情報処理学会論文誌 Vol.63 No.4 1008–1018 (Apr. 2022)

深層強化学習による営業活動意思決定支援システム

中山 義人[1,2,a)] 森 雅広[2] 斎藤 忍[3] 成末 義哲[1] 森川 博之[1]

受付日 2021年8月3日, 採録日 2022年1月11日

概要：営業活動における意思決定から，営業担当者個人の経験や直感といった属人的要素を取り除くことにより，営業活動を大幅に効率化するための手段が求められている．筆者らはこの課題に対し，機械学習を用いた営業意思決定支援システムの構築を試みている．これまでの検討では，営業活動の意思決定プロセスに部分観測マルコフ決定過程（POMDP）を適用することにより，受注確率の高い営業プロセスの規則性を抽出し，それにより構築したプロセスモデルに基づいて次に実施すべきアクティビティのリコメンドを行う営業活動意思決定支援システムを開発し，実際の企業への適用評価を行った．本論文では，さらなるリコメンド精度の向上を目的として，新たに深層強化学習によりプロセスモデルを構築し直し，前論文で構築したPOMDPによるプロセスモデルと，熟練営業担当者によるアンケート比較による評価を行った．その結果，深層強化学習に対する熟練営業担当者の経験合致度の評価は，POMDPによるプロセスモデルよりも高いことを示すことができた．

キーワード：営業活動，意思決定支援，深層強化学習，環境モデル，プロセスモデル

Construction of Decision-making Support System in Sales Activities by Deep Reinforcement Learning

Yoshihito Nakayama[1,2,a)] Masahiro Mori[2] Shinobu Saito[3]
Yoshiaki Narusue[1] Hiroyuki Morikawa[1]

Received: August 3, 2021, Accepted: January 11, 2022

Abstract: In the decision-making of sales activities, the means for greatly increasing the efficiency of sales activities are required by eliminating individuals' factors such as experience and intuition. For this purpose, we are developing a decision support system of sales activity using a machine learning model. In the previous studies, we have developed the system based on a process model created by extracting the regularity of sales process with a high order acceptance probability, and then evaluated it in a real company. In this paper, for the purpose of further improving the recommendation accuracy, we reconstruct the process model by deep reinforcement learning and evaluated it through a questionnaire by skilled sales staff. As a result, it is shown that the evaluation of the experience matching degree of the expert is high.

Keywords: sales activity, decision-making, deep reinforcement learning, environment model, process model

1. はじめに

近年，社会情勢の急激な変化にともない，企業の意思決定に関わる要因や影響が複雑化するとともに，そのスピードも求められる時代になってきた [1]．このような時代では，あらかじめルールの決まっていない非定型業務を対象とした意思決定のシーンにおいても，膨大な可能性を探索しながらその中から有効な意思決定の候補を高速に見つける技術の追求が重要な技術課題となっている [1]．また，この技術課題の解決には，機械学習技術によるアプローチが不可欠である．

このように機械学習技術の適用により，これまで属人性

[1] 東京大学
The University of Tokyo, Bunkyo, Tokyo 113–0033, Japan
[2] 株式会社 NTT データイントラマート
NTT DATA INTRAMART Corporation, Minato, Tokyo 107–0052, Japan
[3] 日本電信電話株式会社
NTT Corporation, Chiyoda, Tokyo 100–8116, Japan
[a)] nakayamay41@gmail.com

の高かった意思決定を支援する仕組みが求められているなかで，筆者らは，企業活動のなかでもより複雑性が高く，かつ非定型の業務領域である営業活動の意思決定に注目してきた．この営業活動領域に機械学習技術を適用することで，営業活動の意思決定を支援する試みはこれまでも例がある．たとえば，顧客の行動データに基づいた有望顧客選別の効率化 [2], [3] や，販売見込みに基づいた売上予測 [4] などがあげられる．一方，これらの事例は営業活動における部分的な業務への適用に限られている．

また，事例 [5] では，銀行の営業活動で活用されているリコメンドについて報告されている．顧客属性を事前にクラスタリングしたうえで，営業タイミングに合わせた類似商談をリコメンドするという，特定時点のアクティビティに基づいたリコメンドである．同様に，事例 [6] では，どのお客様を訪問すればよいか，その商談ではどのような課題をヒアリングしどのような商材の提案を行うと効果的かを，1 人 1 人に合った形でリコメンドしている．しかし，このシステムも営業活動のスタート段階で利用する限定的な意思決定支援に限られている．

近年では，実企業での失敗事例も報告されている．それらの失敗原因は，営業活動における意思決定にはあらかじめ明確なルールが存在せず，営業担当者の属人的な経験に基づいて実施されるため，意思決定パターンがきわめて多岐にわたることや，顧客反応を見ながら意思決定が長期間にわたり連続で実施されるなかで顧客要求も時間につれて変化するなどの問題にある [7]．

以上の事例からも，初期訪問から受注に至るまでの一連のアクティビティの流れである営業活動全般を対象にした意思決定支援については研究事例の報告がないのが現状である．

営業活動では，営業担当者が顧客とのコミュニケーションを通じて営業知識を動員し，場面に応じて顧客に対する適切な対応方針（営業のアクティビティ）を選択することが重要である．そのため，この意思決定を支援するシステムには，顧客との間でつねに変化する場面に応じた適切な対応方針のリコメンドを，営業担当者に対して提示することが求められる．このような意思決定支援システムが，営業活動のプロセス全般を通じてリコメンドを繰り返していくことで，営業担当者の活動を受注に誘導し，結果として高い売上を実現することが可能となる（図 1）．さらには，受注成果の高い営業担当者の意思決定が，組織全員で再現可能となることで組織力の強化にもつなげることができる．以上のような考えに基づき，これまでに筆者らは，受注確率の高い営業プロセスが反映されたプロセスモデルの構築，およびそれに基づいた営業活動意思決定支援システムを構築している [8], [9]．

具体的には，既存の業務システムから出力されるイベントログからプロセスモデルを構築するプロセスマイニング

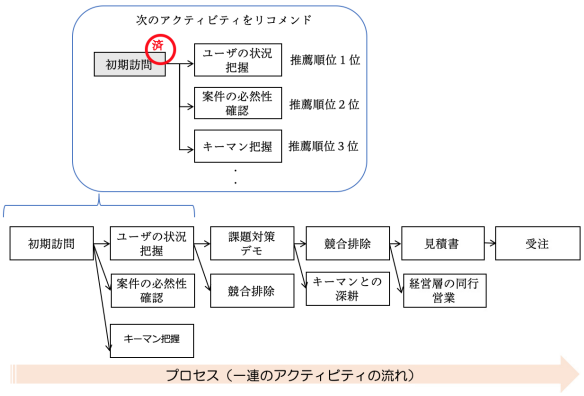


図 1 営業活動意思決定支援システムのイメージ
Fig. 1 Decision support system of sales activity.

手法 [10], [11] を応用し，社内に蓄積された営業日報データから，各アクティビティの選択が受注結果に与える影響を受注確率として推定することで，受注確率の高い営業活動プロセスの規則性を機械学習により抽出しプロセスモデルを構築した．

前論文では，プロセスモデル構築アルゴリズムとして，部分観測マルコフ決定過程 [12]（以下，POMDP）を適用した営業活動意思決定支援システムを，営業活動を営む企業の営業担当者による実適用を通して評価を行った結果を示した [13]．

本論文では，さらなるリコメンド精度の向上を目的として，前論文で構築した POMDP に代えて，新たに環境モデルとセルフプレイを用いた深層強化学習 [14] によりプロセスモデルを構築し直し，POMDP によるプロセスモデルと比較評価を行う．探索系のアルゴリズムによるプロセスモデルの構築となり，リコメンドの選択肢が拡大することで，営業の納得性の高いリコメンドとなることが期待される．

これまでに筆者らが実施した関連研究の調査においては，営業活動意思決定支援システムの学習アルゴリズムとして深層強化学習が適用された報告はなく，今回が初めての試みとなる．

本論文の構成は以下のとおりである．2 章「背景」では，筆者らの既存研究（POMDP によるプロセスモデル構築アルゴリズム）の概要を述べる．3 章「深層強化学習による営業活動意思決定支援システム」では，深層強化学習によるプロセスモデル構築の概要について解説し，さらに 4 章「評価」では，熟練営業担当者を対象にしたアンケート評価の結果を示した．最後に 5 章「考察」，6 章「まとめ」を述べる．

2. 背景（POMDP による営業活動意思決定支援システム）

筆者らは，蓄積した営業日報データから，受注確率の高い営業プロセスを抽出しプロセスモデルを構築し，それに基づいたリコメンドを提供するシステムを研究している．

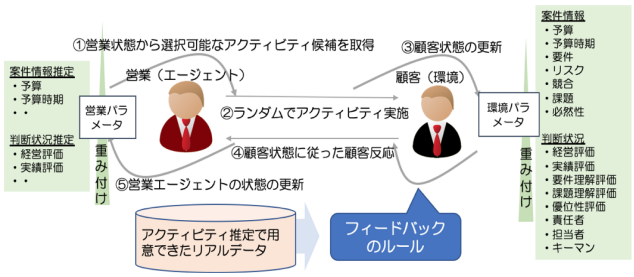


図 2　案件シミュレータの詳細
Fig. 2　Details of case simulator.

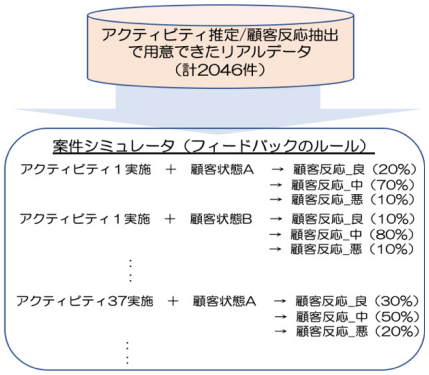


図 3　フィードバックのルール詳細
Fig. 3　Detail of feedback rules.

以下にその概要と評価結果を再掲する．

2.1　プロセスモデル構築アルゴリズムの特徴

営業活動におけるプロセスモデル構築アルゴリズムに求められる特徴を以下に示す．

特徴 1：営業活動を確率的状態遷移で表現

営業活動意思決定支援システムの目的である受注確率の高いリコメンドの提示には，受注確率を含んだプロセスモデルの構築が必要となる．アルゴリズムには，営業プロセスを通じたアクティビティの選択を確率的に表現できることが求められる．

特徴 2：営業活動における時間連続の意思決定を支援

一般的に営業活動は長い系列となる．アルゴリズムには，営業活動の時間的推移を考慮できることが求められる．

特徴 3：顧客状態によりリコメンドが変更可能

長い営業活動のプロセスを通じて顧客の状態は変化していくことから，リコメンドもそれに合わせて変更していく必要がある．アルゴリズムには，営業担当者から直接は観測できない顧客状態を推定し，高い精度のリコメンドを生成できることが求められる．

そこで，上記 3 つの特徴を満たすプロセスモデル構築アルゴリズムとして，既存研究 [13] では POMDP を採用した．採用理由は以下のとおりである．営業活動における顧客の内部状態は遷移するが，その遷移確率は営業担当者からのアクティビティにより変化する．すなわち，営業活動における顧客の内部状態はマルコフ決定過程であるといえる．さらに，顧客状態は直接観測できないことから，POMDP によってモデル化が可能であると考えた [12]．

2.2　案件シミュレータ

POMDP では，前処理として大量の教師データを必要とする．一方，実際の営業活動（日報メールなど）から取得できるデータには限りがある．そのため，既存研究 [13] では，大量の教師データを準備するため，営業をエージェント，顧客を環境と位置付けた案件シミュレータを開発し，ルールベースの営業日報データを自動生成した．

案件シミュレータの詳細を述べる（**図 2**）．営業パラメータ，顧客パラメータの全項目が正規分布で初期設定された状態から始まり，その後，営業状態に基づくルールベースのアクティビティ候補の中からランダムにアクティビティが選択され実施される．次に顧客側では，正規分布で重み付けされた顧客特性に従い顧客状態が更新される．たとえば，アクティビティとして「プレゼン実施」が選択された際には，「実績重視」の特性を持つ顧客ではそのパラメータがより加算して更新される．その後，更新された顧客状態に即した顧客反応が後述するフィードバックルールから選択され，最後に，顧客反応に対応したルールベースの重み付けにより営業パラメータが更新される．このため，営業側でも顧客状態の推定値を保有できる．

次に，案件シミュレータのなかで重要な位置付けとなるフィードバックルールについて述べる．ここでは，計 2046 件の実際の営業活動から取得した日報データを用いて，アクティビティと顧客状態の区分ごとに顧客反応のフィードバックルールを，熟練営業担当者が設定した（**図 3**）．取得した営業日報データは，顧客反応までかなり詳細に記述されていたことから，人によるルール設定ではあるが，営業実態をほぼ正確に投影したルールといえる．以上の案件シミュレータにより自動作成した案件データ総数 10,000 件と，営業日報データであるアクティビティデータ総数 267,000 件のデータセットを教師データにして，POMDP によるプロセスモデルを構築した．

2.3　プロセスモデルの構築手順

形式的には，POMDP は隠れマルコフモデルに行動および行動を変更する動機を与える報酬を付与したものと解釈できる [15]．そこで，前論文においては，POMDP の近似手法として，隠れマルコフモデルによる状態観測に対して，方策学習による行動制御を組み合わせたプロセスモデルを構築した．具体的には，営業をエージェント，顧客を環境と位置付けた隠れマルコフモデルと方策学習によるプロセスモデルを以下の手順で構築する（**図 4**）．

(1)　価値関数の計算

案件シミュレータにより確保できた営業日報の教師デー

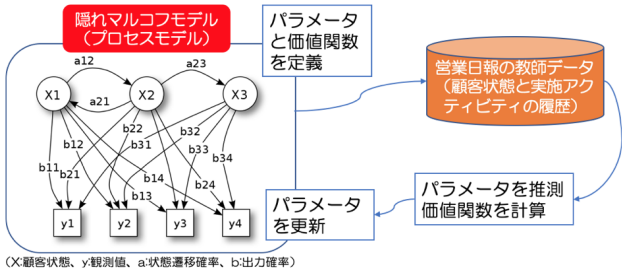


図 4　隠れマルコフモデルによるプロセスモデルの構築
Fig. 4　Process model with hidden Markov model.

表 1　隠れマルコフモデルのパラメータ
Table 1　Parameters of hidden Markov model.

分類	項目	説明	設定値
HMM	系列数	学習もしくは評価用の系列の数。実運用では、「案件データ数」に相当。	10000
HMM	イテレーション	HMMの学習は、勾配降下法に近いイメージで行われる。本値は勾配降下法の更新ループの繰り返し数に相当。	100
HMM	状態数	HMMの隠れ状態の数。顧客と営業の判断基準の種類数だけ必要。	300
MDP	成功報酬	MDPの最適方策の学習で使用する受注状態になった際に得られる報酬。	100000
MDP	失敗報酬	MDPの最適方策の学習で使用する失注状態になった際に得られる報酬。	5000
MDP	減衰率	MDPの最適方策の学習では系列長が長くなるたびに報酬を減衰する。その減衰率。	0.9
HMM→MDP変換	アクション実行確率の足切	HMMの誤差で、本来は発生しないアクションが非常に低確率で発生する事がある。これをアクションの選択肢に入れると、本来は選べないアクションが選べてしまう。 HMMの学習精度が高いほど低い値を設定する。	0.02
HMM→MDP変換	遷移確率の足切	顧客アクションを省く自然遷移の過程ですべての遷移を考慮すると状態の遷移パターンが発散し、MDPの学習が非常に遅くなる。このため、一定確率以下の遷移は無視するように設定。	0.000001

タから，隠れマルコフモデルのパラメータである状態遷移確率（顧客状態の遷移確率）と出力確率（顧客状態におけるアクティビティ選択確率）を推測し価値関数を計算する．ここで価値関数は推測した顧客状態で，営業が次のアクティビティを選択後，最適方策をとり続ける際の受注率の期待値である．

(2)　方策の学習

価値反復法により価値関数の値を最大化するアクティビティの組み合わせ方策を学習する．

(3)　パラメータの更新

学習結果に基づいて，隠れマルコフモデルの出力確率パラメータを更新する．

以上を方策の価値が収束するまで繰り返し実施することでプロセスモデルを構築し，それに基づいた営業活動意思決定支援システムを構築した．**表 1** にモデル構築時のパラメータを掲載する．顧客応答が「受注/失注」，アクティビティが「提案辞退」に到達したら終了，さらにイテレーション数が 100 に至ったら強制終了，減衰率により早く終わるほど高い報酬となるよう設定した．顧客状態数は試行錯誤により 300 とした．

2.4　アルゴリズムの評価

構築した POMDP によるプロセス構築アルゴリズムは，実企業の営業現場で評価を実施している．実企業とは，筆者らの一部が所属する（株）NTT データイントラマートの法人営業グループであり，法人企業向けの自社製ソフトウェアを販売している．これまで法人営業のベテラン社員を中心とした属人的な営業で継続的に成果をあげてきている．若手社員の人数も増加していることから，属人性からの脱却とそのノウハウの継承が課題であった．

6 カ月の営業活動期間を 2 つに分け，前半 3 カ月を予備実験（ターム 1），後半 3 カ月を本実験（ターム 2）として，営業活動意思決定支援システムを実際の営業活動に適用した．ターム 1 の予備実験の終了後に，ターム 1 運用期間における営業担当者の実際のアクティビティ選択結果をフィードバックしてプロセスモデルを再構築し，意思決定支援システムのリコメンド精度を改善した．ターム 2 の本実験では，その精度改善した意思決定支援システムを適用して評価した．その結果，営業担当者による予備実験の実運用結果を反映してプロセスモデルの学習は進み，ターム 2 の本実験ではターム 1 の予備実験よりも精度の高いリコメンドが提示されるようになった．つまりプロセスモデルの学習向上により，営業活動意思決定支援システムのリコメンドが優秀な営業担当者の思考に近づき，営業担当者にとってより納得性の高いリコメンドへと精度が向上したことを確認できた [13]．

3.　深層強化学習による営業活動意思決定支援システム

本論文では，リコメンド精度のさらなる向上を狙い，2.1 節で述べた 3 つの特徴を満たしたプロセスモデル構築アルゴリズムとして，深層強化学習を新たに採用した．本章では，その着眼点について述べたのちに，このアルゴリズムを内包した営業活動意思決定支援システムの内容を示す．

3.1　深層強化学習による改善の着眼点

既存研究のプロセスモデル構築アルゴリズムである POMDP は，意思決定支援システムのリコメンド精度において，以下の改善点が存在する．

2.2 節で述べたように，案件シミュレータから生成される教師データは，フィードバックのルールによる顧客反応とアクティビティ（営業活動）の組合せとなる．しかし，そこから構築される既存研究（POMDP）の学習モデルでは，ルールには存在しない想定外のアクティビティが提示された際に，顧客からは現実の営業活動と乖離した顧客反応を返してしまうことが想定される．

以下に乖離の例を示す．本例では，営業活動のスタートである「初期訪問」のアクティビティから始まり，「要件取得」「課題取得」「課題対応」へとフィードバックのルールに沿ったやりとりを続けながら，順調に顧客反応が良い状態へと推移したとする．学習モデルでは，フィードバック

のルールに基づき，受注率をさらに高めていくためのアクティビティとして「予算取得」や「ステークホルダ取得」などが想定される．ここでもし，営業活動の後半で実施することの多い「提案プレゼンテーション」などのアクティビティ提示が実施されると，フィードバックのルールでは活動不足の状態となるため，顧客反応は悪くなることが予想される．一方，現実世界では，それまで順調に営業活動が推移しているので，多少の活動不足の状況における情報追加であっても顧客反応は悪くならず，良好な顧客反応が持続することが多い．そのため，想定している順番が前後逆転したアクティビティの実施のみで顧客反応が悪くなることは，現実の営業活動からは乖離した反応となる．このように，既存研究（POMDP）の学習モデルでは，フィードバックのルールには存在しない，すなわち事前に営業担当者が想定していないアクティビティに対して実態に即さない顧客状態に至ってしまうというリスクがある．

そこで本論文では，案件シミュレータにより生成される教師データに深層学習を実施した．教師データに含まれる特徴を多段的により深く学習することが可能となり，想定外の営業アクティビティに対してもより精度の高い顧客反応を推定できる．これにより，深層強化学習による学習モデルは，現実の営業活動のやりとりにより近づいたリコメンドが期待できる．

3.2 アプローチ

深層強化学習によるアルゴリズムは 2 つのモデルを用いる（**図 5**）．顧客状態をシミュレートした環境モデルと，営業担当者に受注確率の高いアクティビティをレコメンドするプロセスモデルである．1 つ目の環境モデルは，案件の背景，要件，顧客のパーソナリティといった顧客の環境を深層学習することで，現実の顧客反応に即したモデルを構築することになる．この環境モデルには，営業活動が確率的かつ時間推移で表現されている（2.1 節の特徴 1，特徴 2）．

また 2 つ目のプロセスモデルでは，環境モデルを用いた営業活動のシミュレーション（セルフプレイ [16] を用いた強化学習）によりモデルを構築することになる．案件ごとの営業日報のデータセットから深層学習による環境モデルの構築を行い，最適なアクティビティ選択を環境モデルと営業エージェントによるセルフプレイ探索で学習する仕組みである [17]．そのため，2.1 節の特徴 3 である顧客状態の変化にともなうリコメンド変更が可能となる．

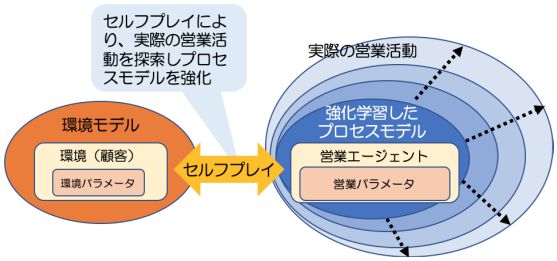


図 5 深層強化学習によるプロセスモデルの構築
Fig. 5 Process model with deep reinforcement learning.

以上で述べた深層強化学習によるモデルの構築を通じて，本システムでは営業の納得性の高いアクティビティのリコメンドを可能とする．以下に本システムの構築ステップを述べる．

(1) 教師データの準備

深層強化学習によるプロセスモデルの構築では大量の教師データが必要となる．既存研究 [13] と同様，案件シミュレータから自動生成された同一の教師データを用いる．

(2) 環境モデルの構築

案件シミュレータから生成された教師データは，フィードバックルールによる顧客反応とアクティビティ（営業活動）の組合せであるため，そのバリエーションが人の考えたルールに沿ったものしかない．そのため，ある顧客状態で，あらかじめ用意されたルール以外のアクティビティが生じると，想定したバリエーションから外れてしまい，顧客反応がそれまでの流れと大きく変化してしまう．その結果として，営業担当者にとって，リコメンドの納得性が減少するという問題が起きる．

本システムのアプローチでは，環境モデルを深層学習で構築するため，環境の特徴を多段的により深く学習することが可能となり，想定外のアクティビティに対してもより精度の高い顧客応答を推定して提示することができる．そのため，ルール以外のアクティビティが生じた際にも，比較的現実に近い顧客反応の提示へと安定した応答が期待される．環境モデルの構築方法については 3.3 節で述べる．

(3) セルフプレイによる強化学習とプロセスモデル構築

その後，構築した環境モデルと営業エージェントの間のセルフプレイで強化学習を実施することで，探索と活用を繰り返しながら環境の制御を学習しプロセスモデルを構築する．最終的にこのプロセスモデルを組み込むことで，営業担当者へより精度が高いリコメンドが可能な営業意思決定支援システムが構築できる．この詳細については 3.4 節で述べる．

3.3 環境モデルの構築

環境自体を深層学習してそれを学習に活用するという環境モデルの構築により，環境を案件シミュレータの生成する案件データに近づけていく．具体的には，案件シミュレータにより自動生成された 10,000 件の案件データ（営業のアクティビティを入力，それに対する顧客反応を出力としたアクティビティ系列）を教師データとして，長期的な依存関係を学習することのできる深層学習アルゴリズムである LSTM（Long Short Term Memory）[18] で環境モデルを構築する（**図 6**）．RNN（Recurrent Neural Network）

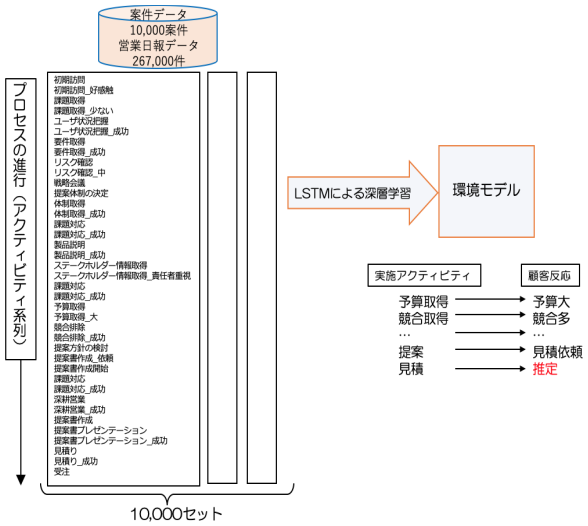


図 6　LSTM による環境モデルの構築

Fig. 6　Environmental modeling with LSTM.

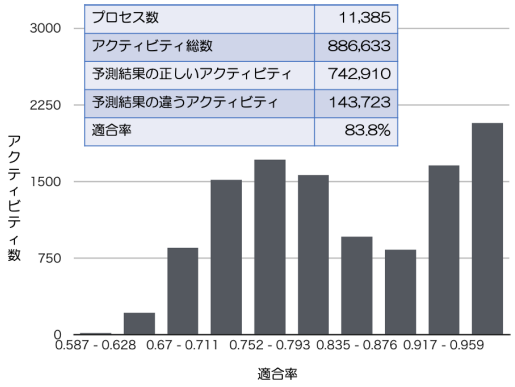


図 7　構築された環境モデルによる予測結果

Fig. 7　Prediction result of environmental modeling.

の拡張として登場した LSTM は，特別な種類の RNN であり，長期的な依存関係を学習できる．ここでは，アクティビティ系列の最終状態は，「受注/失注/提案辞退/強制終了」のいずれかになる．LSTM による深層学習で環境モデルを構築することにより，環境モデルが現実の顧客反応を網羅した状態に近づいていく．

構築された環境モデルによる予測結果を**図 7** に掲載する．LSTM による深層学習で，入力されたアクティビティからそれに対応する顧客反応を予測できるようになるが，その正解率は最終的に 83.8%となり，予測結果と違うアクティビティは情報の少ない案件初期段階において発生していたことから，適切に環境モデルの学習が行われ要件を満たす精度が達成できたと判断した．

3.4　セルフプレイによる強化学習とプロセスモデル構築

作成された環境モデルを利用して，営業エージェントとの間のシミュレーションをセルフプレイを用いた強化学習で行い，営業パラメータのポリシと価値を試行錯誤で獲得する．これにより，現実の営業活動を探索することが可能

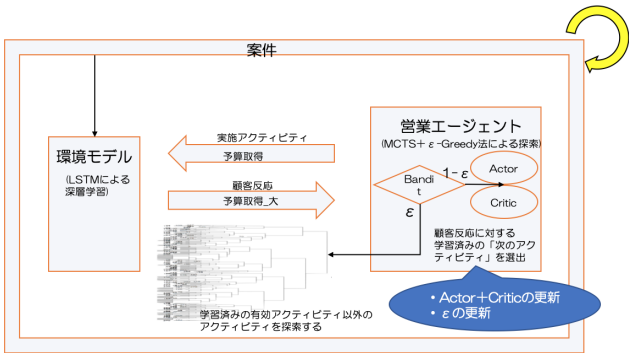


図 8　セルフプレイによる強化学習

Fig. 8　Reinforcement learning with self-play.

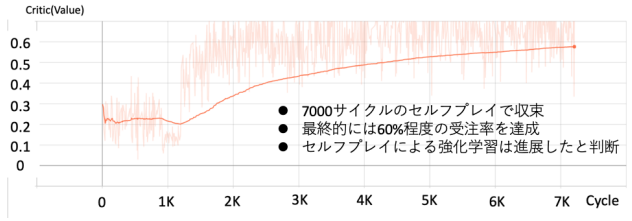


図 9　セルフプレイの収束結果

Fig. 9　Change in order rate using self-play.

となり，営業エージェントの学習モデルを強化する．現実の営業活動との乖離が少ない学習モデルが構築でき，営業活動場面でより現実の営業活動に沿ったリコメンドが可能となる手法を確立できる．ここではゲーム分野で多く用いられる強化学習のアルゴリズムである MCTS（モンテカルロ木探索）アルゴリズム [19] を使った．ゲーム分野での強化学習では，対戦相手がどのような相手でも勝てる確率が高い手を探索することができる．

具体的には，顧客と営業の対戦という形で，10 案件で 1 サイクルを 10,000 サイクル，双方を対戦させ，学習モデルの精度を向上させる（**図 8**）．このセルフプレイを繰り返すことで営業モデルをシミュレートで強化することができ，結果として MCTS による探索で作成されたプロセスモデルは理想的なリコメンドが可能な仕組みとなる．**図 9** に示すように，7,000 サイクルのセルフプレイが 1 週間程度で受注率 60%程度に収束することができた．あわせてセルフプレイの初期値（営業エージェントのパラメータ）についても**表 2** に掲載する．

セルフプレイの実施過程ではかなりの試行錯誤があり，それについても述べる．初めに MCTS でセルフプレイを実施したところ，セルフプレイは終了しなかった（1 週間に 5 案件，探索のみの繰返し）．そこで，MCTS+ε-Greedy 法 [20]（学習に応じて徐々に挙動を改善する手法）によるセルフプレイを実施したところ，途中終了や無限ループする案件が過半となった．そこで，教師データの受注案件からアクティビティのみを抜き出して模倣学習 [21] を実施し，初期方策を事前に学習した後，MCTS+ε-Greedy によ

表 2 セルフプレイの初期値
Table 2 Initial value of self-play.

区別	項目	設定値
	イテレーション回数	10,000
Actor-Critic	Actor-Criticの設定(学習率)	5e-3
	Actor-Criticの設定(報酬の減退率)	0.99
	Actor-Criticの設定(先読み数)	8
	Actor-Criticの設定(価値ロスの影響度)	0.5
	Actor-Criticの設定(エントロピー影響度)	0.01
ε-Greedy	ε-Greedyの設定(開始時の確率)	0.2
	ε-Greedyの設定(終了時の確率)	0.0
	ε-Greedyの設定(ステップ数) ※ランダム選択の確率が開始時の確率から終了時の確率までステップ数のイテレーション回で変動する	100,000
MCTS	MTCSの設定数(シミュレーション回数) ※環境の状態を一定数にするために，連続量の状態を量子数の状態に変換する	1,000
	MCTSの設定(量子数)	10
	MCTSの設定(UCTの定数)	10
	MCTSの設定(確率計算時の通過回数の累乗数)	1

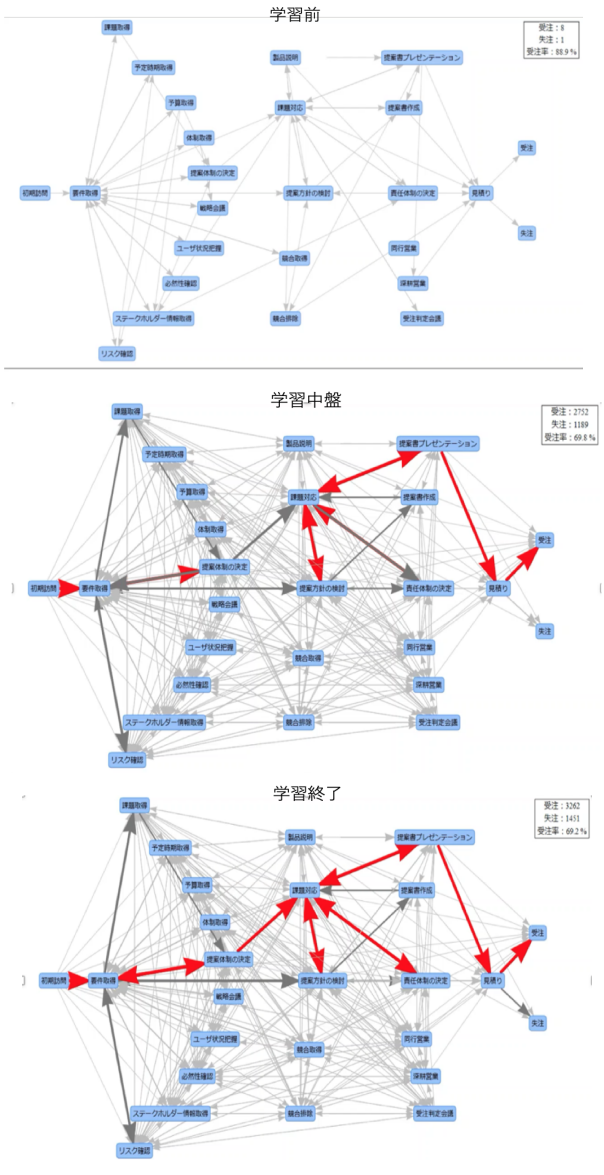


図 10 セルフプレイにより学習が進む過程
Fig. 10 Learning process through self-play.

るセルフプレイを実施した．結果として，Actor-Critic [22] で方策を学習し収束することができた．図 10 に，セルフプレイにより徐々に学習が進む過程を動画でビジュアル化した際の流れを示す．

4. 評価

本論文で提案した深層強化学習による営業活動意思決定支援システムを，既存研究 [13] と同じ熟練営業担当者 20 名に再度協力してもらい評価してもらった．具体的には，前論文で実適用評価した POMDP によるプロセスモデルをリコメンド精度比較の対象とし，それぞれのアルゴリズムによるプロセスモデルからリコメンドされたアクティビティの推移を，営業経験 10 年以上の熟練営業担当者 20 名に確認してもらい，その後，アンケートを実施してリコメンドの納得度を測定した．

質問は，POMDP と深層強化学習によるプロセスモデルのそれぞれについて，営業プロセスの展開推移が自分の営業経験と照らして，合致するか否かで評価した．具体的には，1 人ごとに，2 つのアルゴリズムからそれぞれ 10 案件合計 20 案件のリコメンドについて，ランダムにブラインドテストしてもらい，合致しない（= 0），少し合致（= 1），おおむね合致（= 2），合致する（= 3）の中からどの適合度に該当するかを判断してもらった．それぞれのアルゴリズムの適合度の平均値を縦軸にした 20 名の営業担当者の評価結果が図 11 である．

営業担当者 20 名中 18 名が深層強化学習によるプロセスモデルのリコメンド結果に高い評価を与えている．また 20 名の平均値でも深層強化学習によるプロセスモデルは 2.20 と「おおむね合致（= 2）」以上のスコアとなった．一方で，POMDP によるプロセスモデルの評価平均値は，1.61 であった．

あわせて，2 つのプロセスモデルのアンケート結果について，平均の差の検定を実施した．独立 2 群の差を検定するときのノンパラメトリックな方法であるマン・ホイットニーの u 検定 [23] で検定を実施した．検定の結果，統計量 U = 356 を算出し，有意水準 0.05（n1 = 20 かつ n2 = 20

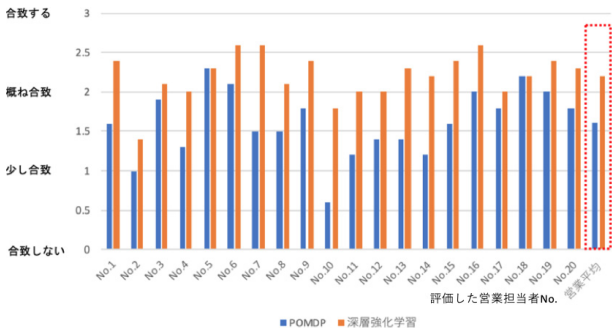


図 11 深層強化学習のプロセスモデルの評価
Fig. 11 Evaluation of deep reinforcement process model.

のとき限界値は 127）で有意差があることを確認できた．このことは，高い統計的有意性を持って，熟練営業担当者の深層強化学習に対する評価（経験合致度）は POMDP よりも高いことが説明可能であることを示している．

5. 考察

5.1 プロセスモデルの比較と営業担当者の評価

既存研究（POMDP）と本論文で提案した深層強化学習，それぞれのアルゴリズムによるプロセスモデルからリコメンドされたアクティビティ推移を詳しく調べることで，営業担当者の評価に与える影響を考察した．その結果，以下 2 点の特徴的な違いが観察された．それぞれの内容について，両手法の比較結果と営業担当者の評価コメントを示す．

(1) リコメンドの提示範囲の比較と営業担当者の評価

それぞれのプロセスモデルからリコメンドされるアクティビティの提示範囲を測定し比較した．構築したプロセスモデルにおいては，合計で 38 種類のアクティビティがリコメンド可能となっているが，深層強化学習では 1 案件平均で 38 種類のアクティビティをすべて使って提案していた．一方，既存研究（POMDP）では，1 案件平均で 27 種類のアクティビティの提案にとどまっていた（**表 3** (1) 提案アクティビティ数の比較を参照）．

また，アンケートに参加した熟練営業担当者に対して，個別にヒアリングを実施した際の代表的な意見を以下に記す．

- 深層強化学習によるプロセスモデルでは，過去の受注確率を超えたモデルが構築できるため，顧客反応に対して提示されるアクティビティが，営業経験上から理解しやすい．
- 同じく深層強化学習によるプロセスモデルでは，提示されたリコメンドは想定したアクティビティとは異なるものの，受注までの流れを追って確認すると納得性がある．

以上のような意見は，強化学習によるセレンディピティ（意外性のある提案）が熟練営業担当者に評価されたと考えられる．深層強化学習では，より良いアクティビティを探して可能性を探りつくした結果を提示するモデルになっているため，営業担当者にとってもセレンディピティが大きくなると推察される．

(2) リコメンドの内容の比較と営業担当者の評価

それぞれのプロセスモデルからリコメンドされたアクティビティ推移の内容を詳しく調べると，「提案辞退」のリコメンドがその違いとして特徴的であったことから，「提案辞退」のリコメンドを中心に内容を分析した．まず，既存研究（POMDP）では深層強化学習に比べて「提案辞退」というリコメンドで終了している案件が多いことが注目できる．評価する営業担当者 1 人あたりの「提案辞退」の案件数は，既存研究（POMDP）では 10 案件のうち平均 3.8 件，深層強化学習では 10 案件のうち平均 1.7 件と 2 倍以上の差が確認できた（表 3 (2) 提案辞退案件数の比較を参照）．

「提案辞退」は，顧客と営業の状態乖離が大きく，どんなアクティビティを実施してもいい顧客反応がなく，徐々に悪い状況に入り込んでしまうことから，以後はどのアクティビティを選んでも受注できないと判断された場合に提示される．既存研究（POMDP）は過去の学習データに依存しているため，学習データにない応答パターンになると「提案辞退」になりやすいと推察される．

さらに，「提案辞退」で終了した案件を抽出し，1 案件あたりのアクティビティ系列数を確認すると，既存研究（POMDP）では平均 12.6 系列，深層強化学習は平均 41.2 系列と 3 倍以上の差が確認できた（表 3 (3) 提案辞退案件の系列数の比較を参照）．このことは，既存研究（POMDP）では深層強化学習に比べて比較的早期に「提案辞退」が出やすい傾向があることを示している．そのため，営業担当者にとって既存研究（POMDP）は深層強化学習と比べても「すぐに諦める」という印象に映ったのではないかと推察される．

それぞれのアルゴリズムによるプロセスモデルからリコメンドされたアクティビティ推移から，「提案辞退」で終了となった案件を一部抽出し可視化してみた．**図 12** では，1 案件を横 1 列として，プロセスモデルからリコメンドされたアクティビティ種類が色の違いで可視化されている（左端から案件はスタートし右に向かってアクティビティが推移する）．この図からも，既存研究（POMDP）では深層強化学習に比べて比較的早期に「提案辞退」が出やすい傾向があることを確認することができる．

この分析後，アンケートに参加した熟練営業担当者に対して，個別にヒアリングを実施した際の代表的な意見を以下に記す．

- 既存研究（POMDP）では，「提案辞退」で終わる案件が多いことが気になる．
- 既存研究（POMDP）では，営業活動の継続中であるにもかかわらず突然に「提案辞退」が出て，提案をす

表 3 プロセスモデルによるリコメンドの比較
Table 3 Comparison of recommendation of process models.

（1）提案アクティビティ数の比較

	POMDP	深層強化学習
提案アクティビティ数（1 案件平均）	27	38

（2）提案辞退案件数の比較

	POMDP	深層強化学習
提案辞退案件数 （評価営業担当者 1 人あたり平均）	3.8案件 /10案件	1.7案件 /10案件

（3）提案辞退案件の系列数の比較

	POMDP	深層強化学習
提案辞退案件の系列数 （1 案件あたり平均）	12.6	41.2

図 12 提案辞退で終了した案件の可視化
Fig. 12 Visualization of cases that ended by decline of proposal.

ぐに諦めることに違和感を感じた．

以上 2 点の特徴的な違いと営業担当者のコメントから，熟練担当者が既存研究（POMDP）に対して低く評点し，深層強化学習では高く評点したと推察できる十分な理由を分析することができた．

このように熟練営業担当者によるアンケートを通じて，深層強化学習のプロセスモデルは，2.20 と「おおむね合致（= 2）」以上のスコアとなったことから，営業活動意思決定支援システムの実適用を予備実験としてスタート可能と判断した．以後は本実験を繰り返すことにより，実用に向けてさらに精度を向上させてくこととなる．

5.2 アンケート評価の妥当性

熟練営業担当者による案件ごとの評価を 1 つ 1 つ確認すると，深層強化学習に良い評点をしていない 2 名の営業担当者（図 11 の営業担当者 No.5 と No.18）に限らず，すべての営業担当者において，既存研究（POMDP）であっても経験合致度 2.0（おおむね合致）以上の高い評価を与えているものもあれば，逆に深層強化学習で経験合致度 1.0（少し合致）を下回る低い評点を与えているものもある．ただ 20 名の平均値を取ると，深層強化学習は 2.20 と「おおむね合致（= 2）」以上のスコアである一方で，既存研究（POMDP）によるプロセスモデルの評価平均値は，1.61 であった．

今回は，営業担当者による主観的なアンケート評価であるため，営業担当者の案件ごとの評価理由まではつかめなかったが，ほとんどの熟練営業担当者にも納得性が高い結果となったことから，営業人材の底上げとして新人営業担当者の育成には有効であると考えられる．しかし，今後は熟練営業担当者にとっても，納得性が高いだけではなく，利用することでさらに効果が上がるようなリコメンド創出が求められる．それには受注率などの定性的な評価手法を追加で取り入れて，多面的に評価をしていく必要がある．

5.3 営業意思決定支援システムの適用可能性の評価

構築した営業支援意思決定支援システムの適用可能性を述べる．本論文の適用環境では，ソフトウェアを販売する法人営業に限定して実施したものであるが，長期にわたるアクティビティを通じて受注率を高めていくタイプの案件

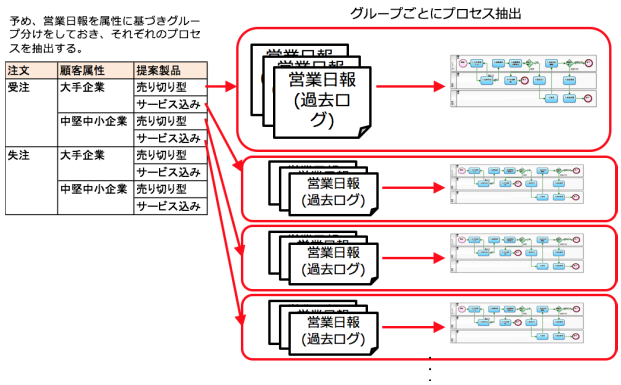


図 13 グループ分類ごとのプロセス抽出
Fig. 13 Process extraction by group classification.

型営業においては，本論文で提示したプロセスモデルの構築手法はそのまま適用可能である．

具体的には，受注確率の高いアクティビティ選択のプロセスは，販売する品目の種類や複雑性，さらには対象とする顧客規模などによっても異なるといわれている [24]．しかし，営業活動のアクティビティ自体を，販売する品目，種類，複雑性，顧客規模などに合わせて新たに設定し，本論文の構築手法に従いプロセスモデルを再構築することで対応可能である．**図 13** に示すように，営業対象となる顧客の属性情報（企業規模）と提案対象製品（売り切り製品か，サービス込みの製品か）によりあらかじめグループ分類しておいた営業日報ごとに，このプロセスモデルを構築する作業を繰り返していくことになる．これにより適用環境ごとに複数の理想的な営業活動のプロセスモデルが生成できる．

特に，本論文で想定している顧客訪問を前提とした営業活動の標準アクティビティは，現在のコロナ環境下においてはリモート営業のアクティビティへと大きく変化するが，これらの変化に対しても同様に対応可能である．

一方，営業のパターンの分類によると（**図 14** 参照），本手法が適用できないのは，発注頻度の高い商品を中心に特定の顧客を定期的に巡回訪問するルート営業（選択アクティビティに変化がないタイプ），公共系に多い入札案件（アクティビティ選択よりも提示価格などの特定のアクティビティの中身が受注率に大きく影響するタイプ），リアル店舗や EC サイトなどの販売業務（長期のプロセスがない

本手法の適用対象外

	案件営業	ルート営業	入札案件	販売営業
内容	車や家、システム開発などの案件をベースにした営業	発注頻度の高い商品を中心に特定の顧客を定期的に巡回訪問する	公共系に多く、価格競争入札と総合評価入札、プロポーサル入札などがある	リアル店舗やECサイトなどの販売業務
特性	長期に渡るアクティビティ選択を通じて受注率を高めていくタイプ	選択アクティビティに変化がないタイプ	アクティビティ選択よりも特定のアクティビティの中身が受注率に大きく影響するタイプ	長期のプロセスがないタイプ

図 14 本手法の適用対象外の営業タイプ
Fig. 14 Sales type not covered by this method.

タイプ）である．

あわせて学習時間についても適用可能性を考察した．検証案件数 100,000 の実測値として，顧客状態数 300/アクティビティ数 37 のとき 25.0 時間，顧客状態数 300/アクティビティ数 38 のとき 26.2 時間となった．そのため，アクティビティ 1 件程度の追加では，学習時間の変化が実務利用でネックになることはない．ただし，計算上はアクティビティ増加率の 2 乗に比例するので，アクティビティ数が顕著に増加した際には学習時間の考慮が必要となる．一方で，学習モデルの構築は毎日実施するわけではなく，営業トレンドの乖離が大きくなった場合に再学習をすることになるため，通常の営業活動では多くても月 1 回程度となり，また仮にアクティビティ数が 2 倍になったとしても，4 日程度の学習時間となる．ちなみに，構築されたプロセスモデルからのリコメンドは即座に返却される．

6. まとめ

本論文では，営業日報などの非構造化データを入力情報として，あらかじめ明確なルールが存在しない非定型プロセスから，受注確率の高い営業プロセスの規則性を抽出・構築したプロセスモデルに基づき，営業意思決定支援システムを提案した．深層強化学習により構築したプロセスモデルについて，熟練営業担当者へのアンケート評価を行った．提案システムは，アクティビティを選択するつど，受注までのプロセスを予測して示すことになる．そのため，仮に理想から外れたアクティビティを選択しても，その後に可能な限りのリカバリアクションが提示されることになる．ディシジョンツリーで表現されるルールベースの意思決定支援システムでは，いったん間違った選択肢をとるとリカバリができないことと比べて大きな優位点となる．

しかしながら依然として課題もある．環境モデルの深層学習においては，教師データとなる案件データは十分なデータ量を確保しているものの，いまだすべての判断パターンを包含しているわけでなく，また非現実的なやり取りも含まれている．そこで今後は，深層学習による環境モデルの作成時に，人間参加型深層学習（human-in-the-loop）[25] を取り入れることで，より人間と機械知能を組み合わせて，効果的な機械学習アルゴリズムを生成する予定である．

参考文献

[1] 戦略プロポーザル：複雑社会における意思決定・合意形成を支える情報科学技術，国立研究開発法人科学技術振興機構研究開発戦略センター，CRDS-FY2017-SP-03 (2018).
[2] Huang, Y., Zhu, F., Yuan, M., Deng, K., Li, Y., Ni, B. and Dai, W.: Telco churn prediction with big data, *Proc. 2015 ACM SIGMOD International Conference on Management of Data*, pp.607–618, DOI: 10.1145/2723372.2742794 (2015).
[3] 河村一輝，諏訪博彦，小川祐樹，荒川 豊，安本慶一，太田敏澄：飲食店向け不動産営業を支援する申込み顧客推薦モデルの提案，人工知能学会論文誌，Vol.32, No.1, p.WII-O_1–10, DOI: 10.1527/tjsai.WII-O (2017).
[4] Yan, J., Zhang, C., Zha, H., Gong, M., Sun, C. and Huang, J.: On machine learning towards predictive sales pipeline analytics, *Proc. 29th AAAI Conference on Artificial Intelligence*, pp.1945–1951 (Feb. 2015).
[5] 広瀬洋平：熟練の営業を AI が伝授 みずほ，提案力底上げ，日本経済新聞オンライン，入手先 ⟨https://www.nikkei.com/article/DGKKZO56067130W0A220C2EE9000/⟩（参照 2021-10-25）.
[6] 日立製作所：「営業 × AI」の力で DX 時さいのスマートセールスを推進，日立製作所事例紹介（オンライン），入手先 ⟨https://www.hitachi.co.jp/products/it/ws_sol/web_concierge/case/otsuka-shokai/⟩（参照 2021-10-25）.
[7] 多田和市，吉野次郎，高橋 芳，竹居智久，奥平 力，広野彩子：AI バブル 失敗の法則，日経ビジネス，5 月 22 日号，pp.28–29 (2019).
[8] 中山義人，森 雅広，成末義哲，森川博之：営業活動における意思決定のプロセス発見手法：プロセスマイニングの応用アプローチ，情報システム学会誌，Vol.14, No.1, pp.26–38, DOI: 10.19014/jissj.14.1_26 (2018).
[9] Nakayama, Y., Mori, M., Narusue, Y. and Morikawa, H.: The process discovery Approaches for decision making in sales activities, *Proc. SCIS&ISIS 2018*, Okayama, Japan, pp.1394–1399, DOI: 10.1109/SCIS-ISIS.2018.00217 (2018).
[10] Van Der Aalst, W.: *Process Mining: Discovery, Conformance and Enhancement of Business Processes*, Springer, Heidelberg (2011).
[11] 飯島 正，田端啓一，斎藤 忍：プロセスマイニング・サーベイ（第 01 回：概要と基本概念），情報システム学会誌，Vol.11, No.2, pp.20–53, DOI: 10.19014/jissj.11.2_20 (2017).
[12] 澁谷長史：《第 4 回》部分観測マルコフ決定過程と強化学習，計測と制御，Vol.52, No.4, pp.374–380 (2013).
[13] 中山義人，森 雅広，斎藤 忍，成末義哲，森川博之：非定型業務における意思決定支援システムの適用ステップの提案と実践，情報処理学会論文誌トランザクションデジタルプラクティス，Vol.2, No.4, pp.1–12 (2021).
[14] Racanière, S., Reichert, D., Weber, T., Wierstra, V.D., Buesing, L., Battaglia, P., Pascanu, R., Li, Y., Heess, N., et al.: Imagination-augmented agents for deep reinforcement learning, *Advances in Neural Information Processing Systems*, Vol.30, pp.5692–5699 (2017).
[15] 木村 元：部分観測マルコフ決定過程下での強化学習（〈特集〉強化学習），人工知能学会誌，Vol.12, No.6, pp.822–830 (1997).
[16] Heinrich, J. and Silver, D.: Deep reinforcement learning from self-play in imperfect-information games, arXiv preprint arXiv:1603.01121 (2016).
[17] 中山義人，森 雅広，成末義哲，森川博之：営業活動の意思決定プロセス強化における環境モデルに基づくアプローチ，情報処理学会研究報告グループウェアとネットワークサービス，Vol.2019-GN-107, pp.1–5 (July 2019).

[18] Hochreiter, S. and Schmidhuber, J.: Long shortterm memory, *Neural Computation*, Vol.9, No.8, pp.1735–1780 (1997).
[19] Browne, C.B., Powley, E., et al.: A survey of monte carlo tree search methods, *IEEE Trans. Computational Intelligence and AI in Games*, Vol.4, No.1, pp.1–43 (2012).
[20] Tokic, M.: Adaptive ε-greedy exploration in reinforcement learning based on value differences, *Proc. Annual Conference on Artificial Intelligence*, pp.203–210 (2010).
[21] Ho, J. and Ermon, S.: Generative adversarial imitation learning, *Advances in Neural Information Processing Systems*, pp.4565–4573 (2016).
[22] Konda, V.R. and Tsitsiklis, J.N.: Actor-critic algorithms, *Advances in Neural Information Processing Systems*, pp.1008–1014 (2000).
[23] Study channel：Mann-Whitney の U 検定, Study Channel On Line, 入手先 ⟨https://www.study-channel.com/2015/07/mann-whitney-u-test.html⟩（参照 2021-10-25）.
[24] Viio, P.: *Strategic sales process adaptation: Relationship orientation of the sales process in a business-to-business context*, Svenska handelshögskolan (2011).
[25] Karwowski, W. et al.: *International encyclopedia of ergono.mics and human factors-3 Volume Set*, CRC Press (2006).

中山 義人 （正会員）

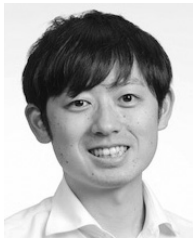
東京大学，NTT データイントラマート．1992 年東京大学応用生命工学系大学院修士課程修了．同年（株）NTT データ入社．2001 年 (株) NTT データイントラマート代表取締役社長，2021 年東京大学大学院工学系研究科博士課程修了，プロセスマネジメントの研究に従事．電気通信協会 ICT 事業奨励特別賞（2011 年），情報システム学会ベストペーパー特別賞（2017 年），電子情報通信学会 LOIS 研究賞（2019 年），FIT2019 FIT 論文賞（2019 年）受賞．

森 雅広

NTT データイントラマート．2001 年埼玉大学卒業．2009 年 (株) NTT データイントラマート入社 デジタルビジネス推進室所属．機械学習を用いたプロセスモデルの構築とプロセスオートメーションの研究開発に従事．

斎藤 忍 （正会員）

日本電信電話株式会社．2001 年慶應義塾大学大学院修士課程修了，同年（株）NTT データに入社．2015 年日本電信電話株式会社に転籍．現在，コンピュータ&データサイエンス研究所に所属．2016～2018 年カリフォルニア大学アーバイン校客員研究員．ソフトウェア工学，要求工学に関する研究開発に従事．2007 年慶應義塾大学大学院博士課程修了．博士（工学）．



成末 義哲 （正会員）

東京大学．2012 年東京大学工学部電子情報工学科卒業．2017 年同大学大学院情報理工学系研究科博士課程修了．博士（情報理工学）．同年同大学院工学系研究科助教．現在，同講師．無線工学，サイバーフィジカルシステム，エッジ AI 等の研究に従事．2013 年電気電子情報学術振興財団原島学術奨励賞，2019 年 IEEE CCNC Best Paper Award，2020 年 ACM IMWUT Distinguished Paper Award 等受賞．電子情報通信学会，IEEE 各会員．

森川 博之 （正会員）

東京大学．1987 年東京大学工学部電子工学科卒業．1992 年同大学大学院博士課程修了．博士（工学）．現在，同大学院工学系研究科教授．モノのインターネット/ビッグデータ/DX，無線通信システム，情報社会デザイン等の研究に従事．本会論文賞，電子情報通信学会論文賞（3 回），ドコモモバイルサイエンス賞，総務大臣表彰，志田林三郎賞，大川出版賞等受賞．OECD デジタル経済政策委員会（CDEP）副議長，Beyond 5G 新経営戦略センター長，総務省情報通信審議会部会長，スマートレジリエンスネットワーク代表幹事，情報社会デザイン協会代表幹事等．