機械学習による新規アイテムの需要予測と

意思決定最適化に関する研究

提出先　　大阪大学大学院情報科学研究科

提出年月　2023 年 6 月

出 水　宰

研究業績

学術論文 (英文)

- Tsukasa Demizu, Yusuke Fukazawa, and Hiroshi Morita. Inventory management of new products in retailers using model-based deep reinforcement learning. *Expert Systems with Applications*, Vol. 229, p. 120256, 2023. https://doi.org/10.1016/j.eswa.2023.120256

学術論文 (和文)

- 出水宰, 深澤佑介, 森田浩. 深層学習による時間減衰を考慮したインフィード広告の CTR 予測. 情報処理学会論文誌, Vol. 62, No. 1, pp. 292–301, 2021.

- 出水宰, Rubén Manzano, Sergio Gómez, 深澤佑介. モバイルネットワーク特徴量のクラスタリングによる Contextual Bandit Algorithm. 情報処理学会論文誌, Vol. 60, No. 1, pp. 38–47, 2019.

国際会議

- Tsukasa Demizu, Norihiro Katsumaru, Hiroyuki Suzuki. KDD Cup 2020 Reinforcement Learning Competition Track Task 2, 3rd Prize: Learning to Dispatch and Reposition (LDR) Competition 3rd Place Solution. *ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2020.

- Keiichi Ochiai, Tsukasa Demizu, Shin Ishiguro, Shohei Maruyama, Akihiro Kawana. KDD Cup 2019 Regular Machine Learning Competition Track Task 2, 1st Prize: Simulating the Effects of Eco-Friendly Transportation Selections for Air Pollution Reduction. *ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019.

- Tsukasa Demizu, Shunji Umetani, Hiroshi Morita. Optimal electric power management in a residential building using photovoltaic and storage battery. *ASME International Symposium on Flexible Automation*, 2012.

国内会議

- 出水宰. ドコモ R&D におけるデータサイエンスの活用事例. 日本オペレーションズ・リサーチ学会 2021 年春季研究発表会, 2021.

- 出水宰, 深澤佑介, 森田浩. 深層学習による時間減衰を考慮したインフィード広告の CTR 予測. 研究報告モバイルコンピューティングとパーベイシブシステム (MBL), Vol. 2019, No. 11, pp. 1–7, 2019. (優秀論文賞 受賞)

- 出水宰, Rubén Manzano, Sergio Gómez, 深澤佑介. モバイルネットワーク特徴量を用いた Contextual Bandit Algorithm. 研究報告モバイルコンピューティングとパーベイシブシステム (MBL), Vol. 2018, No. 19, pp. 1–8, 2018. (優秀発表賞 受賞)

- 出水宰, 梅谷俊治, 森田浩. 太陽光発電・蓄電池を用いた住宅規模での電力運用最適化. 日本オペレーションズ・リサーチ学会 2013 年春季研究発表会, 2013.

- 出水宰, 梅谷俊治, 森田浩. 太陽光発電・蓄電池を用いた住宅規模での電力最適運用計画. 京都大学数理解析研究所 RIMS 研究集会, 2012.

- 出水宰, 福嶋悠大, 梅谷俊治, 森田浩, 飯田亨, 小林美佐世. ユーザー入力情報の不確実性を考慮した電気自動車の最適充電スケジューリング. 第 55 回システム制御情報学会研究発表講演会, 2011. (学会賞奨励賞 受賞)

記事掲載

- 出水宰. 強化学習によるアドテク分野での研究報告. ビジネスコミュニケーション 12 月号 第 56 巻 第 12 号 pp. 20–21, 2019
- 出水宰. 深層学習によるアドテク分野での研究報告. ビジネスコミュニケーション 9 月号 第 56 巻 第 9 号 pp. 24–25, 2019

その他

- Keiichi Ochiai, Tsukasa Demizu, Shin Ishiguro, Shohei Maruyama, Akihiro Kawana. Simulating the Effects of Eco-Friendly Transportation Selections for Air Pollution Reduction. *arXiv preprint arXiv:2109.04831*, 2021. https://doi.org/10.48550/arXiv.2109.04831
- 出水宰, 梅谷俊治, 森田浩. 太陽光発電・蓄電池を用いた住宅規模での電力最適運用計画. 数理解析研究所講究録. Vol. 1829, pp. 64–71, 2013.

目 次

図 目 次

表 目 次

第 1 章

序論

1.1 研究背景

事業活動を通じて得られるデータの蓄積と活用が，多くの企業で進んでいる．特に，インターネットやスマートフォン，IoT (Internet of Things) の普及に伴い，2010 年代からはビッグデータというワードで代表されるように，大量かつ多様なデータが高頻度で生成されている．これらのデータを活用し，価値ある示唆の獲得や，将来傾向の予測ができれば，事業におけるオペレーション改善や，より良い経営戦略策定が可能になる．

このようなデータ活用の手段の一つとして，機械学習を用いることが挙げられる．機械学習は，観測したデータを基にして事象の規則性といった知識を，統計的かつ自動的に獲得する学習手法である．これによって，モノの分類や予測，行動決定を行うことが可能である．そのため，機械学習の適用領域は，オンライン・オフラインを含めて多岐にわたっている．例えば，オンライン領域では，Web 広告や e コマース，ソーシャルネットワークサービス (SNS) が挙げられ，オフライン領域では，サプライチェーン・マネージメントや移動交通制御，エネルギー管理など様々である．

オンライン領域の例として挙げた Web 広告では，特に多様な形式のデータを扱うことが可能である．Web 広告に紐付く画像データをはじめ，広告文言に含まれるテキストデータ，配信対象ユーザの属性情報や広告主の情報といったテーブル形式のデータなどがある．こうした Web 広告に紐付くマルチモーダルな情

報は，配信後の広告効果に大きく寄与していると考えられる．特に，Web 広告の配信効果を測る指標として，クリック率 (Click through rate, 以下 CTR) が用いられる．蓄積された各 Web 広告の CTR 実績，及びそれらに紐付くマルチモーダルな情報から，教師あり学習によって CTR を予測するモデルの構築が可能となる．このモデルを新たな Web 広告に適用し，CTR の予測を行うことで，複数の Web 広告の中から効果の高い広告を選定するといった配信計画の意思決定にも期待できる．

オフライン領域のサプライチェーン・マネジメントでは，供給側から需要側への商品・材料の流れや，需要側から供給側への需要情報の流れ，さらに物流に伴うキャッシュフローといった多くの要素が含まれる．これらの要素を適切に管理し，サプライチェーン全体の利益最大化を最小のコストとリードタイムで行うことを目指している．特に，需要に合わせて適切な供給量を決定して在庫量を保つ在庫管理は，サプライチェーンにおける重要な部分である．しかし，実際には需要側の急な変動や，供給側における材料供給の遅延など，あらゆる局面で不確実な環境に置かれている．こうした不確実性を伴う環境下では，将来時点におけるサプライチェーンの状態を推定し，それに応じた供給量を都度，決定すべきである [1]．この供給量の最適決定問題は，マルコフ決定過程でモデリングすることができ，強化学習のフレームワークに落とし込むことが可能である．例えば，小売店での在庫管理を考えた場合であれば，各商品・各店舗における販売量や在庫量といったデータを観測しながら，次回の発注量を決定し，その運用評価を使って発注量決定の方策を更新するというものである．日々の発注量決定が最適化されることで，在庫切れに伴う販売機会損失の回避や，逆に，過剰な発注による在庫コスト高騰の抑止も期待できる．

1.2 研究目的

1.1 節で述べた通り，Web 広告の配信計画や商品の在庫管理といった意思決定に，データを基にした機械学習の活用が考えられる．本研究の目的は，特に新規の Web 広告や商品といったアイテムを市場に投入する際に，それらの時系列的な需要変動を予測した上で意思決定を最適化する手法を確立することである．最終消費者の様々なニーズを汲み取るため，こうした新規アイテムは日々，市場

に投入されており，ここでの意思決定をより効果的かつ迅速に行うことは，事業活動において非常に重要である．本研究で対象とする新規アイテムは (1) オンラインで次々と配信される Web 広告，及び (2) オフラインで販売や在庫の管理がなされる小売商品を想定している．それぞれの具体的なケースを示しながら，課題と研究目的について説明する．

1.2.1 新規 Web 広告の CTR 予測に関する研究

本研究では，Web 広告のうちインフィード広告と呼ばれる種類の広告を対象にする．インフィード広告とは，Web ニュースメディアや SNS におけるフィード型コンテンツの間に，同様の体裁で表示される Web 広告である．その特徴としては，ユーザの目に止まりやすいという視認性の高さや，広告然とした雰囲気も薄いといった点が挙げられる．しかし，インフィード広告は，その高頻度な表示のために，配信開始以後の CTR について時間的な減衰が速いといった傾向がある．新たなインフィード広告をメディア上で配信する際に，どれ程の CTR で効果が見え始め，その後の CTR の減衰具合を見積もることは一般的には容易でない．そこで，本研究の第一の目的は，ユーザの Web 広告に対する需要の現れである CTR を，その時系列的な変動も含めて配信前に予測する手法を提案することである．この時間減衰を考慮した CTR 予測によって，どの広告を配信するべきか，また，CTR の時間減衰に応じてどの程度の掲載期間としておくべきか，といった配信計画に関する意思決定に活用することができる．

CTR 予測は広告収益に直結する重要な手段であるため，従来より多数の研究が行われている．特に，広告クリエイティブ（広告におけるバナー画像やテキスト）に紐付く多様なデータと，深層学習とを用いた CTR 予測手法が提案されている [2, 3, 4, 5, 6, 7]．しかし，広告配信の事前に，その CTR の時間的な減衰を，広告クリエイティブの情報を用いて陽にモデル化するような CTR 予測手法は従来，提案されていない．そこで本研究では，広告クリエイティブ自体の内容による CTR の時間減衰への影響を表現する目的で，配信設定情報だけでなく画像情報やテキスト情報を加えたマルチモーダルな特徴量を利用する．さらに，CTR を単一の期間ではなく，多期間にわたる系列データとして拡張した上での，深層学習モデルを提案する．

1.2.2　新規小売商品の在庫管理に関する研究

本研究では，スマートフォン製品についての販売店における在庫管理を対象にする．スマートフォンは，製品ライフサイクルが短く，多くの製品バリエーションを持つサプライチェーン上の小売商品の一つである．通常，スマートフォンの販売店では，複数のサプライヤーから出される様々な製品を扱い，かつ毎シーズンで新たなラインナップとして刷新されていく．こうしたライフサイクルの速い環境のなかで，新規商品が発売された日からの，日別，製品別，店舗別といった粒度での在庫管理を考える．しかし，発売以後の販売実績が未だない状況の中で，発売当初からの販売需要をこの粒度で見積もり，きめ細やかに在庫管理することは一般的には容易でない．そこで，本研究の第二の目的は，多商品かつ多店舗における新規商品の在庫管理について，発売以後の販売需要の時系列変化を予測した上で，発売当初からの在庫管理を最適化する手法を確立することである．これによって，新規商品における欠品発生率を抑えつつ不良在庫量も削減することで，各販売店での利益増大に貢献することが可能になる．また，このような最適な発注量を自動的に導出することで，販売店側での発注業務を軽減し，本来の販売促進活動へリソースを配分することにも期待できる．

在庫管理はサプライチェーン上の重要な手段であるため，従来から多数の研究が行われている．数理最適化によるアプローチ [8, 9, 10, 11, 12] や，強化学習によるアプローチ [13, 14] があり，その有用性が報告されている．また近年，強化学習のフレームワークに深層学習を組み入れた深層強化学習を，在庫管理へ適用する事例も報告されている [15, 16, 17, 18]．しかし，これらの既存研究は，定常的に販売されている商品を扱う在庫管理が中心であり，新規商品が発売した当初からそれらの手法をそのまま適用することは難しい．そこで本研究では，この課題を解決するために，学習効率性の高いモデルベース深層強化学習を用いて，過去商品の販売実績データを用いたオフライン環境での需要予測モデルの学習と，新規商品発注についてのオンライン環境でのプランニングを組み合わせた手法を提案する．

図 1.1: 取り組みの全体像

1.3 本論文の構成

本論文は 5 章から構成される．本論文の取り組み全体像を図 1.1 に示す．本章以降の内容は次の通りである．

第 2 章では，新規アイテムに対する需要予測を定義し，その役割を明確にした上で，本研究の方針を示す．

第 3 章では，深層学習による新規 Web 広告の時間減衰を考慮した CTR 予測に関する研究について述べる．まず，Web 広告の CTR 予測に関する従来手法を整理した上で，本研究の位置付けを明確にする．そして，今回対象とするインフィード広告に紐付く，マルチモーダルなデータの具体内容を示し，これらを用いた CTR の時間変化を予測する提案手法について説明する．さらに，提案手法の有効性を確認するために行った，広告配信実績を用いた評価実験の結果を示す．

第 4 章では，モデルベース深層強化学習による新規小売商品の在庫管理に関する研究について述べる．まず，数理最適化や強化学習による在庫管理の従来手法を整理した上で，本研究の位置付けを明確にする．そして，製品ライフサイクルが短いサプライチェーンの一つであるスマートフォンを対象に，新規商品の在庫管理問題の定義と，観測できるデータの具体内容を示す．さらに，提案手法で

ある，モデルベース深層強化学習を用いたアルゴリズムと，商品の需要予測モデルの詳細について説明した上で，商品販売実績を用いた評価実験の結果を示す．

最後に，第 5 章では，本論文のまとめを行い，今後の課題について述べる．

第 2 章

新規アイテムに対する需要予測の定義と課題

本研究では，過去に市場投入されたアイテムに対して，まだ市場で利用可能になっていないアイテムを新規アイテムとして定義している．前章で述べた研究背景，及び研究目的を鑑み，本研究での主題は以下とする．

> 新規アイテムを市場に投入する際に，その需要に対する時系列的な変動を予測した上で，意思決定を最適化する手法を確立すること

本章では，新規アイテムに対する需要予測を定義し，その役割を明確にした上で，本研究の方針を示す．はじめに，需要予測についての定義を行い，需要予測を活用した意思決定について述べる．次に，具体的な対象領域である，オンライン領域における Web 広告と，オフライン領域であるサプライチェーン上の小売商品での従来研究とその課題について述べ，本研究での方針を記す．

2.1 需要予測から意思決定までの流れ

需要予測とは，対象とするアイテムの将来における需要の値やパターンを予測することである．例えば，対象アイテムにおける需要量やその変動，アイテムの属性に応じた需要の傾向差を，前もって予測することが挙げられる．

図 2.1 に，データを基にした需要予測と意思決定までの流れの概要を示す．デー

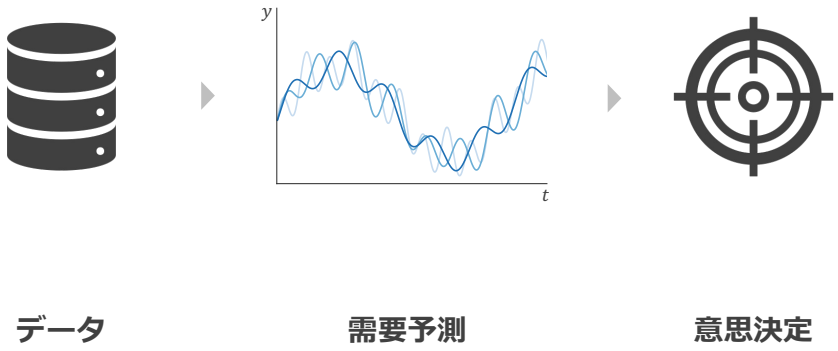


図 2.1: データを基にした需要予測と意思決定までの流れの概要

タとしては，アイテムの属性に関する情報や過去の需要量の実績，天候といった外生情報など様々なものが想定される．これらのデータを基に，対象アイテムの将来にわたる需要予測をする際の一つの手段として，機械学習が挙げられる．需要量の実績を含む蓄積されたデータを教師データとして，教師あり学習により予測モデルを構築することで，将来需要の予測値を算出することが可能になる．この需要予測は，対象アイテムの属する事業分野に応じて，様々な意思決定に活用することができる．例えば，製造分野では，商品需要としての販売量を将来にわたって見積もることで，対象商品の生産計画や部材の調達計画といった意思決定への活用が可能である．また，小売分野では，需要予測を基にした商品発注による在庫管理への活用も考えられる．他にも，Web 広告分野では，Web 広告の需要の現れである CTR を予測することで，より効果の高い広告クリエイティブを選定するといった配信計画への応用も考えられる．このように，対象アイテムに関する意思決定の内容に応じて，需要予測のモデル構築や利用するデータの蓄積が行われる．

ここで，対象とするアイテムが既に市場で利用可能な状態にあるかどうかは，需要予測やその後の意思決定において重要な点になる．既に市場で利用されている状態であれば，対象アイテムの需要を観測することができ，その需要の実績を直接，モデルで学習することが可能である．一方で，まだ市場で利用されてい

ないような新規のアイテムを対象とする場合は，その需要に関するデータは蓄積されていないため，需要予測をする際には大きく不確実性が伴うと考えられる．このように新規アイテムの需要予測は，既存アイテムのものに比べて一般的には難しいとされるが，一方で，不確実性が高いからこそ意思決定へ役立てたいというニーズは多くある．最終消費者の様々な要望を汲み取るため，こうした新規アイテムは日々，市場に投入されており，ここでの意思決定をより効果的かつ迅速に行うことは，事業活動において重要な問題となる．

2.2 諸分野での需要予測と活用に関する従来研究

本節では，具体的な事業分野を挙げながら，各分野におけるアイテムの需要予測とその活用に関連する従来研究を紹介し，その課題について述べる．ここでは，アイテムに対する需要を広義に捉えて，オンライン領域における Web 広告の需要としての CTR と，オフライン領域における小売商品の需要としての販売量に着目する．

2.2.1 オンライン領域の Web 広告需要に関する研究

Web 広告の需要の現れである CTR を予測することは，より効果的な広告買い付けや広告配信に直結する重要な手段であるため，従来より多数の研究が行われている．特に，新規の Web 広告を含め，その CTR 予測の精度向上を目的として，Web 広告のクリエイティブ自体に紐付く多種のデータから，深層学習により予測するアプローチが，近年多く用いられている．

Chen ら [2] は，広告の配信設定情報（ユーザの性別や掲載商品のカテゴリ，Web ページにおける広告の掲載位置）だけでなく，広告クリエイティブの画像情報を利用している．こうした Web 広告のマルチモーダルな特徴量を用いて，各ユーザに Web 広告を配信した際のインプレッション時のクリック有無を予測する，Deep CTR Prediction を提案している．この Chen らの研究を起点に，利用する特徴量の追加や予測モデルの改良，Web 広告種類の拡張といった発展がなされている．Aryafar ら [3] は，広告の配信設定情報と画像情報に加えて，広告のタグやタイトルといったテキスト情報も特徴量に取り入れている．さらに，単一の予測モデルではなく複数の予測モデルを組み合わせたアンサンブルモデ

ルにより，予測精度の飛躍的な向上を達成している．岩崎 [6] は，Web 広告の新たな種類であるインフィード広告に CTR 予測を適用している．インフィード広告は，Web ニュースメディア等のフィード型コンテンツの間に，同様の体裁で表示される広告であり，画像だけでなくテキストも重要な要素となっている．岩崎の手法では，Chen らの Deep CTR Prediction に大きく二つの変更を加えている．一つ目は，広告に紐付くテキスト情報も特徴量に取り入れている点である．二つ目は，各ユーザのインプレッション単位での CTR 予測ではなく，複数ユーザへ配信した際の広告単位での CTR 予測としている点である．広告単位という巨視的な観点で CTR を見積もることで，より高い効果の広告を選択するといった配信計画を立てやすくする利点が挙げられる．こうした変更を加えた CTR 予測手法での，学習性能と配信影響についてを報告している．Park ら [7] も，広告単位で CTR を回帰するアプローチをとっている．Park らの手法では，言語的な情報として，タイトルや説明内容だけでなく，広告クリエイティブの画像に埋め込まれたテキスト表現も，特徴量として抽出している点が特徴である．この画像情報とテキスト情報との空間的な関係性を表現するための注意 (Attention) 機構を取り入れた予測モデルを提案している．

これらの従来研究では，インプレッション単位の微視的，もしくは広告単位の巨視的な観点で CTR 予測の評価がなされている．しかし，実際の CTR は時系列的な変動を伴うものであり，新規 Web 広告の配信後，その CTR は一般的に減衰する傾向にある．従来研究では，こうした CTR の時間的な減衰が考慮されていないという課題があり，広告クリエイティブの情報を用いて CTR の減衰傾向を陽にモデル化する予測手法は提案されていない．

2.2.2　オフライン領域の小売商品需要に関する研究

小売商品の販売需要は，店舗における在庫管理と密接な関わりがある．商品供給量を決定する在庫管理のなかで将来時点の需要量を見積もることは，適正在庫を維持する上で重要な部分となる．この供給量の最適決定問題は，マルコフ決定過程でモデリングすることができ，数理最適化や強化学習を用いたアプローチが従来より，多数研究されている．

Dillon ら [11] は，需要の不確実性を考慮するために確率変数として与えた上

で，在庫管理を確率計画法で定式化している．Van ら [12] は，動的計画法を用いた在庫管理手法を提案し，一定の効果を報告している．しかし，実際の大規模なサプライチェーン上で供給量を決定する場合，これらの厳密解法は計算量の観点で現実的ではない．これに対し，強化学習によって厳密解法の計算量を克服するアプローチが取られている．Kara ら [14] は，生鮮品の在庫システム (Perishable inventory system) に強化学習を適用している．生鮮品の在庫システムでは，食品や化学薬品についての有限の寿命を考慮し，寿命を過ぎたこれらの商品は廃棄されるという特徴がある [14, 19]．Kara ら は，この在庫システムに強化学習における Q 学習 [20] や Sarsa [21] を適用し，その有効性を示している．また，Meisheri ら [16] は，商品数が 100 から 200 程度の比較的大規模な在庫管理に対して，深層強化学習における Advantage actor critic (A2C) [22] や Deep Q-network (DQN) [23] を適用し，手法の有用性を報告している．これらの従来研究では，規定された分布に従う確率変数として将来の需要量を表現しているが，需要量を予測するモデルを構築した上で，在庫管理に適用するアプローチも提案されている．Malik ら [18] は，スーパーマーケットチェーンにおける生鮮品の需要予測モデルを学習し，モデルベース深層強化学習における報酬の推定に用いている．さらに需要についての不確実性に対応するため，需要予測モデルのキャリブレーションを行うことで，従来手法と比較して大幅な性能改善を達成している．

こうした従来研究は，定常的に販売されている既存商品を扱ったものが中心である．在庫管理の対象である商品の販売履歴が既に得られている状態で，需要量の確率分布の規定や，需要予測モデルの学習が行われている．そのため，新規商品の在庫管理を対象にした場合は，これらの手法を発売当初から即座に適用することは難しいという課題がある．Wanke ら [24] は，統計的手法による新規商品に対する在庫管理として，商品需要を三角分布 (Triangular distribution) で与え，古典的な在庫管理方策の (Q, r) モデル [8, 10] で供給量を決定するアプローチを提案している．この手法は，簡潔で実際の在庫管理システムへの適用が容易という利点がある．一方，需要予測の観点では，商品特性や地域特性といった外生変数を考慮していない単純な統計モデルであるため，外生変数を考慮した機械学習モデルに比べて，予測精度が劣る可能性がある．加えて，在庫管理の観点では，(Q, r) モデルのようなヒューリスティック手法では，需要変動が激しい場合に，在庫欠品や過剰発注につながる恐れがあるという課題も挙げられる．

2.3 本研究の方針

前節で述べた諸分野での需要予測と活用に関する従来研究の課題を鑑み，本研究ではそれらの課題を解決する手法を確立する．そのための方針として

- 新規アイテムが市場に投入される前に，その需要の時系列変動をアイテムに基づくデータから予測する
- 得られた時系列の需要予測値から，新規アイテムに関わる意思決定を最適化する

という流れで進める．研究対象とする事業分野は，特に新規アイテムの市場投入の機会が多く提案手法による効果が高いと考えられる，オンライン領域でのWeb 広告分野と，オフライン領域での小売分野とする．

オンライン領域の Web 広告分野では，広告配信以後に CTR の時間減衰が顕著に速いインフィード広告に着目し，その時系列変動を配信の事前に精度高く予測することを主眼にする．尚，対象とする予測期間は配信開始から 1 ヶ月程度先までで，週次での CTR の変動を含めて予測することを想定している．インフィード広告に紐付くマルチモーダルな特徴量から，CTR の時間減衰を予測する深層学習によるモデルを構築する．得られた時系列としての CTR 予測を基に，どの広告を配信するべきかや，CTR の時間減衰に応じてどの程度の掲載期間とすべきかといった配信計画に関わる意思決定への活用を目指す．

オフライン領域の小売分野では，新規商品の発売以後の販売需要の時系列変化を予測した上で，発売当初からの在庫管理に適用することを主眼にする．尚，対象とする予測期間は 1 週間程度先までの日別の販売需要であり，この需要予測を発売開始から日次で繰り返し行うことを想定している．新規商品の販売実績が未だない状況で，過去商品の実績から学習した需要予測モデルを先ず構築し，深層強化学習を用いた発注量決定の方策に組み入れることを考える．特に製品ライフサイクルが短いサプライチェーンの商品の一つであるスマートフォンに着目し，新規商品における欠品発生率を抑えつつ不良在庫量も削減することで利益を増大させる在庫管理手法の確立を目指す．

第 3 章

深層学習による新規Web広告の時間減衰を考慮したCTR予測

3.1 はじめに

インフィード広告は，Web ニュースメディアや SNS におけるフィード型コンテンツの間に，同様の体裁で表示される広告である．そのため，ユーザの目に止まりやすいという視認性の高さや，広告然とした雰囲気も薄いという特徴がある．こうした特徴から，広告主，及びユーザからの高い支持を受けて，インフィード広告の市場は年々増加している．2017 年時点の市場規模は 1,903 億円であったが，2023 年では 3,921 億円になると推計されている [25].

インフィード広告に限らず，Web 広告全般において，広告クリエイティブの効果を測る指標として，インプレッション回数（広告が表示された回数）に対するクリック回数の比率である CTR が用いられる．この CTR を予測することは，ユーザ，及び広告主の双方にとって重要である．例えば，クリック課金型（ユーザが広告をクリックした場合にのみ広告主が予め入札していた金額を支払う形式）の設定では，広告を配信した場合の収益の期待値は，広告主が設定した入札額と CTR によって決まる [26]．ある広告枠に対し，複数の広告候補の中から配信すべき広告を買い付けて，収益の最大化を考えた際に，入札額は既知であるが，CTR は未知であるため予測を行う必要がある．この予測値が実際と乖離してしまうと，効果の低い広告を配信する可能性が高まり，収益の最大化は難し

くなる．このように広告配信において，CTR 予測は収益に直結する重要な意思決定の手段である [26]．

こうした背景から，広告配信における CTR 予測の研究は盛んに行われている．特に，広告クリエイティブに紐付く画像情報やテキスト情報，そして配信設定情報といった，マルチモーダルな特徴量を用いた CTR 予測が提案されている [2, 3]．これら多くの従来研究では，インプレッション単位でユーザにクリックされるかどうかを予測する問題設定として定義され，その下で予測手法が提案されている．一方で，インプレッション単位ではなく，広告単位で CTR がどの程度になるかを事前に予測する問題設定は，ビジネスの観点からも求められている．しかし，CTR は比率としての値であり，インプレッション回数が少ない場合にはばらつきが生じる．例えば，真の CTR が 5%の広告でクリック回数が二項分布に従うと仮定した下で，信頼度 85%で推定値が真値の 1%pt 以内に収まるようにするためには，約 1,000 回のインプレッション回数が必要である [27]．このように，CTR は分散が大きくなりやすい傾向があるため，過学習を避けたよりロバストな予測手法が必要となる．

本章の研究で対象とするインフィード広告の特徴として，その高頻度な表示のために，配信開始以後の CTR の時間的な減衰が速いという点が挙げられる．その減衰速度は，広告クリエイティブ自体の内容や，対象ユーザー属性，掲載面といった配信設定に依存すると考えられる．この CTR の時間的な減衰まで含めて予測できれば，事前に適切な掲載期間を計画するといった意思決定も可能となる．例えば，ある広告について掲載初期の CTR は高いが，急峻な CTR 減衰が予測される場合は，掲載期間を短く設定するといった計画を事前に立てることができる．また，定められた予算と期間の中で，累積クリック数が最大となるような広告の買い付け，及び掲載期間を決定する最適化問題のパラメータとして活用することもできる．そこで本章の研究では，図 3.1 に示すように，広告ごとの時間減衰を考慮した上での CTR 予測手法を提案する．まず，広告クリエイティブの内容による CTR の時間減衰への影響を表現する目的で，配信設定情報だけでなく画像情報やテキスト情報といったマルチモーダルな特徴量を入れ込むことを考える．次に，CTR の時系列的な変化を表現する目的で，Recurrent Neural Network (RNN) による予測モデルを構築する．提案手法の効果を測るため，アドネットワーク上の配信履歴を用いたオフライン検証で CTR 予測を実施する．

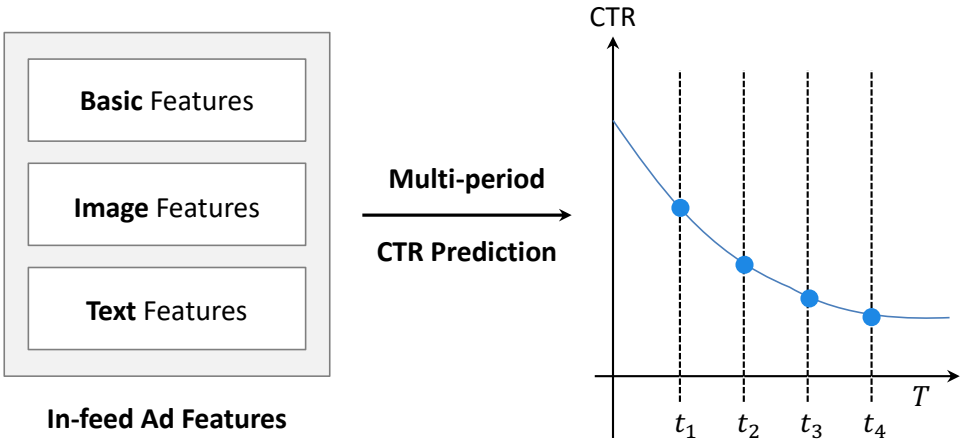


図 3.1: インフィード広告における多期間 CTR 予測の概要図

本章の研究の貢献は以下の通りである．

- マルチモーダルな特徴量から広告単位の CTR をロバストに予測するため，従来研究 [2, 6] に Dropout 及び L2 正則化を導入するネットワーク構造の改良を行い，実データによる精度検証で，マルチモーダルな特徴量による CTR 予測の精度向上を確認した．
- CTR の時間減衰を推定するために，多期間にわたる CTR の時系列変化を抽象的に表現可能な RNN 型のネットワーク構造を提案した．
- 実データによる精度検証で，ベースライン（多期間の CTR を独立したものとみなし同時に予測するモデル）との比較を行い，提案手法による精度向上を確認した．

以下，3.2 節では，Web 広告の CTR 予測に関する従来手法について説明し，本章の研究の位置付けを明確にする．3.3 節では，対象とするインフィード広告に紐付くマルチモーダルな情報を示し，CTR の時間減衰を考慮した多期間 CTR 予測問題の設定を述べる．3.4 節では，提案手法として従来手法の改良点と，多期間の CTR 予測へ拡張した RNN モデルを説明する．3.5 節では，それらの有効性を確認するために行った，アドネットワーク上の配信履歴を用いた評価実験について述べる．最後に，3.6 節では，本章の研究のまとめと課題を述べる．

3.2 関連研究

本節では，広告配信における CTR 予測のアプローチを分類し，それぞれのアプローチに関連する研究について述べ，本研究の位置付けについてを明確にする．

3.2.1 インプレッション単位での CTR 予測

第一のアプローチは，あるユーザに対する広告の 1 インプレッションあたりの結果 (click or non-click) から，二値の分類問題として学習するものである．広告表示の最小単位に着目し，対象とする広告とユーザの双方の情報を用いて CTR を予測する．例えば，インプレッション時の d 次元の特徴量を $x \in \mathbb{R}^d$ としたとき，クリックされるかどうかの確度を表す $\hat{y}$ はロジスティック回帰を用いて次のように求められる．

$$\hat{y} = \frac{1}{1 + \exp(-z)} \tag{3.1}$$

$$z = f(x) \tag{3.2}$$

ここで，$f(.)$ は特徴量 x についての重み付き線形和をとる関数とする．N 回のインプレッションについて，それぞれの特徴量 $x_i \in \mathbb{R}^d$ とそのクリック結果 $y_i \in \{0, 1\}$, (0: not-click, 1: click) が訓練データとして与えられた場合，次に示す誤差関数 L の最小化によって $f(.)$ を学習する．

$$L = \frac{1}{N} \sum_{i=1}^{N} \big(-y_i \log \hat{y}_i - (1 - y_i) \log (1 - \hat{y}_i) \big) \tag{3.3}$$

Chen ら [2] は，インプレッションにおける特徴量 x として，広告の配信設定（ユーザの性別や掲載商品のカテゴリ，Web ページにおける広告の掲載位置）といった基本的な特徴量 (basic features) に加えて，広告クリエイティブの画像情報 (image features) を用い，End-to-End で CTR を学習する Deep CTR Prediction を提案した．

Aryafar ら [3] も，マルチモーダルな特徴量表現として，画像に加えて，広告に紐付くタグやタイトルといったテキスト情報を取り入れた CTR 予測を行った．画像情報については，ImageNet [28] で事前学習された ResNet101 [29] を使い，

テキスト情報については，Hashing Trick [30] を使って，より低次元の空間に情報を圧縮した分散表現として各々の特徴量を用いている．また，マルチモーダルな特徴量を単に結合するだけでなく，アンサンブルモデルとして学習することで，飛躍的な精度向上を達成した．

Zhang ら [4] は，ユーザの過去の広告クリック挙動との関係に注目し，RNN [31] による CTR 予測を提案した．この研究では，ユーザが前回の広告クリックで訪問したページ上での滞在時間や，前回クリックからの経過時間といった情報と，CTR との関係を解析していた．その上で，ユーザの過去の挙動を系列データとして入力に加えた RNN モデルを構築し，同一データでの Logistic Regression モデルや Neural Network モデルからの精度向上を示した．

インプレッション単位での CTR を予測する際に，CTR が極端に低い値の場合は，訓練データ $y_i \in \{0, 1\}$ について不均衡な分類問題となってしまう．Deng ら [5] は，こうした状況を是正するために Generative Adversarial Network (GAN) [32] から着想を得た Disguise Adversarial Network (DAN) を提案した．分類問題における不均衡データの是正方法として，SMOTE [33] により，既にある少数の正例データ (1: click) から線形な組合せで人工的にデータを生成する方法がある．これに対して DAN では，大量の負例データ (0: not-click) を非線形な変換を経て，一部を「偽造された」正例データとして増幅させる．ディスプレイ広告とモバイル広告の双方の CTR 予測タスクにおいて，DAN を利用することで SMOTE と比較しての精度向上が確認された．

3.2.2 広告単位での CTR 予測

第二のアプローチは，複数ユーザへのインプレッションを経た後の，広告単位での結果を用いて，CTR を直接予測するものである．ここでの訓練データとして，L 個の広告の，それぞれの特徴量 $X_i \in \mathbb{R}^d$ とその CTR 結果である連続量 $Y_i \in [0, 1]$ を用いた学習を行う．

Shi ら [34] は，検索連動型広告を対象として，CTR の指標だけでなく，1 クリックあたりの費用であるクリック単価 (Cost per click, 以下 CPC) も含めた予測を行った．用いた予測モデルとしては，Linear Regression, Random Forest, そして Gradient Boosting を挙げており，CTR 予測，及び CPC 予測の精度と，各

モデルでの重要特徴量の解析をしていた．本橋ら [35] は，バナー広告の CTR に対するトレンド効果や曜日効果，祝日効果といった影響に着目し，これらを考慮した状態空間モデルを構築した．また，実際の配信データを用いた実証分析では，提案モデルの有用性を示し，CTR の長期的な傾向の変化や曜日によって異なる振舞いを捉えることができていた．

Chen ら [2] の Deep CTR Prediction に対し，岩崎 [6] は広告クリエイティブのテキスト情報を付与した予測モデルを提案し，インフィード広告である Facebook Ads[1] に対しての実験を行った．実験では，CTR を求める回帰問題として定義し，広告の基本的な特徴量に加え，高次元な画像やテキスト特徴量を付与した場合の学習性能と配信影響について分析している．また，坂田ら [36] は，広告における CTR の数値が閾値よりも高いかどうかの結果（effective or ineffective）を分類する問題と定義し，画像特徴量の寄与について示していた．

Park ら [7] は，言語的な情報として，タイトルや説明内容だけでなく，広告クリエイティブの画像に埋め込まれたテキスト表現も用いて CTR の予測を行った．画像に埋め込まれたテキストの検出は OCR によってなされており，抽出後に BERT [37] によってベクトル化している．また，画像情報とテキスト情報との空間的な関係性を反映するための注意 (Attention) 機構を取り入れた Multistep Modality Fusion Network (M2FN) を提案した．

3.2.3 本研究の位置付け

表 3.1 に，提案手法を含めた，Web 広告の CTR 予測に関する研究の比較を示す．広告配信の事前に，その CTR の時間的な減衰を，広告クリエイティブの情報を用いて陽にモデル化するような CTR 予測手法は，従来，提案されていない．広告配信ではなく，Web 上のニュース記事配信における CTR の時間的な減衰を，時系列モデルとして扱った既存研究は報告されている．Agarwal ら [38] は，自社のポータルサイト上に掲載されるニュース記事である Yahoo! Front Page Today Module[2] において，それらの掲載位置という空間的情報と掲載時刻からの時間的情報を基にした CTR 予測手法を提案した．この研究では，ニュース記事における CTR の減衰は，Wu ら [39] が提唱したように，ユーザへの繰り返し

[1] https://www.facebook.com/business/ads

[2] https://webscope.sandbox.yahoo.com/catalog.php?datatype=r&did=49

の表示による一種の疲弊から発生するものと仮定されている．また，ニュース記事における CTR 予測の時間スケールは 1 時間単位で扱われており，繰り返しの表示に伴う CTR の減衰だけでなく，1 日における時間帯ごとの傾向も考慮されている．これらの時間的情報を組み入れた Dynamic Gamma-Poisson モデルを提案し，シミュレーションを通じて良好な予測性能を示した．しかし，Agarwal らのモデルで，ニュース記事についてのマルチモーダルな情報（見出しにある画像やテキスト）は扱われていないため，記事自体の内容による影響は考慮されていない．また，CTR の予測を行うタイミングとしても，ニュース記事を配信した後に動的に予測することを想定しており，配信の事前に予測することは想定されていない．

そこで，本研究では，新規広告を入稿する際に時系列変動を含めて CTR を予測し，その広告を配信するべきか，また，CTR の時間減衰に応じてどの程度の掲載期間にすべきかという意思決定に役立てることを目標とする．そのため，インプレッション単位の CTR 予測ではなく，3.2.2 項で述べた広告単位での CTR を回帰するアプローチを取ることにする．また，対象の広告は，近年その重要度が増しているインフィード広告としている．インフィード広告は，CTR の時間減衰が速いという性質を持つが，広告クリエイティブ自体の内容による CTR の時間減衰への影響を表現する目的で，配信設定情報だけでなく画像情報やテキスト情報を加えたマルチモーダルな特徴量を利用する．さらに，CTR を一点ではなく，多期間にわたる減衰傾向までを考慮した系列データとして扱えるように拡張する．

表 3.1: Web 広告の CTR 予測に関する研究の比較

研究	予測対象	利用特徴量			時系列考慮	主な手法
		配信設定情報	画像情報	テキスト情報		
Chen ら (2016)[2]	インプレッション単位	✓	✓			Deep CTR Prediction
Aryafar ら (2017)[3]	インプレッション単位	✓	✓	✓		Ensemble model
岩崎 (2018)[6]	広告単位	✓	✓	✓		CNN model
Park ら (2019)[7]	広告単位	✓	✓	✓		Attention model
Agarwal ら (2009)[38]	広告単位	✓			✓	Dynamic Gamma-Poisson
提案手法 (2021)[40]	広告単位	✓	✓	✓	✓	CNN-RNN model

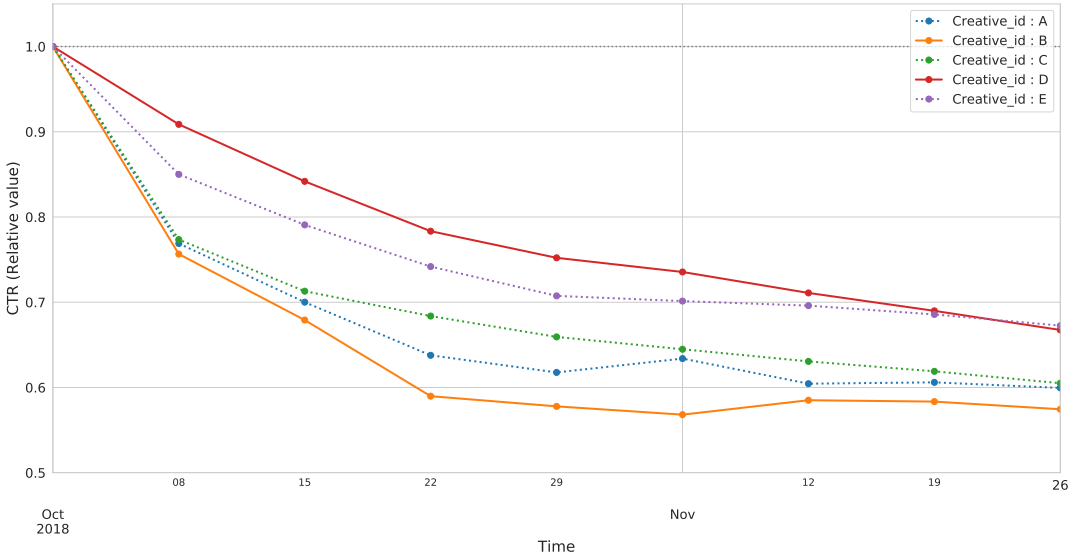


図 3.2: インフィード広告における CTR の時間減衰の実例

3.3 問題設定

3.3.1 CTR の時間減衰

インフィード広告は，ユーザの視認性が高いため，同じ広告クリエイティブを使い続けたときの CTR の減衰速度が，他の Web 広告に比べ速いという特徴がある [41]．図 3.2 に，同期間で配信された，ある 5 つのインフィード広告についての，広告毎の CTR の時間的な減衰の様子を示す．5 つの広告のそれぞれについて，配信を開始した初週の CTR を基準値とし，週毎の相対的な CTR をプロットしている．どの広告も翌週には CTR が大きく低下し，初週から 25%程度も低下する広告も存在する．このように，インフィード広告は CTR の時間減衰が速いという特徴を持つが，その速さは広告クリエイティブにより異なっている．例えば図 3.2 中の広告クリエイティブ B は急激に CTR が減少しているが，広告クリエイティブ D の減少は比較的緩やかである．このように，広告クリエイティブによって CTR の時間減衰には速さの違いが存在する．これは，クリエイティブ自体の内容や，広告対象となるユーザ属性や掲載面といった配信設定に依存するものと考えられる．

こうしたインフィード広告の CTR において，時間減衰まで含めて予測できれば，広告買い付けだけでなく，配信の事前に配信スケジュールの計画が可能とな

り，更なる収益向上が期待できる．例えば，それぞれの広告について多期間にわたる CTR を予測し，ある程度の減衰が起きるところで，別の広告配信に切り替える，というオペレーションが事前に計画可能となり，長期的な意思決定に役立てることができる．また，予測された多期間の CTR をパラメータとして，定められた予算と期間の中で，累積クリック数が最大となるような広告の買い付け，及び掲載期間を決定する最適化問題としての定式化に活用することも可能となる．

3.3.2 利用する特徴量

本研究では，インフィード広告の時間減衰まで含めた CTR 予測を目的とする．CTR 予測モデルの入力として用いられる特徴量は，3.2 節で述べたように様々なものが提案されている．本研究で用いるインフィード広告の特徴量を表 3.2 に示す．CTR 予測モデルの入力として，画像情報やテキスト情報といった高次元な特徴量を用い，クリエイティブ自体が与える影響を捉えることを目指す．また，広告クリエイティブにおける関係図を図 3.3 に示す．広告クリエイティブごとに，その上位階層として，広告グループやキャンペーンが存在する．また，広告クリエイティブごとに画像とテキストの組合せが存在している．広告の配信設定に関する情報（Basic）の一部を表 3.3 に示す．ここで示すように，広告自体の情報や，配信されるユーザ属性などの幅広い設定情報を用いる．

3.3.3 多期間 CTR 予測問題

広告クリエイティブ l の特徴量 X_l は，表 3.2 で示したものより，$X_l = [B_l, I_l, T_l]$ と表す．ここで B_l, I_l, T_l は，それぞれ Basic, Image, Text の特徴量であり，ベクトルとして表現される．次に，予測対象とする多期間の CTR を定義する．広告クリエイティブ l について，その配信日を基準とした週 t の CTR を以下で表す．

$$\mathrm{CTR}_{l,t} = \frac{k_{l,t}}{n_{l,t}} \tag{3.4}$$

ここで，$k_{l,t}$ は，広告クリエイティブ l についての配信日から週 t 初日までの累計クリック回数，$n_{l,t}$ は，配信日から週 t 初日までの累計インプレッション回数とする．本研究では $X_l = [B_l, I_l, T_l]$ から多期間 $t = 1, \cdots, T$ における $\mathrm{CTR}_{l,t}$ を

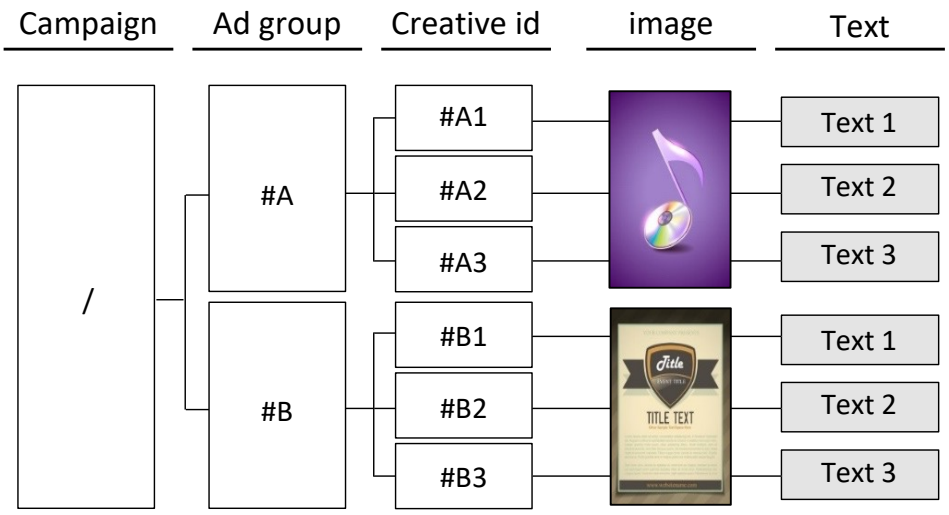


図 3.3: インフィード広告における階層構造

表 3.2: インフィード広告における特徴量の種類

Features	Details
Basic	広告の配信設定に関する情報
Image	クリエイティブの画像情報（1200×628 pixel）
Text	クリエイティブのテキスト情報

予測する問題を考える．

3.4 提案手法

本節では，3.3.2 項で述べた Basic, Image や Text といった広告クリエイティブに紐付く特徴量から CTR を予測する提案モデルを説明する．3.4.1 項で従来研究 [2, 6] のネットワーク構造に改良を加えた予測モデルを述べる．3.4.2 項でそれを多期間の CTR 予測へ拡張した RNN モデルについて述べる．

3.4.1 CTR 予測モデル

広告クリエイティブの Basic, Image そして Text を特徴量として，クリエイティブ自体が CTR に与える影響を捉えることを考える．従来研究 [2] は，3.2.1

表 3.3: Basic 特徴量の例

Features	Details
広告主	広告主の ID
広告種別	ターゲティング型，リターゲティング型といった広告の種類
掲載面	ページ上部からの掲載位置
対象ユーザ属性	対象とするユーザの性別，年齢，居住地に関する設定情報
配信可能時間帯	曜日及び時間帯の配信可否情報
CPC 上限値	1 クリックあたりに支払うコスト (Cost Per Click; CPC) の上限額
日予算	1 日あたりの設定予算額

項で述べたように，インプレッション単位の CTR 予測タスクにおいて，マルチモーダルな特徴量を End-to-End に学習する手法を提案した．しかし，広告単位の CTR 予測タスクにいては，インプレッション回数が少ないような場合，1 クリックが与える影響に左右されて CTR がばらつくため，よりロバストな予測が求められる．また，扱うデータ量にも差が生じやすく，従来研究 [2] が扱うインプレッション単位の予測タスクは，広告がユーザに表示される度にデータが蓄積されるが，広告単位の予測タスクは，広告ごとに 1 レコードの情報となるため，データ量は少なくなる傾向にある．そこで，広告単位の CTR 予測タスクにおいて，訓練データでの過学習を避けて，より安定的な学習を目指すため，従来研究 [2, 6] のネットワーク構造へ新たに Dropout [42] 及び L2 正則化を追加した．Dropout は，多層ネットワークのユニットを確率 p でランダムに選出して無効化し，残りのユニットで重みを更新する方法である．ユニットの選出は重みの更新のたびにランダムに行い，通常，$p = 0.5$ 程度で設定される [43]．Dropout では，学習時にネットワークの自由度を小さくすることで，訓練データへの過学習を防いで汎化性能を高める効果が期待できる [43]．L2 正則化は，誤差関数にネットワークの重みの二乗ノルムを加える正則化手法である．これにより，学習時に重みの値をより小さくすることで，モデルの複雑性を抑制して同じく汎

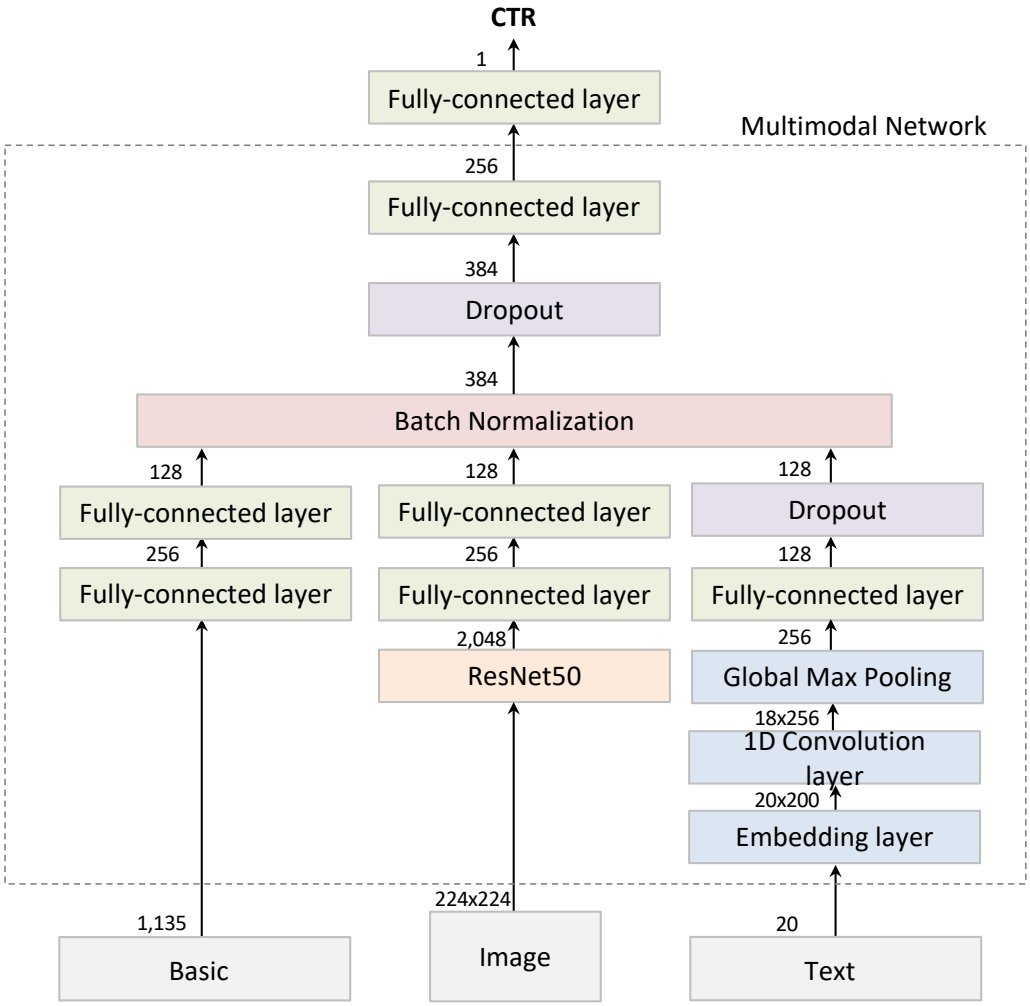


図 3.4: CTR 予測モデル

化性能を高める効果が期待できる．

改良したモデルを図 3.4 に示す．Basic, Image そして Text からのスケールを吸収するために Batch Normalization をかけた結合の後に，Dropout を加えて過学習の回避を行っている．さらに，その後の全結合層に L2 正則化を取り入れた．最終の全結合層で sigmoid 関数による活性化を経た後，数値として CTR を出力する．次にそれぞれの特徴量の加工，及びネットワーク構造について述べる．

Basic 特徴量

広告の配信設定に関する情報の Basic は，表 3.3 に示したように大部分がカテゴリカルデータである．こうしたカテゴリカルな変数は，one-hot encoding によ

りバイナリー変数を用いて表現している．すなわち，$[0, 1, 0, \ldots, 0, 0]$ のように，一つの成分のみ 1 で残りの成分は全て 0 であるようなベクトルとして表す．また CPC 上限値や日予算額といった計量値は，標準化（Standardization）によって平均 0，分散 1 の特徴量に変換をしている．このようにして得られた約 1,000 次元の Basic 特徴量から，2 層の全結合層を経て 128 次元の特徴量を取り出す．

Image 特徴量

Image 特徴量は，3.3.2 項で述べたように，1200×628 pixel の画像を基としている．まず，この画像を 224×224 pixel にリサイズする．次に，ImageNet で事前学習した重みを用いた ResNet50 モデル [29] から 128 次元の特徴量を取り出す．ここで，ResNet50 モデルの最後の畳み込み層の出力では，Global max pooling を適用している．

Text 特徴量

Text 特徴量は，まず，形態素解析で最大 20 シーケンスの単語リストへと分割する．ここで，20 シーケンスに満たない場合はゼロパディングによって先頭部分を穴埋めする．次に，日本語版 Wikipedia の本文全文から学習したエンティティベクトル [44] を用いて，分散表現 [45] を得る．その後，1 次元方向の畳み込み層と Global max pooling 層，全結合層を経た後に，Dropout をかけて 128 次元の特徴量を取り出す．

3.4.2　多期間 CTR 予測への拡張

本節では，インフィード広告で顕著に現れる CTR の時間減衰を推定する目的で，多期間の CTR 予測への拡張を考える．まず自然な拡張として図 3.5(a) のように，3.4.1 項で示した CTR 予測モデルの最終層の出力の次元を，1 次元から多次元にした Deep Neural Network (DNN) モデルが考えられる．このモデルは，多期間の CTR を独立したものとみなし，それらを同時に予測していることになる．しかし，広告クリエイティブ l の週 t の CTR である $\mathrm{CTR}_{l,t}$ は，$\mathrm{CTR}_{l,t-1}$ の影響を受けるため，このように独立して扱うことは現実に即していないと考えられる．

表 3.4: 計算機環境

OS	Ubuntu 16.04.4 LTS
CPU	Intel(R) Xeon(R) CPU E5-1660 v4 3.20GHz
GPU	NVIDIA TITAN X 12GB 1.53GHz

表 3.5: 配信データの概要

Period	# Total impressions	# Ads
2018/10/01 - 2019/02/20	1,814,634,077	28,470

これに対して，多期間にわたる CTR の時系列変化を，マルチモーダルな特徴量から抽象的に表現することを目指した提案モデルを図 3.5(b) に示す．これは，3.4.1 項の CTR 予測モデルの最終層において，Long short-term memory (LSTM) を使って繰り返すことにより，多次元の出力を系列データとして取り出す RNN モデルである．この提案モデルは，入力としての Basic, Image そして Text がそれぞれ一時点での情報であり，出力が複数時点での CTR 情報であることから，Image Captioning [46] のような one-to-many 型と捉えることができる．

3.5 評価実験

本節では，提案手法である多期間 CTR 予測モデルを評価するために，実データを用いた実験について説明する．評価実験の目的は，(1) CTR 予測モデルの精度評価，(2) 多期間 CTR 予測における提案モデルの精度評価の 2 点である．また，本評価実験での計算機環境を表 3.4 に示す．

3.5.1 データセット

評価実験で用いたインフィード広告の配信データの概要を表 3.5 に示す．これらは，株式会社 NTT ドコモが保有するアドネットワーク[3]に蓄積されたデータで，特定のメディア（広告枠）における広告情報と配信実績である．また，CTR 予測における従来研究 [27] と同様に，インプレッション回数が少なくとも 100 回以上

[3] https://www.ntt.com/business/services/docomo_ad_network.html

の広告を選択している．インプレッション回数が 100 回に満たないような極端に少ない状態で算出した CTR は，真値と大きく異なる可能性があるという理由で，予測の対象外としている．また，表 3.5 に記載の期間 (2018/10/01-2019/02/20) を通じて，インプレッション回数が 100 回未満の広告の割合は 4%程度と僅かであり，除外した場合の広告配信における実効的な影響もほぼ無いと言える．

配信実績は，広告クリエイティブの 1 インプレッション単位の結果（click or not-click）が保持されており，広告の配信開始以降の任意の期間での CTR 算出が可能である．今回の評価実験のうち，広告クリエイティブごとの CTR は以下のように算出している．

- CTR 予測の精度比較の場合：広告の配信日から終了までの全期間における CTR
- 多期間 CTR 予測の精度比較の場合：式 (3.4) で示した，配信日から週 t 初日までの $\mathrm{CTR}_{l,t}$

予測モデルの精度は，実際の運用を想定し，新規の広告クリエイティブの場合で検証した．すなわち，訓練データとテストデータは，広告の配信日で分割し，テストデータの広告クリエイティブは，訓練データ中に一切含まれない．また，予測モデルの学習中の評価を行うために，訓練データからランダムに分割してバリデーションデータを作成した．各データセットの量に関して，訓練データ，バリデーションデータ，テストデータの比率はそれぞれ 70%，15%，15%としている．

3.5.2 CTR 予測の精度評価

マルチモーダル特徴量による効果

3.4.1 項で述べた Basic, Image そして Text といった，広告クリエイティブに基づく特徴量を用いた CTR 予測モデルの精度評価を表 3.6 に示す．精度評価の Baseline は，次のように算出している．まず，訓練データ中の全ての広告クリエイティブ l における，全期間 T での CTR の平均値 $\bar{y} = \frac{\sum_l k_l}{\sum_l n_l}$ を算出する．ここで，k_l 及び n_l はそれぞれ，広告クリエイティブ l の全期間 T におけるクリック回数とインプレッション回数を示す．次に，この算出された一点の値である $\bar{y}$

表 3.6: CTR 予測における精度比較

Model	MSE ($\times 10^{-5}$)	% Improve
Avg. CTR (Baseline)	3.07	—
Basic	2.29	25.4%
Basic + Image	1.97	35.8%
Basic + Image + Text	1.82	**40.7**%

を，テストデータにおける個々の広告クリエイティブの CTR の予測値として用いる．この Baseline に対し，特徴量として Basic のみのモデル，Basic に Image を追加したモデル，Basic に Image と Text の双方を追加したモデルで，平均二乗誤差 (Mean Squared Error; MSE) とその改善率を求めた．表 3.6 に示すように，広告クリエイティブに関する特徴量である Basic, Image そして Text を追加していくことで，Baseline から誤差を段階的に改善できていることが分かる．特に，全ての特徴量を用いたモデル (Basic + Image + Text) の場合，Baseline から誤差を 40%以上も改善できている．このことから，広告クリエイティブの CTR 予測を行う上で，画像情報やテキスト情報が有用であることが示唆される．

次に，CTR 予測をする際に寄与した特徴量を調べるために，SHAP value [47] を用いて重要度を算出し，上位 50 個の特徴量を抽出した．モデルごとの主な特徴量を表 3.7 に示す．Basic のみのモデルは，上位 50 個の重要特徴量のうち大部分が広告主に関する特徴量である．一方で，全ての特徴量を用いたモデル (Basic + Image + Text) では，広告主に関する特徴量の重要度が減少し，代わりに Image や Text 由来の特徴量の重要度が増している．このことから，広告主といった過去の配信結果を基にした特徴量に依存せず，Image や Text といった広告クリエイティブ自体の特徴を反映していることが分かる．

これによる効果として，新規の広告主に対する CTR 予測精度の向上が考えられる．そこで，表 3.6 のテストデータのうち，訓練データ中に含まれていない新規の広告主のみに限定した場合で，同様に CTR 予測の精度比較を行った．その結果を表 3.8 に示す．新規広告主の場合においても，Basic のみのモデルに比べて，全ての特徴量を用いたモデル (Basic + Image + Text) により予測精度が向上できている．尚，新規広告主の場合では，Basic における広告主に関する特

表 3.7: 上位 50 個の重要特徴量

Model	主な重要特徴量
Basic	広告主に関する特徴量（29 個）
Basic + Image + Text	広告主に関する特徴量（6 個）， Image 由来の特徴量（15 個）， Text 由来の特徴量（12 個）

表 3.8: 新規広告主の場合の CTR 予測における精度比較

Model	MSE ($\times 10^{-5}$)
Basic	2.69
Basic + Image + Text	2.43

徴量は欠損値となっている．また，テストデータにおける新規広告主の割合は，22%程度を占めている．これらの結果から，新規の広告主に対しても，Image や Text といった広告クリエイティブ自体の特徴を，予測に用いることは有効であると言える．

ネットワーク構造の改良による効果

3.4.1 項で述べた Dropout と L2 正則化の追加といったネットワーク構造の改良による CTR 予測の精度評価を表 3.9 に示す．ここでは全ての特徴量 (Basic + Image + Text) を用いており，図 3.4 で示した提案手法の場合 (w/ dropout + L2 normalization) と，従来手法の場合 (w/o dropout + L2 normalization) との精度比較を行っている．表 3.9 に示すように，Dropout と L2 正則化の追加といったネットワーク構造の改良によって，CTR 予測の精度が大きく向上していることが分かる．また，それぞれの場合の学習曲線を図 3.6 に示す．図 3.6 (a) は提案手法の場合，図 3.6 (b) は従来手法の場合の学習曲線であり，横軸は学習のエポック数，縦軸は誤差 (MSE) を表す．また，破線と実線は，それぞれ訓練データとバリデーションデータにおける結果を示している．学習における最適化アルゴリズムは Adam [48] を用いており，学習率は 10^{-5} とした．従来手法では，バリデーションデータと訓練データでの誤差に乖離があり，ハイバリアンスな

表 3.9: Dropout 及び L2 正則化の有無による精度比較

Model	MSE ($\times 10^{-5}$)
w/ dropout+L2 normalization	1.82
w/o dropout+L2 normalization	3.01

状態となっているが，提案手法では，バリデーションデータの誤差も順調に減少していることが分かる．このことから，提案手法であるマルチモーダルな特徴量を用いた CTR 予測モデルへの Dropout と L2 正則化の追加によって，過学習を防ぎ，CTR 予測のロバスト性を向上できたと言える．

3.5.3 多期間 CTR 予測における提案モデルの評価

3.4.2 項で述べた，CTR 予測を多期間に拡張した DNN モデル，及び提案モデルである RNN モデルにて，新規広告の配信開始から 4 週目までの CTR 予測を行った．対象とする広告は，累計インプレッション回数が 1,000 回以上のものを選択している．これは，初週における CTR のばらつきを軽減し，CTR の時間的な減衰をより精緻に見るためである．また，3.1 節の例で述べたように，インプレッション回数が 1,000 回あれば，CTR の真値が 5%の広告について，信頼度 85%で推定値がその 1%以内に収まっていると言える．

モデル比較のため，訓練データにおける全広告で算出した各期間の平均 CTR を予測値とするモデル (Avg. CTR) での精度も求めた．表 3.10 に各モデルの多期間 ($t = 1, 2, 3, 4$) の CTR に対する誤差 (MSE)，表 3.11 に Avg. CTR モデルからの MSE 改善率を示す．表 3.10 から，提案手法である RNN モデルが初週の week 1 から含めて全ての期間で最も良い精度であることが分かる．また，表 3.11 から，短期先の予測 (week 1) では，DNN モデルと RNN モデルの改善率の差は 0.6%と限定的であるが，より長期先の予測 (week 4) では，改善率の差が 1.6%になり提案手法による効果が現れていることが分かる．これは DNN モデルにおける多出力の CTR は独立した予測値になっているが，RNN モデルでは多期間の CTR を系列データとして学習させているため，CTR の時間変化を効果的に捉えられたためだと考えられる．これより，評価実験で示した 4 週間先，またはそれ以上といった，長期的な広告掲載が見込まれる場合において，どの程度の CTR

表 3.10: 多期間 CTR 予測モデルの精度比較

Model	MSE ($\times 10^{-5}$)			
	week 1	week 2	week 3	week 4
Avg. CTR	2.28	1.71	1.60	1.57
DNN	1.36	0.99	0.91	0.89
RNN	**1.35**	**0.98**	**0.90**	**0.86**

減衰が起きるかを，掲載事前に精度良く予測できるという点で，DNN モデルに比べ提案手法に優位性があると言える．より正確な CTR 予測ができることで，クリックされやすい広告を効果的に選択することが可能になり，クリック数の増加に繋がるという利点がある．今回，提案手法による 4 週間先の CTR 予測の DNN モデルに対する精度改善が 1.6%であり，大規模なアドネットワークでよりインプレッション数が多い場面ほど，広告クリックによる送客数は増えていくと言える．

次に，提案手法である RNN モデルの計算時間について説明する．3.5.1 項で示した本データセットに対する RNN モデルの学習時間は 48 分であり，新規広告でのテストデータに対する推論時間は 1 分未満である．実際の広告配信に提案手法を用いることを想定した場合，十分許容できる計算時間であり，さらに，日々の配信結果から RNN モデルを日次で再学習する運用も可能と考えられる．

今回の評価実験で使用した広告配信の実績データは，特定のメディア上で掲載された広告であり，また，掲載期間も表 3.5 に示した特定の期間 (2018/10/01-2019/02/20) であるため，データの選択バイアスを含む可能性がある．しかしながら，広告の内容は他メディアでも掲載されている一般的なものであり，メディアや季節に依存する広告が占める割合は少なく，その影響は軽微であると考えられる．今後，さらに汎用性を高めるためには，複数メディアでの長期間のデータセットを用い，広告の掲載先に関する空間的な特徴量や，季節性・イベントに関する時間的な特徴量を組み入れるといった工夫が考えられる．

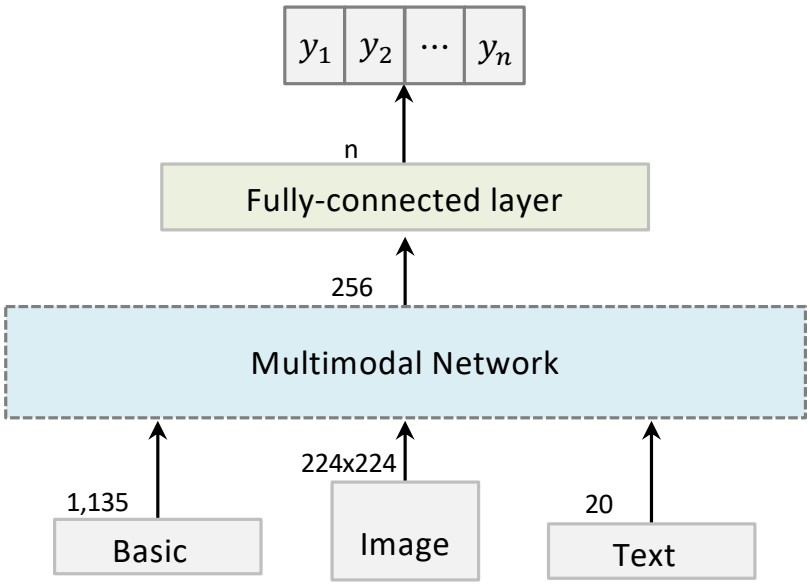
表 3.11: Avg. CTR からの MSE 改善率比較

Model	% Improve			
	week 1	week 2	week 3	week 4
Avg. CTR	—	—	—	—
DNN	40.2%	42.4%	42.9%	43.5%
RNN	40.8%	42.6%	43.6%	**45.1**%

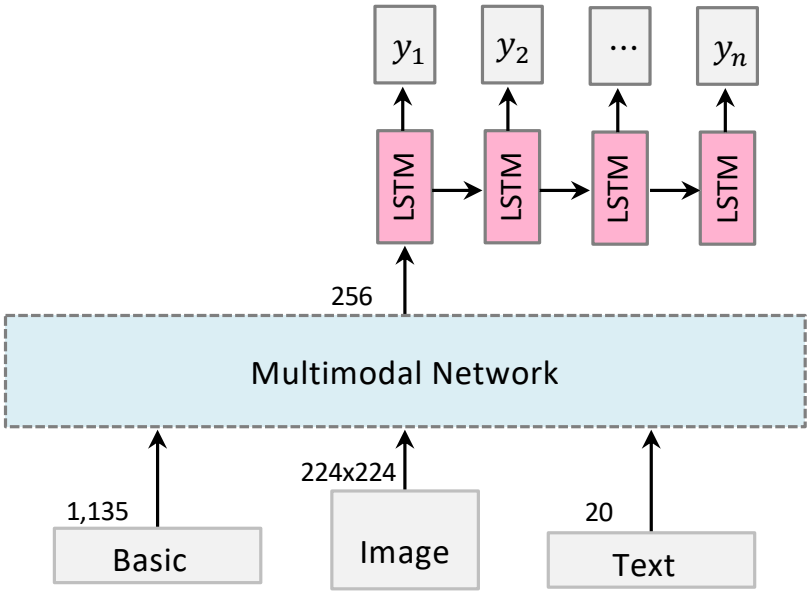
3.6 まとめ

本章の研究では，インフィード広告における CTR の時間減衰を考慮した多期間の CTR 予測手法を提案した．まず，マルチモーダルな特徴量から広告単位の CTR をロバストに予測するための CTR 予測モデルを構築した．次に，多期間にわたる CTR の時系列変化を抽象的に表現可能な RNN 型のネットワーク構造を提案した．提案手法の効果を測るために，アドネットワーク上の配信履歴を用いたオフライン検証で多期間 CTR 予測を実施し，提案手法の有効性を示した．

今後の課題として，週次といった時間単位だけでなく，インプレッション回数に応じた CTR の減衰を捉えることで，意思決定により反映されやすい予測にすることが考えられる．また，CTR の時間減衰を推定した上で，最適な配信計画を離散最適化問題として求解した場合の効果を明らかにすることが課題として挙げられる．さらに，多期間 CTR 予測において各期間の重要特徴量がどのように変化していくかを SHAP value [47] を用いて解析し，クリエイティブ自体の改善に繋げていくことも考えられる．

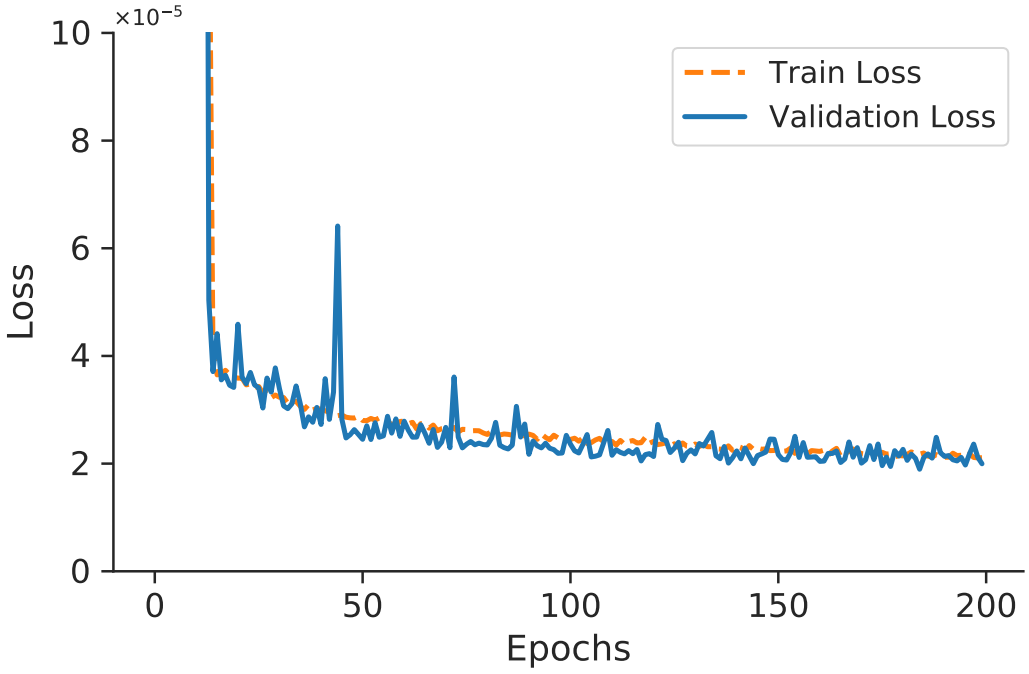


(a) DNN model

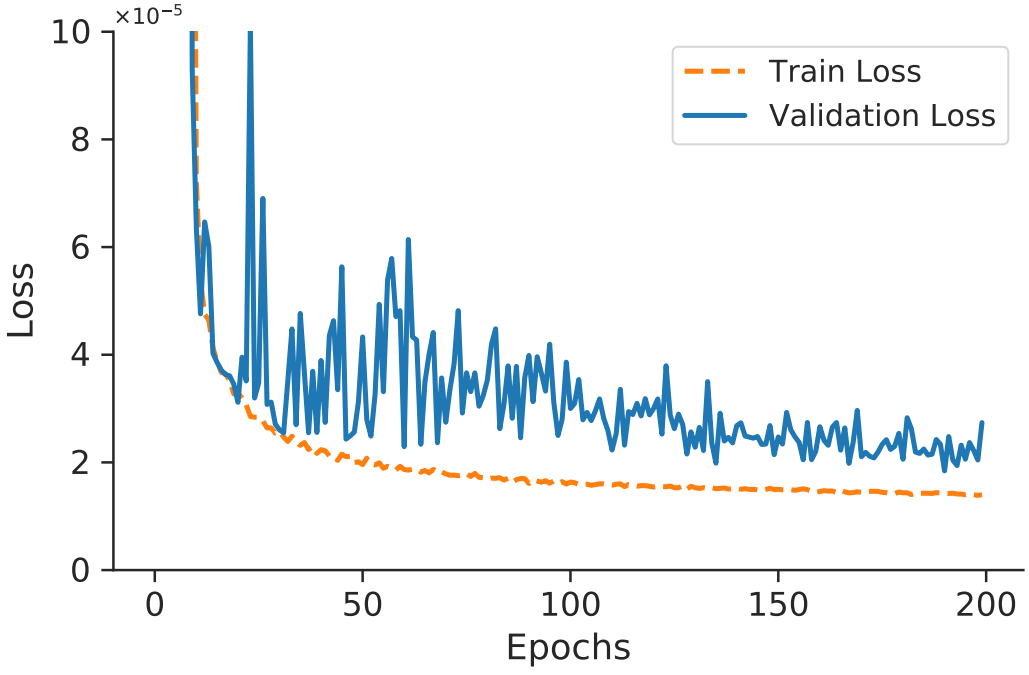


(b) RNN model

図 3.5: 多期間 CTR 予測モデル



(a) w/ dropout + L2 normalization



(b) w/o dropout + L2 normalization

図 3.6: 学習曲線

第 4 章

モデルベース深層強化学習による新規小売商品の在庫管理

4.1 はじめに

商品の販売期間を通じて，需要に合わせて適切な供給量を決定して在庫量を一定に保つ在庫管理は，サプライチェーンにおける重要な要素である．日々の発注量を最適化することで，在庫切れに伴う販売機会損失の回避や，逆に，過剰な発注による在庫コスト高騰の抑止が可能になる．しかし，実際には，需要側の急な変動や，供給側における原材料供給の遅延など，あらゆる局面で不確実な環境に置かれている．また，扱う商品によっては，製品ライフサイクルが短い場合や，多くの製品バリエーションを持つ場合もあり，環境変化の速さや需要の多様性に対応する必要もある．本章の研究では，このようなサプライチェーンの商品の一つであるスマートフォンに着目し，多商品かつ多店舗における新規商品の在庫管理問題を扱う．

通常，スマートフォンの販売店では，複数のサプライヤーから出される様々な製品を扱い，かつ毎シーズンで新たなラインナップとして刷新されていく．こうした変化の速い環境のなかで，新規商品が発売された日からの在庫運用の最適化を目指す．この最適化の目的としては，大きく二つ挙げられる．一つは，新規商品における欠品発生率を抑えながら不良在庫量を削減することで，各販売店での利益を最大化することである．もう一方は，適切な供給量を自動的に導

出することで，販売店側での発注業務を削減し，販売促進活動へリソースを割けるようにすることである．

この多商品かつ多店舗における在庫管理問題において，新規商品の在庫量の決定粒度は，日別，製品別，店舗別といった最小の単位を想定する．しかしながら，これらの単位で在庫量を適切に保つためには，以下の課題が存在する．

1. 商品に対する購買需要は確率的であり，将来の需要を見積もりながら供給量を決定する必要がる．

2. 季節性，商品性，地域性によって需要傾向は異なるために，それらに合わせたきめ細やかな在庫管理が求められる．

3. 新規商品の在庫管理においては，発売実績が未だない状況のなかでも，発売当初からその需要を見越して配備する必要がある．

従来より，数理最適化や強化学習に基づく在庫管理手法は多く研究されている．例えば，在庫管理を確率計画問題として定式化したもの [11] や，動的計画法を用いたものがあり [12]，一定の効果が示されている．しかし，実際の大規模なサプライチェーン上で供給量を決定する場合，これらの厳密解法は，計算量の観点で現実的ではない．これに対して，強化学習によって厳密解法の計算量を克服するアプローチが取られている [13]．また，近年，深層学習を組み入れた深層強化学習が多くの分野で発展しており，在庫管理への適用事例も報告されている [15]．しかし，これらの既存研究では，定常的に販売されている商品を扱う在庫管理が中心であり，新規商品が発売された当初から適用することは不可能である．

そこで本章の研究では，新規商品の発売当初から最適な在庫管理を行う問題を解決するために，学習効率性の高いモデルベース深層強化学習を用いて，オフライン環境での需要予測モデルの学習と，オンライン環境でのプランニングを組み合わせた手法を提案する．まず，学習フェーズは，過去商品の販売実績データを使って，商品需要を含めた環境のモデル化を行う．ここで，需要予測モデルとして，確率的な予測が可能な Bayesian neural network (BNN) [49] や，メタ学習の手法である MAML (Model-Agnostic Meta-Learning) [50, 51] を取り入れ，新規商品の需要を精度高く予測するモデルを構築する．次に，プランニングフェー

ズでは，統計的な見積もりから需要量に対するバッファを加えた上で，Random shooting [52, 53] を用いて新規商品における日別，製品別，店舗別の最適発注量を決定する．提案手法の効果を測るため，実際の販売実績を用いた在庫管理シミュレーションによる評価実験を行い，収益性観点の総報酬，効率性観点の在庫回転率，顧客満足度観点の欠品発生率といった複数の在庫管理指標を確認する．

本章の研究の貢献は以下である．

- 過去商品の販売実績で学習した需要予測モデルと，新規商品の発売開始後の発注プランニングとを組み合わせた，モデルベース深層強化学習のアルゴリズムを考案し，新規商品が発売された当初から適用可能な在庫管理手法を提案した．
- 実際の販売実績を用いた在庫管理シミュレーションを行い，提案手法により総報酬，在庫回転率，欠品発生率の全ての指標において，従来手法からの改善を確認した．
- 在庫管理シミュレーションでは，異なる需要傾向を持つ複数の新規商品，及び複数の販売店で評価を行い，それぞれで在庫量をおよそ一定に保つことができ，安定した在庫管理ができていることも確認した．

以下，4.2 節では，強化学習の適用事例と，在庫管理における既存手法について説明し，本章の研究の位置付けを明確にする．4.3 節では，新規商品に対する発注量決定問題の定式化と，対象とする在庫管理指標について述べる．4.4 節では，モデルベース深層強化学習による提案手法について説明する．4.5 節では，実際の販売実績を用いた在庫管理シミュレーションの結果について述べる．最後に，4.6 節では，本章の研究のまとめと課題を述べる．

4.2 関連研究

本節では，まず，強化学習の概要と，本研究のアプローチであるモデルベース強化学習について説明する．次に，在庫管理における強化学習の既存研究と，それらを新規商品の在庫管理へ適用する際の課題について述べ，本研究の位置付けを明確にする．

4.2.1 強化学習の概要

強化学習 (Reinforcement learning; RL) は，エージェントが環境との試行錯誤を通じて，最適な制御を獲得する機械学習の一つである [54]．強化学習では，マルコフ決定過程で各時刻 t における行動 a_t，状態 s_t，報酬 r_t を伴い，時間発展の中で学習を進める．エージェントは，この報酬を最大化するような方策 $\pi(a_t|s_t)$ や，状態 s_t で行動 a_t をとった際の行動価値関数 $Q(a_t, s_t)$，状態価値関数 $V(s_t)$ などを環境から学習することになる．

強化学習は多くの分野で適用が進んでいる．例えば，ロボティックス [55] や移動交通領域 [56]，ヘルスケア領域 [57, 58]，マーケティング領域 [59, 60]，無線通信領域 [61] など，多岐にわたっている．特に，移動交通領域の例としては，オンデマンド交通プラットフォーム上のタクシー車両の配車決定への適用が挙げられる．Xu ら [56] が提案した手法では，各タクシー車両をエージェントとして，タクシー車両の乗降履歴から時空間的な状態価値関数 $V(s_t)$ を学習し，その更新には強化学習における TD 学習を用いている．

4.2.2 モデルベース強化学習

強化学習のアプローチはモデルフリー型 (Model-free RL) とモデルベース型 (Model-based RL) に大別される [62]．モデルとは，遷移関数 (Transition function) と報酬関数 (Reward function) を意味していて，エージェントがこれらを利用するかどうかで，そのタイプが決まる．モデルフリー型は，環境モデルを利用せずに，方策を学習するアプローチである．遷移関数や報酬関数が未知な環境でも適用できるため，モデルフリー型が多く用いられている．一方，モデルベース型は，状態遷移や報酬を予測する関数を用いることで，エージェントが状況の先読みを行い，行動空間の中から適切な選択が可能となる．そのため，モデルベース型のメリットしては，学習の効率性が良いという点が挙げられる．一方で，デメリットとしては，エージェントの行動選択による報酬の良し悪しは，方策の学習の部分だけでなく，環境モデルの学習にも依存する点が挙げられる．

従来は, 環境をモデル化できるケースが少なかったが, 表現力の高い深層学習の登場によりモデル化できるケースが増えており, 近年, モデルベース型の手法が多く提案されている [63, 64, 65]．一つのアプローチとしては，Dyna algorithm [66]

を用いたものである．Dyna algorithm では，(1) モデルフリー型で方策を学習しながら，その過程で得られた環境からのデータを使ってモデルを学習し，(2) 次のステップで，モデルを使って追加で方策を学習する，を繰り返すことで学習の効率性を高めている．これを応用したアルゴリズムとして，ME-TRPO [67] が挙げられる．ME-TRPO では，方策の更新で非線形最適化手法を利用した TRPO [68] を用い，環境モデルは複数のニューラルネットワークモデルのアンサンブルを駆使することで，サンプル効率性とモデルバイアスの排除を行っている．

モデルベース型の他のアプローチとして，Model Predictive Control (MPC) [69, 63] を用いた Shooting algorithm が挙げられる．例えば，その一手法である Random shooting (RS) [52, 53] では，エージェントが一様分布に従うランダムな行動系列を複数生成し，学習済みモデルを使って各行動系列の報酬を評価する．そして，エージェントは得られた最適な系列のうち最初の行動のみを実行し，毎ステップで再計画を繰り返し実行していく．Wang ら [63] による 18 種類のタスクのベンチマーク環境で実施された評価によると，Random shooting を含む Shooting algorithm は，異なる環境下でも有効かつ頑健であることが報告されている．一方で，デメリットとしては，行動空間，または状態空間が膨大な場合，十分な探索は困難な場合があるといった点が挙げられる．

4.2.3　在庫管理における強化学習の適用事例

サプライチェーンにおける在庫管理においても，強化学習によるアプローチは多く提案されている．

Giannoccaro ら [13] は，3 ステージ (Supplier, Manufacturer and Distributor) の在庫管理におけるマルコフ決定過程での定式化と，強化学習によるアプローチを提案し，従来の動的計画法と比較して，より大規模な問題へも適用可能と述べている．Kara ら [14] は，Perishable inventory system での在庫管理として，強化学習における Q 学習や Sarsa のアルゴリズムを適用し，その有効性を示している．

近年では，強化学習と深層学習とを組み合わせた深層強化学習 (Deep Reinforcement Learning; DRL) を，在庫管理に適用した事例も報告されている [15]．Meisheri ら [16] は，多商品の在庫管理システムに対して，Advantage actor critic

(A2C) と Deep Q-network (DQN) を適用し，比較的大規模な問題 (100 から 200 商品での在庫管理) における手法の有用性を提案している．Gijsbrechts ら [17] は，Dual-sourcing and multi-echelon inventory management system に対して，Asynchronous advantage actor critic (A3C) [70] を適用している．在庫管理への深層強化学習の応用は進んでいるが，そのタイプはモデルフリー型に限られている．モデルフリー型の利点としては，未知の環境下でも適用できる柔軟性の高さが挙げられる．一方，デメリットとしては，学習には大量のデータと多数のエピソードが必要であるため，学習効率性が高くないといった点が挙げられる．

このモデルフリー型の弱点である非効率な学習過程を補うため，近年では，モデルベース型を在庫管理に適用する事例も報告されている．例えば, Malik ら [18] は，Perishable inventory system に対するモデルベース強化学習の適用を報告している．スーパーマーケットチェーンにおける生鮮品の需要予測モデルを学習させ，モデルベース深層強化学習における報酬の推定に用いている．需要についての不確実性に対応するため，深層学習モデルのキャリブレーションを行うことで，ヒューリスティック手法と比較して大幅な性能改善を達成している．

4.2.4 新規商品への適用課題

これらのモデルフリー型，または，モデルベース型の深層強化学習を用いた在庫管理の既存研究において，対象のサプライチェーンは，定常的に販売されている商品を扱ったものが中心である．在庫管理の対象である当該商品の販売履歴が既に得られている状態で，エージェントの方策学習を行っている．そのため，新規商品の在庫管理を対象にした場合は，発売当初から即座に方策を適用することは難しいという問題が挙げられる．

一方で，統計的手法による新規商品に対する在庫管理はいくつか従来手法が報告されている．例えば，Wanke ら [24] は，商品の需要を三角分布 (Triangular distribution) で表し，古典的な在庫管理方策である (Q, r) モデルで商品の供給量を決定するアプローチを提案している．(Q, r) モデルでは，在庫レベルが設定された再発注点 r を下回った段階で，注文点 Q まで発注する方策である．Rojas [71] は，Wanke らの手法を時系列的に拡張し，自己回帰移動平均モデルを適用する在庫管理手法を提案している．

これらの新規商品に対する在庫管理の既存手法の利点は，単純なアプローチであるため，実際の在庫管理システムへの適用が容易であるといった点が挙げられる．一方で，デメリットとしては，需要予測の観点において，商品特性や地域特性などの外生変数を考慮していない単純な統計モデルであるため，外生変数を考慮した機械学習モデルに比べて，予測精度が劣る可能性が挙げられる．加えて，在庫管理の観点において，(Q, r) モデルのようなヒューリスティック手法では，需要の変動が激しい場合に，在庫切れや過剰発注につながる可能性が挙げられる．

4.2.5 本研究の位置付け

表 4.1 に，提案手法を含めた，既存商品・新規商品の在庫管理に関する研究の比較を示す．深層強化学習を用いた既存研究とは対照的に，本研究では新規商品を対象とし，商品の発売直後から適用可能で，欠品発生率を低減しつつ高い収益性と効率性を目指した在庫管理手法を提案する．ここで，発売してから間もない商品の在庫管理を最適化する目的で，モデルベース深層強化学習を用いている．モデルフリー型の深層強化学習では，大量のデータや試行回数を必要とするため，新規商品を扱う在庫管理には不向きと考えられる．これに対し，サンプル効率性の高いモデルベース深層強化学習を用いることで，新規商品の在庫管理における有効性を示す．

提案手法では，環境モデル（遷移関数や報酬関数の推定）として，新規商品の需要予測モデルを利用する．この需要予測モデルは，過去商品の販売実績を用いて学習を行う．需要予測モデルは複数のバリエーションを適用しており，不確実性を考慮した予測に対応した Bayesian neural networks (BNN) , メタ学習の代表的手法である MAML (Model-Agnostic Meta-Learning) をそれぞれ用いている．この需要予測モデルを，実際の環境を模倣したシミュレーターとして用いて，Random shooting (RS) による発注プランニングを通じて，新規商品の在庫管理を最適化する．

表 4.1: 既存商品・新規商品の在庫管理に関する研究の比較

研究	対象在庫	深層強化学習		統計的手法	主な手法
		モデルベース型	モデルフリー型		
Kara ら (2018)[14]	既存商品		✓		Q-learning, SARSA
Meisheri ら (2020)[16]	既存商品		✓		A2C, DQN
Gijsbrechts ら (2021)[17]	既存商品		✓		A3C
Malik ら (2019)[18]	既存商品	✓			Calibrated model-based DRL
Wanke ら (2016)[24]	新規商品			✓	(Q, r) with statistical model
提案手法 (2023)[72]	新規商品	✓			RS with BNN or MAML

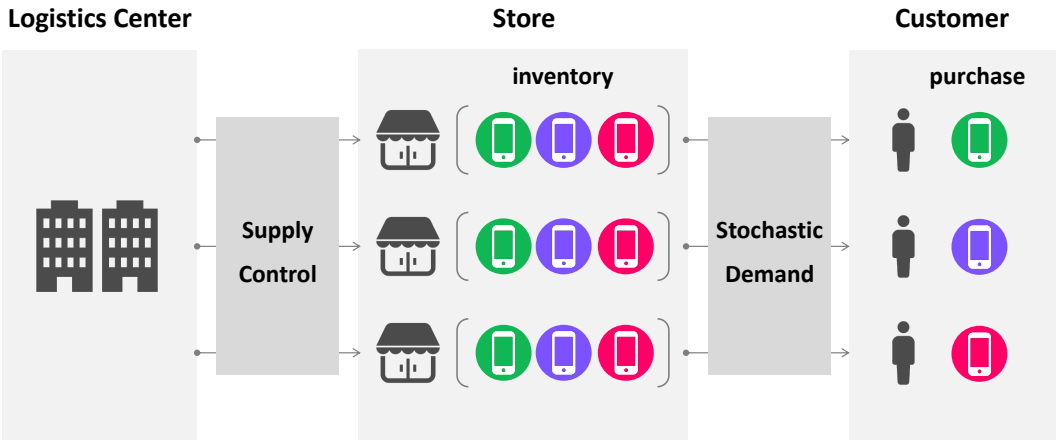


図 4.1: 多商品・多店舗の環境でのスマートフォンの新規商品の在庫管理

4.3 問題設定

本研究では，スマートフォンを例に挙げて，欠品発生率を抑えつつ，余剰在庫数を最小化するような新規商品の供給量決定問題を扱う．本節では，新規商品の供給量決定問題についての定義を示し，マルコフ決定過程による定式化，及び対象とする在庫管理指標について説明する．

4.3.1 新規商品の供給量決定問題

図 4.1 に示すように，対象とするサプライチェーンは，ロジスティックセンター，販売店，最終消費者の連鎖であり，複数の販売店と複数の新規商品を想定している．エージェントは，各店舗と各商品の在庫量を観測しながら，それらの次の供給量を適切に決定し，商品を配備することになる．

在庫管理は，各商品が発売された日から開始される．発売日初日における各店舗，各商品の在庫量は，初期値として任意に与えられる．エージェントは，日々の在庫量の変動を観測しながら，各日の営業終了後に供給量を決定して翌日の営業開始前までに商品が納入されるとする．ここでは簡単のため，商品発注から納入までのリードタイムは 1 日未満としている．

各商品は，異なるサプライヤーから調達していて，商品によって需要傾向は異なっている．また，スマートフォンの製品ライフサイクルは短く，毎シーズンで各サプライヤーから新規商品が発売される．ここで，各商品は，それぞれ複数

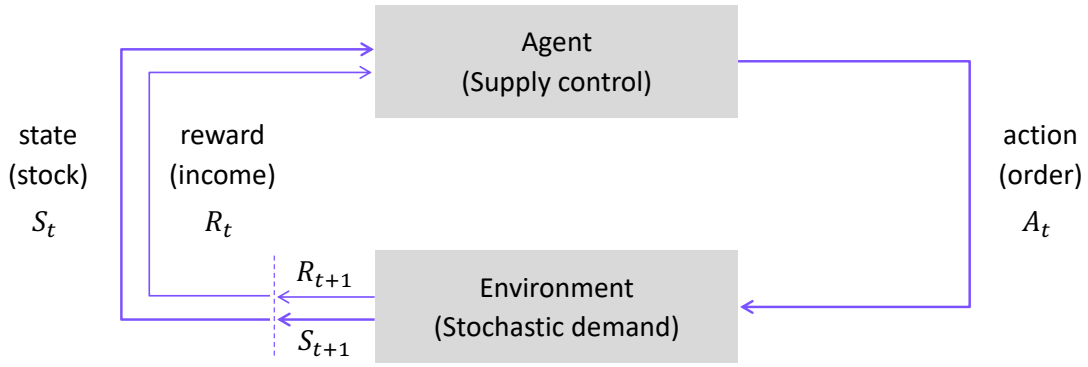


図 4.2: 新規商品の供給量決定問題でのエージェントと環境の相互作用

のカラーバリエーションを持つことが一般的であるが，本研究では簡単のため，色の区別はせずに製品シリーズ単位での在庫量を対象とする．

各店舗は，立地エリアによっても各商品の需要傾向は異なっている．簡単な例として，アーバンエリアでは最新のフラッグシップ商品は発売日と共に高い販売需要を持つが，逆にルーラルエリアでは過去商品の販売需要が強く残っているなどである．

このような特徴を持ったスマートフォンの新規商品の販売需要に対して，日別，商品別，店舗別の単位で，供給量の意思決定を行うものとする．

4.3.2 マルコフ決定過程

本節では，新規商品の供給量決定問題における，マルコフ決定過程での定式化について述べる．ここで，図 4.2 にエージェントと環境との間でのやりとりを示す．環境は，多商品かつ多店舗における確率的な購買需要を持ったサプライチェーンである．エージェントは，各ステップの在庫量を含めた状態 S_t を観測しながら，次の供給量 A_t を決定して在庫管理を行う．以下，このマルコフ決定過程での各要素の詳細を説明する．

時間ステップ

本問題では，時刻 t を離散時間とし，時間ステップは 1 日とする．すなわち，エージェントは毎日の供給量決定を通じて，適切な在庫管理を維持することになる．ここで，1 つのエピソードは有限な時間で打ち切られるとして，有限集合

$\mathcal{T}$ を式 (4.1) のように表す．

$$\mathcal{T} = \{t \mid 0 \leq t \leq T, t \in \mathbb{N}\} \tag{4.1}$$

ここで，定数 T は在庫運用の終了時刻である．

状態変数

状態 s は，時刻 t での商品 i，店舗 j に関する各種の観測値として，式 (4.2) のように表す．

$$s = [q, u_i, v_j, w_t], q \in \mathcal{L}, i \in \mathcal{I}, j \in \mathcal{J}, t \in \mathcal{T} \tag{4.2}$$

ここで，q は時刻 t での商品 i の店舗 j における在庫量を表すスカラー値である．また，u_i は商品 i の特徴量ベクトル，v_j は店舗 j の特徴量ベクトル，w_t は時刻 t の特徴量ベクトルである （各特徴量の具体例は表 4.2 を参照のこと）．また，各集合 $\mathcal{L}, \mathcal{I}, \mathcal{J}$ は次の通りである．$\mathcal{L}$ は，取り得る在庫量の集合であり，式 (4.3) のように表す．

$$\mathcal{L} = \{q \mid 0 \leq q \leq M, q \in \mathbb{N}\} \tag{4.3}$$

ここで，定数 M はとり得る在庫量の最大値である．$\mathcal{I}, \mathcal{J}$ は，それぞれ商品 i，店舗 j の集合を示している．

行動変数

行動 a は，時刻 t における商品 i の店舗 j に対する発注量である．行動として取りうる値は $a \in \mathcal{L}$ と表現され，行動空間は離散的となっている．4.3.1 項で述べた様に，商品発注から納入までのリードタイムは 1 日未満としているため，エージェントが行動決定をした後に，発注量 a が時刻 $[t, t+1)$ で在庫配備されることになる．

遷移関数

状態 $S_t = s$ において行動 $A_t = a$ をとった際に，時間ステップ $t+1$ での状態が $S_{t+1} = s'$ である確率 $P\left(S_{t+1} = s' \mid S_t = s, A_t = a\right)$ が遷移関数である．本問題において遷移関数は未知であるが，時刻 $t+1$ での商品 i の店舗 j における未知の

需要量 d を用いて，時刻 $t+1$ の在庫量 q' は，式 (4.4) のように表すことができる．

$$q' = \max\{\min\{q+a, M\} - d, 0\} \tag{4.4}$$

ここで，$\min\{q+a, M\}$ の項は，在庫配備された直後の時刻 $[t, t+1)$ での商品 i の店舗 j における在庫量であり，時刻 $t+1$ の未知の需要量 d の発生の後に残る在庫量が q' である．式 (4.4) で示した遷移関数は，Szepesvári [73] が在庫管理の例で示している遷移関数と同じものである．

報酬関数

状態 $S_t = s$ において行動 $A_t = a$ をとった際の，$t+1$ での報酬 r' は，遷移関数と同様に未知の需要量 d を用いて，式 (4.5) で表すことができる．

$$\begin{aligned} r' &= R(S_t = s, A_t = a) \\ &= g(\min\{\min\{q+a, M\}, d\}) - c_1(q) - c_2(a) \end{aligned} \tag{4.5}$$

ここで，$g(.)$ は $t+1$ 時点で販売できた台数に対する収入を計算する関数，$c_1(.)$ は t 時点で保持していた台数に対する在庫コストを計算する関数，$c_2(.)$ は t 時点で発注した台数に応じた発注コストを計算する関数である．本研究では，簡単のため，これらの関数 g, c_1, c_2 はいずれも既知で，線形な関数と仮定している．

初期状態

在庫運用の開始時点である時刻 $t = 0$ での商品 i の店舗 j における在庫量は，初期値 q_0 として既知で与えられるとする．

4.3.3 在庫管理の指標

全体の運用期間 T を通じての在庫管理の指標として，本研究では，収益性観点で総報酬，効率性観点で在庫回転率，顧客満足度観点で欠品発生率の 3 つを用いる．以下では，各指標について詳細を述べる．また，各指標は商品 i と店舗 j の組である (i, j) 毎に求めることとする．

総報酬

総報酬は，全期間 T における各時間ステップでの報酬の総和である．商品

と店舗の組 (i, j) に関する時刻 t における報酬 r_t は，式 (4.5) で計算される．従って，1 エピソードでの (i, j) に関する総報酬は，以下で表される．

$$\sum_{t=1}^{T} r_t \tag{4.6}$$

在庫回転率

在庫回転率は，全期間 T において在庫が何回入れ替わったかを表す指標である．在庫回転率が大きいほど，機会損失を防ぎながら在庫量を少なく保てていることを意味し，より効率的な在庫管理がなされていることを示す．在庫回転率は，1 エピソードの終了後に以下で求められる．

$$\frac{\sum_{t=1}^{T} d'_t}{(q_0 + q_T)/2} \tag{4.7}$$

ここで，d'_t は時刻 t での (i, j) の販売量である．(i, j) に関する在庫量を q_{t-1}，発注量を a_{t-1}，次ステップでの需要量を d_t とした場合に，販売量 d'_t は以下で計算される．

$$d'_t = \min\left\{\min\left\{q_{t-1} + a_{t-1}, M\right\}, d_t\right\} \tag{4.8}$$

式 (4.7) の分子は，在庫管理の期間全体での販売量合計である．また，分母は，平均在庫量を表している．本研究では，1 つのエピソードである期間 T における在庫管理の効率性を測るために，平均在庫量は，開始時刻 $t = 0$ と終了時刻 $t = T$ での在庫量で平均化している [74, 75, 76]．

欠品発生率

欠品発生率は，全期間 T のなかで在庫の欠品事象，$\min\left\{q_t + a_t, M\right\} < d_{t+1}$ が発生した日数の比率である．従って，1 エピソードでの (i, j) の欠品発生率は，以下で表される．

$$\sum_{t=1}^{T} \frac{u_t}{T} \tag{4.9}$$

ここで，u_t は，(i, j) に関する各時刻 t での欠品発生の有無を表すバイナリー変数であり，以下で定義される．

$$u_t = \begin{cases} 1, & \text{if } \min\left\{q_{t-1} + a_{t-1}, M\right\} < d_t \\ 0, & \text{otherwise} \end{cases} \tag{4.10}$$

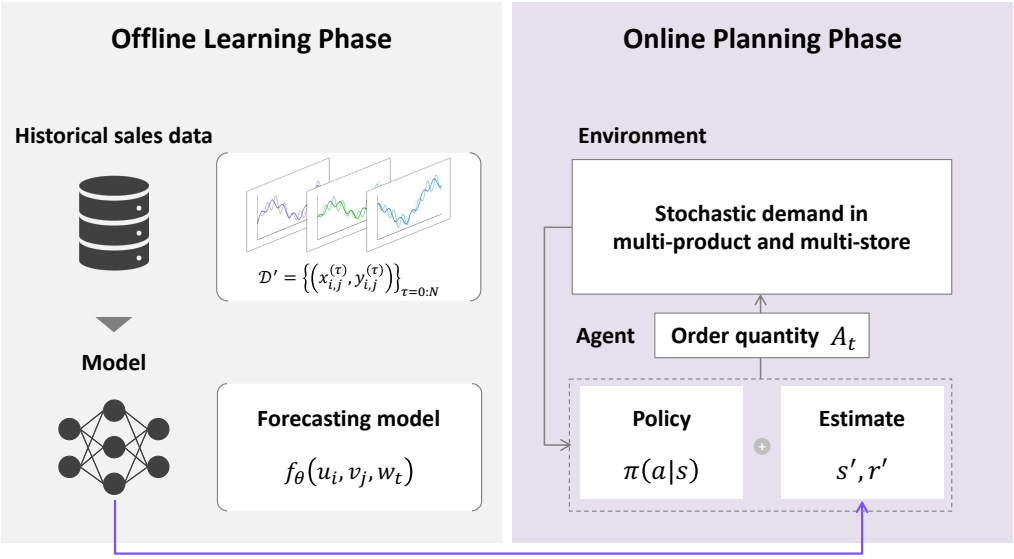


図 4.3: 多商品・多店舗での新規商品の在庫管理に対するモデルベース深層強化学習を用いた提案手法の概略図

4.4 提案手法

本節では，モデルベース深層強化学習を用いた新規商品に対する在庫管理の提案手法について述べる．図 4.3 に提案手法の概略を示す．提案手法は，大きく二つのフェーズがあり，オフライン環境における需要予測モデルの学習フェーズと，オンライン環境における供給決定のプランニングフェーズに分かれている．

4.4.1 商品の需要予測

ここでは商品需要予測におけるモデルの定義や，モデルの学習と推論，利用する特徴量の詳細について説明する．

モデルの定義

各時刻での商品 $i \in \mathcal{I}$ の店舗 $j \in \mathcal{J}$ における未知の需要量 d を予測できれば，式 (4.4)，及び式 (4.5) から，次ステップの遷移状態と報酬を推測することが可能になる．

そこで時刻 t 時点までの観測情報から，時刻 $t+1$ 以降の需要を予測する関数 $f_\theta(.)$ を式 (4.11) で定義する．

$$D = f_\theta(u_i, v_j, w_t) \tag{4.11}$$

ここで，θ は関数 f_θ のパラメータである．また，$D = [d_{t+1}, d_{t+2}, \ldots, d_{t+h}]$ であり，時刻 $t+h$ までのマルチステップで，各時刻での商品 i の店舗 j における需要量を推定している．このうち d_{t+1} を用いることで，時刻 $t+1$ の需要量の推定値 $\hat{d}$ は，以下のように表せる．

$$\hat{d} = d_{t+1} \tag{4.12}$$

モデルの学習と推論

本問題では，新規商品における在庫管理を，強化学習を通じて販売日当日の時刻 $t=0$ から開始する．その際に，時刻 $t+1$ の需要量の推定値 $\hat{d}$ を利用することで，時刻 $t+1$ の在庫量 q' と報酬 r' が推定可能となる．しかし，時刻 $t=0$ の時点で，新規商品に対する十分な販売実績 $\mathcal{D}$ は得られていないため，このままでは，商品需要に対する予測モデル f_θ を学習することは困難である．

そこで，新規商品ではなく，既にデータが蓄積されている過去商品 $i \in \mathcal{I}' \neq \mathcal{I}$ について，過去の期間 $\tau \in \mathcal{T}' \neq \mathcal{T}$ における店舗 $j \in \mathcal{J}'$ での販売実績 $\mathcal{D}'$ を用いて予測モデル f_θ を学習する方法をとる．すなわち，過去の販売実績 $\mathcal{D}' = \{(X_{i,j}^{(\tau)}, Y_{i,j}^{(\tau)})\}_{\tau=0:N}$ を所与として，以下のような予測誤差の最小化問題を解くことで予測モデル f_θ を獲得する．

$$\min_\theta \sum_{i,j,\tau} \left\| f_\theta\left(X_{i,j}^{(\tau)}\right) - Y_{i,j}^{(\tau)} \right\|_2^2 \tag{4.13}$$

ここで，$X_{i,j}^{(\tau)}$ は説明変数であり，$X_{i,j}^{(\tau)} = [u_i, v_j, w_\tau]$ として，過去商品 i，店舗 j，時刻 τ における特徴量ベクトルである．$Y_{i,j}^{(\tau)}$ は目的変数であり，過去商品 i の店舗 j における時刻 $\tau+1$ から $\tau+h$ までの販売実績値のベクトル $Y_{i,j}^{(\tau)} = [y_{\tau+1}, y_{\tau+2}, \ldots, y_{\tau+h}]$ である．

利用する特徴量

需要予測モデル f_θ で用いる特徴量 $X_{i,j}^{(t)} = [u_i, v_j, w_t]$ の詳細について述べる．本研究では，製品ライフサイクルが短い商品の例として，スマートフォンの在庫

表 4.2: スマートフォン商品における需要予測モデル f_θ で用いる主な特徴量

特徴量	詳細	データ型
u_i	販売量の移動平均値 (3, 7, 14, 28 日間)	量的変数
	商品発売日からの経過日数	量的変数
	商品のカラーバリエーション数	量的変数
	商品のサプライヤー ID	カテゴリ変数
	商品の各種スペック情報	カテゴリ変数
v_j	店舗の立地エリア情報	カテゴリ変数
w_t	平日・休日のフラグ情報	カテゴリ変数
	週番号	カテゴリ変数
	曜日	カテゴリ変数
	月・年の中での経過日数に関する三角関数の値 (Sine, Cosine)	量的変数

管理を対象としている．表 4.2 に，商品の需要予測モデルで用いる主な特徴量を示す．u_i は商品 i についての特徴量で，多くの変数を用いている．特に，u_i では商品 i の製品スペック情報といったカテゴリカルデータだけでなく，過去の複数期間 (3, 7, 14, 28 日間) にわたっての販売量の移動平均値も用いている．v_j は店舗 j についての特徴量で，特に立地エリアについてのカテゴリカルな情報を用いている．w_t は時刻 t についてのカレンダー特徴量で，平日休日の区分や週番号，曜日といったカテゴリカルな情報を用いている．また，月や年の中での経過日数を三角関数 $\sin, \cos$ で表した連続値も用いている．これは，月や年の中での周期を表す特徴量としてエンコードされた値であり，月末と月初（または年末と年始）といった周期の最後と最初の値に乖離が生じないようにしたものである．

4.4.2 提案アルゴリズム

提案手法の詳細を Algorithm 1 に示す．過去商品 $i' \in \mathcal{I}' \neq \mathcal{I}$ の期間 $\tau \in \mathcal{T}' \neq \mathcal{T}$ における店舗 $j \in \mathcal{J}'$ での販売実績 $\mathcal{D}'$ を入力として与える．オフライン環境にて，$\texttt{TrainModel}(\mathcal{D}')$ により，商品の需要予測モデル f_θ を学習する．次に，オンライン環境で，エージェントによる新規商品 $i \in \mathcal{I}$ の店舗 $j \in \mathcal{J}$ についての供給

Algorithm 1: Inventory control of new products using model-based DRL

Input: Historical sales dataset $\mathcal{D}'$ for product i' and store j and current state S_t

Output: Order quantity A_t at time t for new product i and store j

`/* Offline Learning Phase */`

Train a product demand forecasting model $f_\theta \leftarrow$ `TrainModel(`$\mathcal{D}'$`)`

`/* Online Planning Phase */`

Set inventory conditions q_0 and M

repeat

Observe the current state $S_t = [q, u_i, v_j, w_t]$ and reward R_t for product i and store j

Predict product demand $D \leftarrow f_\theta(u_i, v_j, w_t)$

Run the agent and determine order quantity $A_t \leftarrow$ `Planning(`S_t, D`)`

return A_t

until $t \leq T$

決定のプランニングを行う．まず，商品 i の店舗 j における各在庫量について初期値 q_0，最大値 M を設定する．各時刻ステップ $t \in \mathcal{T}$ で，状態 $S_t = [q, u_i, v_j, w_t]$ と報酬 R_t を観測する．次に，学習済みモデル f_θ を使って，次の時刻ステップ以降の販売需要 $D = [d_{t+1}, d_{t+2}, \ldots, d_{t+h}]$ を予測する．エージェントは得られた予測値 D と現在の状態 S_t を用いて，環境をシミュレートしながら `Planning(`S_t, D`)` を通じて供給量 A_t を決定する．

4.4.3 不確実性を考慮した需要予測モデル

ここでは，`TrainModel(`$\mathcal{D}'$`)` によるオフライン環境での学習と，オンライン環境での推論について説明する．新規商品の需要という不確実性に対処するために，本研究では二つの深層学習モデルをそれぞれ取り入れた．一つ目のモデルは，BNN により f_θ を確率的モデルとして扱うことである．モデルパラメータの θ を決定論的な値ではなく，確率変数として表現し，深層学習モデルの f_θ をベ

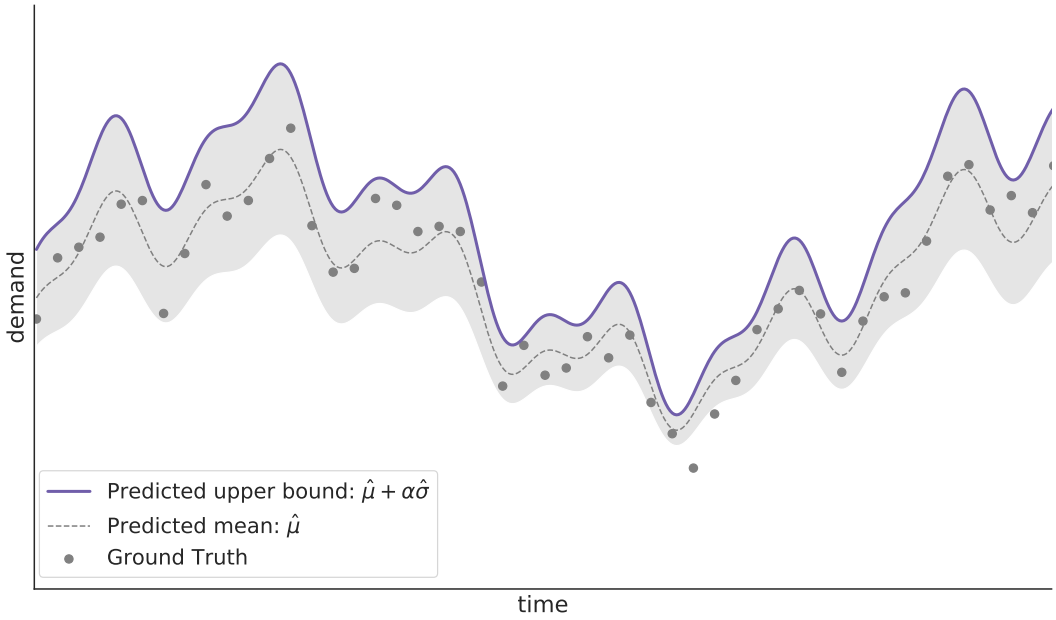


図 4.4: 不確実性を考慮した需要予測の概要図．横軸は時間ステップ，縦軸は需要量を表す．需要の不確実性による商品欠品を防ぐために，予測値の上側信頼限界 $\hat{\mu}_\tau + \alpha\hat{\sigma}_\tau$ を用いる．

イズ推論により学習させる．二つ目のモデルは，メタ学習を用いたアプローチである．データセット $\mathcal{D}'$ から MAML によって θ を学習し，新しい需要予測タスクに即座に適用できることを目的としている．

次に，学習済みモデル f_θ を使ったオンライン環境での推論について述べる．在庫管理において，欠品発生率を抑える目的で，図 4.4 に示すように，予測された需要量に対してバッファーを与えることを考える．すなわち，確率的な需要に対する上側信頼限界を使って，次の時刻ステップ以降の推定需要量 d_τ を式 (4.14) で求める．

$$d_\tau = \hat{\mu}_\tau + \alpha\hat{\sigma}_\tau \tag{4.14}$$

ここで，予測期間は $\tau = t+1, t+2, \ldots, t+h$ とする．また $\hat{\mu}_\tau, \hat{\sigma}_\tau$ は，それぞれ予測値の平均と標準偏差であり，α は信頼区間のパラメータである．

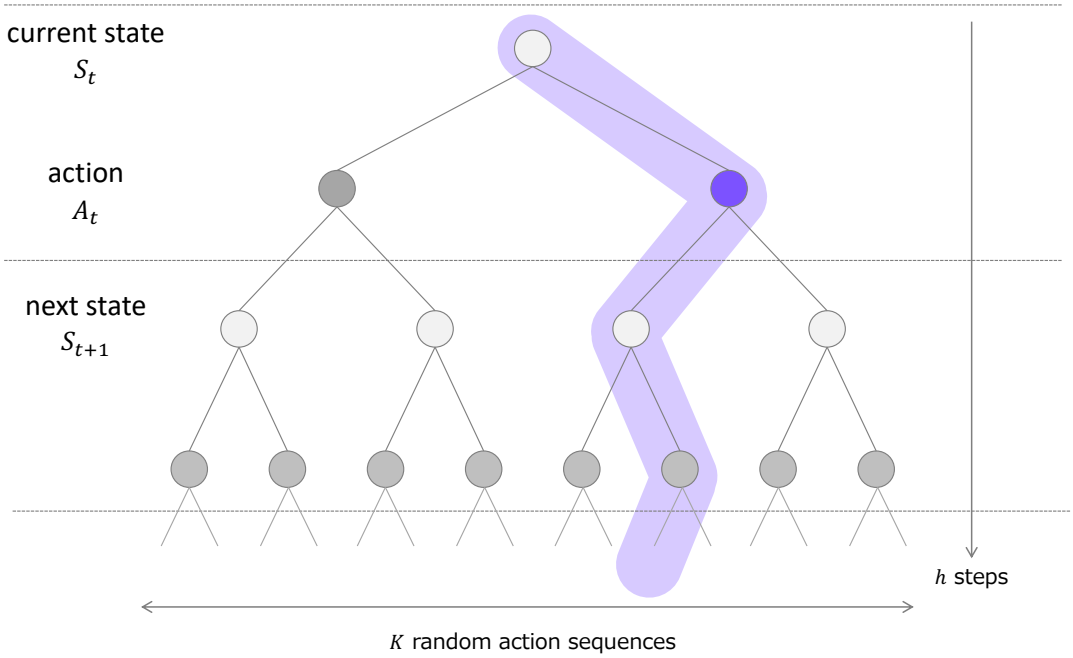


図 4.5: Random shooting によるプランニングの概略図．h ステップ先までのランダム行動系列を K 個生成し，需要予測値 D を基にして将来の状態 S_{t+1} と報酬 R_{t+1} をシミュレートする．

4.4.4 Random shooting によるプランニング

次に，`Planning`(S_t, D) の詳細を説明する．プランニングでは，現在の状態 S_t と，上側信頼限界で見積もった需要量 D を入力として，発注量 A_t を決定する．ここで，モデルベース強化学習のアプローチとして，Random shooting (RS) を用いる．プランニングの概要を図 4.5 に示す．状態 $S_t = [q, u_i, v_j, w_t]$ に含まれる現在庫量 q を起点にして，一様分布に基づき，h ステップ先までのランダムな行動系列 $a_{t:t+h} = [a_t, a_{t+1}, \ldots, a_{t+h}]$ を K 個生成する．各行動系列について，需要予測値 D を基に，式 (4.4) 及び (4.5) を用いて，将来の状態 S_{t+1} と報酬 R_{t+1} をシミュレートする．そして，式 (4.15) に示すように，h ステップ先までの報酬和が最大となる行動系列 $a^*_{t:t+h} = [a^*_t, a^*_{t+1}, \ldots, a^*_{t+h}]$ を導出して，時刻 t での行動である発注量を $A_t = a^*_t$ で決定する．

$$a^*_{t:t+h} = \underset{a_{t:t+h}}{\arg\max} \sum_{\tau=t}^{t+h} R\left(S_\tau = s_\tau, A_\tau = a_\tau\right) \tag{4.15}$$

表 4.3: 計算機環境

OS	Ubuntu 18.04.2 LTS
CPU	Intel(R) Xeon(R) Gold 6152 CPU 2.10GHz
GPU	NVIDIA Tesla V100 SXM2 32GB 1.53GHz

4.5 評価実験

スマートフォンに関する複数商品，及び複数店舗の実際の販売実績を用いた在庫運用のシミュレーションを行い，提案手法である，オフライン環境での学習とオンライン環境でのプランニングを組み合わせた，モデルベース深層強化学習による在庫管理における有効性を示す．

4.5.1 在庫管理シミュレーションの設定

ここでは，在庫管理シミュレーションで用いるデータセットの詳細，及び比較手法についてを説明する．また，本評価実験での計算機環境を表 4.3 に示す．

データセット及びパラメータ

在庫管理シミュレーションでは，株式会社 NTT ドコモが保有するスマートフォンの販売実績を用いている．オンライン環境のプランニングフェーズで用いるデータ $\mathcal{D}$ は，2016 年 11 月以降の 200 日間における新規商品の販売実績である．また，オフライン環境のモデル学習フェーズで用いるデータ $\mathcal{D}'$ は，それ以前に発売された過去商品の販売実績である．オンライン環境での新規商品の在庫管理に関する設定を，表 4.4 に示す．新規商品が発売された $t = 0$ から状態を観測して，期間 $T = 200$ 日目までの各日の発注量をエージェントが決定する．この発注量と，対象店舗における新規商品の日々の販売実績のデータセット $\mathcal{D}$ から，環境側は次の状態 $S_{t+1} = s'$ を返すシミュレーションを行う．

対象商品 $i \in \mathcal{I}, |\mathcal{I}| = 4$ は，それぞれ異なる 4 社のサプライヤーから発売されたスマートフォンの新規商品である．また，対象店舗 $j \in \mathcal{J}, |\mathcal{J}| = 10$ は，日本に拠点をもつ各地の販売店である．シミュレーション上の設定値として，各店舗 j で各商品 i を確保できる在庫量の最大値 M は 19 とする．

表 4.4: 在庫管理シミュレーションの設定値

パラメータ	設定値
在庫管理期間 T	200
対象店舗数 $\|\mathcal{J}\|$	10
対象商品数 $\|\mathcal{I}\|$	4
サプライヤー数	4
在庫量の初期値 q_0	19
在庫量の最大値 M	19

表 4.5: 各製品の販売数の実績値

商品	サプライヤー	販売量の構成比
A	SA	9.4%
B	SB	22.7%
C	SC	54.7%
D	SD	13.2%

各商品 $i \in \mathcal{I}$ について，在庫管理の対象期間での全店舗 $j \in \mathcal{J}$ での販売実績値の構成比を表 4.5 に示す．製品 C が最も需要の高い製品であり，商品によって販売需要が大きく異なっていることが分かる．

次に，在庫管理シミュレーションとは別の，環境側の需要予測モデルの学習用データセット $\mathcal{D}'$ について述べる．表 4.6 に示す通り，学習期間 N は在庫管理期間 T と同じ 200 日間としている．また，過去の商品 $i' \in \mathcal{I}'$ は，$\mathcal{I}$ に含まれる 4 製品のサプライヤー (A, B, C, D) のうち，3 社 (A, B, C) から出された各 2 商品を選択し，合計で 6 商品を扱っている．すなわち，在庫管理シミュレーションでは，過去商品でのデータセット $\mathcal{D}'$ に含まれていなかった，全く新しいサプライヤーである，サプライヤー D の商品の需要予測と在庫管理を実施することになる．また，予測モデルで用いる特徴量 $X_{i,j}^{(t)}$ については，表 4.2 に示したカテゴリカルデータは One-hot encoding を通じてダミー変数化している．これによる特徴量の次元数は，合計で 196 となっている．

表 4.6: 需要予測モデルの設定値

パラメータ	設定値
データ学習期間 N	200
対象商品数 $\|\mathcal{I}'\|$	6
サプライヤー数	3
特徴量 $X_{i,j}^{(t)}$ の次元数	196
上側信頼限界のパラメータ α	3
プランニング時の先読みのステップ数 h	3
プランニング時の行動系列の生成数 K	10,000

比較手法

在庫管理シミュレーションで実施する，提案手法と比較するためのヒューリスティック手法，及びモデルフリー型の深層強化学習の各手法について説明する．

Heuristic

シンプルなルールベースとして，商品 i の店舗 j における販売実績の過去 28 日間の移動平均値 $\tilde{d}$ と安全係数 β により，推奨在庫量を $\beta \times \tilde{d}$ として発注量を決定する．ここでは，$\beta = 6$ として大きな安全係数に設定し，欠品発生を回避する消極的な戦略としている．

Model-free DRL

方策ベースの手法である TRPO [68] をデータセット $\mathcal{D}'$ を通じて方策を学習し，新規商品 $i \in \mathcal{I}$ の在庫管理に適用する．

Model-based DRL (提案手法)

データセット $\mathcal{D}'$ を通じて予測モデルを Bayesian neural network (BNN) または MAML で学習する．新規商品 $i \in \mathcal{I}$ の在庫管理では, Random shooting (RS) によって発注量を決定する．

Oracle

新規商品 $i \in \mathcal{I}$ の実際の販売需要 d を既知として扱い，最小の在庫量を

保ちつつ，販売量を最大にするよう完全な在庫管理を行う．在庫管理のパフォーマンスにおける最大値として扱う．

4.5.2　在庫管理シミュレーションの結果

多商品・多店舗の新規商品の在庫管理に対する提案アルゴリズムの有効性を検証するため，以下の観点からシミュレーション結果を確認する．

(1) 全体結果　総報酬や在庫回転率，欠品発生率を評価指標として，各アルゴリズムの性能を比較する．

(2) 商品別の在庫管理比較　商品別の在庫量がどのように保持されているかを，各アルゴリズムで比較する．

(3) 商品・店舗別の在庫管理比較　商品・店舗別の在庫量がどのように保持されているかを，各アルゴリズムで比較する．

(4) 需要予測モデルの性能検証　それぞれの需要予測モデルでの予測精度や予測傾向を比較する．

(5) 計算コストの評価　需要予測モデルの学習と発注量決定のプランニングに要する計算時間を評価する．

全体結果

アルゴリズム別に，4 商品，10 店舗における新規商品の在庫管理シミュレーションを，10 回試行した際の計 400 エピソードにおける，総報酬，在庫回転率，欠品発生率の各々の中央値を，表 4.7 に示す．ここで，総報酬については全エピソードのなかで正規化 (Min-Max normalization) を行っている．提案手法である，Random shooting と MAML を用いたモデルベース深層強化学習での RS_MAML が，在庫回転率の指標が 14.367 と最も高く，Heuristic での 13.605 から 5%以上の改善ができていた．加えて，RS_MAML での欠品発生率も 0.5%と，欠品発生を回避する消極的な方策である Heuristic と同等に抑えることができていた．また，モデルを BNN にした RS_BNN の場合も同様の傾向であり，特に総報酬では 0.733 と最も高い結果であり，Heuristic の 0.670 から 9%以上の改善ができて

表 4.7: 在庫運用シミュレーション結果．4 商品 10 店舗での在庫管理シミュレーションを 10 回繰り返した場合の各アルゴリズムの指標結果 (各アルゴリズムの総報酬，在庫回転率，欠品発生率の計 400 エピソードの中央値)

アルゴリズム	手法		総報酬	在庫回転率	欠品発生率
	方策	モデル			
Heuristic	Rule	MA	0.670	13.605	**0.5%**
Model-free DRL	TRPO	-	0.640	5.102	25.0%
Model-based DRL	RS	BNN	**0.733**	14.003	1.0%
	RS	MAML	0.710	**14.367**	**0.5%**
Oracle	-	-	0.821	16.276	0%

いた．一方で，方策ベース型の TRPO では在庫回転率が 5.102 と悪く，欠品発生率も 25.0%と非常に高くなっていることが分かる．この結果から，提案手法である RS_MAML や RS_BNN により，新規商品の発売以後の需要を満たしながら発注量決定を行い，より良い在庫管理が実現できていると言える．

商品別の在庫管理比較

この多商品・多店舗での在庫管理シミュレーションについて，4 つの新規商品 (A,B,C,D) 別に時刻 t ごとの全店舗での平均在庫量の時系列推移を図 4.6 に示す．表 4.5 で示した様に，最も販売需要が高いは製品 C であり，この在庫管理において，Heuristic ではおよそ 10 台程度といった，過剰な在庫量を常に配備していることが分かる．一方で，提案手法の RS_MAML や RS_BNN による製品 C の在庫管理では，その半分のおよそ 5 台程度といった，より少ない在庫量で保持できていることが分かる．このように，RS_MAML や RS_BNN は，新規商品ごとの需要を満たしながら供給量を適切に制御していることで，表 4.7 に示した通り，欠品発生率を悪化させることなく総報酬と在庫回転率を同時に改善できていると言える．一方で，方策ベース型の TRPO は，在庫量が 0 に近づいた段階で，発注量を一度に多く取って在庫量を一定期間，確保する戦略を取っている．

次に，各手法における 4 つの新規商品 (A,B,C,D) ごとの全期間 T での平均在庫量の分布を図 4.7 に示す．Heuristic では，先に述べた通り，高需要である製品

C の在庫量が非常に多く，逆に，TRPO では製品 C の在庫量は在庫量が著しく少なくなっている．これに対して，RS_MAML や RS_BNN では，全ての商品でおよそ同程度の在庫量を維持していることが分かる．このことから，提案手法では，各商品の異なる販売需要を考慮した上で，適切な供給決定を行い安定した在庫管理ができていると言える．

商品・店舗別の在庫管理比較

各商品と各店舗の組み合わせごとでの．全期間 T における平均在庫量を図 4.8 に示す．各グラフについて，横軸は各店舗の ID, 縦軸が在庫量を示している．Heuristic や TRPO では，商品だけでなく店舗との組み合わせにおいても，在庫量のバラつきが大きいことが分かる．例えば，Heuristic での製品 C の在庫管理では，店舗によって在庫量が大きく異なっている．これに対して，RS_MAML や RS_BNN では，商品間や店舗間においても在庫量をおよそ一定に保つことができている．この結果から，提案手法によって，商品や店舗での異なる販売需要を吸収して，適切な商品供給ができていることが言える．

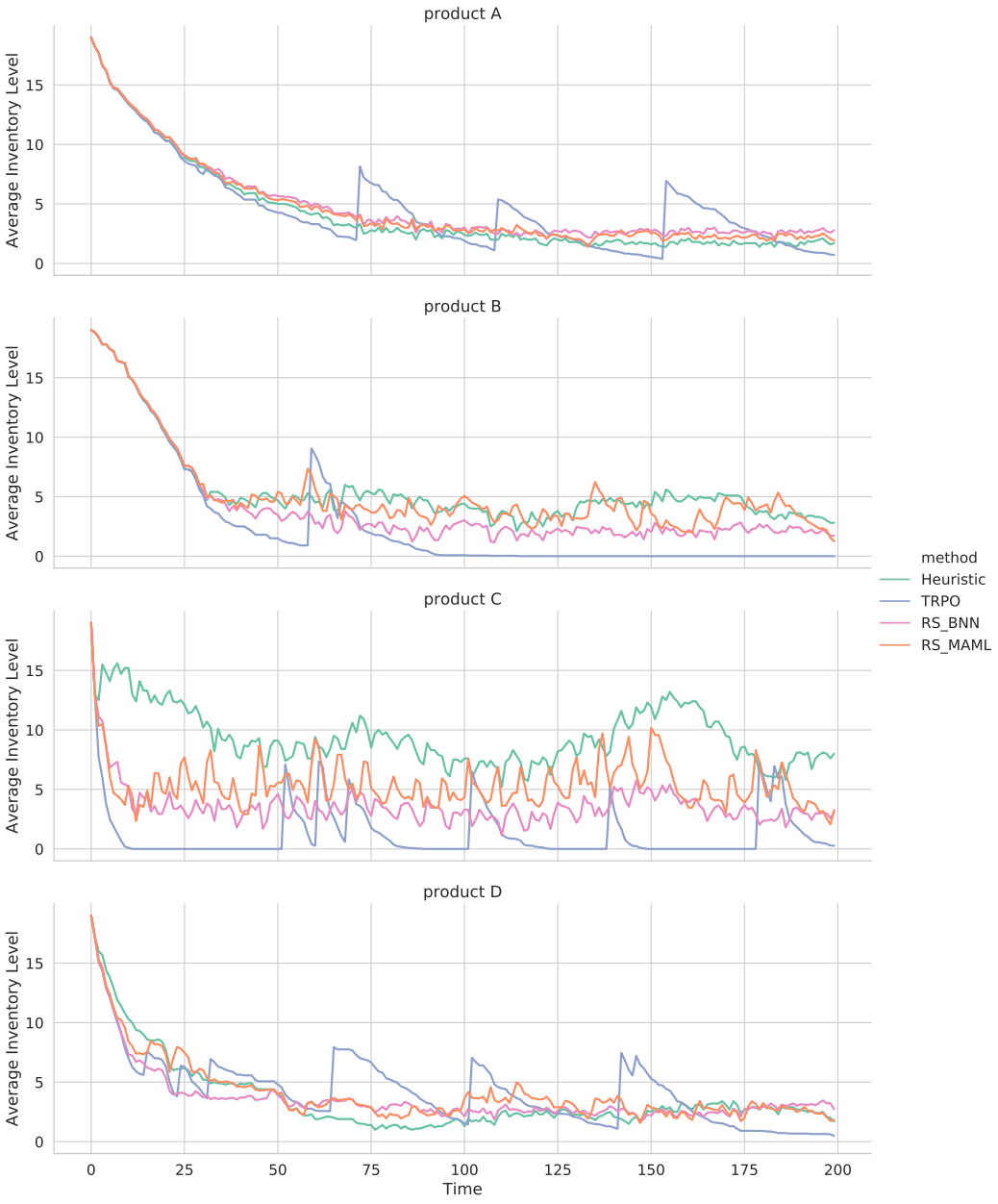


図 4.6: 商品別の平均在庫量の時系列推移．4 つの新規商品 (A,B,C,D) 別に平均在庫量の時系列推移を手法ごとで示している．

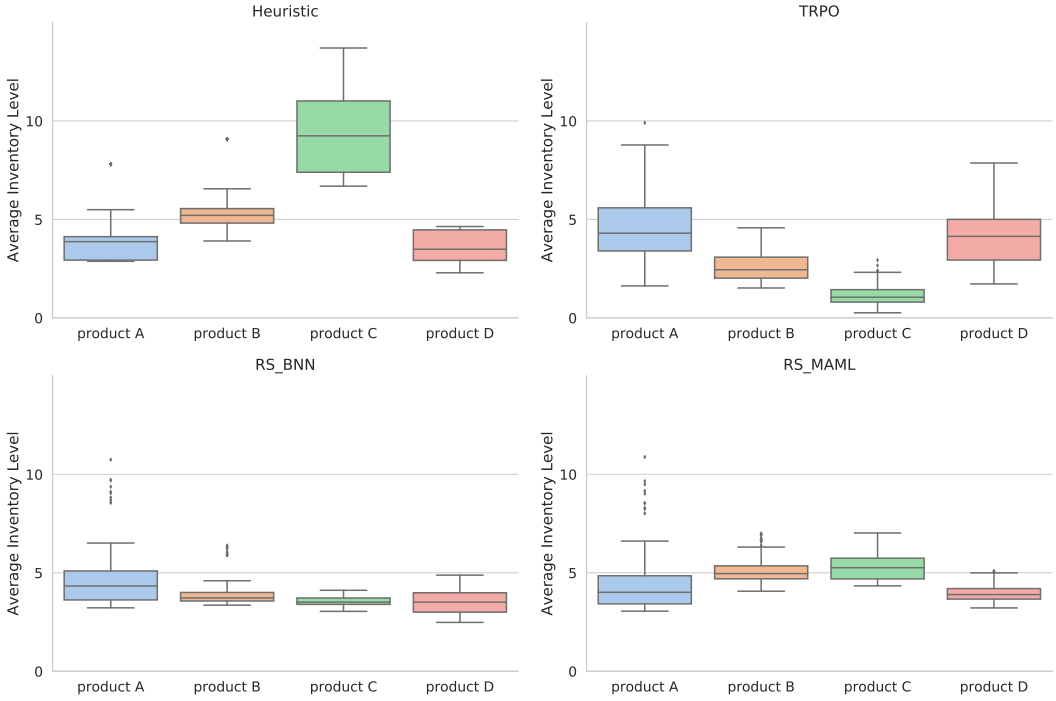


図 4.7: 各商品の平均在庫量分布．各手法における全期間 T での 4 つの新規商品 (A,B,C,D) それぞれの平均在庫量の分布であり，提案手法の RS_MAML や RS_BNN は，全ての商品でおよそ同程度の在庫量を維持できている．

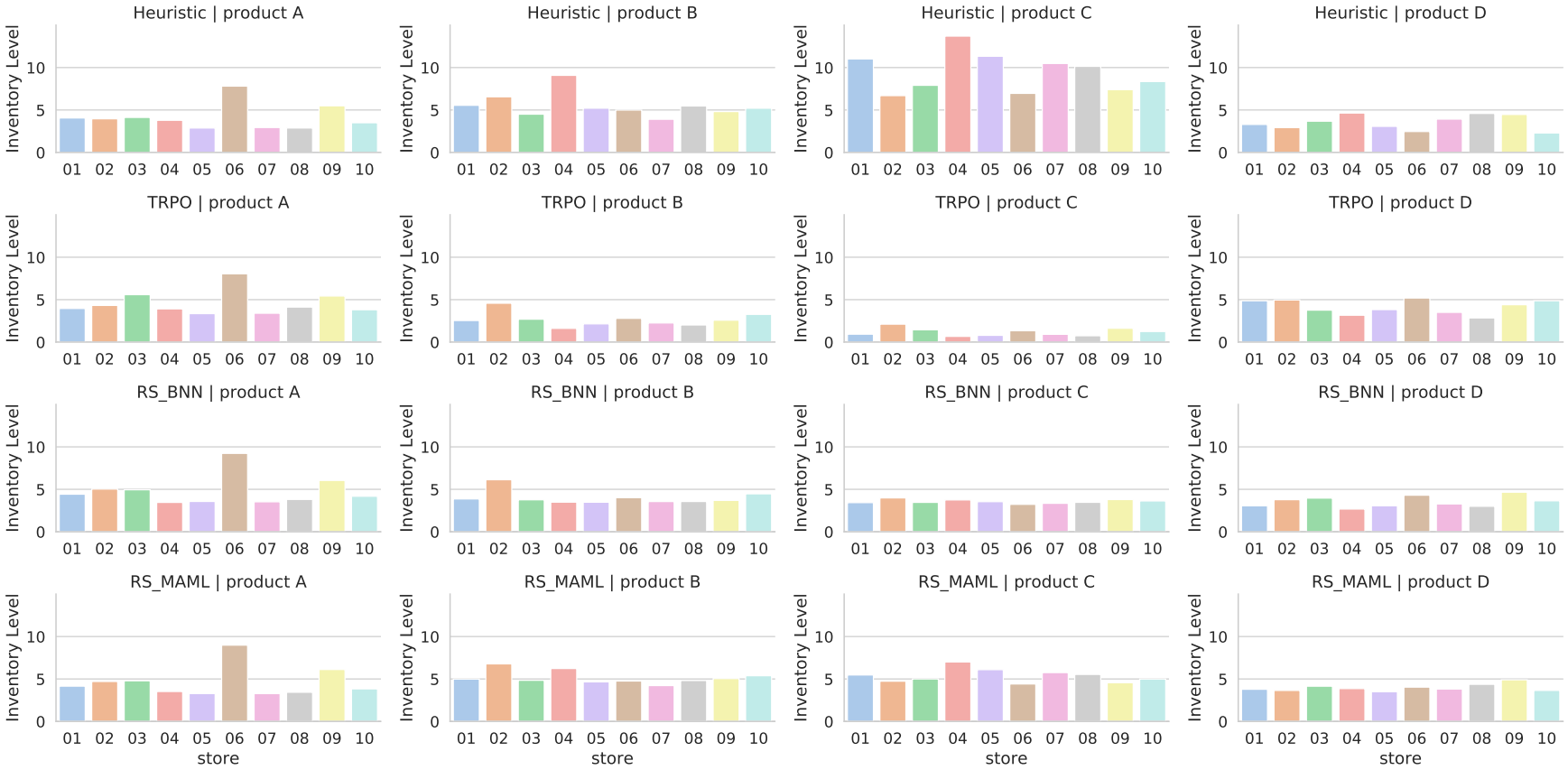


図 4.8: 各商品の店舗別の在庫量分布．各手法における全期間 T での商品と店舗の組み合わせごとの平均在庫量であり，提案手法の RS_MAML や RS_BNN は，商品や店舗に寄らずおよそ同程度の在庫量を維持できている．

表 4.8: 需要予測モデルの精度結果

Model	MAE	RMSE	R^2
MA	0.666	1.001	0.240
BNN	**0.625**	0.969	0.287
MAML	0.636	**0.956**	**0.306**

需要予測モデルの性能検証

提案手法の一部である，需要予測モデルの予測性能結果についても述べる．過去の販売実績であるデータセット $\mathcal{D}'$ によって学習した，BNN モデルと MAML による深層学習モデルについて，予測性能を表 4.8 に示す．各新規商品の各店舗での販売量について，翌日の需要予測値 $\hat{d}$ と実績値 d における平均絶対誤差 (MAE)，二乗平均平方根誤差 (RMSE)，R^2 スコアで比較している．また，比較手法として，Heuristic で用いた，過去 28 日までの販売数の移動平均値で翌日の販売需要を予測した MA を用いている．MA による移動平均値での予測では MAE が 0.666 であることに対して，BNN では 0.625 にまで低減できている．MAML では，RMSE，及び R^2 スコアが最も良い予測となっていて，特に，R^2 スコアでは，MA が 0.240 に対して MAML は 0.306 と大きく改善できている．

次に，商品ごとの時刻 t における需要予測値と実績値の時系列推移を図 4.9 に示す．ここで，横軸は時間ステップ，縦軸は全店，全商品，全期間の総販売量に対する各商品の販売比率として正規化した値を示す．スマートフォンの 4 つの新規商品 (A,B,C,D) の販売実績 (Ground Truth) の時系列パターンは，大きく異なっている．最も需要の高いスマートフォンである製品 C の需要は，$t = 0$ の発売直後から高く，時間が経過しても大きな減少は見られない．一方，製品 D は，発売直後は需要が高いが，その後は急速に減少している．逆に，製品 B の需要は，発売直後は低いが，時間の経過とともに徐々に増加している．

こうした実績に対して，MA による予測では，なだらかな値として傾向は捉えられているものの，急な需要変動は捉え切れていない．一方で，MAML では，需要の急峻なピークを捉えたきめ細やかな予測になっている．これに対して，BNN では，MA と MAML の予測の中間という形で，MA に比べて需要傾向を捕捉しているが，MAML に比べて急峻な需要は予測しきれていない．この予測傾向を

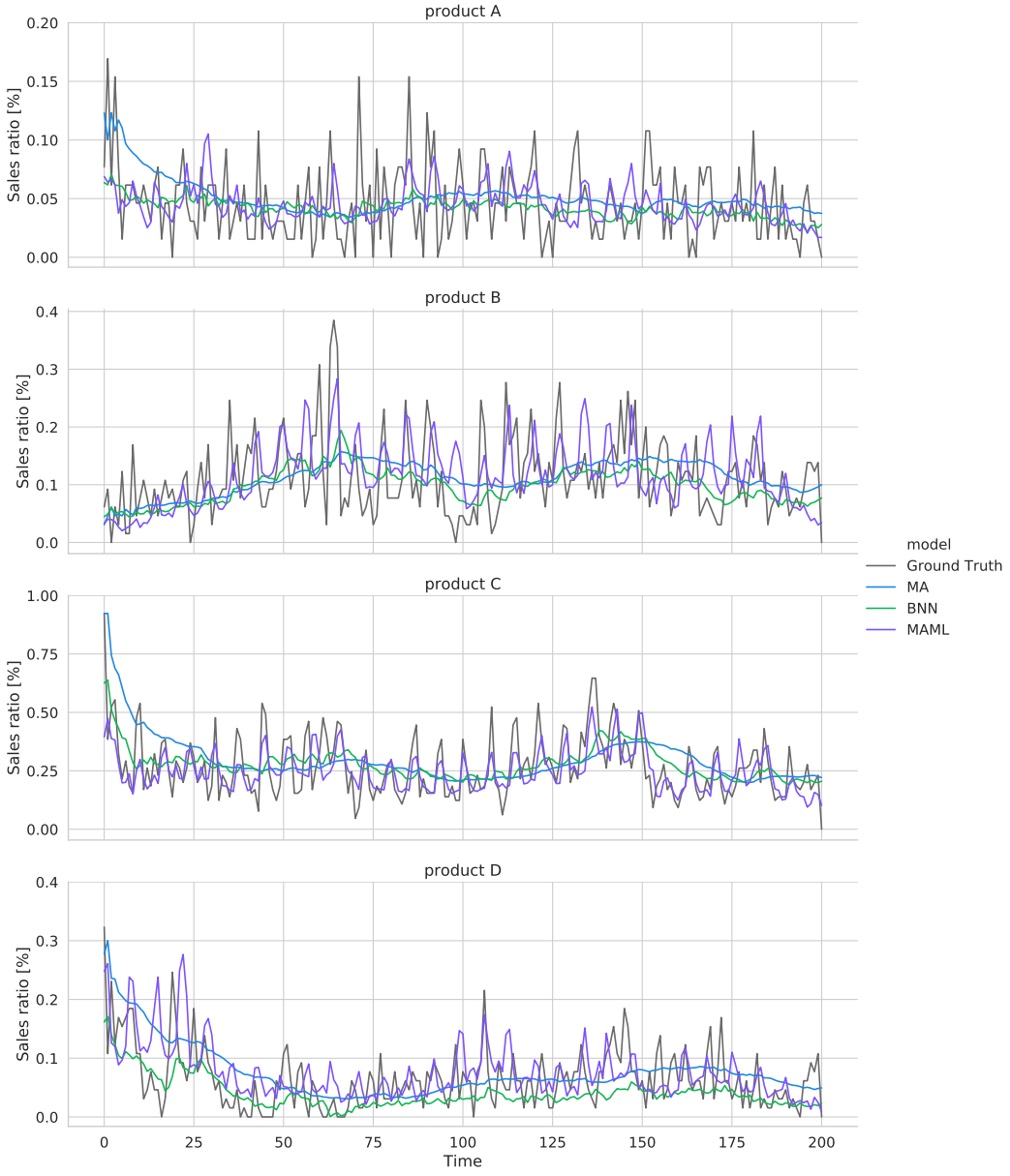


図 4.9: 需要予測値と実績値の時系列推移．横軸は時間ステップ，縦軸は全店，全商品，全期間の総販売量に対する各商品の販売比率として正規化した値である．

受けて，表 4.7 に示した通り，RS_MAML に比べて RS_BNN の欠品発生率が僅かに増加したと考えられる．

計算コストの評価

表 4.3 での計算機環境，及び 4.5.1 項で示したデータセットとパラメータの下での，計算コストについて説明する．提案手法では，オフライン環境での需要予

測モデルの学習とオンライン環境でのプランニングとを組み合わせているため，それぞれで計算時間を測定している．

まず，需要予測モデルの学習時間について，BNN の場合は 46 分，MAML の場合は 92 分である．MAML の学習では，勾配の勾配を求める操作があるため，Hessian での計算コストが大きいという特徴がある [50]．

次に，RS_MAML による 4 商品，10 店舗における 200 日間の在庫管理を 10 回試行した，計 400 エピソードのシミュレーションに要する全体の計算時間は 18 時間である．特に，Random shooting による 1 回の発注量決定のプランニング（商品，店舗，日別の粒度）に要する時間は平均 0.82 秒である．

4.5.3 考察

スマートフォンの新規商品を例に，4 商品，10 店舗での発売後 200 日間の在庫管理シミュレーションを通じて，提案手法の有効性を確認した．特に，店舗・商品ごとの日々の需要を学習済みモデルで推定した上で，適切な供給量をプランニングで決定することで，収益性観点で総報酬，効率性観点で在庫回転率，顧客満足度観点で欠品発生率の全ての指標において，高い在庫管理性能が得られた．これらの結果から，提案手法の適用は，製品ライフサイクルの短い商品の在庫管理が必要なケースで，有効であると言える．

本手法での需要予測モデルは，過去商品での販売実績のみで学習しており，新規商品での販売結果を用いた予測モデルの再学習は行なっていない．これは，実際の在庫管理システムへの適用を想定し，学習済みモデルをオンライン環境で利用するシンプルなアプローチとして検討したためである．ただし，オンライン環境のフェーズで予測モデルの再学習を組み入れることで，予測精度の向上と，それによる在庫管理の性能向上も期待できる．特に，MAML をオンライン環境で再学習する際には，二階の勾配を無視した一次近似によって計算コストを抑えた First-Order MAML (FOMAML) [50] の利用が考えられる．

他の限界として，適用可能な商品範囲が挙げられる．今回は製品ライフサイクルの短い例としてスマートフォンを取り上げて，新規商品として 4 つの在庫管理で提案アルゴリズムの性能を評価した．一方で，一般的な小売店で扱うような，数万から数十万の商品への適用を想定した場合には，データのスパース性を

考慮する必要がある．商品によっては，販売数が極端に少ないものも想定されるために，商品についてのセグメンテーションや，階層型の予測が必要になると考えられる．

4.6 まとめ

本章では，新規商品における多商品かつ多店舗の発売直後からの在庫管理を最適化する，モデルベース深層強化学習による手法を提案した．モデルフリー型の深層強化学習では，大量のデータや試行回数を必要とするため，新規商品を扱う在庫管理には不向きと考えられる．そのため，本研究では学習効率性の高いモデルベース深層強化学習に着目し，オフライン環境でのモデル学習と，オンライン環境での Random shooting によるプランニングを組み合わせた手法を提案した．過去の販売実績から学習した需要予測モデルを，実際の環境を模倣したシミュレーターとして用いて，エージェントによる在庫管理を行うこととしている．商品需要の不確実性に対処するために，深層学習モデルとして，確率的な予測が可能な BNN と，メタ学習の代表的手法である MAML を採用した．

製品ライフサイクルが短いサプライチェーンの例として，スマートフォンの在庫管理を例に挙げて，実際の販売実績を使った在庫管理シミュレーションを行い，提案手法に対する有効性を示した．特に，提案手法では，収益性観点で総報酬，効率性観点で在庫回転率，顧客満足度観点で欠品発生率の全ての指標で，最も良い性能を得た．また，各商品，各店舗において保持在庫量を，およそ一定に保つことができていることも確認した．このことから，商品や店舗での異なる販売需要を吸収して，適切な商品供給ができていることを示した．このように，発売当日から直ぐに適用できる在庫管理手法は，製品ライフサイクルが短い商品を扱うサプライチェーンでは，特に重要であると言える．今後の展望は，オンライン環境での需要予測モデルの再学習による更なる高精度化や，商品セグメンテーションや階層的予測を用いたデータのスパース性への対応による，適用可能な商品数の拡大が考えられる．

第 5 章

結論

5.1 本研究のまとめ

本論文では，新規アイテムとしてオンライン領域での Web 広告と，オフライン領域でのサプライチェーン上の小売商品に着目し，それらを市場投入した以降の需要の時系列推移を予測し，意思決定を最適化する手法を確立した．また，それぞれの領域でのデータセットを基にした評価実験により，それらの有効性を示した．

第 3 章では，深層学習による新規 Web 広告の時間減衰を考慮した CTR 予測の手法を提案した．Web 広告のうち，特に，インフィード広告はその視認性の高さから，CTR の時間的な減衰が速いという特徴があった．その CTR の時間減衰を，インフィード広告に紐付くマルチモーダルな特徴量から予測することで，広告選択や広告掲載の期間決定といった配信計画へ活用することを目的とした．インプレッション単位のクリック有無でなく，広告単位の CTR をロバストに回帰するため，深層学習を用いた従来の CTR 予測モデルに対して，Dropout 及び L2 正則化を導入するネットワーク構造の改良を行った．また，CTR の時間減衰を推定するために，多期間にわたる CTR の時系列変化を抽象的に表現可能な，RNN 型のネットワーク構造を提案した．過去の配信実績を用いた評価実験では，多期間の CTR を独立したものとみなし同時に予測するベースライン手法との比較を行い，提案手法による精度向上を確認した．

第 4 章では，モデルベース深層強化学習による新規小売商品の在庫管理手法を

提案した．製品ライフサイクルが短いサプライチェーン上の小売商品の一つであるスマートフォンを対象とし，新規商品が発売開始されて直ぐに適用ができ，欠品発生率を抑えつつ，利益全体の最大化と余剰在庫数の最小化が可能な在庫運用の実現を目的とした．強化学習を用いた在庫管理手法は従来から提案されているが，定常的に販売されている商品を扱うものが中心であるため，新規商品が発売した当初から適用することは困難であった．この課題を解決するために，学習効率性の高いモデルベース深層強化学習を取り入れ，オフライン環境での過去商品の販売実績を用いた需要予測モデルの学習と，オンライン環境での新規商品の発注プランニングを組み合わせた手法を提案した．需要予測モデルでは，販売需要の不確実性を考慮するため，MAML と BNN の 2 種類の深層学習モデルを用いた．この需要予測モデルを，実際の環境を模倣したシミュレーターとして用い，Random shooting を通じた発注量決定の方式を提案した．実際の販売実績を用いた在庫管理シミュレーションを行い，提案手法により総報酬，在庫回転率，欠品発生率の全ての指標において，従来手法からの改善を確認した．在庫管理シミュレーションでは，異なる需要傾向を持つ複数の新規商品，及び複数の販売店で評価を行い，それぞれで在庫量をおよそ一定に保つことができ，安定した在庫管理ができていることも確認した．

5.2　今後の研究課題

本論文の第 3 章で述べた，時間減衰を考慮した Web 広告の CTR 予測では，インフィード広告のみを対象にしている．しかし，Web 広告市場において近年では動画広告も成長しており，2023 年には 7,209 億円，2026 年には 1 兆 2,451 億円に達すると推計されている [77]．こうした動画広告における CTR 予測や，掲載時期や掲載メディアに応じた CTR 変動を予測することへの拡張も考えられる．また，予測された CTR 変動をパラメータとして，最適な配信計画を決定する離散最適化問題の定式化に活用することも考えられる．

また，本論文の第 4 章で述べた，新規商品の需要予測を通じた在庫管理では，スマートフォンを対象にしている．販売実績を用いた評価実験では，4 商品・10 店舗での在庫管理シミュレーションにて提案手法の有効性を示した．一方で，一般的な小売店で扱うような，数万から数十万の商品への適用を想定した場合は

未検証である．商品によっては，販売数が極端に少ないものも想定されるため，データのスパース性に対する改善が必要と考えられる．また，オフライン環境で学習済みの需要予測モデルを，オンライン環境で再学習して適用することで更なる性能向上が期待できる．ただし，オンライン環境で再学習する場合は，計算時間が在庫管理上のボトルネックにならないようにする必要がある．そのため，計算コストを抑えたモデルの利用が有効と考えられる．また，本研究におけるオフライン環境でのモデル学習では，新規商品と類似の過去商品のデータセットを用いていた．しかし，過去に類似のものがないような全く新しいコンセプトの商品を在庫管理の対象としたい場合も考えられる．その際に，無関係の商品のデータセットをオフライン環境でのモデル学習に用いた時に，新規商品の需要予測精度や在庫管理性能の影響を確認する必要がある．

本論文では，オンライン領域の Web 広告とオフライン領域のサプライチェーン上の小売商品を対象に，それらの新規アイテムの需要予測と意思決定の最適化手法を提案した．今後は，更に他の分野への応用が考えられる．例えば．金融分野であれば，新規の金融商品を市場に投入する際の，需要予測やリスク評価などが挙げられる．他にも，エネルギー分野であれば，新電力サービスを開始する際の，他電力会社からのスイッチング需要の推定や，電力調達を含めた最適なエネルギーマネジメントへの拡張も挙げられる．

謝辞

本研究は，著者が大阪大学大学院情報科学研究科 博士後期課程において，同大学 情報数理学専攻 森田浩教授のご指導のもとに行ったものです．研究を行うにあたって，非常に多くの方々のご指導とご支援をいただきましたことを，心より感謝申し上げます．

森田教授には，研究全体を通して多大なるご助言を賜りました．また，筆者の学生時代から研究の基礎をご指導いただき，技術研究の道にも導いて下さいました．これまでご指導とご鞭撻をいただきましたことを謹んで御礼申し上げます．

本博士論文の審査過程において，ご指導とご助言をいただきました大阪大学大学院情報科学研究科 情報数理学専攻 谷田純教授，沼尾正行教授，山口勇太郎准教授に心より感謝申し上げます．

社会人ドクターとして入学する機会をいただいた株式会社 NTT ドコモ 総務人事部担当部長 津田雅之博士 (前クロステック開発部長)，ドコモ・テクノロジ株式会社 スマートソリューション開発部担当部長 木本勝敏氏 (前 NTT ドコモ サービスイノベーション部担当部長)，上智大学 応用データサイエンス学位プログラム 深澤佑介准教授 (前 NTT ドコモ データプラットフォーム部担当部長) に深く御礼申し上げます．特に，深澤准教授には当時の上長として，また企業研究者の先輩として，研究の進め方や実用化，論文執筆のあらゆる面で懇切丁寧なご指導と数々のご助言を賜り心より感謝の意を表します．

また，社会人ドクターとしての活動もご理解いただき，業務遂行にあたっても多大なるご支援をいただきました NTT ドコモ 常務執行役員 R&D イノベーション本部長 佐藤隆明氏 (前 NTT ドコモ 北陸支社長)，宮城大学事業構想学群 太田賢教授 (前 NTT ドコモ サービスイノベーション部長)，NTT ドコモ 移動機開発部担当部長 青木秀憲氏 (前サービスイノベーション部担当部長)，サービ

スイノベーション部担当部長 大西純氏，担当課長 南部洋平氏に深く感謝いたします．

また，数理最適化に関する多くのご助言やご討論だけなく，学会運営を通じても多大なるご支援をいただきました大阪大学大学院情報科学研究科 数理最適化寄附講座 梅谷俊治教授に深く御礼申し上げます．

また，社会人ドクターの先輩として，研究の進め方や論文執筆の点で多くのご助言をいただきましたNTT ドコモ クロステック開発部主査 落合桂一博士に御礼申し上げます．

また，本研究を進めるにあたり，多くのご討論やご助言をいただきましたNTT ドコモ サービスイノベーション部，及び大阪大学大学院情報科学研究科 情報数理学専攻森田研究室の皆様に深く感謝いたします．

最後に，これまで支え続けてくれた家族に感謝します．研究開発や論文執筆を進める上で家事と育児に多く時間が取れないなかでも，いつも応援してくれていた妻治菜と，笑顔で励ましてくれた息子碧，燈に心から感謝します．

参考文献

[1] 大野勝久. サプライチェーンの最適運用: かんばん方式を超えて. 朝倉書店, 2011.

[2] Junxuan Chen, Baigui Sun, Hao Li, Hongtao Lu, and Xian-Sheng Hua. Deep ctr prediction in display advertising. In *Proceedings of the 24th ACM international conference on Multimedia*, pp. 811–820, 2016.

[3] Kamelia Aryafar, Devin Guillory, and Liangjie Hong. An ensemble-based approach to click-through rate prediction for promoted listings at etsy. In *Proceedings of the ADKDD'17*, pp. 1–6. 2017.

[4] Yuyu Zhang, Hanjun Dai, Chang Xu, Jun Feng, Taifeng Wang, Jiang Bian, Bin Wang, and Tie-Yan Liu. Sequential click prediction for sponsored search with recurrent neural networks. In *Twenty-Eighth AAAI Conference on Artificial Intelligence*, 2014.

[5] Yue Deng, Yilin Shen, Hongxia Jin, et al. Disguise adversarial networks for click-through rate prediction. In *IJCAI*, pp. 1589–1595, 2017.

[6] 岩崎祐貴. 深層学習による Facebook 広告の CTR 予測. 人工知能学会全国大会論文集 第 32 回全国大会 (2018). 一般社団法人人工知能学会, 2018.

[7] Kyung-Wha Park, JungHoon Lee, Sunyoung Kwon, Jung-Woo Ha, Kyung-Min Kim, and Byoung-Tak Zhang. Which ads to show? advertisement image assessment with auxiliary information via multi-step modality fusion. *arXiv preprint arXiv:1910.02358*, 2019.

[8] R. H. Wilson. *A scientific routine for stock control*. Harvard Univ., 1934.

[9] Kenneth J Arrow, Theodore Harris, and Jacob Marschak. Optimal inventory policy. *Econometrica: Journal of the Econometric Society*, pp. 250–272, 1951.

[10] G. (George) Hadley and T. M. (Thomson McLintock) Whitin. *Analysis of inventory systems*. Prentice-Hall quantitative methods series. Prentice Hall, 1963.

[11] Mary Dillon, Fabricio Oliveira, and Babak Abbasi. A two-stage stochastic programming model for inventory management in the blood supply chain. *International Journal of Production Economics*, Vol. 187, pp. 27–41, 2017. http://doi.org/10.1016/j.ijpe.2017.02.006.

[12] Benjamin Van Roy, Dimitri P Bertsekas, Yuchun Lee, and John N Tsitsiklis. A neuro-dynamic programming approach to retailer inventory management. In *Proceedings of the 36th IEEE Conference on Decision and Control*, Vol. 4, pp. 4052–4057. IEEE, 1997. http://doi.org/10.1109/CDC.1997.652501.

[13] Ilaria Giannoccaro and Pierpaolo Pontrandolfo. Inventory management in supply chains: a reinforcement learning approach. *International Journal of Production Economics*, Vol. 78, No. 2, pp. 153–161, 2002. http://doi.org/10.1016/S0925-5273(00)00156-0.

[14] Ahmet Kara and Ibrahim Dogan. Reinforcement learning approaches for specifying ordering policies of perishable inventory systems. *Expert Systems with Applications*, Vol. 91, pp. 150–158, 2018. http://dx.doi.org/10.1016/j.eswa.2017.08.046.

[15] Robert N Boute, Joren Gijsbrechts, Willem van Jaarsveld, and Nathalie Vanvuchelen. Deep reinforcement learning for inventory control: A roadmap. *European Journal of Operational Research*, 2021. http://doi.org/10.1016/j.ejor.2021.07.016.

[16] Hardik Meisheri, Vinita Baniwal, Nazneen N Sultana, Harshad Khadilkar, and Balaraman Ravindran. Using reinforcement learning for a large

variable-dimensional inventory management problem. In *Adaptive Learning Agents Workshop at AAMAS*, 2020.

[17] Joren Gijsbrechts, Robert N Boute, Jan A Van Mieghem, and Dennis Zhang. Can deep reinforcement learning improve inventory management? performance on dual sourcing, lost sales and multi-echelon problems. *Manufacturing & Service Operations Management*, 2021. http://doi.org/10.1287/msom.2021.1064.

[18] Ali Malik, Volodymyr Kuleshov, Jiaming Song, Danny Nemer, Harlan Seymour, and Stefano Ermon. Calibrated model-based deep reinforcement learning. In *International Conference on Machine Learning*, pp. 4314–4323. PMLR, 2019. https://doi.org/10.48550/arXiv.1906.08312.

[19] 髙橋勇人, 星野満博. 強化学習を用いた腐敗性を有する在庫問題の最適化について. 数理解析研究所講究録, No. 2214, pp. 27–38, 2022.

[20] Christopher John Cornish Hellaby Watkins. Learning from delayed rewards. 1989.

[21] Satinder P Singh and Richard S Sutton. Reinforcement learning with replacing eligibility traces. *Machine learning*, Vol. 22, No. 1, pp. 123–158, 1996. http://doi.org/10.1007/BF00114726.

[22] Vijay Konda and John Tsitsiklis. Actor-critic algorithms. *Advances in neural information processing systems*, Vol. 12, , 1999.

[23] Arash Tavakoli, Fabio Pardo, and Petar Kormushev. Action branching architectures for deep reinforcement learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 32, 2018. http://doi.org/10.1609/aaai.v32i1.11798.

[24] Peter Wanke, Henrique Ewbank, Víctor Leiva, and Fernando Rojas. Inventory management for new products with triangularly distributed demand and lead-time. *Computers & Operations Research*, Vol. 69, pp. 97–108, 2016. http://doi.org/10.1016/j.cor.2015.10.017.

[25] 株式会社サイバーエージェント. インフィード広告市場調査を実施. https://www.cyberagent.co.jp/news/detail/id=21333, 2018. 参照 2018-02-14.

[26] 田頭幸浩, 山本浩司, 小野真吾, 塚本浩司, 田島玲. オンライン広告における CTR 予測モデルの素性評価. 第 5 回データ工学と情報マネジメントに関するフォーラム (DEIM2013), 2013.

[27] Matthew Richardson, Ewa Dominowska, and Robert Ragno. Predicting clicks: estimating the click-through rate for new ads. In *Proceedings of the 16th international conference on World Wide Web*, pp. 521–530, 2007.

[28] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255. Ieee, 2009.

[29] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[30] Kilian Weinberger, Anirban Dasgupta, John Langford, Alex Smola, and Josh Attenberg. Feature hashing for large scale multitask learning. In *Proceedings of the 26th annual international conference on machine learning*, pp. 1113–1120, 2009.

[31] Tomáš Mikolov, Martin Karafiát, Lukáš Burget, Jan Černockỳ, and Sanjeev Khudanpur. Recurrent neural network based language model. In *Eleventh annual conference of the international speech communication association*, 2010.

[32] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pp. 2672–2680, 2014.

[33] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, Vol. 16, pp. 321–357, 2002.

[34] Lihui Shi and Bo Li. Predict the click-through rate and average cost per click for keywords using machine learning methodologies. In *Proceedings of the International Conference on Industrial Engineering and Operations ManagementDetroit, Michigan, USA*, 2016.

[35] 本橋永至, 磯崎直樹, 長尾大道, 樋口知之. 状態空間モデルによるインターネット広告のクリック率予測. オペレーションズ・リサーチ: 経営の科学, Vol. 57, No. 10, pp. 574–583, 2012.

[36] 坂田隼人, 栗田啓大, 山崎俊彦. Convolution neural network による広告画像効果の推定. 人工知能学会全国大会論文集 第 32 回全国大会 (2018). 一般社団法人人工知能学会, 2018.

[37] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[38] Deepak Agarwal, Bee-Chung Chen, and Pradheep Elango. Spatio-temporal models for estimating click-through rate. In *Proceedings of the 18th international conference on World wide web*, pp. 21–30, 2009.

[39] Fang Wu and Bernardo A Huberman. Novelty and collective attention. *Proceedings of the National Academy of Sciences*, Vol. 104, No. 45, pp. 17599–17601, 2007.

[40] 出水宰, 深澤佑介, 森田浩. 深層学習による時間減衰を考慮したインフィード広告の CTR 予測. 情報処理学会論文誌, Vol. 62, No. 1, pp. 292–301, 2021.

[41] 株式会社サイバーエージェント. 大量のクリエイティブを自動精査し、インフィード広告の効果最大化に貢献. https://www.cyberagent.co.jp/news/detail/id=12643, 2016. 参照 2016-10-04.

[42] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, Vol. 15, No. 1, pp. 1929–1958, 2014.

[43] 岡谷貴之. 深層学習 (MLP 機械学習プロフェッショナルシリーズ). 講談社, 2015.

[44] Masatoshi Suzuki, Koji Matsuda, Satoshi Sekine, Naoaki Okazaki, and Kentaro Inui. Neural joint learning for classifying wikipedia articles into fine-grained named entity types. In *Proceedings of the 30th Pacific Asia Conference on Language, Information and Computation: Posters*, pp. 535–544, 2016.

[45] Geoffrey E Hinton. Distributed representations. 1984.

[46] Oriol Vinyals, Alexander Toshev, Samy Bengio, and Dumitru Erhan. Show and tell: A neural image caption generator. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3156–3164, 2015.

[47] Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In *Advances in neural information processing systems*, pp. 4765–4774, 2017.

[48] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[49] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pp. 1050–1059. PMLR, 2016. https://doi.org/10.48550/arXiv.1506.02142.

[50] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *International conference on machine learning*, pp. 1126–1135. PMLR, 2017. https://doi.org/10.48550/arXiv.1703.03400.

[51] Sebastian Pineda Arango, Felix Heinrich, Kiran Madhusudhanan, and Lars Schmidt-Thieme. Multimodal meta-learning for time series regression. In *International Workshop on Advanced Analytics and Learning on Temporal Data*, pp. 123–138. Springer, 2021. http://doi.org/10.1007/978-3-030-91445-5_8.

[52] Arthur George Richards. *Robust constrained model predictive control*. PhD thesis, Massachusetts Institute of Technology, 2005.

[53] Anil V Rao. A survey of numerical methods for optimal control. *Advances in the Astronautical Sciences*, Vol. 135, No. 1, pp. 497–528, 2009.

[54] Richard S Sutton and Andrew G Barto. Reinforcement learning: An introduction. 2011. http://doi.org/10.1109/TNN.1998.712192.

[55] Jens Kober, J Andrew Bagnell, and Jan Peters. Reinforcement learning in robotics: A survey. *The International Journal of Robotics Research*, Vol. 32, No. 11, pp. 1238–1274, 2013. http://doi.org/10.1177/0278364913495721.

[56] Zhe Xu, Zhixin Li, Qingwen Guan, Dingshui Zhang, Qiang Li, Junxiao Nan, Chunyang Liu, Wei Bian, and Jieping Ye. Large-scale order dispatch in on-demand ride-hailing platforms: A learning and planning approach. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 905–913, 2018. http://doi.org/10.1145/3219819.3219824.

[57] Hideki Asoh, Masanori Shiro1 Shotaro Akaho, Toshihiro Kamishima, Koiti Hasida, Eiji Aramaki, and Takahide Kohro. An application of inverse reinforcement learning to medical records of diabetes treatment. In *ECMLP-KDD2013 workshop on reinforcement learning with generalized feedback*, 2013.

[58] Lu Wang, Wei Zhang, Xiaofeng He, and Hongyuan Zha. Supervised reinforcement learning with recurrent neural network for dynamic treatment

recommendation. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 2447–2456, 2018. http://doi.org/10.1145/3219819.3219961.

[59] Georgios Theocharous, Philip S Thomas, and Mohammad Ghavamzadeh. Personalized ad recommendation systems for life-time value optimization with guarantees. In *Twenty-Fourth International Joint Conference on Artificial Intelligence*, pp. 1806–1812, 2015.

[60] Igor Halperin. Inverse reinforcement learning for marketing. *arXiv preprint arXiv:1712.04612*, 2017. http://doi.org/10.2139/ssrn.3087057.

[61] Vijaya Yajnanarayana, Henrik Rydén, and László Hévizi. 5g handover using reinforcement learning. In *2020 IEEE 3rd 5G World Forum (5GWF)*, pp. 349–354. IEEE, 2020. http://doi.org/10.1109/5GWF49715.2020.9221072.

[62] Yuxi Li. Deep reinforcement learning: An overview. *arXiv preprint arXiv:1701.07274*, 2017. https://doi.org/10.48550/arXiv.1806.08894.

[63] Tingwu Wang, Xuchan Bao, Ignasi Clavera, Jerrick Hoang, Yeming Wen, Eric Langlois, Shunshi Zhang, Guodong Zhang, Pieter Abbeel, and Jimmy Ba. Benchmarking model-based reinforcement learning. *arXiv preprint arXiv:1907.02057*, 2019. https://doi.org/10.48550/arXiv.1907.02057.

[64] Thomas M Moerland, Joost Broekens, and Catholijn M Jonker. Model-based reinforcement learning: A survey. *arXiv preprint arXiv:2006.16712*, 2020. https://doi.org/10.48550/arXiv.2006.16712.

[65] Fan-Ming Luo, Tian Xu, Hang Lai, Xiong-Hui Chen, Weinan Zhang, and Yang Yu. A survey on model-based reinforcement learning. *arXiv preprint arXiv:2206.09328*, 2022. https://doi.org/10.48550/arXiv.2206.09328.

[66] Richard S Sutton. Dyna, an integrated architecture for learning, planning, and reacting. *ACM Sigart Bulletin*, Vol. 2, No. 4, pp. 160–163, 1991. http://doi.org/10.1145/122344.122377.

[67] Thanard Kurutach, Ignasi Clavera, Yan Duan, Aviv Tamar, and Pieter Abbeel. Model-ensemble trust-region policy optimization. *arXiv preprint arXiv:1802.10592*, 2018. https://doi.org/10.48550/arXiv.1802.10592.

[68] John Schulman, Sergey Levine, Pieter Abbeel, Michael Jordan, and Philipp Moritz. Trust region policy optimization. In *International conference on machine learning*, pp. 1889–1897. PMLR, 2015. https://doi.org/10.48550/arXiv.1502.05477.

[69] Eduardo F Camacho and Carlos Bordons Alba. *Model predictive control*. Springer science & business media, 2013.

[70] Volodymyr Mnih, Adria Puigdomenech Badia, Mehdi Mirza, Alex Graves, Timothy Lillicrap, Tim Harley, David Silver, and Koray Kavukcuoglu. Asynchronous methods for deep reinforcement learning. In *International conference on machine learning*, pp. 1928–1937. PMLR, 2016. https://doi.org/10.48550/arXiv.1602.01783.

[71] Fernando Rojas. A methodology for stochastic inventory modelling with arma triangular distribution for new products. *Cogent Business & Management*, Vol. 4, No. 1, p. 1270706, 2017. http://doi.org/10.1080/23311975.2016.1270706.

[72] Tsukasa Demizu, Yusuke Fukazawa, and Hiroshi Morita. Inventory management of new products in retailers using model-based deep reinforcement learning. *Expert Systems with Applications*, Vol. 229, p. 120256, 2023. https://doi.org/10.1016/j.eswa.2023.120256.

[73] Csaba Szepesvári. Algorithms for reinforcement learning. *Synthesis lectures on artificial intelligence and machine learning*, Vol. 4, No. 1, pp. 1–103, 2010.

[74] Ayad K Ali. Inventory management in pharmacy practice: a review of literature. *Archives of pharmacy practice*, Vol. 2, No. 4, p. 151, 2011.

[75] Unleashed. How to calculate average inventory - the complete guide. https://www.unleashedsoftware.com/blog/your-complete-guide-to-average-inventory, 2022. Accessed February 23, 2023.

[76] Amazon. What is inventory turnover: How to calculate and optimize the turnover rate for your business. https://sell.amazon.com.sg/blog/inventory-turnover, 2022. Accessed February 23, 2023.

[77] 株式会社サイバーエージェント. 2022 年国内動画広告の市場調査を実施. https://www.cyberagent.co.jp/news/detail/id=28533, 2023. 参照 2023-02-13.