

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2023年8月17日(17.08.2023)



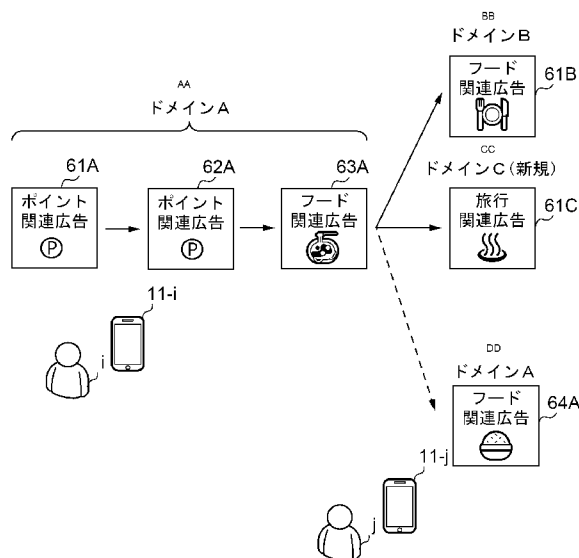
(10) 国際公開番号

WO 2023/152950 A1

- (51) 国際特許分類:
G06Q 30/02 (2012.01)
- (21) 国際出願番号: PCT/JP2022/005602
- (22) 国際出願日: 2022年2月14日(14.02.2022)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (71) 出願人: 楽天グループ株式会社 (RAKUTEN GROUP, INC.) [JP/JP]; 〒1580094 東京都世田谷区玉川一丁目14番1号 Tokyo (JP).
- (72) 発明者: 石川 詩 苑 (ISHIKAWA Shion); 〒1580094 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 Tokyo (JP).
- (74) 代理人: 小 西 恵, 外 (KONISHI Kay et al.); 〒1070052 東京都港区赤坂2-21-8 Tokyo (JP).
- (81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,

(54) Title: INFORMATION PROCESSING DEVICE, INFORMATION PROCESSING METHOD, PROGRAM, AND LEARNING MODEL

(54) 発明の名称: 情報処理装置、情報処理方法、プログラム、および学習モデル



61A, 62A Point-related advertisement
61B, 63A, 64A Food-related advertisement
61C Travel-related advertisement
AA, DD Domain A
BB Domain B
CC Domain C (new)

(57) Abstract: An information processing device (100) comprises: an acquisition means that acquires a content similarity, which is the similarity between target content and one or more other pieces of content, and user similarity, which is the similarity between a target user and one or more other users; an estimation means that, on the basis of the content similarity and the user similarity, estimates the prior distribution of expected remuneration to be obtained as a result of an execution process by the target user on the target content; and a derivation means that uses the prior distribution to derive the posterior



WO 2023/152950 A1

ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG,
US, UZ, VC, VN, WS, ZA, ZM, ZW.

- (84) 指定国(表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

添付公開書類 :

一 国際調査報告 (条約第21条(3))

distribution of the expected remuneration.

(57) 要約 : 情報処理装置 (100) は、対象のコンテンツと1以上の他のコンテンツとの類似度であるコンテンツ間類似度と、対象のユーザと1以上の他のユーザとの類似度であるユーザ間類似度と、を取得する取得手段と、前記コンテンツ間類似度と前記ユーザ間類似度とに基づいて、前記対象のユーザによる前記対象のコンテンツに対する実行処理により得られる期待報酬の事前分布を推定する推定手段と、前記事前分布を用いて、前記期待報酬の事後分布を導出する導出手段と、を有する。

明 細 書

発明の名称：

情報処理装置、情報処理方法、プログラム、および学習モデル

技術分野

[0001] 本発明は、情報処理装置、情報処理方法、プログラム、および学習モデルに関し、特に、種々のアイテムをレコメンドするレコメンデーションシステムに応用可能な強化学習の技術に関する。

背景技術

[0002] 機械学習の適用分野として、種々のアイテムをレコメンドするレコメンデーションシステムが知られている。従来、このようなシステムでは、レコメンデーション効果を高めるために、アイテムの取引履歴などを基にユーザ間の類似性を判定することで、一方のユーザが購買したアイテムを、当該ユーザに類似する他のユーザに適したアイテムとして特定する協調フィルタリングを用いる手法が活用されている。しかしながら、当該手法は、取引履歴などの履歴情報が乏しい場合に最適化が困難となるコールドスタート問題に直面するという課題がある。

[0003] コールドスタート問題による影響を低減させるために、モデル間で知識を転移させる手法が、例えば、非特許文献1において開示されている。レコメンデーションのためのモデルは、特定のドメインに特化し、各モデルは高いに独立していることが一般的であったが、当該文献には、ドメイン間の類似性に基づいて知識を転移させるアルゴリズムが開示されている。

先行技術文献

非特許文献

[0004] 非特許文献1：Liu, B., Wei, Y., Zhang, Y., Yan, Z., Yang, Q.: Transferable contextual bandit for cross-domain recommendation. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 32 (2018). URL:<https://ojs.aaai.org/index.php/AAAI/article/view/11699>

発明の概要

発明が解決しようとする課題

[0005] 上記文献に記載の手法では、レコメンデーションの対象となるユーザの観点が考慮されておらず、コールドスタート問題による影響が十分に低減されないという課題があった。

[0006] 本発明は上記課題に鑑みてなされたものであり、機械学習におけるコールドスタート問題による影響を低減するための技術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

[0007] 上記課題を解決するために、本発明による情報処理装置の一態様は、対象のコンテンツと1以上の他のコンテンツとの類似度であるコンテンツ間類似度と、対象のユーザと1以上の他のユーザとの類似度であるユーザ間類似度と、を取得する取得手段と、前記コンテンツ間類似度と前記ユーザ間類似度とに基づいて、前記対象のユーザによる前記対象のコンテンツに対する実行処理により得られる期待報酬の事前分布を推定する推定手段と、前記事前分布を用いて、前記期待報酬の事後分布を導出する導出手段と、を有する。

[0008] 前記情報処理装置において、前記取得手段は、前記対象のコンテンツと前記1以上の他のコンテンツの特徴を用いて、前記コンテンツ間類似度を取得し、前記対象のユーザと前記1以上の他のユーザの特徴を用いて前記ユーザ間類似度を取得しうる。

[0009] 前記情報処理装置において、前記推定手段は、前記対象のユーザによる前記他のコンテンツに対する実行処理により得られた第1の報酬を用いて、前記事前分布を推定しうる。

[0010] 前記第1の報酬は、時間の経過による報酬の割引により、前記対象のユーザによる前記他のコンテンツに対する過去の実行処理より最近の実行処理により得られた報酬が高くなるように構成されうる。

[0011] 前記情報処理装置において、前記推定手段は、前記他のユーザによる前記対象のコンテンツに対する実行処理により得られた第2の報酬を用いて、前

記事前分布を推定しうる。

- [0012] 前記第2の報酬は、時間の経過による報酬の割引により、前記他のユーザによる前記対象のコンテンツに対する過去の実行処理より最近の実行処理により得られた報酬が高くなるように構成されうる。
- [0013] 前記情報処理装置において、前記導出手段により導出された前記期待報酬の事後分布に基づいて、前記対象のコンテンツを前記対象のユーザに提供するかを判定する判定手段をさらに有しうる。
- [0014] 前記コンテンツは、有形または無形の商品またはサービスに関する広告であり、前記実行処理は、広告の表示処理であり、前記報酬は、前記広告に対するクリックの有無を示しうる。
- [0015] 上記課題を解決するために、本発明による情報処理装置の別の態様は、複数のコンテンツ間の類似度と、複数のユーザ間の類似度を取得する取得手段と、前記複数のコンテンツ間の類似度と前記複数のユーザ間の類似度を用いて、前記複数のコンテンツのうち、期待報酬が最大のコンテンツを、前記複数のユーザのうちの1以上のユーザに適したコンテンツとして決定する決定手段と、を有する。
- [0016] 上記課題を解決するために、本発明による情報処理方法の一態様は、対象のコンテンツと1以上の他のコンテンツとの類似度であるコンテンツ間類似度と、対象のユーザと1以上の他のユーザとの類似度であるユーザ間類似度と、を取得する取得工程と、前記コンテンツ間類似度と前記ユーザ間類似度とに基づいて、前記対象のユーザによる前記対象のコンテンツに対する実行処理により得られる期待報酬の事前分布を推定する推定工程と、前記事前分布を用いて、前記期待報酬の事後分布を導出する導出工程と、を有する。
- [0017] 上記課題を解決するために、本発明によるプログラムの一態様は、情報処理をコンピュータに実行させるための情報処理プログラムであって、該プログラムは、前記コンピュータに、対象のコンテンツと1以上の他のコンテンツとの類似度であるコンテンツ間類似度と、対象のユーザと1以上の他のユーザとの類似度であるユーザ間類似度と、を取得する取得処理と、前記コン

テンツ間類似度と前記ユーザ間類似度とに基づいて、前記対象のユーザによる前記対象のコンテンツに対する実行処理により得られる期待報酬の事前分布を推定する推定処理と、前記事前分布を用いて、前記期待報酬の事後分布を導出する導出処理と、を含む処理を実行させるためのものである。

[0018] 上記課題を解決するために、本発明による情報処理方法の別の態様は、複数のコンテンツ間の類似度と、複数のユーザ間の類似度を取得する取得工程と、前記複数のコンテンツ間の類似度と前記複数のユーザ間の類似度を用いて、前記複数のコンテンツのうち、期待報酬が最大のコンテンツを、前記複数のユーザのうちの1以上のユーザに適したコンテンツとして決定する決定工程と、を有する。

[0019] 上記課題を解決するために、本発明によるプログラムの別の態様は、情報処理をコンピュータに実行させるための情報処理プログラムであって、該プログラムは、前記コンピュータに、複数のコンテンツ間の類似度と、複数のユーザ間の類似度を取得する取得処理と、前記複数のコンテンツ間の類似度と前記複数のユーザ間の類似度を用いて、前記複数のコンテンツのうち、期待報酬が最大のコンテンツを、前記複数のユーザのうちの1以上のユーザに適したコンテンツとして決定する決定処理と、を含む処理を実行させるためのものである。

[0020] 上記課題を解決するために、本発明による学習モデルの一態様は、対象のコンテンツと1以上の他のコンテンツとの類似度であるコンテンツ間類似度と、対象のユーザと1以上の他のユーザとの類似度であるユーザ間類似度とに基づいて、前記対象のユーザによる前記対象のコンテンツに対する実行処理により得られる期待報酬の事前分布を推定し、前記事前分布を用いて、前記期待報酬の事後分布を導出するように構成される。

発明の効果

[0021] 本発明によれば、機械学習におけるコールドスタート問題による影響を低減することが可能となる。

図面の簡単な説明

[0022] [図1]図1は、情報処理システムの構成例を示す。

[図2]図2は、情報処理装置10の機能構成例を示す。

[図3]図3は、実施形態による学習モデルのアルゴリズムを示す。

[図4]図4は、情報処理装置10とユーザ装置11のハードウェア構成例を示す。

[図5]図5は、情報処理装置10により実行される処理のフローチャートを示す。

[図6]図6は、実施形態による学習モデルを適用した適用例を示す。

発明を実施するための形態

[0023] 以下、添付図面を参照して、本発明を実施するための実施形態について詳細に説明する。以下に開示される構成要素のうち、同一機能を有するものには同一の符号を付し、その説明を省略する。なお、以下に開示される実施形態は、本発明の実現手段としての一例であり、本発明が適用される装置の構成や各種条件によって適宜修正または変更されるべきものであり、本発明は以下の実施形態に限定されるものではない。また、本実施形態で説明されている特徴の組み合わせの全てが本発明の解決手段に必須のものとは限らない。

[0024] [情報処理システムの構成]

図1に、本実施形態による情報処理システムの構成例を示す。本情報処理システムは、その一例として、図1に示すように、情報処理装置10と、任意の複数のユーザ1～Mにより使用される複数のユーザ装置11-1～11-M ($M > 1$) とを含んで構成される。なお、以下の説明において、特に説明がない限り、ユーザ装置11-1～11-Mをユーザ装置11と総称しうる。また、以下の説明において、ユーザ装置とユーザという語は同義に使用されうる。

[0025] ユーザ装置11は、例えば、スマートフォンやタブレットといったデバイスであり、LTE (Long Term Evolution) 等の公衆網や、無線LAN (Local Area Network) 等の無線通信網

を介して、情報処理装置10と通信可能に構成されている。ユーザ装置11は、液晶ディスプレイ等の表示部（表示面）を有し、ユーザ1～Nは、当該液晶ディスプレイに装備されたGUI（Graphic User Interface）により各種操作を行うことができる。当該操作は、指やスタイラス等によりタップ操作、スライド操作、スクロール操作等、画面に表示された画像等のコンテンツに対する各種の操作を含む。

なお、ユーザ装置11は、図1に示すような形態のデバイスに限らず、デスクトップ型のPC（Personal Computer）や、ノート型のPCといったデバイスであってもよい。その場合、ユーザ1～Mによる操作は、マウスやキーボードといった入力装置を用いて行われうる。また、ユーザ装置11は、表示面を別に備えてもよい。

[0026] 情報処理装置10は、有形または無形の商品やサービス（例えば、旅行商品）等のアイテムをレコメンドするためのコンテンツをユーザ装置11に提供し、ユーザ装置11は、ユーザ装置11の表示部に当該コンテンツを表示可能に構成される。本実施形態では、情報処理装置10は、コンテンツとして、各種アイテムに関する広告の画像（広告画像。以下、単に広告とも称する）をユーザ装置11に提供し、ユーザ装置11は、ユーザ装置11の表示部に当該広告を表示可能に構成される。情報処理装置10は、当該広告の提供のために、各種ウェブサイトを提供する。なお、各種ウェブサイトの運営は、情報処理装置10により行われてもよいし、不図示のサーバ装置により行われてもよい。各種ウェブサイトは、例えば、電子商取引サイトや、レストラン予約サイトや、ホテル予約サイト等を含むことができる。

[0027] [情報処理装置10の機能構成]

情報処理装置10は、ユーザ装置11-1～11Mのユーザ1～Mのそれぞれの属性（属性を表す情報）を、ユーザ特徴として取得することができる。また、情報処理装置10は、提供する広告に関する複数の特徴を、広告特徴として取得することができる。情報処理装置10は、取得したユーザ特徴と広告特徴を用いて、後述する学習モデルのアルゴリズムを実行し、ユーザ

装置 11-1 ~ 11-M のうちの任意の 1 つ以上のユーザ装置に適した広告を決定して、当該任意の 1 つ以上のユーザ装置に提供する。当該学習モデルと当該学習モデルを用いた処理については、後述する。

[0028] 本実施形態による情報処理装置 10 の機能構成の一例を図 2 に示す。本実施形態による情報処理装置 10 は、その機能構成の一例として、ユーザ特徴取得部 101、コンテンツ特徴取得部 102、パラメータ設定部 103、推定部 104、および提供部 105 を備える。

[0029] ユーザ特徴取得部 101 は、ユーザ装置 11-1 ~ 11-M のユーザ 1 ~ M のそれぞれの属性を、ユーザ特徴として取得する。当該ユーザ特徴は、性別、年齢、年収、学歴、居住地といった人口統計学的属性（デモグラフィック属性）や、趣味、趣向といった心理学的属性（サイコグラフィック属性）や、過去のインターネットでの検索履歴、閲覧履歴、購買履歴といった行動学的属性（ビヘイビオラル属性）や、特定のアプリケーションによる登録情報等の少なくとも一部を含みうる。

[0030] コンテンツ特徴取得部 102 は、ユーザに提供するコンテンツ（本実施形態では広告）の属性を、コンテンツ特徴として取得する。本実施形態では、当該コンテンツは広告であり、当該コンテンツ特徴（広告特徴）は、広告の対象となるアイテム（有形または無形の商品やサービス（例えば、旅行商品）等）の属性、広告を構成する画像の特徴等を含みうる。

コンテンツ特徴取得部 102 は、過去にユーザに提供したコンテンツだけでなく、将来に提供予定のコンテンツのコンテンツ特徴も取得可能に構成される。

[0031] パラメータ設定部 103 は、推定部 104 により実行される学習モデルのアルゴリズムで必要となる所定のパラメータを設定する。当該パラメータについては後述する。当該パラメータは、予め情報処理装置 10 に設定されていてもよいし、情報処理装置 10 の操作者により入力されてもよい。

[0032] 推定部 104 は、後述する本実施形態による学習モデルのアルゴリズムを実行し、コンテンツに対する実行処理から得られる期待報酬を推定し、任意

のユーザに適したコンテンツを推定する。本実施形態では、推定部104は、当該学習モデルのアルゴリズムを実行し、広告の表示処理により得られる期待報酬を推定し、ユーザ装置11-1~11-Mのうちの任意の1つ以上のユーザ装置に表示するのに適した広告を決定する。また、推定部104は、任意のコンテンツに対して、当該コンテンツが任意のユーザに適しているかを判定することができる。

[0033] 提供部105は、推定部104により決定された広告を、ユーザ装置11に提供する。これにより、ユーザ装置11は、提供された広告を表示部に表示することが可能となる。

[0034] [学習モデルのアルゴリズム]

次に、本実施形態による学習モデルのアルゴリズムについて説明する。本実施形態による学習モデルは、バンディットアルゴリズム (Bandit Algorithm) のためのモデルである。バンディットアルゴリズムは、強化学習 (Reinforcement Learning) のアルゴリズムとして知られており、累積報酬を最大化することを目的としている。具体的には、バンディットアルゴリズムは、アームに対する活用 (Exploitation) と探索 (Exploration) のバランス (活用と探索の割合) を調節することにより、期待報酬を最大にするようにアームを引くことを目的としている。なお、強化学習の分野では、アームは、一般的にはアクションと呼ばれ、以下の説明においてもアクションという用語を用いる。

[0035] 本実施形態による学習モデルのアルゴリズムは、マルチドメインかつマルチユーザで知識 (報酬) を転移させることを特徴とする。当該アルゴリズムでは、アクションとして広告の表示を用い、広告の表示に対する累積報酬を最大化することを目的とする。また、当該アルゴリズムにおいて、各ドメインは、有形または無形の商品やサービス (例えば、旅行商品) 等のアイテムを扱うウェブサイトとする。例えば、電子商取引サイト、レストラン予約サイト、ホテル予約サイトはそれぞれ異なるドメインに対応する。

[0036] 広告は、映画や製品と異なり、新たな販売キャンペーンが開始されたときに作成され、キャンペーンが終了したときに削除されるコンテンツである。したがって、新たに作成された広告の比率は、新たに作成される映画や製品の比率に比べて高くなるため、広告の場合はコールドスタート問題が顕著となりうる。本実施形態では、コールドスタート問題による影響を低減するための学習モデルとして、バンディットアルゴリズムための方策の1つである、公知のトンプソンサンプリング方策を基にした新たな学習モデルを説明する。以下、数式を用いて、本実施形態による学習モデルのアルゴリズムを説明する。

[0037] まず、 N 個の利用可能なソースを仮定する。各ソースは、広告を表示するウィジェット (`w i d g e t`) に対応する。ウィジェットは、端末装置 (本実施形態では、図1における任意のユーザ装置11に対応) の表示部の画面上に小さく広告 (例えば、バナー広告) を表示する機能を有するアプリケーションソフトである。本実施形態では、各ソースにより、複数の広告からなる広告のセットが表示されることを想定する。当該複数の広告は、一定時間ごとに切り替えて表示されるように構成されてもよいし、カルーセル機能を用いて表示されるように構成されてもよい。カルーセル機能は、1つの広告表示枠に対して、ユーザが主体的に操作することにより複数の広告の表示を切り替えることができる機能である。

[0038] ここで、任意のソース s における広告のセット (ソース s により表示可能な広告のセット) を、 A_s とする。また、 M ($M > 0$) ユーザの各ユーザは、それぞれ d_u 種類 ($d_u > 0$) の特徴 (`f e a t u r e`) を有し、 M ユーザの特徴を表すセット (ユーザ特徴セット) を、 X とする。よって、 X は、 $M \times d_u$ のサイズの行列で表される。また、ソース s において、 K_s 個 ($K_s > 0$) の広告のそれぞれが d_a 種類 ($d_a > 0$) の特徴を有し、 K_s 個の広告の特徴を表すセット (広告特徴セット) を、 Y^s とする。よって、 Y^s は、 $K_s \times d_a$ のサイズの行列で表される。

[0039] さらに、時間ステップ t でのソース s におけるユーザ i を、ユーザ i_t^s と

して示す。時間ステップ t でのソース s におけるユーザ特徴：

$$x_{i_t^s} \in X$$

と、時間ステップ t でのソース s における広告特徴：

$$y_{a_t^s} \in Y^s$$

を、観察する。

そして、ユーザは、時間ステップ t でのソース s における広告：

$$a_t^s \in A_s$$

を見て、それによる報酬 (reward)

$$r_{i_t^s a_t^s}$$

を観察する。当該報酬は、ユーザ i_t^s が広告 a_t^s をクリックしたかどうか（広告に対するクリックの有無）を示す、暗黙的な (implicit) 報酬を表す。

よって、全体の観察は、

$$O_t^s = (x_{i_t^s}, y_{a_t^s}, r_{i_t^s a_t^s})$$

として示される。

なお、当該報酬は、広告をクリックし、かつ、コンバージョン（商品購入や資料請求といった最終成果）に至ったかを示す指標に対応するように構成されてもよい。

[0040] 本実施形態による学習モデルは、累積報酬

$$\sum_s \sum_{t=0}^T r_{i_t^s a_t^s}$$

を最大化するときに表示する広告 a_t^s を決定することを目的としている。ここで、 T は、時間ステップ t が取りうる最大値である。累積報酬期待値の最大化は、ユーザ i の総リグレットを最小化するものとして、式 (1) のように表すことができる。

[数1]

$$\text{minimize } E[\text{regret}_i(T)] = \sum_{s=0}^N E \left[\max_{a_t^s \in A_s} \sum_{t=0}^T r_{i_t^s}^* a_t^s - \sum_{t=0}^T r_{i_t^s} a_t^s \right] \quad (1)$$

ここで、 r^* は、ユーザ i_t^s に適した広告表示（すなわち、アクション）から得られる報酬である。

[0041] 本実施形態による学習モデルは、全てのソース s の観察：

$$O = \{O^s\}_{s=1, \dots, N}$$

から、広告を取り出すポリシーを学習するモデルである。また、本実施形態による学習モデルは、ソース間の接続を利用して、ソース間（すなわち、広告間）とユーザ間のそれぞれにおいて、知識を転移させる。これにより、当該ポリシーは、より一般化されたユーザの挙動（behavior）を認識することになる。本実施形態による学習モデルは、ソース間とユーザ間とで、知識としての報酬を転移させる。報酬の転移の度合い（degree）は、一方のオブジェクトの特徴と、対象（target）としての他方のオブジェクトの特徴との類似度に基づく。

[0042] 本実施形態では、ユーザ間の報酬の転移の度合い（すなわち、ユーザ間の類似度）は、コサイン類似度（cosine similarity）を利用して、式（2A）のように表される。

[数2A]

$$S_{user}(x_i, x_j) = \frac{x_i \cdot x_j}{|x_i| |x_j|} \quad (2A)$$

ここで、 x_i と x_j はそれぞれ、ユーザ i に対するユーザ特徴とユーザ j に対するユーザ特徴を示す。前述したように、各ユーザは d_u 種類の特徴を有することから、 x_i は各種類の特徴を示してもよい。あるいは、 x_i は、 d_u 種類の特徴から生成された特徴ベクトルであってもよい。 x_j についても同様である。

[0043] 同様にして、広告間の報酬の移転の度合い（すなわち、広告間の類似度）は、式（2B）のように表される。

[数2B]

$$S_{ad}(y_i, y_j) = \frac{y_i \cdot y_j}{|y_i| |y_j|} \quad (2B)$$

ここで、 y_i と y_j はそれぞれ、広告*i*に対する広告特徴と広告*j*に対する広告特徴を示す。前述したように、各広告は d_a 種類の特徴を有することから、 y_i は各種類の特徴を示してもよい。あるいは、 y_i は、 d_a 種類の特徴から生成された特徴ベクトルであってもよい。 y_j についても同様である。

[0044] なお、広告*i*と広告*j*は、同じドメインから選択されてもよいし、異なるドメインから取得されてもよい。前述のように、ドメインが、有形または無形の商品やサービス（例えば、旅行商品）等のアイテムを扱うウェブサイトである場合、広告*i*が電子商取引サイトにおける広告であって、広告*j*がレストラン予約サイトにおける広告であってもよい。

[0045] 現実世界のデータセットにおいて、特にユーザの数は膨大であり、ユーザの全てのペアの間の類似度を計算することは困難である。よって、実装時には、上記の類似度を効率的に得るために、公知の局所性鋭敏型ハッシュ（locality sensitive hashing）を利用してもよい。

[0046] 前述のように、本実施形態による学習モデルは、トムソンサンプリング方策を基にしている。トムソンサンプリング方策は、期待報酬を最大にするアームを引くために、事前分布を基に導出した、各ラウンドにおけるアームの事後分布から期待される複数のサンプルスコアのうち、最も高いスコアを有するアームを選択する手法である。ベルヌーイバンディットの場合、尤度関数は、ベルヌーイ分布によって定式化され、事前分布は、自然共役事前確率分布（natural conjugate prior）として、ベータ分布によって表される。

ベータ分布関数は、式（3）のように表すことができる。

[数3]

$$p(\theta_k) = \frac{\Gamma(\alpha_k + \beta_k)}{\Gamma(\alpha_k)\Gamma(\beta_k)} \theta_k^{\alpha_k - 1} (1 - \theta_k)^{\beta_k - 1} \quad (3)$$

ここで、 Γ はガンマ関数を表す。本実施形態による学習モデルに照らすと、 θ_k は、広告 k の表示（すなわち、アクション）により報酬がもたらされる確率であり、 α_k と β_k はそれぞれ、広告 k に対する正報酬と負報酬を表すパラメータである。

[0047] オリジナルのトンプソンサンプリングでは、 $\alpha_k = 1$ かつ $\beta_k = 1$ という一様分布のケースを想定していたが、本実施形態による学習モデルでは、履歴データを利用して、事前分布を推定する。具体的には、前述のユーザと広告の類似度関数を利用する。これにより、事前分布のより良い推定を提供することを可能にする。

まず、時間ステップ t での対象のユーザ i と対象の広告 k に対する事前推定のための正報酬 (α) と負報酬 (β) を表すパラメータを、以下の式 (4) のように定式化する。

[数4]

$$\begin{aligned}\alpha_{ik}^0(t) &= \sum_{l \neq k} S_{ad}(y_k, y_l) s_{il}(t) + \sum_{j \neq i} S_{user}(x_i, x_j) s_{jk}(t) \\ \beta_{ik}^0(t) &= \sum_{l \neq k} S_{ad}(y_k, y_l) f_{il}(t) + \sum_{j \neq i} S_{user}(x_i, x_j) f_{jk}(t)\end{aligned}\quad (4)$$

ここで、 $s_{il}(t)$ は、割引を意識した (discount-aware) 累積正報酬であり、式 (5A) のように表される。

[数5A]

$$s_{il}(t) = \sum_{\tau=0}^t \gamma^{t-\tau} s_{il\tau} \quad (5A)$$

式 (5A) において、 $s_{il\tau}$ は、時間 τ において対象のユーザ i と他の広告 l に対する報酬が観察される場合に 1 であり、それ以外の場合は 0 である、バイナリ変数である。また、 γ は、割引率を示す。割引率 γ に $(t - \tau)$ の乗数が掛けられることにより、時間 τ が大きく時間 t に近いほど、割引率は低くなる。すなわち、 $s_{il}(t)$ は、ユーザの挙動による報酬の時間変化に対応し、ユーザによる過去の挙動より、時間 t に近い時間の（すなわち、最近の）挙動が大きく反映される。

[0048] 同様に、 $f_{jk}(t)$ は、 $f_{jk\tau}$ と割引率 γ よる割引を意識した累積負報酬として (5B) 式のように定義される。ここで、 $f_{jk\tau}$ は、推奨 (recommendation) の失敗の数を表し、他のユーザ j が時間 τ において対象の広告 k を見ているが当該広告をクリックしなかった場合に、1 になる。

[数5B]

$$f_{jk}(t) = \sum_{\tau=0}^t \gamma^{t-\tau} f_{jk\tau} \quad (5B)$$

[0049] このように、対象のユーザ i に対する対象の広告 k に対する事前推定のための正報酬 (α) と負報酬 (β) を表すパラメータは、ユーザ間の類似度 (対象のユーザと 1 以上の他のユーザとの類似度 ($\sum S_{user}$)) と広告間の類似度 (対象の広告と 1 以上の他の広告間との類似度 ($\sum S_{ad}$)) によって推定することができる。

したがって、ユーザ間の類似度と広告間の類似度に基づいて、報酬を転移させる。また、ユーザの好みは時間とともに変化するものであるから、累積正報酬 $s_{ij}(t)$ に割引率 γ を導入し、ユーザの最近の挙動からの報酬に、高い値を与えることができる。本実施形態では、上記 (4)、(5A)、(5B) 式で示したように、対象のユーザ i による他の広告 l に対する実行処理 (広告表示) で得られた報酬と、他のユーザ j による対象の広告 k に対する実行処理の報酬を、時間の経過によって割引き (あるいは逓減) する。そして、これらの割引き後の報酬を用いて事前分布が推定される。なお、割引率の設定はオプションであってもよい。

[0050] 事前分布の推定後に、当該事前分布を用いた事後分布を導出する。オリジナルのトンプソンサンプリング方策と同様に、事後分布をベータ分布によって定式化する。オリジナルのトンプソンサンプリング方策の場合、一様分布の事前分布を用いて、事後ベータ分布のパラメータは、 $\alpha_k = s_k + 1$ 、 $\beta_k = f_k + 1$ であった。

本実施形態では、式 (4) の事前知識を用いて、事後ベータ分布のパラメータを式 (6) のように定式化する。

[数6]

$$\begin{aligned}\alpha_{ik}(t) &= \lambda(s)\alpha_{ik}^0(t) + gs_k(t) + s_{ik}(t) + 1 \\ \beta_{ik}(t) &= \lambda(f)\beta_{ik}^0(t) + gf_k(t) + f_{ik}(t) + 1\end{aligned}\quad (6)$$

ここで、 $\lambda(s)$ と $\lambda(f)$ は、事前知識の重要性を調整するハイパーパラメータであり、

$$\lambda(s) = \frac{\lambda}{s_{ik}(t) + 1}$$

$$\lambda(f) = \frac{\lambda}{f_{ik}(t) + 1}$$

である。

また、 g は、ユーザ間と広告間で報酬を転移させた、グローバルな報酬の重要性を調整するハイパーパラメータである。

また、オリジナルのトンプソンサンプリング方策と同様に、 $s_k(t)$ と $f_k(t)$ が、ユーザ間の報酬の平均として組み込まれている。これは、現実世界のケースでは、有効な類似するユーザがほとんど発生せず、ユーザと広告の相互作用 (interaction) は、疎 (sparse) であるからである。最後の項の「1」は、履歴的報酬が利用可能でなかった場合のエラーを避けるための擬似カウントである。

[0051] このように、本実施形態による学習モデルのアルゴリズムは、オリジナルのトンプソンサンプリング方策をベースとして、ユーザ類似度とコンテンツの類似度を動的に用いて推論を行うことから、動的協調フィルタリングトンプソンサンプリング方策 (Dynamic collaborative filtering Thompson sampling) と称することができる。

[0052] 図3に、本実施形態による学習モデルのアルゴリズム (ポリシー) をアルゴリズム1として示す。当該アルゴリズムは、情報処理装置10の推定部104により実行される。任意の対象ユーザ*i*に対する図3のアルゴリズムの処理を順に説明する。ここでは、図3の各処理1~10を、S1~S10と示す。

[0053] まず、ハイパーパラメータ λ と g 、割引率 γ 、ユーザ間の類似度 (S_{user})、広告間の類似度 (S_{ad}) を入力する。また、過去の観察 O を入力する。その後、 $S1 \sim S10$ の処理が行われる。

$S1$: 時間ステップ t が $0 \sim T$ の間、

$S2$: 時間ステップ t でのソース s におけるユーザ i のユーザ特徴 (ユーザのコンテキスト特徴) と、広告 (アクション) のセット A_s と、それらの広告特徴 (広告のコンテキスト特徴) のセット Y_s を観察する

$S3$: A_s に含まれる全ての広告 k ($k \in A_s$) に対して、

$S4$: 式 (4) に従い、時間ステップ t でのユーザ i と広告 k に対する事前推定のための正報酬 (α) と負報酬 (β) を表すパラメータを計算する

$S5$: 式 (6) に従い、事後ベータ分布の正報酬 (α) と負報酬 (β) を表すパラメータを計算する

$S6$: 事後ベータ分布のパラメータ α と β を用いたベータ分布から、 θ_k をサンプリングする

$S7$: $S4 \sim S6$ の処理のループの終了

$S8$: θ_k の最大値を与える広告 k の表示を行い、報酬 r を観察する

$S9$: 観察 O を追加

$S10$: $S2 \sim S9$ の処理のループの終了

[0054] このように、ユーザの類似性と広告の類似性に基づいて複数のドメイン (本実施形態では複数のウェブサイト) 間で報酬を転移させ、観察 O を繰り返すことにより、任意のユーザ i に適した広告表示を継続して行うことが可能となる。また、当該転移の手法により、履歴情報が少ない新たな広告について報酬を評価することができ、コールドスタート問題による影響を低減させることが可能となる。

なお、前述の事前分布と事後分布のために用いる関数は、ベータ分布関数に限定されない。例えば、ガウス分布関数も用いることが可能である。

[0055] [情報処理装置 10 のハードウェア構成]

図4は、本実施形態による情報処理装置10のハードウェア構成の一例を示すブロック図である。

本実施形態による情報処理装置10は、単一または複数の、あらゆるコンピュータ、モバイルデバイス、または他のいかなる処理プラットフォーム上にも実装することができる。

図4を参照して、情報処理装置10は、単一のコンピュータに実装される例が示されているが、本実施形態による情報処理装置10は、複数のコンピュータを含むコンピュータシステムに実装されてよい。複数のコンピュータは、有線または無線のネットワークにより相互通信可能に接続されてよい。

[0056] 図4に示すように、情報処理装置10は、CPU41と、ROM42と、RAM43と、HDD44と、入力部45と、表示部46と、通信I/F47と、システムバス48とを備えてよい。情報処理装置10はまた、外部メモリを備えてよい。

CPU (Central Processing Unit) 41は、情報処理装置10における動作を統括的に制御するものであり、データ伝送路であるシステムバス48を介して、各構成部(42~47)を制御する。

[0057] ROM (Read Only Memory) 42は、CPU41が処理を実行するために必要な制御プログラム等を記憶する不揮発性メモリである。なお、当該プログラムは、HDD (Hard Disk Drive) 44、SSD (Solid State Drive) 等の不揮発性メモリや着脱可能な記憶媒体(不図示)等の外部メモリに記憶されていてもよい。

RAM (Random Access Memory) 43は、揮発性メモリであり、CPU51の主メモリ、ワークエリア等として機能する。すなわち、CPU41は、処理の実行に際してROM42から必要なプログラム等をRAM43にロードし、当該プログラム等を実行することで各種の機能動作を実現する。

[0058] HDD44は、例えば、CPU41がプログラムを用いた処理を行う際に必要な各種データや各種情報等を記憶している。また、HDD44には、例

例えば、CPU 41がプログラム等を用いた処理を行うことにより得られた各種データや各種情報等が記憶される。

入力部45は、キーボードやマウス等のポインティングデバイスにより構成される。

表示部46は、液晶ディスプレイ（LCD）等のモニターにより構成される。表示部56は、入力部55と組み合わせて構成されることにより、GUI（Graphical User Interface）として機能してもよい。

[0059] 通信I/F47は、情報処理装置10と外部装置との通信を制御するインタフェースである。

通信I/F47は、ネットワークとのインタフェースを提供し、ネットワークを介して、外部装置との通信を実行する。通信I/F47を介して、外部装置との間で各種データや各種パラメータ等が送受信される。本実施形態では、通信I/F47は、イーサネット（登録商標）等の通信規格に準拠する有線LAN（Local Area Network）や専用線を介した通信を実行してよい。ただし、本実施形態で利用可能なネットワークはこれに限定されず、無線ネットワークで構成されてもよい。この無線ネットワークは、Bluetooth（登録商標）、ZigBee（登録商標）、UWB（Ultra Wide Band）等の無線PAN（Personal Area Network）を含む。また、Wi-Fi（Wireless Fidelity）（登録商標）等の無線LAN（Local Area Network）や、WiMAX（登録商標）等の無線MAN（Metropolitan Area Network）を含む。さらに、LTE/3G、4G、5G等の無線WAN（Wide Area Network）を含む。なお、ネットワークは、各機器を相互に通信可能に接続し、通信が可能であればよく、通信の規格、規模、構成は上記に限定されない。

[0060] 図4に示す情報処理装置10の各要素のうち少なくとも一部の機能は、CPU41がプログラムを実行することで実現することができる。ただし、図

4に示す情報処理装置10の各要素のうち少なくとも一部の機能が専用のハードウェアとして動作するようにしてもよい。この場合、専用のハードウェアは、CPU41の制御に基づいて動作する。

[0061] [ユーザ装置11のハードウェア構成]

図1に示すユーザ装置11のハードウェア構成は、図4と同様でありうる。すなわち、ユーザ装置11は、CPU41と、ROM42と、RAM43と、HDD44と、入力部45と、表示部46と、通信I/F47と、システムバス48とを備えうる。ユーザ装置11は、情報処理装置10により提供された各種情報を、表示部46に表示し、GUI（入力部45と表示部46による構成）を介してユーザ1から受け付ける入力操作に対応する処理を行うことができる。

[0062] [処理の流れ]

図5に、本実施形態による情報処理装置10により実行される処理のフローチャートを示す。図5に示す処理は、情報処理装置10のCPU41がROM42等に格納されたプログラムをRAM43にロードして実行することによって実現されうる。図5に示す処理を、図6に示す適用例に沿って説明する。図6は、図1に示した情報処理システムを参照した、本実施形態による学習モデルを適用した適用例を示す。

[0063] S51において、ユーザ特徴取得部101は、広告提供の対象となるユーザ（対象ユーザ）を特定する。ユーザ特徴取得部101は、対象ユーザとして、すでに広告を提供していたユーザだけでなく、広告を提供していなかった新たなユーザも特定することができる。

図6の例では、ユーザ特徴取得部101は、ユーザ1~Mのうち、すでに広告を提供していたユーザiを対象ユーザとして特定する。

[0064] S52において、ユーザ特徴取得部101は、対象ユーザを含む複数のユーザ（ユーザ1~M）のユーザ特徴を取得する。ユーザ特徴には、過去のインターネットでの検索履歴等、時間と共に変化する属性も含まれるため、ユーザ特徴取得部101は、定期的に、または任意のタイミングで、ユーザ

特徴を取得してもよい。

[0065] S 5 3 において、ユーザ特徴取得部 1 0 1 は、S 5 2 で取得したユーザ特徴を用いて、対象のユーザと、当該対象のユーザ以外の他のユーザ間との間の類似度 (S_{user}) を算出する。S 5 3 の処理は、前述の式 (2 A) に関する処理に対応する。

[0066] S 5 4 において、コンテンツ特徴取得部 1 0 2 は、コンテンツとしての広告の特徴を取得する。S 5 4 では、コンテンツ特徴取得部 1 0 2 は、複数のウェブサイトにおける複数の広告の特徴を取得する。また、コンテンツ特徴取得部 1 0 2 は、新たに作成され、まだあらゆるユーザに提供していない広告についての特徴を取得することができる。

図 6 の例では、コンテンツ特徴取得部 1 0 2 は、ドメイン A (例えば、電子商取引サイト) における広告の特徴と、ドメイン B (例えば、レストラン予約サイト) における広告の特徴を取得する。さらに、コンテンツ特徴取得部 1 0 2 は、新たに作成されてユーザに提供されていないドメイン C (例えば、ホテル予約サイト) における広告の特徴を取得する。ドメイン C における広告は新規な広告であり、ユーザに提供されていない。そのため、ドメイン C の広告に関する履歴情報は蓄積されていない。

[0067] S 5 5 において、コンテンツ特徴取得部 1 0 2 は、複数のコンテンツ間 (本例では広告間) の類似度 (S_{ad}) を算出する。S 5 5 の処理は、前述の式 (2 B) に関する処理に対応する。

図 6 の例では、ドメイン A における広告の特徴、ドメイン B における広告の特徴、およびドメイン C における広告の特徴に対して、あらゆる 2 つの組み合わせの広告の特徴の類似度を算出する。

なお、S 5 1 ~ S 5 5 の処理の順序は図 5 の順序に限定されない。

[0068] S 5 6 において、推定部 1 0 4 は対象ユーザに適したコンテンツを推定により決定する。まず、推定部 1 0 4 は、アルゴリズム 1 を実行するために、パラメータ設定部 1 0 3 から、ハイパーパラメータ λ と g 、割引率 γ を取得し、ユーザ特徴取得部 1 0 1 とコンテンツ特徴取得部 1 0 2 それぞれから、

ユーザ間の類似度 (S_{user})、広告間の類似度 (S_{ad}) を取得する。続いて、推定部 104 は、対象ユーザをユーザ i として設定し、ユーザ i に対するアルゴリズム 1 を実行する。これにより、推定部 104 は、対象ユーザに適する広告を決定する。

なお、推定部 104 は、S56 の処理の対象となった各広告について、得られた期待報酬に基づき、対象ユーザに提供するか否かを判定してもよい。

[0069] 図 6 の例では、ユーザ i は、ポイントを獲得するために、電子商取引サイト (ドメイン A) 上でポイント関連広告 (広告 61A、62A) をクリックすることに関心があった。最近、ユーザ i が美味しい食べものを食べることに興味を持ち、ドメイン A 上でフード関連広告 (広告 63A) をクリックするようになった。このように、ユーザ i は、時間と共に好みが変わる。このようなユーザ i による動的な挙動は、アルゴリズム 1 における $s_{ij}(t)$ に反映される。すなわち、過去の挙動より最近の挙動に価値が置かれるように事前分布における正報酬 (α) のパラメータが決定される。よって、ドメイン A 上でフード関連広告 (広告 63A) がクリックされていることから、推定部 104 は、広告 61A や広告 62A より、広告 63A に特徴がより類似する広告に、高い報酬期待値を与える (式 (1))。図 6 の例の場合、ドメイン B (レストラン予約サイト) におけるフード関連広告 (広告 61B) や、ドメイン C (ホテル予約サイト) における旅行関連広告 (広告 61C) に高い報酬期待値が与えられる。ドメイン C における広告は、ユーザに提供されていないため、当該ドメインに関する履歴情報は蓄積されていないが、広告間の類似性とユーザ間の類似性を用いることにより、推定した報酬を利用することができる。最終的に、推定部 104 は、広告 61B と広告 61C が、ユーザ i に適した広告として決定することができる。

[0070] S57 において、提供部 105 は、S56 において決定された広告を、対象ユーザに提供する。図 6 の例では、提供部 105 は、広告 61B と広告 61C を、ユーザ i に提供する。すなわち、提供部 105 は、ユーザ装置 11-i の表示部に広告 61B と広告 61C を表示させるように制御する。

[0071] 上記説明では、図6を参照して、S51において、対象ユーザとしてユーザ*i*が特定される例を説明したが、続いて、ユーザ*j*が特定される例について説明する。本例では、ユーザ*j*は、ユーザ*i*のユーザ特徴と類似したユーザ特徴を有するものとする。この場合、推定部104は、上記アルゴリズムを実行することにより、ユーザ*i*に適する広告と類似する広告を、ユーザ*j*に適した広告と決定することができる。図6の例では、推定部104が、ドメインAにおける広告64Aを、ユーザ*j*に適した広告として決定し、提供部105がユーザ*j*に提供することができる。

[0072] このように、コンテンツ（広告）やユーザに対する事前の履歴情報が少ない、もしくは存在しない場合であっても、コンテンツ間の類似度およびユーザ間の類似度を用いることにより、任意のユーザに適したコンテンツを推定して決定することができる。

[0073] <その他の実施形態>

上記実施形態では、コンテンツとして広告を用いて説明したが、あらゆるコンテンツに対して、本実施形態を適用可能である。例えば、コンテンツとして、映画、書籍、または各種商品を用いてもよい。

[0074] また、上記実施形態では、コンテンツのレコメンデーションシステムを前提に説明したが、複数のドメイン間で報酬を転移させると共に、サービス提供対象（例えばユーザ）間で報酬を転移させる本実施形態による学習モデルは、あらゆる分野に対して適用可能である。例えば、ファイナンス分野において、複数のポートフォリオから顧客に最適なポートフォリオを選択するために、本実施形態による学習モデルを適用可能である。また、ヘルスケア分野において、治療方法や薬を患者に提供するために、本実施形態による学習モデルを適用可能である。また、ダイアログシステム分野において、会話エージェントによるシステム構築のために、すなわち、複数の会話システム（それぞれはドメインに対応）を統一して1つのシステムを構築するために、本実施形態による学習モデルを適用可能である。

[0075] なお、上記において特定の実施形態が説明されているが、当該実施形態は

単なる例示であり、本発明の範囲を限定する意図はない。本明細書に記載された装置及び方法は上記した以外の形態において具現化することができる。また、本発明の範囲から離れることなく、上記した実施形態に対して適宜、省略、置換及び変更をなすこともできる。かかる省略、置換及び変更をなした形態は、請求の範囲に記載されたもの及びこれらの均等物の範疇に含まれ、本発明の技術的範囲に属する。

符号の説明

[0076] 1～M：ユーザ、10：情報処理装置、11-1～11-M：ユーザ装置、
101：ユーザ特徴取得部、102：コンテンツ特徴取得部、103：パラ
メータ設定部、104：推定部105：提供部

請求の範囲

- [請求項1] 対象のコンテンツと1以上の他のコンテンツとの類似度であるコンテンツ間類似度と、対象のユーザと1以上の他のユーザとの類似度であるユーザ間類似度と、を取得する取得手段と、
- 前記コンテンツ間類似度と前記ユーザ間類似度とに基づいて、前記対象のユーザによる前記対象のコンテンツに対する実行処理により得られる期待報酬の事前分布を推定する推定手段と、
- 前記事前分布を用いて、前記期待報酬の事後分布を導出する導出手段と、
- を有することを特徴とする情報処理装置。
- [請求項2] 前記取得手段は、前記対象のコンテンツと前記1以上の他のコンテンツの特徴を用いて、前記コンテンツ間類似度を取得し、前記対象のユーザと前記1以上の他のユーザの特徴を用いて前記ユーザ間類似度を取得することを特徴とする請求項1に記載の情報処理装置。
- [請求項3] 前記推定手段は、前記対象のユーザによる前記他のコンテンツに対する実行処理により得られた第1の報酬を用いて、前記事前分布を推定することを特徴とする請求項1または2に記載の情報処理装置。
- [請求項4] 前記第1の報酬は、時間の経過による報酬の割引により、前記対象のユーザによる前記他のコンテンツに対する過去の実行処理より最近の実行処理により得られた報酬が高くなるように構成されることを特徴とする請求項3に記載の情報処理装置。
- [請求項5] 前記推定手段は、前記他のユーザによる前記対象のコンテンツに対する実行処理により得られた第2の報酬を用いて、前記事前分布を推定することを特徴とする請求項1から4のいずれか1項に記載の情報処理装置。
- [請求項6] 前記第2の報酬は、時間の経過による報酬の割引により、前記他のユーザによる前記対象のコンテンツに対する過去の実行処理より最近の実行処理により得られた報酬が高くなるように構成されることを特

徴とする請求項5に記載の情報処理装置。

[請求項7] 前記導出手段により導出された前記期待報酬の事後分布に基づいて、前記対象のコンテンツを前記対象のユーザに提供するかを判定する判定手段をさらに有することを特徴とする請求項1から6のいずれか1項に記載の情報処理装置。

[請求項8] 前記コンテンツは、有形または無形の商品またはサービスに関する広告であり、前記実行処理は、広告の表示処理であり、前記報酬は、前記広告に対するクリックの有無を示すことを特徴とする請求項1から7のいずれか1項に記載の情報処理装置。

[請求項9] 複数のコンテンツ間の類似度と、複数のユーザ間の類似度を取得する取得手段と、

前記複数のコンテンツ間の類似度と前記複数のユーザ間の類似度を用いて、前記複数のコンテンツのうち、期待報酬が最大のコンテンツを、前記複数のユーザのうちの1以上のユーザに適したコンテンツとして決定する決定手段と、

を有することを特徴とする情報処理装置。

[請求項10] 対象のコンテンツと1以上の他のコンテンツとの類似度であるコンテンツ間類似度と、対象のユーザと1以上の他のユーザとの類似度であるユーザ間類似度と、を取得する取得工程と、

前記コンテンツ間類似度と前記ユーザ間類似度とに基づいて、前記対象のユーザによる前記対象のコンテンツに対する実行処理により得られる期待報酬の事前分布を推定する推定工程と、

前記事前分布を用いて、前記期待報酬の事後分布を導出する導出工程と、

を有する情報処理方法。

[請求項11] 情報処理をコンピュータに実行させるための情報処理プログラムであって、該プログラムは、前記コンピュータに、

対象のコンテンツと1以上の他のコンテンツとの類似度であるコン

テンツ間類似度と、対象のユーザと1以上の他のユーザとの類似度であるユーザ間類似度と、を取得する取得処理と、

前記コンテンツ間類似度と前記ユーザ間類似度とに基づいて、前記対象のユーザによる前記対象のコンテンツに対する実行処理により得られる期待報酬の事前分布を推定する推定処理と、

前記事前分布を用いて、前記期待報酬の事後分布を導出する導出処理と、を含む処理を実行させるためのものである、

情報処理プログラム。

[請求項12] 複数のコンテンツ間の類似度と、複数のユーザ間の類似度を取得する取得工程と、

前記複数のコンテンツ間の類似度と前記複数のユーザ間の類似度を用いて、前記複数のコンテンツのうち、期待報酬が最大のコンテンツを、前記複数のユーザのうちの1以上のユーザに適したコンテンツとして決定する決定工程と、

を有することを特徴とする情報処理方法。

[請求項13] 情報処理をコンピュータに実行させるための情報処理プログラムであって、該プログラムは、前記コンピュータに、

複数のコンテンツ間の類似度と、複数のユーザ間の類似度を取得する取得処理と、

前記複数のコンテンツ間の類似度と前記複数のユーザ間の類似度を用いて、前記複数のコンテンツのうち、期待報酬が最大のコンテンツを、前記複数のユーザのうちの1以上のユーザに適したコンテンツとして決定する決定処理と、を含む処理を実行させるためのものである、

、

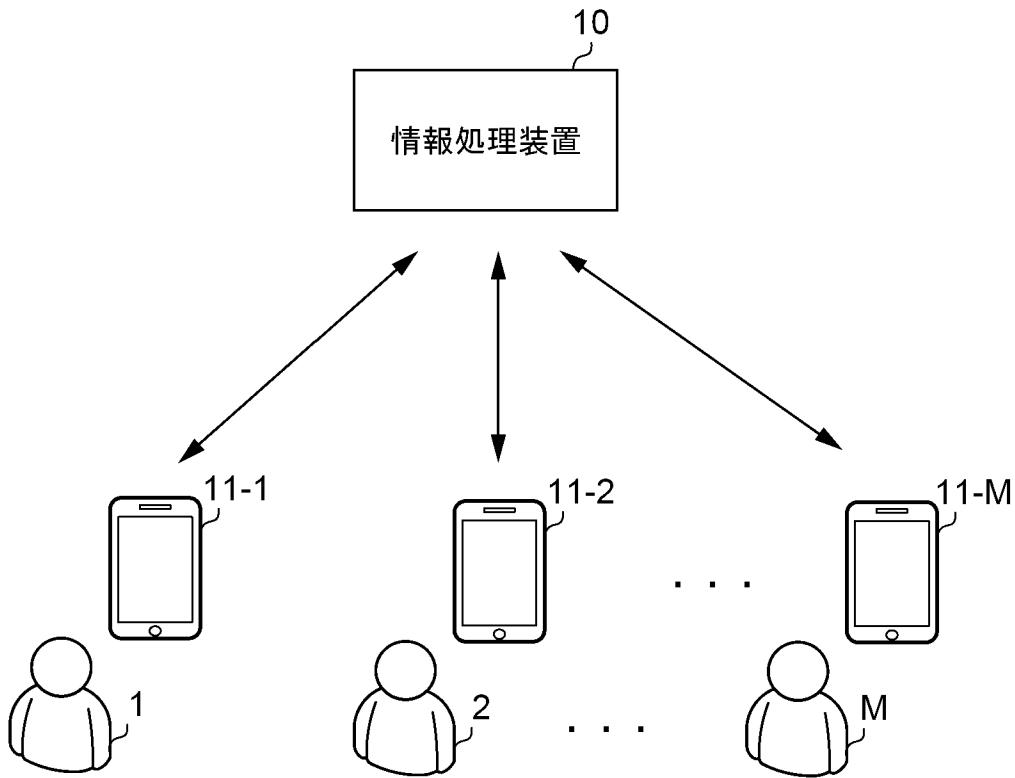
情報処理プログラム。

[請求項14] 対象のコンテンツと1以上の他のコンテンツとの類似度であるコンテンツ間類似度と、対象のユーザと1以上の他のユーザとの類似度であるユーザ間類似度とに基づいて、前記対象のユーザによる前記対象

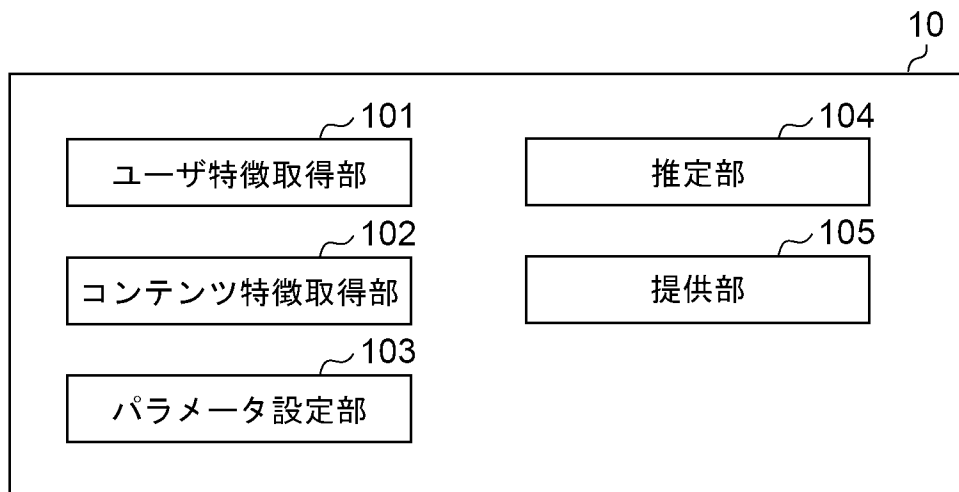
のコンテンツに対する実行処理により得られる期待報酬の事前分布を推定し、

前記事前分布を用いて、前記期待報酬の事後分布を導出するように構成された学習モデル。

[図1]



[図2]

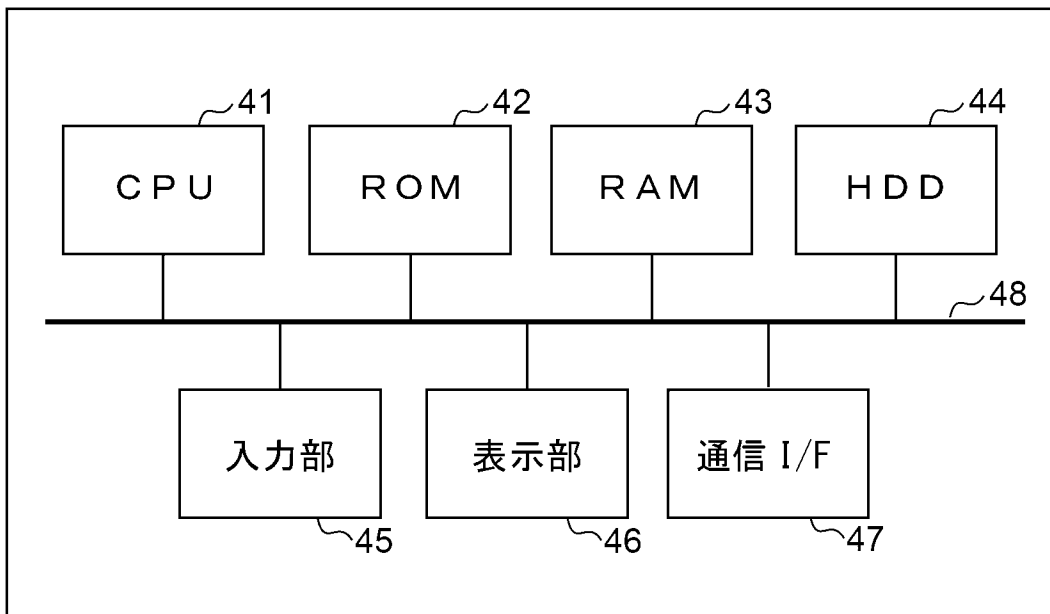


[図3]

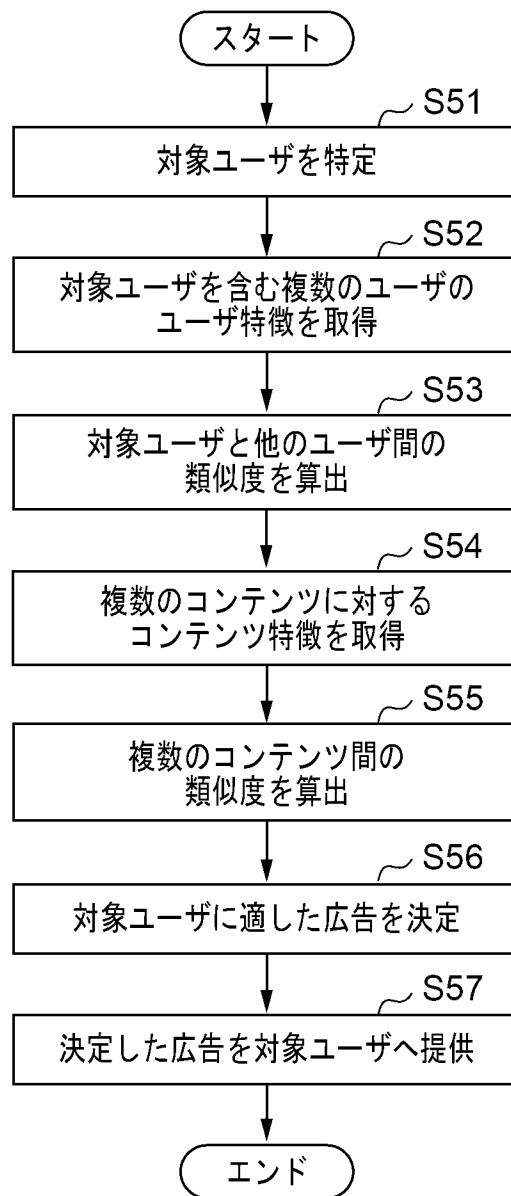
Algorithm 1 Dynamic collaborative filtering Thompson sampling**Input:** $\lambda, g, \gamma \in \mathbb{R}_+^0, S_{user}, S_{ad}$, Source observations \mathbf{O}

-
- 1: **for** $t = 0, \dots, T$ **do**
 - 2: Observe user i_t^s and context $x_{i_t^s}$, action sets \mathbf{A}_s and their contexts \mathbf{Y}^s
 - 3: **for** $k \in \mathbf{A}_s$ **do**
 - 4: Calculate $\alpha_{i_t^s k}^0(t)$ and $\beta_{i_t^s k}^0(t)$ according to Eq. 4
 - 5: Calculate $\alpha_{i_t^s k}(t)$ and $\beta_{i_t^s k}(t)$ according to Eq. 6
 - 6: Sample θ_k from the $\mathcal{B}(\alpha_{i_t^s k}(t), \beta_{i_t^s k}(t))$
 - 7: **end for**
 - 8: Play action $k = \arg \max_k \theta_k$ and observe reward $r_{i_t^s k t}$
 - 9: Add observation $O_t^s \leftarrow (x_{i_t^s}, k, r_{i_t^s k t})$
 - 10: **end for**
-

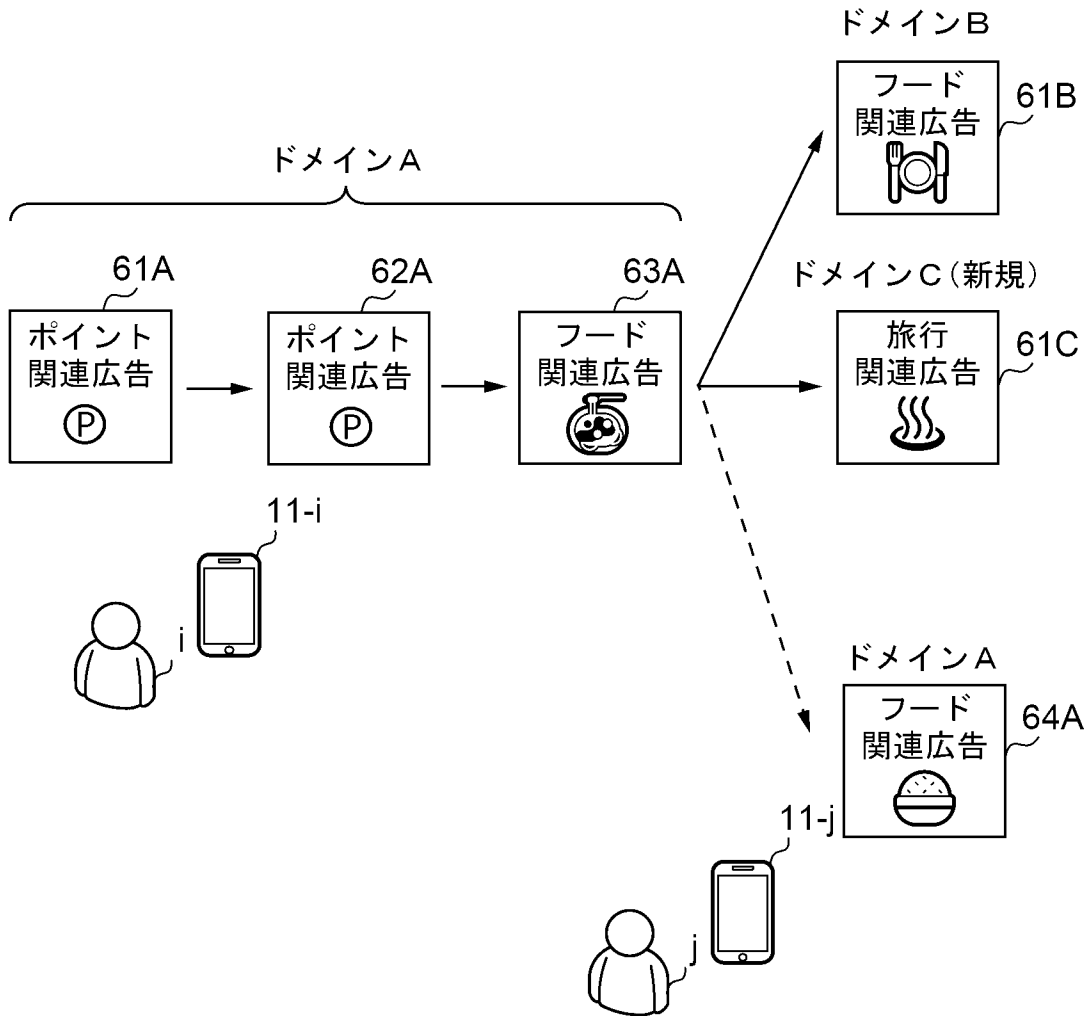
[図4]



[図5]



[図6]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2022/005602

A. CLASSIFICATION OF SUBJECT MATTER		
<i>G06Q 30/02</i> (2012.01)i FI: G06Q30/02 398		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED		
Minimum documentation searched (classification system followed by classification symbols) G06Q10/00-99/00; G06F16/00-16/958		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Published examined utility model applications of Japan 1922-1996 Published unexamined utility model applications of Japan 1971-2022 Registered utility model specifications of Japan 1996-2022 Published registered utility model applications of Japan 1994-2022		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2017/0061481 A1 (STAPLES, INC.) 02 March 2017 (2017-03-02) entire text, all drawings	1-14
A	US 2018/0211303 A1 (WAL-MART STORES, INC.) 26 July 2018 (2018-07-26) entire text, all drawings	1-14
A	US 2017/0278114 A1 (XEROX CORP.) 28 September 2017 (2017-09-28) entire text, all drawings	1-14
A	US 2017/0098236 A1 (YAHOO! INC.) 06 April 2017 (2017-04-06) entire text, all drawings	1-14
A	US 2019/0080348 A1 (YAHOO HOLDINGS, INC.) 14 March 2019 (2019-03-14) entire text, all drawings	1-14
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family		
Date of the actual completion of the international search 09 May 2022		Date of mailing of the international search report 17 May 2022
Name and mailing address of the ISA/JP Japan Patent Office (ISA/JP) 3-4-3 Kasumigaseki, Chiyoda-ku, Tokyo 100-8915 Japan		Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/JP2022/005602

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
US	2017/0061481	A1	02 March 2017	CA 2940205 A1 entire text, all drawings	
US	2018/0211303	A1	26 July 2018	(Family: none)	
US	2017/0278114	A1	28 September 2017	(Family: none)	
US	2017/0098236	A1	06 April 2017	(Family: none)	
US	2019/0080348	A1	14 March 2019	(Family: none)	

A. 発明の属する分野の分類（国際特許分類（IPC）） G06Q 30/02(2012.01)i FI: G06Q30/02 398		
B. 調査を行った分野 調査を行った最小限資料（国際特許分類（IPC）） G06Q10/00-99/00; G06F16/00-16/958 最小限資料以外の資料で調査を行った分野に含まれるもの 日本国実用新案公報 1922-1996年 日本国公開実用新案公報 1971-2022年 日本国実用新案登録公報 1996-2022年 日本国登録実用新案公報 1994-2022年		
国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）		
C. 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
A	US 2017/0061481 A1 (STAPLES, INC.) 02.03.2017 (2017-03-02) 全文, 全図	1-14
A	US 2018/0211303 A1 (WAL-MART STORES, INC.) 26.07.2018 (2018-07-26) 全文, 全図	1-14
A	US 2017/0278114 A1 (XEROX CORP.) 28.09.2017 (2017-09-28) 全文, 全図	1-14
A	US 2017/0098236 A1 (YAHOO! INC.) 06.04.2017 (2017-04-06) 全文, 全図	1-14
A	US 2019/0080348 A1 (YAHOO HOLDINGS, INC.) 14.03.2019 (2019-03-14) 全文, 全図	1-14
<input type="checkbox"/> C欄の続きにも文献が列挙されている。 <input checked="" type="checkbox"/> パテントファミリーに関する別紙を参照。		
* 引用文献のカテゴリー “A” 特に関連のある文献ではなく、一般的な技術水準を示すもの “E” 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの “L” 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す） “O” 口頭による開示、使用、展示等に言及する文献 “P” 国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献	“T” 国際出願日又は優先日後に公表された文献であって出願と抵触するものではなく、発明の原理又は理論の理解のために引用するもの “X” 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの “Y” 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの “&” 同一パテントファミリー文献	
国際調査を完了した日 09.05.2022	国際調査報告の発送日 17.05.2022	
名称及びあて先 日本国特許庁(ISA/JP) 〒100-8915 日本国 東京都千代田区霞が関三丁目4番3号	権限のある職員（特許庁審査官） 後藤 昂彦 5L 6291 電話番号 03-3581-1101 内線 3562	

国際調査報告
特許ファミリーに関する情報

国際出願番号

PCT/JP2022/005602

引用文献	公表日	特許ファミリー文献	公表日
US 2017/0061481 A1	02.03.2017	CA 2940205 A1 全文, 全図	
US 2018/0211303 A1	26.07.2018	(ファミリーなし)	
US 2017/0278114 A1	28.09.2017	(ファミリーなし)	
US 2017/0098236 A1	06.04.2017	(ファミリーなし)	
US 2019/0080348 A1	14.03.2019	(ファミリーなし)	