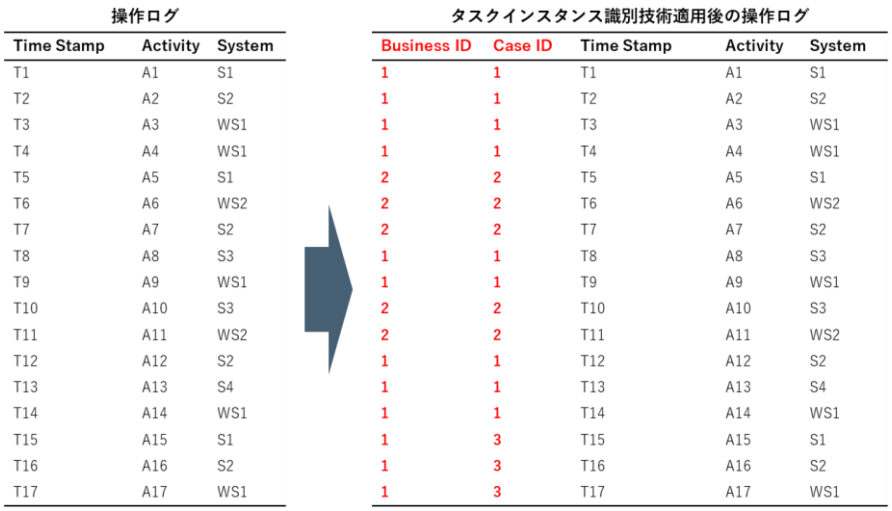


図 3　タスクインスタンスの識別



操作ログ

Time Stamp	Activity	System
T1	A1	S1
T2	A2	S2
T3	A3	WS1
T4	A4	WS1
T5	A5	S1
T6	A6	WS2
T7	A7	S2
T8	A8	S3
T9	A9	WS1
T10	A10	S3
T11	A11	WS2
T12	A12	S2
T13	A13	S4
T14	A14	WS1
T15	A15	S1
T16	A16	S2
T17	A17	WS1

タスクインスタンス識別技術適用後の操作ログ

Business ID	Case ID	Time Stamp	Activity	System
1	1	T1	A1	S1
1	1	T2	A2	S2
1	1	T3	A3	WS1
1	1	T4	A4	WS1
2	2	T5	A5	S1
2	2	T6	A6	WS2
2	2	T7	A7	S2
1	1	T8	A8	S3
1	1	T9	A9	WS1
2	2	T10	A10	S3
2	2	T11	A11	WS2
1	1	T12	A12	S2
1	1	T13	A13	S4
1	1	T14	A14	WS1
1	3	T15	A15	S1
1	3	T16	A16	S2
1	3	T17	A17	WS1

図 4　タスクインスタンス識別技術

ⓒ2020 Information Processing Society of Japan

4. タスクインスタンス識別技術

本研究では，前章で示したタスクインスタンス識別に関する問題解決に向けて，操作ログから分析対象業務に関する作業を推定するタスクインスタンス識別技術を提案する．タスクインスタンス識別技術は操作ログに対し，操作ログの各作業に対し，作業が目的とする業務種別と案件番号を自動推定する技術である．本研究では，図 4 に示す通り，タスクインスタンス識別技術において業務種別を識別する情報を Business ID，案件番号を識別する情報を Case ID と定義する．

4.1 アルゴリズム

タスクインスタンス識別技術の実現に当たり，タスクインスタンス識別技術を適用できる業務及び利用するシステムを利用した業務の前提を以下に示す．

1. 分析対象となる業務では業務遂行に必要となる作業を PC 上で実施する
2. PC 上では業務に関する作業を実施するために業務用システムとその他のシステムが利用される
3. 業務用システムとは，業務を実施するために利用される情報システムである
4. 業務用システムと業務は 1 対 1 の関係で紐づいており，利用された業務用システム名から業務種別を特定できる
5. その他のシステムとは PC 上で利用される業務用システム以外の情報システムである
6. その他システムは業務用システムへ投入する情報やファイルの作成・編集のために利用される
7. 業務用システムを用いた作業の 1 つ前のレコードにその他のシステムを用いた作業が実施されていた場合，その作業は一つ後の作業（業務用システムを用いた作業）の前準備のために実施された作業である
8. 業務終了時には必ず業務用システム上で特定の画面（終了操作）へ遷移する
9. 終了操作は業務終了時にのみ実施される

以上の仮定をもとにタスクインスタンス識別技術を提案するにあたり，操作ログとともに事前調査情報として，操作ログを収集する PC 上で実施されるすべての業務とその業務用システム名と終了操作名を収集する．これらの情報は，業務が実施される現場で取得可能なデータであり，シャドウイング・ヒアリング・文書調査で調査すべき情報と比較して取得が容易な情報であると推測する．

上記の仮定と事前調査情報をもとに，操作ログに記録される作業が属する Business ID と Case ID を推定する例を図 5 に示す．図 5 では操作ログの 5 行目，6 行目に示す作業が属する Case ID と Business ID を上記の仮定をもとに推定する例を示している．図 5 の例に示す考え方をもとにタスクインスタンス識別技術を実現するアルゴリズムを図 6 に示す．図 6 に示すアルゴリズムでは，操作ログと事前調査情報及び 4.1 節に示す前提をもとに，操作ログに含まれる終了操作の位置から案件を終了した際の作業と実施された業務を判定し，業務に関連する作業と関係していない作業を識別している．

また，本アルゴリズムでは 1 つの作業（操作ログ上の 1 行）に対し，複数の Business ID・Case ID が付与される場合がある．これは作業に用いられているシステム名や直後に実施された作業から正しい Business ID・Case ID を一意に特定できなかったためである．この場合，該当する作業が関連している可能性のある Business ID・Case ID をすべて付与することとしている．これは，4.1 節に示す前提およびアルゴリズムが操作ログ上に記録されている作業の前後関係や利用したシステムのみを利用してタスクインスタンス識別技術を実現しているためである． Business ID・Case ID を一意に特定できない作業に対し，より関連しうる可能性の高い Business ID・Case ID を推定する手段として，時刻情報を用いて実施時刻の近い作業へ重みをつける，Activity をテキストデータとして扱い，テキストとして類似する Activity を有する作業同士を同じ Business ID・Case ID へ属する可能性が高い作業として推定する手段が考えられる．しかし，例えば業務を途中で中断し，別の業務を実施した，PC から離席した時間があるなど同じ Business ID・Case ID に属する作業であっても作業の実施時刻に間が存在する場合や，業務を実施する上ために PC 上で編集していたドキュメントのタイトルなど，Activity として取得されるテキストデータが途中で変更された，業務とは関連性の低いタイトルであった場合などに時刻情報や Activity のテキストデータとしての信頼性は大きく損なわれる．したがって，タスクインスタンス識別技術では時刻情報や Activity 名を利用する手段を用いず，作業に対し関連する可能性のある Business ID・Case ID をすべて付与することとした．

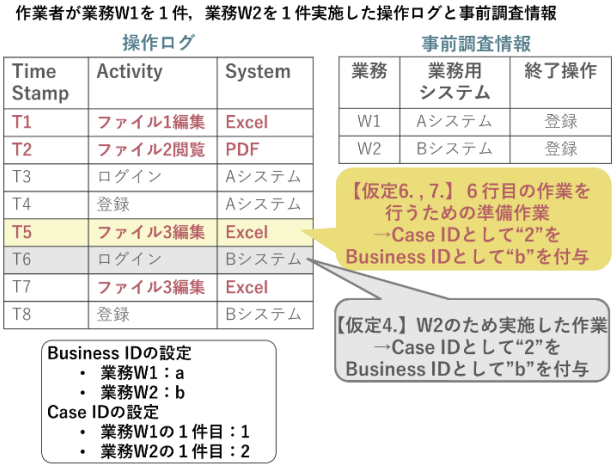


図 5　提案手法の考え方

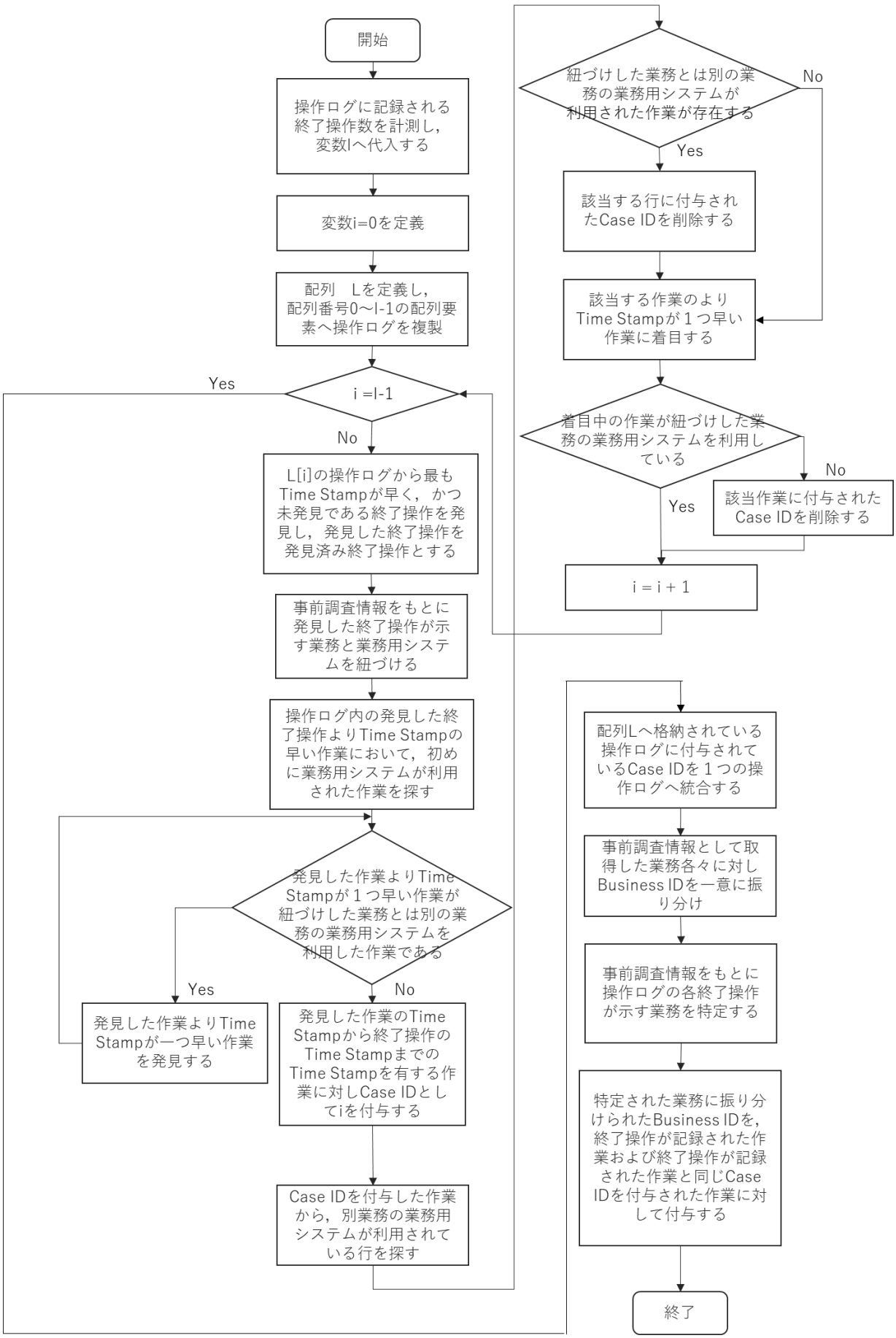


図 6　タスクインスタンス識別アルゴリズム

5. 実験

本章ではタスクインスタンス識別技術の有効性を検証する実験について説明する．

5.1 対象業務とデータセット

実験対象となる分析対象業務およびデータセットについて説明する．実験に利用する分析対象業務は企業で実際に実施された「見積」,「契約」,「出荷」の 3 種類の社内決裁の審査業務である．業務実施者は決裁を起案したユーザが業務用システムに添付した資料と記入した情報を確認し，記載内容に問題があるか否かを判断するタスクを PC 上で実施する．本実験では分析対象業務の操作ログを 10 日間収集した．

3 種類すべての審査業務において図 7 のタスクが実施されるが，種別ごとに業務用システム上に異なるウィンドウが表示されるため,「決裁画面で基本項目確認」作業および「処理結果記録」作業で業務種別を判別できる．

図 7 に示すタスクでは業務用システムを服す複数のシステムが PC 上で利用されている．実施された 3 種類の社内決裁の審査はすべて図 7 に示す手順に従って実施される．タスクは三名の業務実施者から操作ログとして収集した．取得したデータは提案技術の適用に必要となる操作ログと業務用システム名,業務終了時に実施される作業名である．

5.2 評価方法

タスクインスタンス識別技術の評価に当たり，再現率・精度について説明する．タスクインスタンス識別技術の評価では，業務用システム以外のシステムを利用した作業に対して正しく Case ID を付与できているかを評価する．これは，4.1 節に示す仮定から，業務用システムを用いた作業と業務終了時に実施される作業は提案技術を適用せずとも該当する Business ID・Case ID を判定できると考えられるため，純粋にタスクインスタンス識別技術を用いて Business ID・Case ID を推定できたか否かを評価すべき対象は，操作ログ内の業務用システム以外のシステムを利用した作業のレコードであると考えたためである．ここから，再現率・精度を下記の式によって算出した．

$$再現率 = \frac{正しくIDを付与できたレコード数}{業務用システム以外のシステムを利用したレコード総数}$$

$$精度 = \frac{正しくIDを付与できたレコード}{業務用システム以外のシステムを利用したレコード総数 - ID{:}-1 が付与されたレコード}$$

再現率では業務用システム以外のシステムを利用した作業のレコードの内，正しく Case ID を付与できたレコードの割合を計算している．

精度では，操作ログ中に提案技術が想定していない作業手順が記録されていると想定し，それを削除したレコードへ正しく Case ID を付与できた割合を計算している．
ここで，提案技術が想定していない作業手順とは，定義されていた終了操作を実施せず終了した業務を指す．こういった業務が発生する原因としては，業務実施中に何らかの要因で業務を続行できず，途中で中断し終了操作が実施されなかったことが挙げられる．

また，タスクインスタンス識別アルゴリズム上では Business ID を付与する際，4.1 節に示す 4.「業務用システムと業務は 1 対 1 の関係で紐づいており，利用された業務用システム名から業務種別（Business ID）を特定できる」から，正しく Case ID が付与できた場合それをもとに正しい Business ID できると言える．これにより，Case ID が正しく付与できたか否かをもとに再現率・精度を導出する式を用いている．

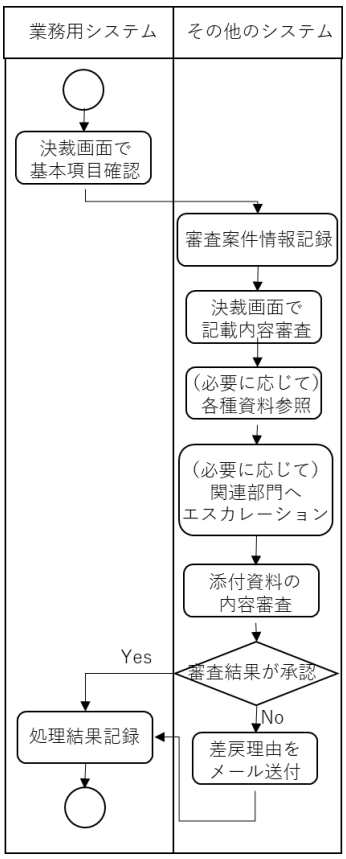


図 7　実験対象となるタスク

6. 実験結果

実験結果を表 2 に示す.

表 2　実験結果

ファイル	業務実施者	レコード総数	業務用システム レコード数	その他システム レコード数	再現率	精度
ログ 1	01	1381	448	933	0.809218	0.810956
ログ 2	01	1150	364	786	0.910941	0.913265
ログ 3	01	1271	377	894	0.928412	0.932584
ログ 4	02	907	247	660	0.872727	0.874052
ログ 5	02	1117	354	763	0.97903	0.981603
ログ 6	02	1546	430	1116	0.968638	0.971249
ログ 7	02	648	152	496	0.971774	0.983673
ログ 8	03	288	119	169	0.715976	0.728916
ログ 9	03	46	18	28	0.892857	0.892857
ログ 10	03	712	266	446	0.737668	0.737668
ログ 11	03	269	90	179	0.865922	0.865922

表 2 の 1 列目は実験に使用した操作ログを示しており，例えばログ 1 は業務実施者 01 の実施したタスクの操作ログであり，レコード総数 1381 行の内，業務用システムを用いて実施された Activity が記録されているレコード数が 448 行，その他のシステムを用いて実施された Activity が記録されているレコード数が 933 行であり，4.2 節の導出式をもとに計算した再現率は 0.809218，精度は 0.810956 であることを示している.

見積の審査業務に関するタスクは業務実施者 01 によって 18 回，業務実施者 02 によって 9 回，業務実施者 03 によって 7 回実施され，契約の審査業務に関するタスクは業務実施者 01 によって 13 回，業務実施者 02 によって 4 回，業務実施者 03 によって 7 実施され，出荷の審査業務に関するタスクは業務実施者 01 によって 4 回，業務実施者 02 によって 1 回，業務実施者 03 によって 1 回実施されている．これらのタスクは収集したログ 1～11 の操作ログ内に実施した時刻順に記録されている.

表 2 より実験の結果再現率は平均 0.87756，精度は平均 0.881159 となったことがわかる.

7. 考察

6 章の結果を元に，提案手法で正しい Business ID・Case ID を付与できなかった作業の詳細および提案手法の有効性について考察する.

正しい Business ID・Case ID を付与できなかった 1 割のレコードには終了操作以降に実施された，業務実施のために確認していたファイルを閉じるための作業やタスクに関する記録を残す作業などが記録されていた．これらの作業はタスクを遂行する主たる目的（案件情報の審査）に関わる作業ではないものの，タスクを完了する上で必要不可欠である. したがってこれらの作業に対し正確な Business ID・Case ID を付与することは提案技術の重要性を高めるためにも重要な課題であると言える.

一方で，タスクインスタンスの識別およびタスクマイニングは業務を実施する上で業務実施者が担っているタスクを知り，業務分析及び改善に生かすための技術である．業務分析を実施し，非効率的な作業の改善を目指すうえで，タスクのファイルやウィンドウを閉じる作業やメモを残す作業が実施されたか否かについては重要な情報とは言い難い．したがって，評価結果として得られた再現率と精度はタスクマイニングおよび業務分析･改善を目的とした場合，有効に活用できる数値であると考える.

8. まとめ

本研究では，業務を実施する際に業務実施者が利用した PC から取得した操作ログを活用してタスクマイニングを行う際，操作ログから，各作業が何を目的として実施されたかを示す情報を取得することが困難であり，タスクインスタンスを識別することが困難であるという問題を解決するため，操作ログの各レコードに対し業務種別を識別する Business ID と案件を識別する Case ID を自動付与するタスクインスタンス識別技術を提案した.

ⓒ2020 Information Processing Society of Japan

実験の結果, 再現率・精度ともに 9 割近く正しい Case ID, Business ID を付与できた．この結果はタスクマイニングを効果的に活用するために有効であると考える．

今後の検討としては，終了操作の後に実施される作業への正しい ID 付与が考えられる．

また，本提案は 3.1 節に示す限定的な業務に対してのみ有効な手段であり，例えば終了操作に該当する Activity が存在しないタスクや複数種類の業務で用いられる業務用システムなどの存在も考えられる．提案手法の適用範囲を広げることは重要な検討事項であると言える．

9. 謝辞

本提案を検討・評価するにあたり，研究の趣旨を理解し，実験に参加していただいた被験者の皆様に心から感謝いたします.

参考文献

[1] W.M.P. van der Aalst, A. Adriansyah, de Medeiros, et al., "Process Mining Manifest, "Business Process Management Workshops. BPM 2011. Lecture Notes in Business Information Processing, Vol 99., pp.169-194, Berlin, Heidelberg, 2011
[2] W.M.P. van der Aalst, "Process Mining Data Science in Action Second Edition," Springer, 2016
[3] “タスクマイニング – 労働生産性向上のための 3 つの分析始点” . https://www.process-mining.jp/2020/04/20/basic-analysis-of-taskmining/，(参照 2020-05-22).
[4] “What is Task Mining?” . https://www.celonis.com/process-mining/what-is-task-mining, (参照 2020-05-22)
[5] 應治沙織，斎藤忍，鈴木源吾, “作業空間情報の利用による業務構造理解技術の提案”，信学技報, vol. 118, no. 471, SS2018-77, pp. 151-156, 2019 年 3 月.
[6]Benedikt Schmidt , Johannes Kastl , Todor Stoitsev , Max Mühlhäuser , “Hierarchical Task Instance Mining in Interaction Histories”, Proceedings of the 29th annual international conference on Design of communication (SIGDOC), 2011
[7] Ana Pajic, Dragana Becejski-Vujaklija , “Process Mining on ERP systems: Logical Structure of The Artifact-Centric Approach of Log Extraction”, ICDSST 2015 PROCEEDINGS – THE 1 ST INT. CONFERENCE ON DECISION SUPPORT SYSTEMS, 2015
[8] W.M.P. van der Aalst, H.A. Reijers, A.J.M.M. Weijters, B.F. van Dongen, A.K. Alves de Medeiros, M. Song, H.M.W. Verbeek, "Business process mining: An industrial application," Information Systems, 2007

ⓒ2020 Information Processing Society of Japan