(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7440703号
(P7440703)

(45)発行日　令和6年2月28日(2024.2.28)

(24)登録日　令和6年2月19日(2024.2.19)

(51)国際特許分類
Ｇ０６Ｑ　30/02　(2023.01)

ＦＩ
Ｇ０６Ｑ　30/02

請求項の数　13　（全19頁）

(21)出願番号　特願2023-501077(P2023-501077)
(86)(22)出願日　令和4年2月14日(2022.2.14)
(86)国際出願番号　PCT/JP2022/005602
(87)国際公開番号　WO2023/152950
(87)国際公開日　令和5年8月17日(2023.8.17)
審査請求日　令和5年1月6日(2023.1.6)
早期審査対象出願

(73)特許権者　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　100109380
弁理士　小西 恵
(74)代理人　100109036
弁理士　永岡 重幸
(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
審査官　佐藤 敬介

最終頁に続く

(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

(57)【特許請求の範囲】

【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、前記期待報酬に係る報酬は、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間と前記複数のユーザ間において報酬を転移させることによりコンテンツごとに得られた期待報酬が、前記複数のコンテンツのうちで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】
情報処理装置によって実行される情報処理方法であって、
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】
情報処理装置によって実行される情報処理方法であって、
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間と前記複数のユーザ間において報酬を転移させることによりコンテンツごとに得られた期待報酬が、前記複数のコンテンツのうちで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間と前記複数のユーザ間において報酬を転移させることによりコンテンツごとに得られた期待報酬が、前記複数のコンテンツのうちで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。
【発明の詳細な説明】
【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】
【文献】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間と前記複数のユーザ間において報酬を転移させることによりコンテンツごとに得られた期待報酬が、前記複数のコンテンツのうちで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間と前記複数のユーザ間において報酬を転移させることによりコンテンツごとに得られた期待報酬が、前記複数のコンテンツのうちで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間と前記複数のユーザ間において報酬を転移させることによりコンテンツごとに得られた期待報酬が、前記複数のコンテンツのうちで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（ｗｉｄｇｅｔ）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（ｆｅａｔｕｒｅ）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×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×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）

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

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、

$$O_t^s = \left(x_{i_t^s}, y_{a_t^s}, r_{i_t^s a_t^s t}\right)$$

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

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

を最大化するときに表示する広告ａｔ^{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】

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

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：

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

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】

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

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】

$$S_{ad}(y_i, y_j) = \frac{\mathrm{y}_i \cdot \mathrm{y}_j}{|\mathrm{y}_i||\mathrm{y}_j|} \qquad (2B)$$

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、トンプソンサンプリング方策を基にしている。トンプ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】

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

ここで、Γはガンマ関数を表する。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】

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

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】

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

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－τ）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】

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

【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（ΣS_{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（ΣS_{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しうるものであるから、累積正報酬s_{il}（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】

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

ここで、λ（ｓ）とλ（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～Ｓ１０の処理が行われる。
　Ｓ１：時間ステップｔが０～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～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～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、例えば、ＣＰＵ４１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、時間と共に変化しう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】

Ｓ５３において、ユーザ特徴取得部１０１は、Ｓ５２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。Ｓ５３の処理は、前述の式（２Ａ）に関する処理に対応する。

【００６６】

Ｓ５４において、コンテンツ特徴取得部１０２は、コンテンツとしての広告の特徴を取得する。Ｓ５４では、コンテンツ特徴取得部１０２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１０２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１０２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１０２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】

Ｓ５５において、コンテンツ特徴取得部１０２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。Ｓ５５の処理は、前述の式（２Ｂ）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、Ｓ５１～Ｓ５５の処理の順序は図５の順序に限定されない。

【００６８】
Ｓ５６において、推定部１０４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１０４は、アルゴリズム１を実行するために、パラメータ設定部１０３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１０１とコンテンツ特徴取得部１０２それぞれから、ユーザ間の類似度（Ｓｕｓｅｒ）、広告間の類似度（Ｓａｄ）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化しうる。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$s_{il}(t)$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を参照いて、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。

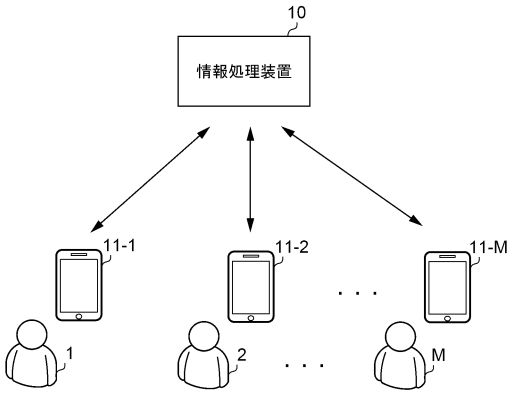
【符号の説明】

【００７６】

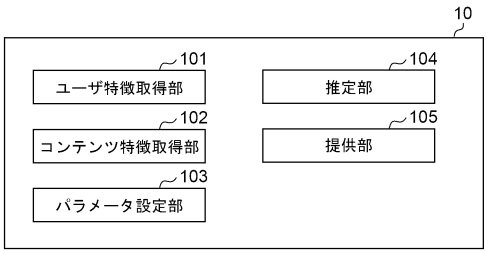
１～Ｍ：ユーザ、１０：情報処理装置、１１－１～１１－Ｍ：ユーザ装置、１０１：ユーザ特徴取得部、１０２：コンテンツ特徴取得部、１０３：パラメータ設定部、１０４：推定部１０５：提供部

【図面】

【図１】



【図２】



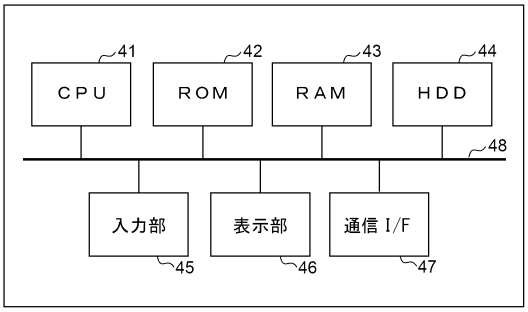
【図３】

Algorithm 1 Dynamic collaborative filtering Thompson sampling

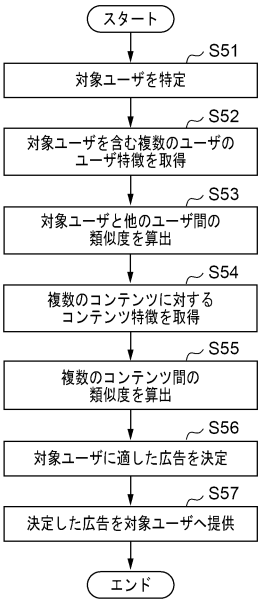
Input: $\lambda, g, \gamma \in \mathbb{R}_+^0, S_{user}, S_{ad}$, Source observations $\mathbf{O}$

for $t = 0, ..., T$ **do**
  Observe user i_t^s and context $x_{i_t^s}$, action sets $\mathbf{A_s}$ and their contexts $\mathbf{Y^s}$
  for $k \in \mathbf{A_s}$ **do**
    Calculate $\alpha_{i_t^s k}^0(t)$ and $\beta_{i_t^s k}^0(t)$ according to Eq. 4
    Calculate $\alpha_{i_t^s k}(t)$ and $\beta_{i_t^s k}(t)$ according to Eq. 6
    Sample θ_k from the $\mathcal{B}(\alpha_{i_t^s k}(t), \beta_{i_t^s k}(t))$
  end for
  Play action $k = \arg\max \theta_k$ and observe reward $r_{i_t^s kt}$
  Add observation $O_t^k \leftarrow (x_{i_t^s}, k, r_{i_t^s kt})$
end for

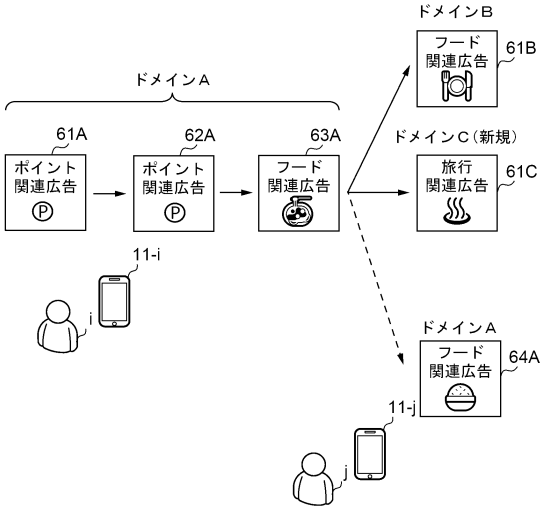
【図４】



【図５】


スタート
S51
対象ユーザを特定
S52
対象ユーザを含む複数のユーザの
ユーザ特徴を取得
S53
対象ユーザと他のユーザ間の
類似度を算出
S54
複数のコンテンツに対する
コンテンツ特徴を取得
S55
複数のコンテンツ間の
類似度を算出
S56
対象ユーザに適した広告を決定
S57
決定した広告を対象ユーザへ提供
エンド

【図６】


ドメインA
61A
ポイント
関連広告
62A
ポイント
関連広告
63A
フード
関連広告
ドメインB
フード
関連広告
61B
ドメインC（新規）
旅行
関連広告
61C
11-i
i
ドメインA
フード
関連広告
64A
11-j
j

フロントページの続き

(56)参考文献 米国特許出願公開第２０１７／００６１４８１（ＵＳ，Ａ１）
米国特許出願公開第２０１８／０２１１３０３（ＵＳ，Ａ１）
米国特許出願公開第２０１７／０２７８１１４（ＵＳ，Ａ１）
米国特許出願公開第２０１７／００９８２３６（ＵＳ，Ａ１）
米国特許出願公開第２０１９／００８０３４８（ＵＳ，Ａ１）
特開２０１８－１５６３０６（ＪＰ，Ａ）
米国特許出願公開第２０１９／０００７７２０（ＵＳ，Ａ１）
米国特許出願公開第２０１６／０３５０８０２（ＵＳ，Ａ１）

(58)調査した分野 (Int.Cl.，ＤＢ名)
Ｇ０６Ｑ １０／００－９９／００