

(19)日本国特許庁(JP)

(12) 公開特許公報(A)

(11)公開番号

特開2024-88222

(P2024-88222A)

(43)公開日 令和6年7月2日(2024.7.2)

(51) 国際特許分類

G O 6 Q 30/0242(2023.01)

FI

G 0 6 Q 30/0242

テーマコード（参考）

5 L 0 3 0

5 L 0 4 9

審査請求 未請求 請求項の数 11 O L (全17頁)

(21)出願番号 特願2022-203285(P2022-203285)

(22)出願日 令和4年12月20日(2022.12.20)

(71)出願人 399037405

楽天グループ株式会社

東京都世田谷区玉川一丁目14番1号

(74)代理人	100105957
---------	-----------

弁理士 恩田 誠

(74)代理人	100068755
---------	-----------

弁理士 恩田 博宣

(72)発明者 石川 詩苑

東京都世田谷区玉川一丁目14番1号

楽天グループ株式会社内

(72)發明者 劉 雲青

東京都世田谷区玉川一丁目14番1号

楽天グループ株式会社内

Fターム (参考)	5L030	BB08
-----------	-------	------

5L049 BB08

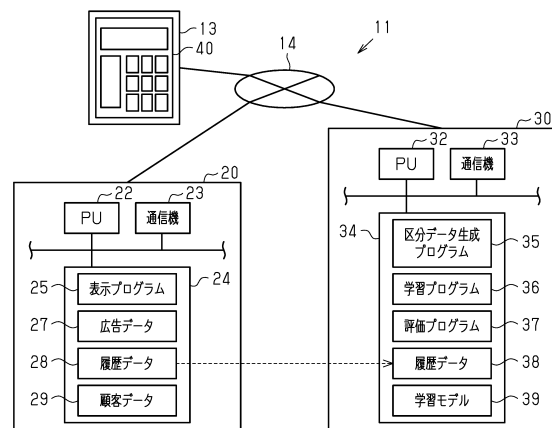
(54) 【発明の名称】 施策評価装置、施策評価方法、及び施策評価プログラム

(57) 【要約】

【課題】オフライン評価におけるバイアスに対処することができる施策評価装置、施策評価方法、及び施策評価プログラムを提供する。

【解決手段】施策評価装置 30 は、1 以上のプロセッサ 32 と 1 以上のメモリ 34 とを備える。メモリ 34 には、第 1 施策を実施したときの記録である複数のデータセットを含む履歴データ 38 が格納されている。プロセッサ 32 は、第 1 施策に関する少なくとも 1 つの要件が不変であることに基づいて複数のデータセットを区分することにより、複数の区分データを生成することと、機械学習のアルゴリズムを各区分データにより訓練することにより、第 2 施策に係る学習モデルを区分毎に生成することと、各区分データから近似した推定量を用いて、第 2 施策をオフライン評価することと、を実行するように構成される。

【選択図】 図1



【特許請求の範囲】

【請求項 1】

1 以上のプロセッサと 1 以上のメモリとを備え、

前記メモリには、第 1 施策を実施したときの記録である複数のデータセットを含む履歴データが格納されており、

前記プロセッサは、

前記第 1 施策に関する少なくとも 1 つの要件が不変であることに基づいて前記複数のデータセットを区分することにより、複数の区分データを生成することと、

機械学習のアルゴリズムを前記各区分データにより訓練することにより、第 2 施策に係る学習モデルを区分毎に生成することと、

前記各区分データから近似した推定量を用いて、前記第 2 施策をオフライン評価することと、

を実行するように構成される、施策評価装置。

【請求項 2】

前記アルゴリズムは多腕バンディットアルゴリズムであり、

前記各データセットは、特徴量と、前記特徴量に対応する行動と、前記行動の結果と、に関するデータを含む、

請求項 1 に記載の施策評価装置。

【請求項 3】

前記推定量の算出は、前記第 1 施策の結果を示す指標値を、前記第 1 施策と前記第 2 施策の行動選択確率の比を用いて重み付けすることを含む、

請求項 2 に記載の施策評価装置。

【請求項 4】

前記第 1 施策及び前記第 2 施策は、複数の候補画像のうちの 1 つを表示枠に表示することを含み、

前記複数の候補画像が前記各区分において不変である、

請求項 1 ～ 3 のうち何れか一項に記載の施策評価装置。

【請求項 5】

前記各データセットは、ユーザからのウェブサイトの表示要求の都度取得された複数のデータを含み、

前記ウェブサイトは 1 以上の広告枠を含み、

前記第 1 施策及び前記第 2 施策は、前記各広告枠に広告画像を表示することを含み、

前記特徴量が前記ユーザの属性であり、前記行動が各候補画像であり、前記行動の結果が表示された前記広告画像を前記ユーザがクリックしたか否かであり、

前記複数の候補画像が前記各区分において不変である、

請求項 2 又は 3 に記載の施策評価装置。

【請求項 6】

前記第 1 施策及び前記第 2 施策は、複数の対象候補から 1 以上の対象を選択することを含み、

前記複数の対象候補が前記各区分において不変である、

請求項 1 ～ 3 のうち何れか一項に記載の施策評価装置。

【請求項 7】

前記第 1 施策及び前記第 2 施策は、1 つの表示枠に順次に表示される複数の画像の表示順序を決定することを含み、

前記複数の画像が前記各区分において不変である、

請求項 1 ～ 3 のうち何れか一項に記載の施策評価装置。

【請求項 8】

前記第 1 施策及び前記第 2 施策は、複数の物品又は複数のサービスからユーザに推薦する 1 以上の物品又はサービスを選択することを含み、

前記各区分において、前記複数の物品又は前記複数のサービスの価格と、前記複数の物

10

20

30

40

50

品又は前記複数のサービスの品揃えと、のうち少なくとも一方が不変である、
請求項 1 ～ 3 のうち何れか一項に記載の施策評価装置。

【請求項 9】

前記各データセットは、検索画面におけるユーザからの検索要求の都度取得された複数のデータを含み、

前記第 1 施策及び前記第 2 施策は、検索結果画面に表示する複数の検索結果の表示順位を決定することを含み、

前記各区分において、複数の検索対象と前記複数の検索結果のうち少なくとも一方が不変である、

請求項 1 ～ 3 のうち何れか一項に記載の施策評価装置。

10

【請求項 10】

1 以上のコンピュータに、

第 1 施策を実施したときの記録である複数のデータセットを含む履歴データを取得することと、

前記第 1 施策に関する少なくとも 1 つの要件が不変であることに基づいて前記複数のデータセットを区分することにより、複数の区分データを生成することと、

機械学習のアルゴリズムを前記各区分データにより訓練することにより、第 2 施策に係る学習モデルを区分毎に生成することと、

前記各区分データから近似した推定量を用いて、前記第 2 施策をオフライン評価することと、

20

を実行させることを含む、施策評価方法。

【請求項 11】

1 以上のコンピュータに、

第 1 施策を実施したときの記録である複数のデータセットを含む履歴データを取得することと、

前記第 1 施策に関する少なくとも 1 つの要件が不変であることに基づいて前記複数のデータセットを区分することにより、複数の区分データを生成することと、

機械学習のアルゴリズムを前記各区分データにより訓練することにより、第 2 施策に係る学習モデルを区分毎に生成することと、

前記各区分データから近似した推定量を用いて、前記第 2 施策をオフライン評価することと、

30

を実行させるための、施策評価プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本開示は、施策評価装置、施策評価方法、及び施策評価プログラムに関する。

【背景技術】

【0002】

例えばウェブページのバナー広告など、1 つの広告枠に表示する広告の内容を決定する場合に、機械学習モデルを用いて広告効果を予測することがある。例えば、特許文献 1 は、広告デザイン案である画像を入力した場合に、広告効果の予測結果を出力する学習モデルを開示している。

40

【0003】

実際に広告を出すにあたっては、広告用画像を決定した後に、その画像をどのようなタイミングでどのようなユーザに表示するか、といった施策 (policy) を決定する必要がある。この施策を評価するために、例えばオンライン広告であれば、一時的にその広告を実装するオンライン実験を行うことがある。ただし、広告の実装にはコストがかかるだけでなく、施策が適切でなければ、ユーザの心証が悪くなったり、売上が低下したりするなどのリスクを伴う。

【0004】

50

これに対して、非特許文献 1 は、リスクの低い施策の評価手法である、オフライン評価について開示している。こうした施策の評価は、広告に限らず、各種のマーケティング又は顧客とのコミュニケーションを行う場合など、様々な分野において使用可能である。

【先行技術文献】

【特許文献】

【0005】

【特許文献 1】特開 2022-032204 号公報

【非特許文献】

【0006】

【非特許文献 1】齋藤優太，他 3 名，“Open Bandit Dataset and Pipeline: Towards Realistic and Reproducible Off-Policy Evaluation”，[online]，令和 3 年 10 月 26 日（v 5），arXiv Labs，[令和 4 年 12 月 16 日検索]，インターネット<URL：<https://arxiv.org/abs/2008.07146>>

【発明の概要】

【発明が解決しようとする課題】

【0007】

上述のオフライン評価では、過去に施策を実施したときのデータを用いて、過去の施策とは異なる新たな施策の性能の評価を行う。しかしながら、当該評価はバイアスの影響を受けることがある。

【0008】

本開示は、オフライン評価におけるバイアスに対処することができる施策評価装置、施策評価方法、及び施策評価プログラムを提供することを目的とする。

【課題を解決するための手段】

【0009】

本開示の一態様に係る施策評価装置は、1 以上のプロセッサと 1 以上のメモリとを備え、前記メモリには、第 1 施策を実施したときの記録である複数のデータセットを含む履歴データが格納されており、前記プロセッサは、前記第 1 施策に関する少なくとも 1 つの要件が不変であることに基づいて前記複数のデータセットを区分することにより、複数の区分データを生成することと、機械学習のアルゴリズムを前記各区分データにより訓練することにより、第 2 施策に係る学習モデルを区分毎に生成することと、前記各区分データから近似した推定量を用いて、前記第 2 施策をオフライン評価することと、を実行するように構成される。

【0010】

本開示の一態様に係る施策評価方法は、1 以上のコンピュータに、第 1 施策を実施したときの記録である複数のデータセットを含む履歴データを取得することと、前記第 1 施策に関する少なくとも 1 つの要件が不変であることに基づいて前記複数のデータセットを区分することにより、複数の区分データを生成することと、機械学習のアルゴリズムを前記各区分データにより訓練することにより、第 2 施策に係る学習モデルを区分毎に生成することと、前記各区分データから近似した推定量を用いて、前記第 2 施策をオフライン評価することと、を実行させることを含む。

【0011】

本開示の一態様に係る施策評価プログラムは、1 以上のコンピュータに、第 1 施策を実施したときの記録である複数のデータセットを含む履歴データを取得することと、前記第 1 施策に関する少なくとも 1 つの要件が不変であることに基づいて前記複数のデータセットを区分することにより、複数の区分データを生成することと、機械学習のアルゴリズムを前記各区分データにより訓練することにより、第 2 施策に係る学習モデルを区分毎に生成することと、前記各区分データから近似した推定量を用いて、前記第 2 施策をオフライン評価することと、を実行させるためのプログラムである。

【図面の簡単な説明】

【0012】

【図 1】 図 1 は実施形態に係る施策評価装置を含むシステムの構成を示す図である。

【図 2】 図 2 は広告が表示されるウェブサイトの一例を示す図である。

【図 3】 図 3 は実施形態に係る施策評価方法を示すフローチャートである。

【発明を実施するための形態】

【0013】

本開示の施策評価装置及び施策評価方法の例を、以下に図面を参照しつつ説明する。

〔システムの概要〕

図 1 に示すように、本開示に係る施策評価システム 11 は、ウェブサーバ 20 及び施策評価装置 30 を含む。ウェブサーバ 20 と施策評価装置 30 は、ネットワーク 14 を通じて互いに通信する。ウェブサーバ 20 及び施策評価装置 30 の各々は、コンピュータの一例である。

10

【0014】

ウェブサーバ 20 は 1 以上の端末 13 とネットワーク 14 を通じて互いに通信する。端末 13 は、例えば、スマートフォン、パーソナルコンピュータ、タブレットなどの情報処理装置である。端末 13 はユーザによって操作されるユーザ端末である。ウェブサーバ 20 は、商品又はサービスを提供又は推薦するためのウェブサイト 40 を提供してもよい。ウェブサイト 40 は、商品又はサービスを検索するため検索窓（図示は省略するが、例えば、テキストボックス）を有してもよい。ウェブサーバ 20 は、1 以上の端末 13 からの要求に応じて、種々の情報又は処理結果を各端末 13 に提供する。

【0015】

20

ネットワーク 14 は、例えば、インターネット、WAN（Wide Area Network）、LAN（Local Area Network）、プロバイダ端末、無線通信網、無線基地局、専用回線等を含む。図 1 に示す各装置の全ての組み合わせが相互に通信可能である必要はなく、ネットワーク 14 は、一部にローカルなネットワークを含んでもよい。

【0016】

ウェブサーバ 20 は、プロセッサ 22、メモリ 24、および通信機 23 を備える。通信機 23 は、ネットワーク 14 を介して他の装置、例えば端末 13 及び施策評価装置 30 との通信を可能とする。メモリ 24 には、広告を表示するための表示プログラム 25 及び広告データ 27 と、履歴データ 28 が格納される。メモリ 24 には、さらに、広告のターゲットとなり得る顧客データ 29 が格納されてもよい。

30

【0017】

顧客データ 29 は、例えば、ウェブサイト 40 を利用するユーザの登録情報を含む。登録情報は、例えば、各ユーザの氏名、年齢、性別、住所、アカウント、メールアドレス、及び決済情報を含むが、これらに限られない。決済情報の例は、クレジットカード番号、デビットカード番号、又は引き落とし口座の番号である。顧客データ 29 には、各ユーザのウェブサイト 40 及び他の店舗での購買履歴を含んでもよい。購買履歴は、例えば、購入した商品の品目、購入の日時、及び購入金額を含む。

【0018】

履歴データ 28 は、各ユーザによるウェブサイト 40 での閲覧履歴、購入履歴、又は利用履歴を含んでもよい。履歴データ 28 は、ウェブサイト 40 に含まれるリンク（例えば、広告又は商品画面へのリンク）に対する各ユーザによるアクセス履歴（例えば、クリック履歴）を含む。履歴データ 28 は顧客データ 29 と連動することにより、ユーザの登録情報の少なくとも一部をユーザの属性情報として含んでもよい。履歴データ 28 に含まれるユーザの属性情報は、例えば、各ユーザの年齢、性別、閲覧履歴、利用履歴、及びアクセス履歴などである。

40

【0019】

施策評価装置 30 は、1 以上のプロセッサ 32、1 以上のメモリ 34、および通信機 33 を備える。施策評価装置 30 は、例えばサーバのようなコンピュータである。通信機 33 は、ネットワーク 14 を介して他の装置、例えばウェブサーバ 20 との通信を可能とする。メモリ 34 には、区分データ生成プログラム 35、学習プログラム 36、評価プログ

50

ラム 37 が格納されている。メモリ 34 は、さらに、履歴データ 38 及び 1 以上の学習モデル 39 が格納されるように構成される。施策評価装置 30 のプロセッサ 32 は、メモリ 34 に格納されている学習プログラム 36 を実行することで、学習モデル 39 を生成する。

【0020】

ウェブサーバ 20 は、施策評価装置 30 の要求に応じて、あるいは、データ送付のためのプログラムに従って、メモリ 24 に格納された履歴データ 28 を施策評価装置 30 に送信する。施策評価装置 30 は、受信した履歴データ 28 を履歴データ 38 としてメモリ 34 に格納する。履歴データ 28 は、持ち運び可能な別のメモリを介して、施策評価装置 30 のメモリ 34 に履歴データ 38 として格納されてもよい。

10

【0021】

プロセッサ 22, 32 は、例えば、CPU、GPU、および TPU のような演算ユニットを含む。プロセッサ 22, 32 は、各種ソフトウェア処理を実行するように構成される処理回路である。処理回路は、ソフトウェア処理の少なくとも一部を処理する専用のハードウェア回路（たとえば ASIC 等）を備えてもよい。すなわち、ソフトウェア処理は、1 又は複数のソフトウェア処理回路及び 1 又は複数の専用のハードウェア回路の少なくとも一方を備えた処理回路（processing circuitry）によって実行されればよい。

【0022】

メモリ 24, 34 は、コンピュータ可読媒体である。メモリ 24, 34 は、例えば、RAM（Random Access Memory）、HDD（Hard Disk Drive）、フラッシュメモリ、ROM（Read Only Memory）などの非一過性の記憶媒体を含む。プロセッサ 22, 32 は、与えられる信号に応じて、又は、予め定められた条件が成立したことに応じて、それぞれメモリ 24, 34 に格納されているプログラムに含まれる一連の命令を実行する。

20

【0023】

【広告システム】

図 2 に示すように、ウェブサイト 40 は、例えば、各種商品又はサービスを提供するための情報を提示する。ウェブサイト 40 の一例は、複数の商品を販売するショッピングサイトである。ウェブサイト 40 で提供される商品又はサービスの例は、旅行プラン、宿泊施設、切符、チケット、書籍、雑誌、音楽、動画、映画、保険、又は証券であるが、これらに限られない。

30

【0024】

ウェブサイト 40 は、1 以上の広告枠 41（41A, 41B, 43C, 43D）を含んでもよい。ウェブサイト 40 は、おすすめの商品を表示する複数の推薦枠 42 を含んでもよい。広告枠 41 及び推薦枠 42 は表示枠の一例である。1 つの広告枠 41A には、1 つの広告画像（例えば、バナー広告）が表示されてもよい。あるいは、1 つの広告枠 41B に複数の候補画像が順次に表示されてもよい。複数の広告枠 41A ~ 41D がある場合、表示の方法及び内容が互いに異なってもよい。また、同じ種類の広告枠 41B（又は 41C）が複数あってもよい。

【0025】

ウェブサイト 40 に表示される広告画像又は商品画像は、規定のルールに従って変更されてもよい。例えば、ある期間毎に画像が変更されてもよいし、閲覧するユーザの属性に応じて画像が変更されてもよい。一例として、販売促進イベントの期間（例えば、1 ヶ月）のうちに、1 つのバナー広告枠に表示する複数（例えば 5 つ）の候補画像を用意しておき、規定された順序でそれら画像を順次に表示するようにしてもよい。さらに、複数の候補画像（例えば、割引クーポン）のいくつかは、イベント期間中に別の候補画像（別の割引クーポン）に入れ替えられてもよい。

40

【0026】

広告データ 27 は、各広告枠 41 の種類、各広告枠 41 に対応する 1 以上の候補画像、各広告枠 41 に対する表示変更履歴、及び、期間毎の広告施策を含む。表示変更履歴は、

50

広告の単位毎（例えばイベント毎）に、使用される又は使用候補となる１以上の広告画像、各広告画像が使用開始される日時、及び各広告が表示される広告枠４１に係るデータを含む。

【００２７】

履歴データ３８は複数のデータセットを含み、各データセットは、ユーザからウェブサーバ２０へのウェブサイト４０（及びウェブサイト４０に含まれる広告）の表示要求の都度取得された複数のデータを含む。これら複数のデータは、各広告枠４１について、例えば、表示を要求したユーザの属性、表示時刻、各時刻に表示された広告、及び、その広告がクリックされたか否か、に係るデータを含む。さらに、各データセットは、クリックしたユーザがその広告に係る商品を購入したか否かに係るデータを含んでもよい。

10

【００２８】

〔施策の策定〕

施策（Policy）とは、ユーザに何らかのコンテンツを推薦したり、何らかの行動を促したりするための手法又は計画を指す。例えば、バナー広告を用いた広告施策では、広告に係る商品の購入をユーザに促すために、ユーザにバナー広告（リンクを伴う広告画像）をクリックさせることを目的とする。あるいは、メールマーケティングにおいては、メールを受信したユーザに、メール本文中に含まれるリンクをクリックさせることを目的とする。

【００２９】

クリック率は、「クリック先のリンクに表示される商品を購入させる」という目標を達成するための指標の一例である。他に、クリック当たりのコンバージョン率（クリック経由のコンバージョン率）、又は広告コストあたりの収益を指標とすることもできる。

20

【００３０】

実施済みの広告施策におけるクリック率は、履歴データ３８に基づいて算出することができる。クリック率は、広告施策、すなわち、広告を表示する場所及びタイミング、又はターゲットとなるユーザの属性などにより変化し得る。そのため、新たな施策の策定にあたっては、より広告効果を高めるために、広告内容の最適化が望まれる。

【００３１】

新たな施策を最適化するために、機械学習モデルを使用することができる。プロセッサ３２は、学習プログラム３６の実行により、例えば、強化学習の一例である多腕バンディット（マルチアームバンディット、Multi-armed Bandit）アルゴリズムを用いて、施策を最適化するように構成される。多腕バンディットアルゴリズムは、複数のアームと呼ばれる候補の中から、最も良いものを逐次的に探す問題である。学習プログラム３６は、一例として、トンプソンサンプリング（Thompson Sampling）を用いる。

30

【００３２】

例えば、特徴量（特徴量ベクトル）を $x \in X$ 、取り得る m 個の行動を $a \in A = \{1, 2, \dots, m\}$ として、多腕バンディット問題を解くことにより、一定期間における報酬 Y を最大化する行動 a を探索する。特徴量 x は、例えば、ユーザの年齢、性別、購入履歴のようなユーザ属性である。行動 a は、例えば、広告の対象となる商品である。報酬 Y は、例えば、クリック率であってもよいし、ある商品の広告を出した場合の売上であってもよい。

40

【００３３】

施策 π は、特徴量 $x \in X$ から行動空間 A 上の確率分布への写像 $\pi: X \rightarrow \Delta(A)$ として定義される。 $\pi(a|x)$ は、ある x というベクトルで特徴付けられるデータに対して、 a という行動を選択する確率（行動選択確率）である。 $\pi(a|x)$ は、どんな状況でどの行動をとるべきかを司る意思決定のための施策であるとも言える。例えば、 $\pi(a|x)$ は、 x （あるユーザ属性）を入力した場合に、どの商品を広告の対象にすべきか（どの広告画像を表示すべきか）を示す関数となり得る。

【００３４】

〔施策の評価〕

50

プロセッサ 32 は、評価プログラム 37 を実行することにより、新たな施策の評価を行うように構成される。新たな施策を実装することなく評価することをオフライン施策評価（OPE: Off-Policy Evaluation、オフポリシー評価又はオフラインポリシー評価）という。評価に供される施策を評価施策 π （Evaluation policy）と呼び、評価施策の比較対象となる施策を行為施策 π_b （Behavior policy）と呼ぶことがある。

【0035】

本開示では、行為施策を第 1 施策と呼び、評価施策を第 2 施策と呼ぶことがある。第 1 施策は、一般的には実施済みの過去の施策であって、本例では、第 1 施策を実施したときの記録が履歴データ 38 である。履歴データ 38 は複数のデータセットを含む。ただし、第 1 施策と第 2 施策が共に実施済みの施策であってもよい。

【0036】

施策 π の性能は以下の目的変数 $V(\pi)$ として定義される。

【0037】

【数 1】

$$V(\pi) = \mathbb{E}_{(Y(\cdot), X)} \left[\sum_{a \in \mathcal{A}} Y(a) \pi(a|X) \right]$$

目的変数 $V(\pi)$ は、施策 π を実施した場合の指標の期待値である。例えば報酬 Y がクリックの有無である場合、 $V(\pi)$ は施策 π の実施によって期待されるクリック率になる。

【0038】

第 2 施策 π の性能は、第 1 施策 π_b を実施した時の履歴データ 38 を用いて評価することができる。例えば、履歴データ 38 は以下の数式 2 で示される T 個のデータセットを含む。

【0039】

【数 2】

$$\mathcal{D} = \{(x_t, a_t, Y_t)\}_{t=1}^T$$

例えば、ある属性 (x) を有するユーザに、ある商品 (a) のバナー広告を出した場合に、そのバナーをクリックしていれば Y_t が「1」、クリックしていなければ Y_t が「0」になる。

【0040】

ここで、ユーザがある行動 a を選択した場合、履歴データ 38 には、その行動 a に基づく報酬 Y_t のデータしか含まれない。例えば、第 1 の商品と第 2 の商品のうち、第 1 の商品の購入をユーザに薦めた場合（行動 a = 第 1 の商品）の結果（クリックの有無）はわかるが、第 1 の商品を薦めなかった場合の結果はわからない。言い換えると、過去の施策で第 1 の商品を推薦対象としてその結果である報酬 Y を得た場合に、新たな施策で第 2 の商品を推薦対象にすると、第 2 の商品を薦めた場合（行動 a = 第 2 の商品）のデータは、履歴データ 38 に含まれていない。そのため、履歴データ 38 に基づいて新たな施策 π について真の性能を得ることは困難である。

【0041】

新たな施策 π の性能は、真の性能である $V(\pi)$ に対して、履歴データ 38 から近似した推定量として、以下の数式 3 で代替し得る。

10

20

30

40

50

【 0 0 4 2 】

【 数 3 】

$$\hat{V}_{IPW}(\pi; \mathcal{D}) = \frac{1}{T} \sum_{t=1}^T Y_t \frac{\pi(a_t | x_t)}{\pi_b(a_t | x_t)}$$

これは、IPW (Inverse Probability Weighting) と呼ばれる手法である。IPWでは、第1施策の結果を示す指標値（観測されている目的変数、例えばクリック率）を第1施策と第2施策の行動選択確率の比を用いて重み付けすることにより、推定量を算出する。

10

【 0 0 4 3 】

数式3に代えて、以下の数式4によって推定量を得ることもできる。数式4は、DR (Doubly Robust) と呼ばれる手法である。

【 0 0 4 4 】

【 数 4 】

$$\hat{V}_{DR}(\pi; \mathcal{D}) = \hat{V}_{DM}(\pi; \mathcal{D}) + \frac{1}{T} \sum_{t=1}^T (Y_t - \mu(x_t, a_t)) \frac{\pi(a_t | x_t)}{\pi_b(a_t | x_t)}$$

20

DRでは、目的変数の期待値のモデル μ の推定誤差を、IPWのような方法によって補正している。DRによる推定量の算出は、第1施策の結果を示す指標値を第1施策と第2施策の行動選択確率の比を用いて重み付けすることを含む。数式3及び数式4など、オフライン評価のための推定量を得る方法については、例えば、上記非特許文献1に開示されている。

【 0 0 4 5 】

一例として、広告枠41Aにバナー広告を表示する場合に、広告対象とするユーザ属性を最適化した第2施策を評価する場合について説明する。バナー広告は、複数（例えば5つ）の候補画像を含み、これら候補画像は、施策が規定する法則に沿って広告枠41Aに表示される。例えば、第1施策（行為施策 π_b ）は、5つの候補画像を広告枠41Aに表示する割合を規定する。第1施策において、5つの候補画像は、例えば、一定の順番で、あるいはランダムに、同じ比率で表示されたものとする。この場合、履歴データ38に基づいて算出したクリック率は、第1施策の性能を示す。

30

【 0 0 4 6 】

本例の第2施策（評価施策 π ）は、第1施策と同じ5つの候補画像を1つずつ同じ広告枠41Aに表示する場合に、ユーザ属性に基づいてどの候補画像を表示させるかを規定する。そのため、第2施策では、ユーザに応じて、5つの候補画像が表示される比率が変化する。本例では、第2施策を実施した場合の推定量（例えば、クリック率）を算出する。推定量としてクリック率を算出した場合、クリック当たりのコンバージョン率と、コンバージョン当たりの収益に基づいて、新たな施策を適用した場合の期待利益を見積もることが可能になる。

40

【 0 0 4 7 】

ここで、数式3又は数式4により算出される推定量は、第1施策の実施中に広告内容が不変であることを前提としている。しかし、実際には、第1施策の実施中に広告内容に係る1以上の要件が変化していることがある。例えば、履歴データ38を記録した期間中に、1回又は複数回、5つの候補画像のうちの一部が別の候補画像と変更されていたとする。この場合の変化する要件は、広告に使用される候補画像のラインナップである。

50

【0048】

また、ここで、数式3又は数式4により算出される推定量は、上記広告内容（ラインナップ）が不変であることに起因してバイアスが生じる。例えば、5つの候補画像のうち、クリック率が最も低い1つの画像が、より広告効果の高い1つの画像と変更されることにより、その変更の前後で、変更された画像及び残る4つの画像のクリック率が変動する可能性がある。

【0049】

上記バイアスに対処するため、本開示では、履歴データ38に含まれる複数のデータセットを、第1施策に関する少なくとも1つの要件（例えば、複数の候補画像のラインナップ）が不変であることに基づいて区分する。そして、区分（セグメンテーション）により生成した区分データを用いて、区分（セグメント）毎に推定量を算出する。

【0050】

例えば、プロセッサ32が、区分データ生成プログラム35を実行して、履歴データ38に含まれるn個のデータセットをK個に区分して、K個の区分データを作成する。このとき、プロセッサ32は、複数のデータセットを、第1施策に関する少なくとも1つの要件が不変である区分に基づいて、複数の区分データを生成するように構成された、区分データ生成部として使用される。さらに、プロセッサ32は、区分データ生成プログラム35又は別の分割プログラム（図示略）を実行して、各区分データを訓練用データと評価用データとに分割する。

【0051】

各データセットは、例えば、広告を表示した時刻（タイムスタンプ）と、特徴量x（ユーザの属性）と、対応する行動a（各時刻の表示画像）と、行動の結果r（画像がクリックされたか否か）、に関するデータを含む。例えば、4週間に亘って実施される広告イベントにおいて、5つの候補画像のうちの1つが一週間毎に変更される場合、5つの候補画像が不変の1週間分の区分データが4つ作成される。

【0052】

区分データについて、k番目の区分は以下のように定義される。

【0053】

【数5】

$$[n_k, n_{k+1}) := \{n_i \in \mathbb{N} \mid k \leq i < k+1\}$$

また、k番目の区分データ $\mathcal{D}_k$ は以下のように定義される。

【0054】

【数6】

$$\mathcal{D}_k := \{(x_i, a_i, r_i)\}_{i=n_k}^{n_{k+1}}.$$

また、第1施策に係る π^b は、以下の数式7のように算出される。

【0055】

【数7】

$$\pi_k^b(a_i | x_i) = \frac{|\{a \in \mathcal{D}_k \mid a = a_i\}|}{n_{k+1} - n_k}$$

〔施策評価方法〕

図 3 を参照して、本開示に係る施策評価方法について説明する。

【0056】

本例では、履歴データ 38 は、広告キャンペーンの実施中に取得された複数のデータセットを含む。各データセットは、ユーザからのウェブサイト 40 の表示要求の都度取得された複数のデータを含む。ウェブサイト 40 は 1 以上の広告枠 41（又は 1 以上の推薦枠 42）を含み、第 1 施策及び第 2 施策は、各広告枠 41（又は推薦枠 42）に広告画像（又は商品画像）を表示することに関する。また、特徴量 x がユーザの属性であり、行動 a が複数の候補画像であり、行動の結果 r が表示された広告画像（又は商品画像）をユーザがクリックしたか否かである。第 1 施策の結果を示す指標値はクリック率である。そして、複数の候補画像が各区分において不変であることを前提にしている。

10

【0057】

まず、ステップ S 11 で、プロセッサ 32 が第 2 施策を策定するための機械学習のアルゴリズム（例えば、多腕バンディットアルゴリズム）を取得する。具体的には、プロセッサ 32 が学習プログラム 36 を取得して、メモリ 34 に格納する。

【0058】

ステップ S 12 で、プロセッサ 32 が履歴データ 38 を取得する。具体的には、プロセッサ 32 がウェブサーバ 20 から履歴データ 28 を受信して、履歴データ 38 としてメモリ 34 に格納する。ステップ S 12 は、ステップ S 11 より先に行われてもよいし、ステップ S 11 と同時に行われてもよい。

20

【0059】

ステップ S 13 で、プロセッサ 32 が、区分データ生成プログラム 35 を実行して、履歴データ 38 に含まれる複数のデータセットから、第 1 施策に関する少なくとも 1 つの要件が不変であることに基づいて、複数の区分データを生成する。さらに、プロセッサ 32 は、区分データ生成プログラム 35 又は別の分割プログラムを実行して、各区分データを訓練用データと評価用データとに分割する。

【0060】

ステップ S 14 で、プロセッサ 32 が、ステップ S 11 で得たアルゴリズムにおいて、ステップ S 13 で生成した区分データを用いてシミュレーションし、第 2 施策を最適化する。具体的には、プロセッサ 32 が、区分データを分割した訓練用データを用いて学習プログラム 36 を実行して、機械学習のアルゴリズムを各区分データにより訓練することにより、区分毎に最適化された学習モデル 39 を作成する。このとき、プロセッサ 32 は、第 2 施策に係る学習モデルを区分毎に生成するように構成された、学習モデル生成部として使用される。また、プロセッサ 32 は、生成した学習モデル 39 をメモリ 34 に格納する。

30

【0061】

ステップ S 14 で、プロセッサ 32 が評価プログラム 37 を実行することにより、区分データから近似したオフライン評価（OPE）の推定量を算出する。具体的には、プロセッサ 32 が、上記数式 3 又は数式 4 において、 $D = D_k$ として、区分毎に生成された学習モデルと区分データを分割した評価用データとを用いて、第 2 施策に係る推定量を算出する。この場合、上記数式 3 又は数式 4 において、 π_b は履歴データ 38 全体のデータセットの頻度の分布として計算される。このとき、プロセッサ 32 は、区分データから近似した推定量を用いて、第 2 施策をオフライン評価するように構成された、評価部として使用される。

40

【0062】

〔本開示の作用〕

オフライン評価では、過去の履歴データ 38 を利用することで、オンライン評価に伴うリスク及びコスト増加を避けつつ、新たな施策（第 2 施策）を評価することができる。ただし、第 1 施策を実施する間に、第 1 施策に関する 1 以上の要件（例えば、広告のための画像）が変更されていると、履歴データ 38 には、その要件変更起因するデータの変動

50

が含まれ得る。そのため、第 1 施策の履歴データ 38 を、第 1 施策とは異なる第 2 施策のオフライン評価に用いると、評価のための推定量にバイアスを含むことになる。

【0063】

その点、本開示では、履歴データ 38 に含まれる複数のデータセットを、第 1 施策に関する 1 以上の要件が不変である区分に分けることにより、複数の区分データを生成している。そして、その区分毎に第 2 施策に係る学習モデル 39 を生成した上で、区分データから近似した推定量を用いて第 2 施策をオフライン評価している。そのため、オフライン評価におけるバイアスが対処される。

【0064】

〔本開示の効果〕

本開示によれば、以下の効果を奏することができる。

(1) 第 2 施策に係る学習モデルを区分毎に生成した後、区分データから近似した推定量を用いて第 2 施策をオフライン評価することにより、オフライン評価におけるバイアスに対処することができる。

【0065】

(2) 多腕バンディットアルゴリズムのような強化学習のアルゴリズムを用いて学習モデルを生成することにより、累積報酬を最大化するための方策を策定することができる。

(3) 推定量を算出する際に、第 1 施策の結果を示す指標値を、第 1 施策と第 2 施策の行動選択確率の比を用いて重み付けすることにより、第 1 施策を実施したときの履歴データ 38 を用いて第 2 施策を評価することができる。

【0066】

(4) 各広告枠 41 に表示され得る複数の候補画像が不変であることに基づいて区分データを生成することにより、オフライン評価におけるバイアスに対処することができる。

(5) ユーザ属性を含む履歴データ 38 を第 2 施策のオフライン評価に用いることにより、ユーザ属性に応じた広告を表示することが可能な第 2 施策を策定することができる。これにより、広告のクリック率の上昇が見込める。

【0067】

本実施形態は、以下のように変更して実施することができる。本実施形態及び以下の変更例は、技術的に矛盾しない範囲で互いに組み合わせて実施することができる。

〔変更例 1〕

第 1 施策及び第 2 施策は、ウェブサイト 40 の広告枠 41 又は推薦枠 42 に限らず、任意の表示枠に画像を表示することを含んでもよい。例えば、表示枠は、駅、店舗、施設、又はオフィスなどに、ディスプレイ又はプロジェクターのような映像表示装置を設置して情報を発信するデジタルサイネージにおける表示枠であってもよい。映像表示装置により情報を発信する場合、特徴量 x は、情報を設置する場所の属性であってもよいし、情報を発信する方法の属性であってもよい。また、行動 a が各候補画像であってもよく、行動の結果 r が表示内容に関する問合せ又は販売促進効果であってもよい。

【0068】

あるいは、第 1 施策及び第 2 施策は、検索サイトにおいて、検索結果の候補となる複数の結果候補のうちの 1 つを、検索結果を示すための表示枠に表示することを含んでもよい。この場合、特徴量 x は、検索を行ったユーザの属性であってもよく、行動 a が各検索結果であってもよく、行動の結果 r が表示された検索結果をユーザがクリックしたか否かであってもよい。

【0069】

〔変更例 2〕

第 1 施策及び第 2 施策は、画像に限らず、複数の対象候補から 1 以上の対象を選択することを含んでもよい。一例として、第 1 施策及び第 2 施策は、広告又はマーケティングの対象となる商品、サービス、ユーザ、又はユーザ属性を決定することを含んでもよい。例えば、対象候補は、郵便広告を送るユーザ又は送付住所であってもよいし、ポスティングの対象エリアであってもよいし、メールの送付先であってもよい。あるいは、対象候補は

10

20

30

40

50

、モニター募集、営業又はサービスに関する案内をするための電話営業又はテレフォンアポイントメントの候補とするユーザ、ユーザ属性、又は電話番号であってもよい。その他、対象候補は、ユーザに様々な媒体（紙媒体又は電子媒体）により配布するクーポン又は商品サンプルであってもよい。

【0070】

この場合、特徴量 x が対象候補の属性であってもよく、行動 a が複数の対象候補であってもよく、行動の結果 r がリアクションの有無又はリアクションの良否（すなわち、望ましい応答が得られたか否か）であってもよい。

【0071】

第2変更例によれば、複数の対象候補が不変であることに基づいて区分データを生成することにより、オフライン評価におけるバイアスに対処することができる。

10

〔変更例3〕

第1施策及び第2施策は、1つの表示枠に順次に表示される複数の画像の表示順序を決定することを含んでもよい。この場合、複数の画像が各区分において不変であればいい。この場合の表示枠は、例えば、カルーセル広告の表示枠であってもよいし、バナー広告の表示枠であってもよいし、おすすめ商品の表示枠であってもよいし、更新情報を告知する告知枠であってもよい。これら表示枠は、変更例1のような映像表示装置によって表示される画像に含まれていてもよいし、ウェブサイトに含まれていてもよい。

【0072】

第3変更例によれば、複数の画像が不変であることに基づいて区分データを生成することにより、オフライン評価におけるバイアスに対処することができる。

20

〔変更例4〕

第1施策及び第2施策は、複数の物品又は複数のサービスからユーザに推薦する1以上の物品又はサービスを選択することを含んでもよい。この場合、複数の物品又は複数のサービスの価格が各区分において不変であってもよいし、複数の物品又は複数のサービスの品揃えが各区分において不変であってもよいし、価格と品揃えの両方が不変であってもよい。サービスは、例えば、音楽、映画、動画、テレビ番組、保険商品、金融商品などであってもよい。

【0073】

この場合、特徴量 x がユーザの属性（例えば、年齢、性別、過去の視聴履歴、過去の購入履歴など）であってもよく、行動 a が物品又はサービスの種類であってもよく、行動の結果 r が推薦に対するリアクションの有無、購買の有無、又は視聴時間であってもよい。

30

【0074】

第4変更例によれば、値段又は品揃えのうち少なくとも一方が不変であることに基づいて区分データを生成することにより、オフライン評価におけるバイアスに対処することができる。

【0075】

〔変更例5〕

ウェブサイト40は、複数の商品又は複数のサービスを検索するための検索画面及びその検索結果画面を提示するように構成されてもよい。このとき、履歴データに含まれる各データセットは、検索画面におけるユーザからの検索要求の都度取得された複数のデータを含んでもよい。そして、第1施策及び第2施策は、検索結果画面に表示する複数の検索結果の表示順位（ランキング）を決定することと、検索結果画面に検索意図に沿った順位で複数の検索結果を表示することと、を含む。複数の検索結果は、例えば検索アルゴリズムの変更により変化し得る。この場合、複数の検索結果が各区分において不変であってもよいし、検索対象となる商品又はサービスの品揃えが各区分において不変であってもよいし、検索結果と検索対象の両方が不変であってもよい。

40

【0076】

第5変更例によれば、複数の検索対象と複数の検索結果のうち少なくとも一方が不変であることに基づいて区分データを生成することにより、オフライン評価におけるバイアス

50

に対処することができる。

【0077】

〔変更例6〕

施策評価装置30は、区分データを生成する区分データ生成装置（例えば、コンピュータ）と、第2施策に係る学習モデルを生成する学習装置（例えば、コンピュータ）と、第2施策をオフライン評価する評価装置（例えば、コンピュータ）と、を備える施策評価システムであってもよい。この施策評価システムは、さらに、ウェブサーバ20を備えてもよい。すなわち、施策評価装置30は、物理的に1つの筐体に収容される装置に限らず、履歴データを取得するデータ取得部、区分データ生成部、学習モデル生成部、及び評価部といった複数の機能部を実現するための複数のコンピュータを含む、仮想的な装置であってもよい。

10

【0078】

〔変更例7〕

第2施策を策定するための機械学習のアルゴリズムは、多腕バンディットアルゴリズムに限らない。例えば、他の強化学習に係るアルゴリズムなど、任意のアルゴリズムを利用することができる。

【0079】

〔変更例8〕

施策の結果を示す指標値、特徴量 x 、行動 a 、及び行動の結果 r は、本開示で例示したものに限らず、任意に変更することができる。例えば、施策に係る表示等にアクセスするユーザの属性情報が取得できない場合、特徴量 x は、広告対象の商品又はサービスの属性であってもよいし、表示枠の属性（例えば、大きさ又は表示位置など）、任意の特徴量に変更することができる。

20

【0080】

〔変更例9〕

区分データを生成する場合の条件は任意に変更することができる。例えば、政策又は税率の変更、社会情勢又は行動様式の変化、季節又は天候の変化など、ユーザを取り巻く環境変化を区分の起点としてもよい。また、区分と区分との間に、変化が落ち着くまでの一定のマージンを設けてもよい。こうした環境変化に伴って、ユーザの行動も変化し得るためである。また、区分データを作成するときに、複数のデータセットを複数の要件が不変である区分に分けてもよい。特に、バイアスを多く含むと想定される1以上の要件が不変となるように、区分データを生成するとよい。逆に、要件変更の影響が小さければ、その要件に基づく区分を行わなくてもよい。すなわち、本開示における「少なくとも1つの要件」は、オフライン評価に有意な影響を生じ得る要件を意味する。

30

【0081】

以下に、上記実施形態及び変更例から把握される態様を列挙する。

〔1〕1以上のプロセッサと1以上のメモリとを備え、

前記メモリには、第1施策を実施したときの記録である複数のデータセットを含む履歴データが格納されており、

前記プロセッサは、

前記第1施策に関する少なくとも1つの要件が不変であることに基づいて前記複数のデータセットを区分することにより、複数の区分データを生成することと、

機械学習のアルゴリズムを前記各区分データにより訓練することにより、第2施策に係る学習モデルを区分毎に生成することと、

前記各区分データから近似した推定量を用いて、前記第2施策をオフライン評価することと、

を実行するように構成される、施策評価装置。

40

【0082】

〔2〕前記アルゴリズムは多腕バンディットアルゴリズムであり、

前記各データセットは、特徴量と、前記特徴量に対応する行動と、前記行動の結果と、

50

に関するデータを含む、

上記〔１〕に記載の施策評価装置。

【００８３】

〔３〕前記推定量の算出は、前記第１施策の結果を示す指標値を、前記第１施策と前記第２施策の行動選択確率の比を用いて重み付けすることを含む、

上記〔２〕に記載の施策評価装置。

【００８４】

〔４〕前記第１施策及び前記第２施策は、複数の候補画像のうちの１つを表示枠に表示することを含み、

前記複数の候補画像が前記各区分において不変である、

上記〔１〕～〔３〕のうち何れかに記載の施策評価装置。

【００８５】

〔５〕前記各データセットは、ユーザからのウェブサイトの表示要求の都度取得された複数のデータを含み、

前記ウェブサイトは１以上の広告枠を含み、

前記第１施策及び前記第２施策は、前記各広告枠に広告画像を表示することを含み、

前記特徴量が前記ユーザの属性であり、前記行動が各候補画像であり、前記行動の結果が表示された前記広告画像を前記ユーザがクリックしたか否かであり、

前記複数の候補画像が前記各区分において不変である、

上記〔２〕又は〔３〕に記載の施策評価装置。

【００８６】

〔６〕前記第１施策及び前記第２施策は、複数の対象候補から１以上の対象を選択することを含み、

前記複数の対象候補が前記各区分において不変である、

上記〔１〕～〔３〕のうち何れかに記載の施策評価装置。

【００８７】

〔７〕前記第１施策及び前記第２施策は、１つの表示枠に順次に表示される複数の画像の表示順序を決定することを含み、

前記複数の画像が前記各区分において不変である、

上記〔１〕～〔３〕のうち何れかに記載の施策評価装置。

【００８８】

〔８〕前記第１施策及び前記第２施策は、複数の物品又は複数のサービスからユーザに推薦する１以上の物品又はサービスを選択することを含み、

前記各区分において、前記複数の物品又は前記複数のサービスの価格と、前記複数の物品又は前記複数のサービスの品揃えと、のうち少なくとも一方が不変である、

上記〔１〕～〔３〕のうち何れかに記載の施策評価装置。

【００８９】

〔９〕前記各データセットは、検索画面におけるユーザからの検索要求の都度取得された複数のデータを含み、

前記第１施策及び前記第２施策は、検索結果画面に表示する複数の検索結果の表示順位を決定することを含み、

前記各区分において、複数の検索対象と前記複数の検索結果のうち少なくとも一方が不変である、

上記〔１〕～〔３〕のうち何れかに記載の施策評価装置。

【００９０】

〔１０〕第１施策を実施したときの記録である複数のデータセットを含む履歴データが格納されたメモリと、

前記第１施策に関する少なくとも１つの要件が不変であることに基づいて前記複数のデータセットを区分することにより、複数の区分データを生成するように構成された区分データ生成部と、

10

20

30

40

50

機械学習のアルゴリズムを前記各区分データにより訓練することにより、前記区分毎に第2施策に係る学習モデルを生成するように構成された学習モデル生成部と、

前記各区分データを用いて、前記区分毎に前記第2施策をオフライン評価するように構成された評価部と、

を備える、施策評価装置。

【0091】

〔11〕1以上のプロセッサと1以上のメモリとを備え、

前記メモリには、第1施策を実施したときの記録である複数のデータセットを含む履歴データが格納されており、

前記プロセッサは、

前記第1施策を実施した期間内における環境変化を起点として前記複数のデータセットを区分することにより、複数の区分データを生成することと、

機械学習のアルゴリズムを前記各区分データにより訓練することにより、第2施策に係る学習モデルを区分毎に生成することと、

前記各区分データから近似した推定量を用いて、前記第2施策をオフライン評価することと、

を実行するように構成される、施策評価装置。

【0092】

〔12〕上記〔1〕～〔11〕のうち何れかに記載の施策評価装置と、

第1施策を実施したときの記録である複数のデータセットを含む履歴データが格納されたメモリを備えるサーバと、

を備える、施策評価システム。

【0093】

〔13〕1以上のコンピュータに、

第1施策を実施したときの記録である複数のデータセットを含む履歴データを取得することと、

前記第1施策に関する少なくとも1つの要件が不変であることに基づいて前記複数のデータセットを区分することにより、複数の区分データを生成することと、

機械学習のアルゴリズムを前記各区分データにより訓練することにより、第2施策に係る学習モデルを区分毎に生成することと、

前記各区分データから近似した推定量を用いて、前記第2施策をオフライン評価することと、

を実行させることを含む、施策評価方法。

【0094】

〔14〕1以上のコンピュータに、

第1施策を実施したときの記録である複数のデータセットを含む履歴データを取得することと、

前記第1施策に関する少なくとも1つの要件が不変であることに基づいて前記複数のデータセットを区分することにより、複数の区分データを生成することと、

機械学習のアルゴリズムを前記各区分データにより訓練することにより、第2施策に係る学習モデルを区分毎に生成することと、

前記各区分データから近似した推定量を用いて、前記第2施策をオフライン評価することと、

を実行させるための、施策評価プログラム。

【符号の説明】

【0095】

11…システム、13…端末、14…ネットワーク、20…ウェブサーバ、22…プロセッサ、23…通信機、24…メモリ、25…表示プログラム、27…広告データ、28…履歴データ、29…顧客データ、30…施策評価装置、32…プロセッサ、33…通信機、34…メモリ、35…区分データ生成プログラム、36…学習プログラム、37…評

10

20

30

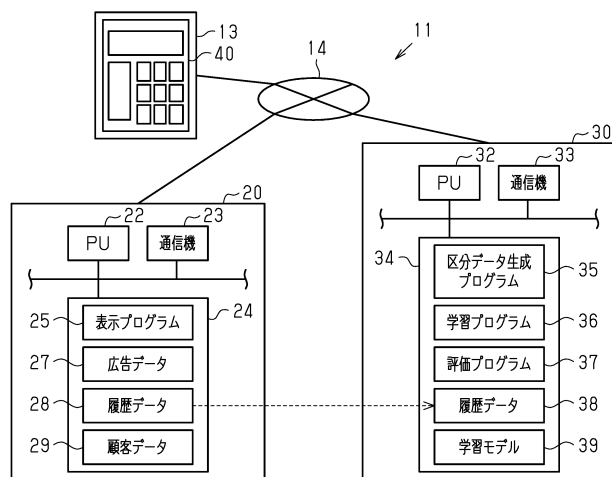
40

50

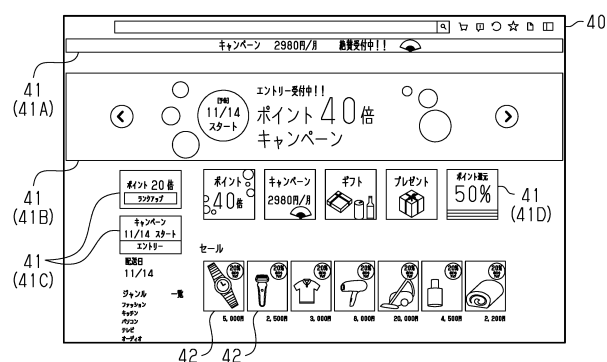
価プログラム、38…履歴データ、39…学習モデル、40…ウェブサイト、41（41A～41D）…広告枠、42…推薦枠。

【凶面】

【図 1】



【図 2】



【圖 3】

