

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号
特許第5691735号
(P5691735)

(45) 発行日 平成27年4月1日(2015.4.1)

(24) 登録日 平成27年2月13日(2015.2.13)

(51) Int.Cl.

F I

G O 6 F 17/30 (2006.01)

G O 6 Q 30/02 (2012.01)

G O 6 F 17/30 3 4 O A

G O 6 F 17/30 2 1 O A

G O 6 Q 30/02 1 5 O

請求項の数 14 (全 44 頁)

(21) 出願番号	特願2011-72324 (P2011-72324)	(73) 特許権者	000002185
(22) 出願日	平成23年3月29日 (2011.3.29)		ソニー株式会社
(65) 公開番号	特開2012-208604 (P2012-208604A)		東京都港区港南1丁目7番1号
(43) 公開日	平成24年10月25日 (2012.10.25)	(74) 代理人	100095957
審査請求日	平成26年2月26日 (2014.2.26)		弁理士 亀谷 美明
		(74) 代理人	100096389
			弁理士 金本 哲男
		(74) 代理人	100101557
			弁理士 萩原 康司
		(74) 代理人	100128587
			弁理士 松本 一騎
		(72) 発明者	上前田 直樹
			東京都港区港南1丁目7番1号 ソニー株 式会社内

最終頁に続く

(54) 【発明の名称】 コンテンツ推薦装置、推薦コンテンツの検索方法、及びプログラム

(57) 【特許請求の範囲】

【請求項1】

対象ユーザが過去に選択した第1のコンテンツに含まれる第1の種類の情報に基づいて第1の特徴量を生成する第1特徴量生成部と、

前記第1のコンテンツを選択した後に前記対象ユーザが選択した第2のコンテンツに含まれる第2の種類の情報に基づいて第2の特徴量を生成する第2特徴量生成部と、

前記第1特徴量生成部により生成された第1の特徴量と、前記第2特徴量生成部により生成された第2の特徴量とに基づき、前記第1のコンテンツと前記第2のコンテンツとの間の関係を示す関係特徴量を生成する関係特徴量生成部と、

前記対象ユーザが新たに選択したコンテンツに含まれる前記第1の種類の情報に対応する第1の特徴量と、前記関係特徴量生成部により生成された関係特徴量とを用いて、前記対象ユーザに推薦するコンテンツを検索する推薦コンテンツ検索部と、
を備える、

コンテンツ推薦装置。

【請求項2】

前記推薦コンテンツ検索部は、

前記第1特徴量生成部により生成された第1の特徴量の中から、前記対象ユーザが新たに選択したコンテンツに含まれる前記第1の種類の情報に対応する第1の特徴量を抽出する第1の処理を実行し、前記関係特徴量生成部により生成された関係特徴量の中から、前記第1の処理で抽出した第1の特徴量に対応する関係特徴量を抽出する第2の処理を実行

し、前記第 2 の処理で抽出した関係特徴量を用いて、前記対象ユーザに推薦するコンテンツを検索する、

請求項 1 に記載のコンテンツ推薦装置。

【請求項 3】

前記第 1 の特徴量は、

前記第 1 の種類の情報を構成する複数の情報要素を含み、前記第 1 のコンテンツを特徴付ける第 1 の特徴ベクトルにより表現され、

前記第 2 の特徴量は、

前記第 2 の種類の情報を構成する複数の情報要素を含み、前記第 2 のコンテンツを特徴付ける第 2 の特徴ベクトルにより表現され、

請求項 1 または 2 に記載のコンテンツ推薦装置。

10

【請求項 4】

前記第 1 特徴量生成部は、

前記対象ユーザが前記第 1 のコンテンツを選択した順序を考慮して前記第 1 の特徴量を生成する、

請求項 1 ～ 3 のいずれか一項に記載のコンテンツ推薦装置。

【請求項 5】

前記第 1 特徴量生成部は、

前記第 1 のコンテンツが有するメタデータの構造を参照して前記第 1 の種類に対応するメタデータが付与された領域から情報を取得し、

前記第 2 特徴量生成部は、

前記第 2 のコンテンツが有するメタデータの構造を参照して前記第 2 の種類に対応するメタデータが付与された領域から情報を取得する、

請求項 1 ～ 4 のいずれか一項に記載のコンテンツ推薦装置。

20

【請求項 6】

前記関係特徴量生成部により生成された複数の前記関係特徴量を前記対象ユーザに提示し、当該対象ユーザに前記関係特徴量を選択させる関係選択要求部をさらに備え、

前記推薦コンテンツ検索部は、

前記対象ユーザにより前記関係特徴量を選択された場合に、前記対象ユーザにより選択された関係特徴量を用いて、前記対象ユーザに推薦するコンテンツを検索する、

請求項 1 ～ 5 のいずれか一項に記載のコンテンツ推薦装置。

30

【請求項 7】

前記推薦コンテンツ検索部は、

前記第 1 のコンテンツと前記第 2 のコンテンツとの間の関係の強さに応じたスコアを算出し、算出したスコアを考慮して前記対象ユーザに推薦するコンテンツを検索する、

請求項 1 ～ 5 のいずれか一項に記載のコンテンツ推薦装置。

【請求項 8】

前記第 1 特徴量生成部は、

前記対象ユーザが新たにコンテンツを選択する以前に前記第 1 の特徴量を生成し、

前記第 2 特徴量生成部は、

前記対象ユーザが新たにコンテンツを選択する以前に前記第 2 の特徴量を生成し、

前記関係特徴量生成部は、

前記対象ユーザが新たにコンテンツを選択する以前に前記関係特徴量を生成する、

請求項 1 ～ 7 のいずれか一項に記載のコンテンツ推薦装置。

40

【請求項 9】

前記推薦コンテンツ検索部は、

前記対象ユーザが新たにコンテンツを選択する以前に、前記第 1 の種類の情報に対応する所定の情報を用いて、当該所定の情報に対応する第 1 の特徴量を抽出する第 1 の処理を実行し、前記関係特徴量生成部により生成された関係特徴量の中から、前記第 1 の処理で抽出した第 1 の特徴量に対応する関係特徴量を抽出する第 2 の処理を実行し、前記第 2 の

50

処理で抽出した関係特徴量のスコアを算出する第3の処理を実行し、

前記対象ユーザが新たにコンテンツを選択した場合に、前記対象ユーザが新たに選択したコンテンツに含まれる前記第1の種類の情報に対応する前記所定の情報を抽出する第3の処理を実行し、前記第3の処理で抽出した所定の情報に対応する関係特徴量のスコアに基づいて、前記対象ユーザに推薦するコンテンツを検索する、

請求項1～8のいずれか一項に記載のコンテンツ推薦装置。

【請求項10】

前記第1及び第2のコンテンツと、前記対象ユーザが新たに選択したコンテンツとは、互いに異なるカテゴリに属する、

請求項1～9のいずれか一項に記載のコンテンツ推薦装置。

10

【請求項11】

前記第1特徴量生成部は、対象ユーザが過去に選択した、時系列に沿った複数の第1のコンテンツに含まれる第1の種類の情報に基づいて第1の特徴量を生成する、

請求項1～10のいずれか一項に記載のコンテンツ推薦装置。

【請求項12】

対象ユーザが過去に選択した第1のコンテンツに含まれる第1の種類の情報に基づいて生成された第1の特徴量と、前記第1のコンテンツを選択した後に前記対象ユーザが選択した第2のコンテンツに含まれる第2の種類の情報に基づいて生成された第2の特徴量とに基づいて生成された、前記第1のコンテンツと前記第2のコンテンツとの間の関係を示す第3の特徴量と、が格納された特徴量格納部と、

20

前記対象ユーザが新たに選択したコンテンツに含まれる前記第1の種類の情報に対応する第1の特徴量と、前記特徴量格納部に格納された第3の特徴量とを用いて、前記対象ユーザに推薦するコンテンツを検索する推薦コンテンツ検索部と、を備える、

コンテンツ推薦装置。

【請求項13】

コンピュータのプロセッサが、

対象ユーザが過去に選択した第1のコンテンツに含まれる第1の種類の情報に基づいて第1の特徴量を生成する第1特徴量生成ステップと、

前記第1のコンテンツを選択した後に前記対象ユーザが選択した第2のコンテンツに含まれる第2の種類の情報に基づいて第2の特徴量を生成する第2特徴量生成ステップと、

30

前記第1特徴量生成ステップで生成された第1の特徴量と、前記第2特徴量生成ステップで生成された第2の特徴量とに基づき、前記第1のコンテンツと前記第2のコンテンツとの間の関係を示す関係特徴量を生成する関係特徴量生成ステップと、

前記対象ユーザが新たに選択したコンテンツに含まれる前記第1の種類の情報に対応する第1の特徴量と、前記関係特徴量生成ステップで生成された関係特徴量とを用いて、前記対象ユーザに推薦するコンテンツを検索する推薦コンテンツ検索ステップと、を含む、

推薦コンテンツの検索方法。

【請求項14】

40

対象ユーザが過去に選択した第1のコンテンツに含まれる第1の種類の情報に基づいて第1の特徴量を生成する第1特徴量生成機能と、

前記第1のコンテンツを選択した後に前記対象ユーザが選択した第2のコンテンツに含まれる第2の種類の情報に基づいて第2の特徴量を生成する第2特徴量生成機能と、

前記第1特徴量生成機能により生成された第1の特徴量と、前記第2特徴量生成機能により生成された第2の特徴量とに基づき、前記第1のコンテンツと前記第2のコンテンツとの間の関係を示す関係特徴量を生成する関係特徴量生成機能と、

前記対象ユーザが新たに選択したコンテンツに含まれる前記第1の種類の情報に対応する第1の特徴量と、前記関係特徴量生成機能により生成された関係特徴量とを用いて、前記対象ユーザに推薦するコンテンツを検索する推薦コンテンツ検索機能と、

50

をコンピュータに実現させるためのプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本技術は、コンテンツ推薦装置、推薦コンテンツの検索方法、及びプログラムに関する。

【背景技術】

【0002】

近年、ネットワークを利用したビジネスが急拡大している。例えば、オンラインで商品を購入することが可能なオンラインストアなどのシステムは広く一般に利用されている。こうしたオンラインストアの多くには、ユーザに商品を推薦する仕組みが設けられている。例えば、ある商品の詳細情報をユーザが閲覧すると、その商品に関連する商品の情報が推薦商品としてユーザに提示される。

10

【0003】

このような仕組みは、例えば、下記の特許文献1に記載されている協調フィルタリングなどの方法を利用して実現される。この協調フィルタリングとは、多くのユーザの嗜好情報に基づき、嗜好の似たユーザの情報を利用して自動的に推薦を行う方法である。この協調フィルタリングを用いると、購入履歴のない新規ユーザに対しても推薦結果を提供することが可能になる。

【0004】

20

また、商品の推薦には、コンテンツベースフィルタリングと呼ばれる方法が利用されることもある。このコンテンツベースフィルタリングとは、コンテンツの属性とユーザの好みとをマッチングさせて関連コンテンツを推薦する方法である。このコンテンツベースフィルタリングは協調フィルタリングに比べ、推薦システムを利用しているユーザ数が少ない状況でも精度良く推薦結果を提供することができる。但し、対象ユーザが好むコンテンツを判別するための情報（例えば、購入履歴やコンテンツのメタ情報など）が少ない状況では、コンテンツベースフィルタリングを用いて精度の良い推薦結果を得るのは難しい。

【先行技術文献】

【特許文献】

【0005】

30

【特許文献1】特開2003-167901号公報

【発明の概要】

【発明が解決しようとする課題】

【0006】

協調フィルタリング及びコンテンツベースフィルタリングは、それぞれ長所と短所とを有している。例えば、コンテンツベースフィルタリングは、ユーザの嗜好を反映させた推薦を実現できるという長所を有する。その反面、コンテンツベースフィルタリングは、ユーザの嗜好に適合した特定の情報ばかりが推薦されるような状況を生じさせ、ユーザにとって新鮮な情報が推薦されなくなってしまうという短所を有する。一方、協調フィルタリングは、他のユーザが嗜好する新しい情報をユーザに提供することができるという長所を有する。しかし、他のユーザが嗜好する新しい情報が推薦対象となるユーザの嗜好に適合しないこともある。つまり、協調フィルタリングは、ユーザの嗜好に適合しない情報をそのユーザに提供する可能性を含むという短所を有する。

40

【0007】

そこで、本技術は、上記のような事情を受けて考案されたものであり、ユーザの嗜好に適合する新しい情報を含んだコンテンツをそのユーザに提供することが可能な、新規かつ改良されたコンテンツ推薦装置、推薦コンテンツの検索方法、及びプログラムを提供することを意図している。

【課題を解決するための手段】

【0008】

50

本技術のある観点によれば、対象ユーザが過去に選択した第1のコンテンツに含まれる第1の種類の情報に基づいて第1の特徴量を生成する第1特徴量生成部と、前記第1のコンテンツを選択した後に前記対象ユーザが選択した第2のコンテンツに含まれる第2の種類の情報に基づいて第2の特徴量を生成する第2特徴量生成部と、前記第1特徴量生成部により生成された第1の特徴量と、前記第2特徴量生成部により生成された第2の特徴量とに基づき、前記第1のコンテンツと前記第2のコンテンツとの間の関係を示す関係特徴量を生成する関係特徴量生成部と、前記対象ユーザが新たに選択したコンテンツに含まれる前記第1の種類の情報と、前記関係特徴量生成部により生成された関係特徴量とを用いて、前記対象ユーザに推薦するコンテンツを検索する推薦コンテンツ検索部と、を備える、コンテンツ推薦装置が提供される。

10

【0009】

また、前記推薦コンテンツ検索部は、前記第1特徴量生成部により生成された第1の特徴量の中から、前記対象ユーザが新たに選択したコンテンツに含まれる前記第1の種類の情報に対応する第1の特徴量を抽出する第1の処理を実行し、前記関係特徴量生成部により生成された関係特徴量の中から、前記第1の処理で抽出した第1の特徴量に対応する関係特徴量を抽出する第2の処理を実行し、前記第2の処理で抽出した関係特徴量を用いて、前記対象ユーザに推薦するコンテンツを検索する、ように構成されていてもよい。

【0010】

また、前記第1の特徴量は、前記第1の種類の情報を構成する複数の情報要素を含み、前記第1のコンテンツを特徴付ける第1の特徴ベクトルにより表現されていてもよい。さらに、前記第2の特徴量は、前記第2の種類の情報を構成する複数の情報要素を含み、前記第2のコンテンツを特徴付ける第2の特徴ベクトルにより表現されていてもよい。

20

【0011】

また、前記第1特徴量生成部は、前記対象ユーザが前記第1のコンテンツを選択した順序を考慮して前記第1の特徴量を生成するように構成されていてもよい。

【0012】

また、前記第1特徴量生成部は、前記第1のコンテンツが有するメタデータの構造を参照して前記第1の種類に対応するメタデータが付与された領域から情報を取得するように構成されていてもよい。さらに、前記第2特徴量生成部は、前記第2のコンテンツが有するメタデータの構造を参照して前記第2の種類に対応するメタデータが付与された領域から情報を取得するように構成されていてもよい。

30

【0013】

また、上記のコンテンツ推薦装置は、前記関係特徴量生成部により生成された複数の前記関係特徴量を前記対象ユーザに提示し、当該対象ユーザに前記関係特徴量を選択させる関係選択要求部をさらに備えていてもよい。この場合、前記推薦コンテンツ検索部は、前記対象ユーザにより前記関係特徴量が選択された場合に、前記対象ユーザにより選択された関係特徴量を用いて、前記対象ユーザに推薦するコンテンツを検索する。

【0014】

また、前記推薦コンテンツ検索部は、前記第1のコンテンツと前記第2のコンテンツとの間の関係の強さに応じたスコアを算出し、算出したスコアを考慮して前記対象ユーザに推薦するコンテンツを検索するように構成されていてもよい。

40

【0015】

また、前記第1特徴量生成部は、前記対象ユーザが新たにコンテンツを選択する以前に前記第1の特徴量を生成するように構成されていてもよい。さらに、前記第2特徴量生成部は、前記対象ユーザが新たにコンテンツを選択する以前に前記第2の特徴量を生成するように構成されていてもよい。そして、前記関係特徴量生成部は、前記対象ユーザが新たにコンテンツを選択する以前に前記関係特徴量を生成するように構成されていてもよい。

【0016】

また、前記推薦コンテンツ検索部は、前記対象ユーザが新たにコンテンツを選択する以前に、前記第1の種類の情報に対応する所定の情報を用いて、当該所定の情報に対応する

50

第1の特徴量を抽出する第1の処理を実行し、前記関係特徴量生成部により生成された関係特徴量の中から、前記第1の処理で抽出した第1の特徴量に対応する関係特徴量を抽出する第2の処理を実行し、前記第2の処理で抽出した関係特徴量のスコアを算出する第3の処理を実行し、前記対象ユーザが新たにコンテンツを選択した場合に、前記対象ユーザが新たに選択したコンテンツに含まれる前記第1の種類の情報に対応する前記所定の情報を抽出する第3の処理を実行し、前記第3の処理で抽出した所定の情報に対応する関係特徴量のスコアに基づいて、前記対象ユーザに推薦するコンテンツを検索するように構成されていてもよい。

【0017】

また、前記第1及び第2のコンテンツと、前記対象ユーザが新たに選択したコンテンツとは、互いに異なるカテゴリに属していてもよい。

10

【0018】

また、本技術の別の観点によれば、対象ユーザが過去に選択した第1のコンテンツに含まれる第1の種類の情報に基づいて生成された第1の特徴量と、前記第1のコンテンツを選択した後に前記対象ユーザが選択した第2のコンテンツに含まれる第2の種類の情報に基づいて生成された第2の特徴量と、前記第1特徴量生成部により生成された第1の特徴量と、前記第2特徴量生成部により生成された第2の特徴量とに基づいて生成された、前記第1のコンテンツと前記第2のコンテンツとの間の関係を示す第3の特徴量と、が格納された特徴量格納部と、前記対象ユーザが新たに選択したコンテンツに含まれる前記第1の種類の情報と、前記特徴量格納部に格納された第3の特徴量とを用いて、前記対象ユーザに推薦するコンテンツを検索する推薦コンテンツ検索部と、を備える、コンテンツ推薦装置が提供される。

20

【0019】

また、本技術の別の観点によれば、対象ユーザが過去に選択した第1のコンテンツに含まれる第1の種類の情報に基づいて第1の特徴量を生成する第1特徴量生成ステップと、前記第1のコンテンツを選択した後に前記対象ユーザが選択した第2のコンテンツに含まれる第2の種類の情報に基づいて第2の特徴量を生成する第2特徴量生成ステップと、前記第1特徴量生成ステップで生成された第1の特徴量と、前記第2特徴量生成ステップで生成された第2の特徴量とに基づき、前記第1のコンテンツと前記第2のコンテンツとの間の関係を示す関係特徴量を生成する関係特徴量生成ステップと、前記対象ユーザが新たに選択したコンテンツに含まれる前記第1の種類の情報と、前記関係特徴量生成ステップで生成された関係特徴量とを用いて、前記対象ユーザに推薦するコンテンツを検索する推薦コンテンツ検索ステップと、含む、推薦コンテンツの検索方法が提供される。

30

【0020】

また、本技術の別の観点によれば、対象ユーザが過去に選択した第1のコンテンツに含まれる第1の種類の情報に基づいて第1の特徴量を生成する第1特徴量生成機能と、前記第1のコンテンツを選択した後に前記対象ユーザが選択した第2のコンテンツに含まれる第2の種類の情報に基づいて第2の特徴量を生成する第2特徴量生成機能と、前記第1特徴量生成機能により生成された第1の特徴量と、前記第2特徴量生成機能により生成された第2の特徴量とに基づき、前記第1のコンテンツと前記第2のコンテンツとの間の関係を示す関係特徴量を生成する関係特徴量生成機能と、前記対象ユーザが新たに選択したコンテンツに含まれる前記第1の種類の情報と、前記関係特徴量生成機能により生成された関係特徴量とを用いて、前記対象ユーザに推薦するコンテンツを検索する推薦コンテンツ検索機能と、をコンピュータに実現させるためのプログラムが提供される。

40

【0021】

また、本技術の別の観点によれば、上記のプログラムが記録された、コンピュータにより読み取り可能な記録媒体が提供される。

【発明の効果】

【0022】

以上説明したように本技術によれば、ユーザの嗜好に適合する新しい情報を含んだコン

50

テンツをそのユーザに提供することが可能になる。

【図面の簡単な説明】

【 0 0 2 3 】

【図 1】四項類推の概念について説明するための説明図である。

【図 2】四項類推に係る処理の流れについて説明するための説明図である。

【図 3】多次元化した四項類推の概要について説明するための説明図である。

【図 4】コンテンツメタデータの構造について説明するための説明図である。

【図 5】本技術の第 1 実施形態に係る推薦システムの構成について説明するための説明図である。

【図 6】本技術の第 1 実施形態に係るコンテンツ特徴量データベースの構造について説明するための説明図である。 10

【図 7】本技術の第 1 実施形態に係るユーザ嗜好データベースの構造について説明するための説明図である。

【図 8】本技術の第 1 実施形態に係る事例データベースの構造について説明するための説明図である。

【図 9】本技術の第 1 実施形態に係る事例データベースの生成方法について説明するための説明図である。

【図 10】本技術の第 1 実施形態に係る事例データベースの生成方法について説明するための説明図である。

【図 11】本技術の第 1 実施形態に係る事例データベースの生成方法について説明するための説明図である。 20

【図 12】本技術の第 1 実施形態に係る事例データベースの生成方法について説明するための説明図である。

【図 13】本技術の第 1 実施形態に係る推薦の処理について説明するための説明図である。

【図 14】本技術の第 1 実施形態に係る嗜好学習の処理について説明するための説明図である。

【図 15】本技術の第 1 実施形態に係る推薦の処理について説明するための説明図である。

【図 16】本技術の第 1 実施形態に係る推薦の処理について説明するための説明図である。 30

【図 17】本技術の第 1 実施形態に係る推薦の処理について説明するための説明図である。

【図 18】本技術の第 1 実施形態に係る推薦の処理について説明するための説明図である。

【図 19】本技術の第 2 実施形態に係る推薦システムの構成について説明するための説明図である。

【図 20】本技術の第 2 実施形態に係る重心データベースの構造について説明するための説明図である。

【図 21】本技術の第 2 実施形態に係る R パターンデータベースの構造について説明するための説明図である。 40

【図 22】本技術の第 2 実施形態に係る推薦の処理について説明するための説明図である。

【図 23】本技術の第 2 実施形態に係る推薦の処理について説明するための説明図である。

【図 24】本技術の第 2 実施形態に係る推薦の処理について説明するための説明図である。

【図 25】本技術の第 2 実施形態に係るクラスタリングの処理について説明するための説明図である。

【図 26】本技術の第 2 実施形態に係るクラスタリングの処理について説明するための説 50

明図である。

【図 27】本技術の第 2 実施形態に係る R パターンの選択について説明するための説明図である。

【図 28】本技術の第 2 実施形態に係る推薦の処理について説明するための説明図である。

【図 29】本技術の第 2 実施形態に係る推薦の処理について説明するための説明図である。

【図 30】本技術の第 3 実施形態に係る推薦システムの構成について説明するための説明図である。

【図 31】本技術の第 3 実施形態に係る推薦リストデータベースの構造について説明するための説明図である。

10

【図 32】本技術の第 3 実施形態に係るオフライン処理（関係 R のスコア算出）について説明するための説明図である。

【図 33】本技術の第 3 実施形態に係るオフライン処理（関係 R のスコア算出）について説明するための説明図である。

【図 34】本技術の第 3 実施形態に係るオフライン処理について説明するための説明図である。

【図 35】本技術の第 3 実施形態に係るオンライン処理について説明するための説明図である。

【図 36】本技術の第 3 実施形態に係るオンライン処理について説明するための説明図である。

20

【図 37】本技術の第 1 ～ 第 3 実施形態に係る技術の応用例（クロスカテゴリ推薦）について説明するための説明図である。

【図 38】本技術の第 1 ～ 第 3 実施形態に係る推薦システムの機能を実現することが可能なハードウェア構成について説明するための説明図である。

【発明を実施するための形態】

【0024】

以下に添付図面を参照しながら、本技術に係る好適な実施の形態について詳細に説明する。なお、本明細書及び図面において、実質的に同一の機能構成を有する構成要素については、同一の符号を付することにより重複説明を省略する。

30

【0025】

[説明の流れについて]

ここで、以下に記載する説明の流れについて簡単に述べる。

【0026】

はじめに、後述する実施形態に係る技術に用いられる四項類推の概念について説明する。まず、図 1 を参照しながら、四項類推の概念について説明する。次いで、図 2 を参照しながら、四項類推に係る処理の流れについて説明する。次いで、図 3 を参照しながら、多次元化した四項類推の概要について説明する。また、図 4 を参照しながら、四項類推を具体的な事例に適用する際に用いるコンテンツメタデータの構造について説明する。

【0027】

40

次に、本技術の第 1 実施形態について説明する。まず、図 5 を参照しながら、本技術の第 1 実施形態に係る推薦システム 100 の構成について説明する。また、図 6 を参照しながら、本技術の第 1 実施形態に係るコンテンツ特徴量データベース 104 の構造について説明する。さらに、図 7 を参照しながら、本技術の第 1 実施形態に係るユーザ嗜好データベース 102 の構造について説明する。そして、図 8 ～ 図 12 を参照しながら、本技術の第 1 実施形態に係る事例データベース 106 の構造及びその生成方法について説明する。次いで、図 13 ～ 図 18 を参照しながら、本技術の第 1 実施形態に係る推薦の処理について説明する。この中で本技術の第 1 実施形態に係る嗜好学習の処理についても説明する。

【0028】

次に、本技術の第 2 実施形態について説明する。まず、図 19 を参照しながら、本技術

50

の第2実施形態に係る推薦システム200の構成について説明する。また、図20を参照しながら、本技術の第2実施形態に係る重心データベース(Rパターンデータベース209)の構造について説明する。さらに、図21を参照しながら、本技術の第2実施形態に係るRパターンデータベース209の構造について説明する。次いで、図22～図29を参照しながら、本技術の第2実施形態に係る推薦の処理について説明する。この中で、本技術の第2実施形態に係るクラスタリングの処理及びRパターンの選択についても説明する。

【0029】

次に、本技術の第3実施形態について説明する。まず、図30を参照しながら、本技術の第3実施形態に係る推薦システム300の構成について説明する。また、図31を参照しながら、本技術の第3実施形態に係る推薦リストデータベース309の構造について説明する。次いで、図31～図34を参照しながら、本技術の第3実施形態に係るオフライン処理について説明する。次いで、図35及び図36を参照しながら、本技術の第3実施形態に係るオンライン処理について説明する。次に、図37を参照しながら、本技術の第1～第3実施形態に係る技術の応用例(クロスカテゴリ推薦)について説明する。次いで、図38を参照しながら、本技術の第1～第3実施形態に係る推薦システムの機能を実現することが可能なハードウェア構成について説明する。

【0030】

最後に、同実施形態の技術的思想について纏め、当該技術的思想から得られる作用効果について簡単に説明する。

【0031】

(説明項目)

1：はじめに

1-1：四項類推とは

1-2：四項類推の多次元化

2：第1実施形態

2-1：システム構成

2-2：オフライン処理の流れ

2-3：オンライン処理の流れ

3：第2実施形態

3-1：システム構成

3-2：オフライン処理の流れ

3-3：オンライン処理の流れ

4：第3実施形態

4-1：システム構成

4-2：オフライン処理の流れ

4-3：オンライン処理の流れ

5：応用例(クロスカテゴリ推薦)

6：ハードウェア構成例

7：まとめ

【0032】

<1：はじめに>

はじめに、本実施形態に係る技術について詳細に説明するに先立ち、四項類推の概念及び本実施形態の概要について簡単に説明する。

【0033】

[1-1：四項類推とは]

まず、図1を参照しながら、四項類推の概念について説明する。図1は、四項類推の概念について説明するための説明図である。

【0034】

四項類推とは、人が前提知識に基づいて物事を類推する過程をモデル化したものである

10

20

30

40

50

。「事例：A B」を前提知識として持つ人に情報Cを与えた場合、その人が情報Cから類推する情報Xはどのようなものであろうか。例えば、Aとして単語「fish」、Bとして単語「scale」を与えると、人は、AとBとの間にある関係Rとして単語「have」や単語「cover」などで表現される概念を想起するであろう。そして、この人に情報Cとして単語「bird」を与え、関係Rに基づいて情報Xを類推させると、この人は、例えば、単語「feather」や単語「wing」などを類推すると考えられる。このような人の類推過程をモデル化したものが四項類推なのである。

【0035】

この四項類推に関し、前提知識として「事例：A B」を与えられた人が類推する「事例：C X」の解Xを推定する技術に注目が集まっている。なお、以下では、「事例：A B」から「事例：C X」を類推する過程を「A : B = C : X」と表現することがある。「A : B = C : X」の解Xを推定する技術としては、例えば、構造写像理論と呼ばれる推定方法が知られている。この推定方法は、図1に示すように、「事例：A B」におけるA（以下、状況A）とB（以下、結果B）との関係Rを「事例：C X」におけるC（以下、状況C）に適用して解X（以下、結果X）を推定するというものである。

【0036】

つまり、上記の構造写像理論は、前提知識を構成する知識領域（以下、基底領域）の構造を、解Xを求める問題の領域（以下、目標領域）に写像する方法とも言える。構造写像理論に関しては、例えば、D. Gentner, "Structure - Mapping : A Theoretical Framework for Analogy", Cognitive Science, 1983などに記載がある。

【0037】

上記の構造写像理論を用いると、基底領域の構造を写像する際に生じる無意味な知識を排除して、ある程度妥当な類推結果Xを得ることができる。例えば、図1に示すように、状況Aとして単語「fish」が与えられた場合に、単語「fish」から類推される「青い」「小さい」などの知識を結果Xの推定時に排除することができる。同様に、結果Bとして単語「scale」が与えられた場合に、「硬い」「透明」などの知識を結果Xの推定時に排除することができる。

【0038】

構造写像理論に基づく結果Xの推定処理は、例えば、図2に示す処理手順で実行される。まず、図2に示すように、状況Aと結果Bとの間の関係Rを推定する処理が実行される（S10）。次いで、ステップS10で推定された関係Rを基底領域から目標領域へ写像する処理が実行される（S11）。次いで、関係Rを状況Cに適用し、結果Xを推定する処理が実行される（S12）。これらステップS10～S12の処理を実行することにより、「事例：A B」に基づいて「事例：C X」の解Xが推定される。

【0039】

以上、四項類推の概念について説明した。ここで説明した四項類推の概念をファジィ理論の視点からシステム化する研究がKanekoらによりなされており、その研究成果が報告されている。例えば、Yosuke Kaneko, Kazuhiro Okada, Shinichiro Ito, Takuya Nomura and

Tomihiko Takagi, "A Proposal of Analogical Reasoning Based on Structural Mapping and Image Schemas", 5th International Conference on Soft Computing and Intelligent Systems and 11th International Symposium on Advanced Intelligent Systems (SCIS & ISIS 10), 2010などの報告がある。これらの報告の中で、Kanekoらは、単語の共起頻度から写像対象となる関係Rを抽出し、単語の品詞情報を構造として利用する推薦システムを提案している。この報告内容も四項類推の概念を理解するための一助となるであろう。

【0040】

[1 - 2 : 四項類推の多次元化]

次に、図3を参照しながら、四項類推を多次元化する方法について説明する。図3は、

四項類推を多次元化する方法について説明するための説明図である。なお、四項類推の多次元化に関する研究成果としては、特願 2011-18787 号に記載の方法がある。

【0041】

図 1 の例は、1 つの基底領域から 1 つの目標領域への構造写像に関するものであった。また、図 1 の例において、状況 A、結果 B、状況 C、結果 X は、それぞれ 1 つの単語で表現されていた。ここでは四項類推の概念を拡張し、図 3 に示すように、複数の基底領域から 1 つの目標領域へと構造を写像する新たな方式について考える。また、ここでは、状況 A、結果 B、状況 C、結果 X をそれぞれ、1 つ又は複数の単語で構成されるワードベクトルにより表現する方式について考える。なお、ここで考える新たな方式のことを「多次元の四項類推」と呼ぶことにする。以下、多次元の四項類推の概念について説明する。

10

【0042】

図 3 に示すように、 n 個の基底領域（基底領域 1 ～ 基底領域 n ）を考える。また、基底領域 k ($k = 1 \sim n$) には、「事例： $A_k \quad B_k$ 」が属するものとする。さらに、状況 A_k 、結果 B_k は、複数の単語を含むワードベクトルで表現されるものとする。また、基底領域 1 ～ 基底領域 n の構造は、1 つの目標領域へと写像されるものとする。さらに、この目標領域には、「事例： $C \quad X_j$ ($j = 1 \sim n$)」が属するものとする。但し、状況 A_k と結果 B_k との関係 R_k は、「事例： $C \quad X_k$ 」における結果 X_k の推定に利用される。

【0043】

例えば、状況 A_k ($k = 1 \sim n$) は、ある人（以下、対象ユーザ）が過去に選択したコンテンツ群から抽出される対象ユーザの嗜好を特徴付けるワードベクトルにより表現される。また、結果 B_k ($k = 1 \sim n$) は、状況 A_k を前提とし、そのコンテンツ群の次に対象ユーザが選択したコンテンツを特徴付けるワードベクトルにより表現される。さらに、関係 R_k ($k = 1 \sim n$) は、状況 A_k と結果 B_k との関係を特徴付けるワードベクトルにより表現される。そして、状況 C は、新たに対象ユーザが選択したコンテンツを含むコンテンツ群から抽出される対象ユーザの嗜好を特徴付けるワードベクトルにより表現される。また、結果 X_k ($k = 1 \sim n$) は、状況 C のワードベクトルと、関係 R のワードベクトルとに基づいて類推されるコンテンツを特徴付けるワードベクトルである。

20

【0044】

つまり、状況 A_1 と結果 B_1 との関係 R_1 、及び状況 C を用いて結果 X_1 が類推される。同様に、関係 R_2 及び状況 C から結果 X_2 が類推され、関係 R_3 及び状況 C から結果 X_3 が類推され、...、関係 R_n 及び状況 C から結果 X_n が類推される。なお、各ワードベクトルは、例えば、TF-IDF と呼ばれるアルゴリズムを利用して生成される。この TF-IDF とは、文書中から特徴的な単語を抽出するためのアルゴリズムである。TF-IDF は、TF-IDF 値と呼ばれる指標を出力する。この TF-IDF 値は、単語の出現頻度を表す TF 値と、逆出現頻度を表す IDF 値との積で表現される。

30

【0045】

例えば、文書 d 中における単語 j の出現回数を N_j 、文書 d 中に含まれる全単語数を N 、全文書数を D 、単語 j が出現する文書数を D_j とすると、TF 値 $t_f(j, d)$ は、下記の式 (1) により表現される。また、IDF 値 $i_d f(j)$ は、下記の式 (2) により表現される。そして、TF-IDF 値 $t_f i_d f(j, d)$ は、下記の式 (3) により表現される。つまり、多くの文書中に出現する単語の TF-IDF 値は下がり、特定の文書中に頻出する単語の TF-IDF 値は上がる。そのため、この指標を用いることにより、個々の文書を特徴付ける単語を抽出することが可能になる。また、TF-IDF 値の高い複数の単語を抽出することにより、文書を特徴付けるワードベクトルが生成される。

40

【0046】

【数 1】

$$tf(j, d) = N_j / N$$

... (1)

$$idf(j) = 1 + \ln(D / D_j)$$

10

... (2)

$$tfidf(j, d) = tf(j, d) \cdot idf(j)$$

... (3)

【 0 0 4 7 】

ここで、料理投稿サイトを情報源として利用する実施例について考えてみたい。料理投稿サイトの多くは、ユーザが自分の創作した料理のレシピを自由に投稿できるように構成されている。また、こうした料理投稿サイトは、この料理投稿サイトを閲覧した他のユーザがレビューを記載できるように構成されている。もちろん、他の情報サイトと同様に、料理投稿サイトには、タイトル、画像、説明文といった欄が設けられている。また、料理投稿サイトの中には、材料、料理の手順、作り方のコツ、料理の生い立ち、登録カテゴリなどの欄を設けているサイトもある。これらの欄は、メタデータにより規定されている。

20

【 0 0 4 8 】

例えば、図 4 に示すように、ある料理投稿サイトは、Title、Image、Description、Ingredients、Cooking Procedure、Knacks of Cooking Procedure、Reviews、History、Categoriesなどのメタデータで構造が規定されている。そのうち、Title、Description、Ingredients、Cooking Procedure、Knacks of Cooking Procedure、Review s、Historyの欄は、多次元の四項類推に利用可能な情報を含む。

30

【 0 0 4 9 】

例えば、図 4 に示すように、Ingredients、Cooking Procedure、Knacks of Cooking Procedureの欄は、状況 A 及び状況 C に関する情報源として利用することができる。また、Title、Description、Reviewsの欄は、結果 B に関する情報源として利用することができる。さらに、Historyの欄は、関係 R に関する情報源として利用することができる。

40

【 0 0 5 0 】

つまり、状況 A 及び状況 C に関する情報源は、ユーザの嗜好（この例では材料、料理手順、料理のコツなど）を表す領域に設定される。一方、結果 B に関する情報源は、この料理投稿サイトに記載の料理を実際に食べてみた結果などが表現されている領域に設定される。また、関係 R に関する情報源は、状況 A と結果 B との関係（この例では、料理投稿サイトに記載の料理が生まれるまでの過程など）が表現されている領域に設定される。このように、メタデータの構造を利用することにより、状況 A、結果 B、状況 C、関係 R に関する情報源を容易に設定することができる。また、各領域に記載の文書から、上記の TF - IDF 値などを利用して、状況 A、結果 B、状況 C に対応するワードベクトルを生成することができる。

50

【 0 0 5 1 】

ここでは料理投稿サイトを情報源として利用する実施例について考えたが、他種類のサイトに関してもメタデータの構造を参照することにより、状況 A、結果 B、状況 C、関係 R に関する情報源を設定することができる。なお、結果 X に関する情報源は、結果 B に関する情報源と同じメタデータが付された領域に設定される。このようにして情報源が設定されると、ユーザが閲覧したサイトの履歴などから抽出されるワードベクトルを利用して、図 3 に示すような多次元の四項類推に基づく結果 X 1 ~ X n の推定が可能になる。

【 0 0 5 2 】

本実施形態に係る技術は、上記の推定に関するものである。但し、本実施形態に係る技術は、多次元の四項類推に基づく結果 X 1 ~ X n の推定に主眼を置くものではなく、関係 R 1 ~ R n を利用してユーザの嗜好に適した推薦コンテンツを検索する技術に関する。また、本実施形態の適用範囲は料理投稿サイトに限定されず、様々なコンテンツに対して適用することが可能である。

10

【 0 0 5 3 】

以上、四項類推の概念及び本実施形態の概要について簡単に説明した。以下、本実施形態に係る技術について詳細に説明する。

【 0 0 5 4 】

< 2 : 第 1 実施形態 >

本技術の第 1 実施形態について説明する。

【 0 0 5 5 】

20

[2 - 1 : システム構成]

まず、図 5 を参照しながら、本実施形態に係る推薦システム 1 0 0 のシステム構成について説明する。図 5 は、本実施形態に係る推薦システム 1 0 0 のシステム構成について説明するための説明図である。

【 0 0 5 6 】

図 5 に示すように、推薦システム 1 0 0 は、主に、嗜好抽出エンジン 1 0 1 と、ユーザ嗜好データベース 1 0 2 と、コンテンツ特徴量抽出エンジン 1 0 3 と、コンテンツ特徴量データベース 1 0 4 と、事例関係抽出エンジン 1 0 5 と、事例データベース 1 0 6 と、推薦エンジン 1 0 7 とにより構成される。

【 0 0 5 7 】

30

なお、嗜好抽出エンジン 1 0 1、コンテンツ特徴量抽出エンジン 1 0 3、事例関係抽出エンジン 1 0 5、推薦エンジン 1 0 7 が有する機能は、図 3 8 に示すハードウェア構成のうち、CPU 9 0 2 などの機能により実現される。また、ユーザ嗜好データベース 1 0 2、コンテンツ特徴量データベース 1 0 4、事例データベース 1 0 6 は、図 3 8 に示すハードウェア構成のうち、ROM 9 0 4、RAM 9 0 6、記憶部 9 2 0、リムーバブル記録媒体 9 2 8 などの機能により実現される。また、推薦システム 1 0 0 の機能は、単一のハードウェアを利用して実現されてもよいし、ネットワーク又は専用回線により接続された複数のハードウェアを利用して実現されてもよい。

【 0 0 5 8 】

(コンテンツ特徴量抽出エンジン 1 0 3、コンテンツ特徴量データベース 1 0 4)

40

まず、コンテンツ特徴量抽出エンジン 1 0 3 及びコンテンツ特徴量データベース 1 0 4 について説明する。

【 0 0 5 9 】

コンテンツ特徴量抽出エンジン 1 0 3 は、図 6 に示すようなコンテンツ特徴量データベース 1 0 4 を構築する手段である。コンテンツ特徴量抽出エンジン 1 0 3 は、まず、コンテンツのメタデータを取得する。そして、コンテンツ特徴量抽出エンジン 1 0 3 は、取得したメタデータの構造を参照してコンテンツを構成する各領域を識別し、TF-IDF 値などに基づいて各領域を特徴付ける一又は複数の単語を抽出する。さらに、コンテンツ特徴量抽出エンジン 1 0 3 は、コンテンツの情報、領域の情報、抽出した単語の情報などをコンテンツ特徴量データベース 1 0 4 に格納する。

50

【0060】

例えば、コンテンツ特徴量データベース104には、図6に示すように、アイテムID、領域ID、特徴量ID、更新回数、及び重要度が格納される。アイテムIDは、コンテンツを識別するための識別情報である。また、領域IDは、コンテンツを構成する各領域を識別するための識別情報である。例えば、図4に示したTitleの欄や、Ingredientsの欄は、領域IDにより識別される。さらに、特徴量IDは、対応する領域を特徴付ける単語を識別するための識別情報である。また、更新回数は、対応する領域の内容が更新された回数を示す情報である。重要度は、対応する単語の重要度を示す情報である。なお、コンテンツ特徴量データベース104は、嗜好抽出エンジン101、事例関係抽出エンジン105、推薦エンジン107により利用される。

10

【0061】

(嗜好抽出エンジン101、ユーザ嗜好データベース102)

次に、嗜好抽出エンジン101及びユーザ嗜好データベース102について説明する。

【0062】

ユーザが機器10を介して情報を入力すると、入力された情報は、嗜好抽出エンジン101に入力される。例えば、ユーザの操作ログは、嗜好抽出エンジン101に入力される。ユーザの操作ログが入力されると、嗜好抽出エンジン101は、入力された操作ログに基づいてユーザの嗜好を抽出する。嗜好抽出エンジン101により抽出されたユーザの嗜好を示す情報は、ユーザ嗜好データベース102に格納される。

20

【0063】

ユーザ嗜好データベース102は、図7に示すような構造を有する。ユーザ嗜好データベース102には、図7に示すように、ユーザID、領域ID、特徴量ID、及び重要度を示す情報が格納される。ユーザIDは、ユーザを識別するための識別情報である。領域IDは、コンテンツを構成する各領域を識別するための識別情報である。特徴量IDは、対応する領域を特徴付ける単語を識別するための識別情報である。また、重要度は、特徴量IDにより特定される単語の重要度を示す情報である。なお、ユーザ嗜好データベース102は、推薦エンジン107で利用される。

【0064】

(事例関係抽出エンジン105、事例データベース106)

次に、事例関係抽出エンジン105及び事例データベース106について説明する。

30

【0065】

事例関係抽出エンジン105は、コンテンツ特徴量データベース104に格納された情報に基づいて事例関係を抽出する。この事例関係とは、状況A、結果B、関係Rの関係を意味する。事例関係抽出エンジン105により抽出された事例関係を示す情報は、事例データベース106に格納される。具体的には、図8に示すように、状況Aに関するワードベクトル、結果Bに関するワードベクトル、関係Rに対応するワードベクトルが事例データベース106に格納される。図8の例では、状況Aに関するワードベクトル及び結果Bに関するワードベクトルの次元数がそれぞれ2に設定されている。以下では、この設定例に基づいて説明を進めるが、ワードベクトルの次元数は3以上であってもよい。

40

【0066】

図6に示すように、コンテンツ特徴量データベース104には、アイテムIDと、領域IDと、特徴量IDとが対応付けて格納されている。そこで、事例関係抽出エンジン105は、状況Aに関する情報源として設定された領域IDに対応する特徴量IDの組(word A1, word A2)を抽出し、状況Aのワードベクトルに設定する。また、事例関係抽出エンジン105は、結果Bに関する情報源として設定された領域IDに対応する特徴量IDの組(word B1, word B2)を抽出し、結果Bのワードベクトルに設定する。さらに、事例関係抽出エンジン105は、関係Rに関する情報源として設定された領域IDに対応する特徴量ID(word R)を抽出する。

【0067】

そして、事例関係抽出エンジン105は、同じアイテムIDに対応する(word A

50

1, word A2)、(word B1, word B2)及びword Rを対応付けて事例データベース106に格納する。図8の例では、状況Aのワードベクトル(1, 5)と結果Bのワードベクトル(2, 1)との組み合わせに関して、関係Rのワードベクトル(8, 3, 10, 15)が抽出されている。このように、事例関係抽出エンジン105は、状況A、結果B、関係Rの事例関係をコンテンツ特徴量データベース104から抽出し、事例データベース106を構築する。なお、事例データベース106は、ユーザ毎に構築されていてもよい。このようにして構築された事例データベース106は、推薦エンジン107により利用される。

【0068】

(推薦エンジン107)

次に、推薦エンジン107について説明する。

【0069】

推薦エンジン107は、機器10を介してユーザにより推薦要求を受けると、ユーザ嗜好データベース102、コンテンツ特徴量データベース104、事例データベース106に格納された情報に基づいて推薦コンテンツを抽出する。そして、推薦エンジン107は、抽出した推薦コンテンツのリストを推薦結果としてユーザに提示する。例えば、あるコンテンツ(以下、新規コンテンツ)をユーザが選択した場合、新規コンテンツの情報が推薦エンジン107に入力される。新規コンテンツの情報が入力されると、推薦エンジン107は、入力された新規コンテンツを含むコンテンツ群から、状況Cに関する情報源として設定された領域IDに対応する特徴量IDの組を抽出する。

【0070】

つまり、推薦エンジン107は、新規コンテンツについて状況Cのワードベクトルを抽出する。状況Cのワードベクトルを抽出すると、推薦エンジン107は、抽出した状況Cのワードベクトルと、事例データベース106に格納されている関係Rのワードベクトルとを用いて推薦コンテンツを抽出する。このとき、推薦エンジン107は、複数の推薦コンテンツを抽出し、各推薦コンテンツに対するスコアを算出する。そして、推薦エンジン107は、スコア付けされた推薦コンテンツのリストを生成し、そのリストを推薦結果としてユーザに提示する。

【0071】

以上、推薦システム100のシステム構成について説明した。次に、推薦システム100により実行される処理の流れについて詳細に説明する。

【0072】

[2-2: オフライン処理の流れ]

まず、図9～図12を参照しながら、推薦システム100により実行される処理のうち、オフライン処理の流れについて説明する。図9～図12は、推薦システム100により実行される処理のうち、オフライン処理の流れについて説明するための説明図である。なお、ここで言うオフライン処理とは、機器10から推薦要求を受けていない状態で実行可能な処理のことを意味する。

【0073】

(図9: オフライン処理の概要)

図9に示すように、オフライン処理として実行される主な処理内容は、事例データベース106の構築である。上記の通り、事例データベース106の構築は、主に事例関係抽出エンジン105の機能を用いて実現される。

【0074】

オフライン処理には、ユーザにより過去に生成されたコンテンツ群が利用される。そのため、図9に示すように、オフライン処理の前にユーザによるコンテンツの生成((1)ユーザ記入)という処理が存在する。図9の例では、コンテンツ1～n+1のn+1個のコンテンツが用意されている。但し、より大きな番号が付されたコンテンツほど新しく生成されたものとする。まず、事例関係抽出エンジン105は、これらn+1個のコンテンツから古い順にn個のコンテンツを状況Aに関する情報源として選択する。また、事例関

10

20

30

40

50

係抽出エンジン 105 は、最も新しいコンテンツを結果 B に関する情報源として選択する。ここでは状況 A に関する情報源として選択された n 個のコンテンツを状況 A 1 と表現し、結果 B に関する情報源として選択されたコンテンツを結果 B 1 と表現する。

【0075】

同様に、 $q = 1, \dots, m - 1$ について、事例関係抽出エンジン 105 は、古い順に $n - q$ 個のコンテンツを状況 A に関する情報源として選択する。また、事例関係抽出エンジン 105 は、 $q + 1$ 番目に新しいコンテンツを結果 B に関する情報源として選択する。 $q = 1, \dots, m - 1$ のそれぞれについて、状況 A に関する情報源として選択した $n - q$ 個のコンテンツを状況 A ($q + 1$) と表現し、結果 B に関する情報源として選択されたコンテンツを結果 B ($q + 1$) と表現する。但し、 m は、状況 A m に対応するコンテンツの数が所定数となるように設定される。このようにして状況 A k ($k = 1, \dots, m$) と結果 B k との組を抽出すると (2)、事例関係抽出エンジン 105 は、 $k = 1, \dots, m$ のそれぞれについて、状況 A k と結果 B k との関係 R_k を特徴付けるワードベクトルを生成する。

【0076】

ここでは一例として、状況 A 1 と結果 B 1 との関係 R_1 を特徴付けるワードベクトルの生成方法について説明する。まず、事例関係抽出エンジン 105 は、状況 A 1 に対応する n 個のコンテンツについて、状況 A に関する情報源として設定された領域 (以下、領域 A) を参照し、その領域を特徴付けるワードベクトルを生成する (3)。例えば、事例関係抽出エンジン 105 は、コンテンツ 1 ~ n の領域 A をそれぞれ特徴付ける n 本のワードベクトルを生成し、これら n 本のワードベクトルを統合して状況 A 1 のワードベクトルに設定する。次いで、事例関係抽出エンジン 105 は、状況 A 1 のワードベクトルから単語 (この例では 2 単語) を抽出する (4)。なお、以下の説明において、ここで抽出された単語の組を状況 A 1 のワードベクトルと呼ぶ場合もある。

【0077】

次いで、事例関係抽出エンジン 105 は、結果 B 1 に対応するコンテンツを対象に、結果 B に関する情報源として設定された領域 (以下、領域 B) を特徴付けるワードベクトルを生成し、そのワードベクトルを結果 B 1 のワードベクトルに設定する (5)。次いで、事例関係抽出エンジン 105 は、結果 B 1 のワードベクトルから単語 (この例では 2 単語) を抽出する (6)。なお、以下の説明において、ここで抽出された単語の組を結果 B 1 のワードベクトルと呼ぶ場合もある。次いで、事例関係抽出エンジン 105 は、状況 A 1 のワードベクトルから抽出した単語を領域 A に含み、かつ、結果 B 1 のワードベクトルから抽出した単語を領域 B に含むコンテンツを検索する (7)。

【0078】

次いで、事例関係抽出エンジン 105 は、検索処理により抽出されたコンテンツを対象に、関係 R に関する情報源として設定された領域 (以下、領域 R) を特徴付けるワードベクトルを生成し、そのワードベクトルを関係 R_1 のワードベクトルに設定する (8)。但し、検索処理により複数のコンテンツが抽出された場合、各コンテンツの領域 R をそれぞれ特徴付ける複数のワードベクトルを生成し、これら複数のワードベクトルを統合して関係 R のワードベクトルに設定する。このようにして生成された関係 R_1 のワードベクトルは、状況 A 1 のワードベクトルから抽出された単語と、結果 B 1 のワードベクトルから抽出された単語とに対応付けて事例データベース 106 に格納される。

【0079】

なお、ワードベクトルから抽出される単語の組み合わせは多数存在する。そのため、上記 (4) (6) (7) (8) の処理は、全ての異なる単語の組み合わせについて実行される。そして、順次、上記 (8) で生成されたワードベクトルが関係 R_1 のワードベクトルに追加される。また、以上説明した処理は、状況 A 1 と結果 B 1 との組み合わせについてだけでなく、状況 A 2, ..., A m 、結果 B 2, ..., B m の全ての組み合わせについて実行される。そして、関係 R_1, \dots, R_m のワードベクトルが生成される。

【0080】

(図 10 ~ 図 12 : オフライン処理の詳細)

10

20

30

40

50

オフライン処理は概ね上記の通りである。以下では、オフライン処理として実行される処理の流れについて、より詳細に説明する。

【0081】

図10に示すように、まず、事例関係抽出エンジン105は、ユーザにより過去に選択されたコンテンツを時系列順に並べ、状況Aと結果Bとの組み合わせを設定する(S101)。つまり、事例関係抽出エンジン105は、状況A1, ..., Am、結果B1, ..., Bmに属するコンテンツを設定する。例えば、事例関係抽出エンジン105は、図11に示すように、コンテンツre1, ..., re4を古い順に並べ、状況A1, ..., A3及び結果B1, ..., B3に属するコンテンツを設定する。

【0082】

次いで、事例関係抽出エンジン105は、状況A_jに属するコンテンツ群から重要度の高いN個(Nは自然数)の単語を抽出し、状況A_jのワードベクトルを生成する(S102)。例えば、事例関係抽出エンジン105は、図11に示すように、状況A1に属するコンテンツre1、re2、re3から重要度の高いN個の単語を抽出して状況A1のワードベクトルを生成する。このとき、事例関係抽出エンジン105は、より新しいコンテンツに含まれる単語の重要度が大きくなるように、各単語の重要度に重みを付ける。この重み付けにより、より新しいコンテンツに含まれる特徴的な単語が、状況A1のワードベクトルに含まれやすくなる。

【0083】

次いで、事例関係抽出エンジン105は、結果B_jに属するコンテンツ群から重要度の高いN個(Nは自然数)の単語を抽出し、結果B_jのワードベクトルを生成する(S103)。例えば、事例関係抽出エンジン105は、図11に示すように、結果B_jに属するコンテンツre4から重要度の高いN個の単語を抽出して結果B_jのワードベクトルを生成する。次いで、事例関係抽出エンジン105は、図12に示すように、状況A_jのワードベクトルから単語(この例では2単語)を抽出すると共に、結果B_jのワードベクトルから単語(この例では2単語)を抽出する(S104)。

【0084】

次いで、事例関係抽出エンジン105は、ステップS104で抽出した単語が共起するコンテンツを検索する。そして、事例関係抽出エンジン105は、検索結果として抽出されたコンテンツの領域Rからスコアの高い単語を抽出して関係R_jのワードベクトルに設定する(S105)。図12の例では、検索結果としてコンテンツre1', ..., re8'が抽出されている。また、コンテンツre1', ..., re8'の領域Rを特徴付ける単語の中から、スコアの高い順に単語(図12の例では5単語)が抽出され、関係R_jのワードベクトルが生成されている。

【0085】

なお、単語r_iのスコアscore_R(r_i)は、下記の式(4)に基づいて算出される。但し、状況A_jのワードベクトルから抽出した単語をa_p(p=1, 2, ...)、結果B_jのワードベクトルから抽出した単語をb_q(q=1, 2, ...)と表現する。また、関数ht(x)は、単語xをキーワードにコンテンツを検索した場合にヒットするコンテンツの数を表す。さらに、関数ht(x₁, x₂)は、単語x₁と単語x₂とが共起するコンテンツを検索した場合にヒットするコンテンツの数を表す。そして、関数ht(x₁, x₂, x₃)は、単語x₁と単語x₂と単語x₃とが共起するコンテンツを検索した場合にヒットするコンテンツの数を表す。また、下記の式(4)に含まれるscore(i, p, q)は、下記の式(5)~(8)により定義される。また、下記の式(6)~(8)に現れるNはコンテンツの総数である。

【0086】

10

20

30

40

【数 2】

$$scoreR(ri) = \sum_{p,q} score(i, p, q)$$

... (4)

$$score(i, p, q) = s1(i, p) \cdot s2(i, q) \cdot s3(i, p, q)$$

10

... (5)

$$s1(i, p) = \log \left\{ 1 + \frac{ht(ap, ri) \cdot N}{ht(ap) \cdot ht(ri)} \right\}$$

20

... (6)

$$s2(i, q) = \log \left\{ 1 + \frac{ht(bq) \cdot N}{ht(bq) \cdot ht(ri)} \right\}$$

30

... (7)

$$s3(i, p, q) = \log \left\{ 1 + \frac{ht(ap, bq, ri) \cdot N}{ht(ap, bq) \cdot ht(ri)} \right\}$$

... (8)

【 0 0 8 7 】

40

次いで、事例関係抽出エンジン 1 0 5 は、状況 A_j、結果 B_j、関係 R_j のワードベクトルを事例データベース 1 0 6 に格納する (S 1 0 6)。このとき、事例関係抽出エンジン 1 0 5 は、ステップ S 1 0 5 で検索した際に用いた単語と、関係 R_j のワードベクトルとを対応付けて事例データベース 1 0 6 に格納する。次いで、事例関係抽出エンジン 1 0 5 は、j = m か否かを判定する (S 1 0 7)。つまり、事例関係抽出エンジン 1 0 5 は、ステップ S 1 0 1 で設定した状況 A と結果 B とに関する全ての組み合わせについてステップ S 1 0 2 ~ S 1 0 6 の処理を実行したか否かを判定する。j = m である場合、事例関係抽出エンジン 1 0 5 は、一連の処理を終了する。j = m である場合、事例関係抽出エンジン 1 0 5 は、j を 1 だけインクリメントし、処理をステップ S 1 0 2 に戻す。

【 0 0 8 8 】

50

以上、オフライン処理の流れについて説明した。

【 0 0 8 9 】

[2 - 3 : オンライン処理の流れ]

次に、図 1 3 ~ 図 1 8 を参照しながら、推薦システム 1 0 0 により実行される処理のうち、オンライン処理の流れについて説明する。図 1 3 ~ 図 1 8 は、推薦システム 1 0 0 により実行される処理のうち、オンライン処理の流れについて説明するための説明図である。なお、ここで言うオンライン処理とは、機器 1 0 から推薦要求を受けた際に実行される処理のことを意味する。

【 0 0 9 0 】

(図 1 3 : オンライン処理の概要)

図 1 3 に示すように、オンライン処理として実行される主な処理内容は、主に推薦エンジン 1 0 7 による推薦コンテンツの検索及び検索結果の提示である。

【 0 0 9 1 】

上記の通り、オンライン処理は、推薦要求を受けた際に実行される。つまり、ユーザが新規コンテンツを選択した場合にオンライン処理が実行される。新規コンテンツが選択されると ((1) ユーザ入力)、推薦エンジン 1 0 7 は、状況 C のワードベクトルを抽出する (2)。このとき、推薦エンジン 1 0 7 は、まず、ユーザ嗜好データベース 1 0 2 からユーザの嗜好を表すワードベクトル (以下、嗜好ベクトル) を抽出し、新規コンテンツの状況 C に関する情報源として設定された領域 (以下、領域 C) を特徴付ける単語を用いて嗜好ベクトルを更新する。次いで、推薦エンジン 1 0 7 は、更新後の嗜好ベクトルを状況 C のワードベクトルに設定する。

【 0 0 9 2 】

次いで、推薦エンジン 1 0 7 は、状況 C のワードベクトルから単語 (この例では 2 単語) を抽出する (3)。次いで、推薦エンジン 1 0 7 は、事例データベース 1 0 6 を参照し、関係 R のワードベクトルから単語 (この例では 1 単語) を抽出する (4)。次いで、推薦エンジン 1 0 7 は、状況 C のワードベクトルから抽出した単語が領域 C に出現し、かつ、関係 R のワードベクトルから抽出した単語が領域 R に出現するコンテンツを検索する (5)。次いで、推薦エンジン 1 0 7 は、検索処理により抽出されたコンテンツを示すアイテム ID のリスト (以下、推薦リスト) を生成する (6)。

【 0 0 9 3 】

状況 C のワードリストから抽出される単語、及び関係 R のワードリストから抽出される単語の組み合わせは多数存在する。そのため、推薦リストの生成処理は異なる組み合わせについて繰り返し実行され、複数の推薦リストが生成される。推薦エンジン 1 0 7 は、複数の推薦リストを統合すると共に、各推薦コンテンツに対してスコアを付与する。そして、推薦エンジン 1 0 7 は、付与したスコアに基づいて推薦対象となる推薦コンテンツの組み合わせを選択し、選択された推薦コンテンツを含む推薦リストを生成する (7)。

【 0 0 9 4 】

(図 1 4 ~ 図 1 8 : オンライン処理の詳細)

上記のように、オンライン処理は、状況 C のワードベクトルを生成する際に用いる嗜好ベクトルの更新処理と、推薦リストの生成処理とに大きく分けられる。

【 0 0 9 5 】

(嗜好ベクトルの更新処理)

まず、図 1 4 を参照しながら、嗜好ベクトルの更新処理について説明する。

【 0 0 9 6 】

図 1 4 に示すように、まず、推薦エンジン 1 0 7 は、ユーザ嗜好データベース 1 0 2 から、推薦要求を行ったユーザ (以下、対象ユーザ) の嗜好ベクトルを抽出する (S 1 1 1)。例えば、推薦エンジン 1 0 7 は、対象ユーザのユーザ ID と、領域 C の領域 ID とに対応する特徴量 ID のうち、重要度の高い所定数の特徴量 ID を抽出して嗜好ベクトルに設定する。次いで、推薦エンジン 1 0 7 は、ステップ S 1 1 1 で抽出した対象ユーザの嗜好ベクトルを短縮する (S 1 1 2)。例えば、推薦エンジン 1 0 7 は、嗜好ベクトルに含

10

20

30

40

50

まれる特徴量 I D のうち、重要度の低い所定数の特徴量 I D を削除する。

【 0 0 9 7 】

次いで、推薦エンジン 1 0 7 は、対象ユーザが選択した新規コンテンツの領域 C を特徴付ける単語を抽出し、抽出した単語に対応する特徴量 I D を嗜好ベクトルに追加する (S 1 1 3)。なお、複数の新規コンテンツが選択された場合には、新しい順に大きな重み付けがされた重要度に基づいて領域 C を特徴付ける単語が抽出され、その単語に対応する特徴量 I D が嗜好ベクトルに追加される (図 1 6 を参照)。次いで、推薦エンジン 1 0 7 は、ステップ S 1 1 3 で特徴量 I D が追加された嗜好ベクトルに同期するようにユーザ嗜好データベース 1 0 2 のレコードを更新し (S 1 1 4)、一連の処理を終了する。

【 0 0 9 8 】

以上の処理により、対象ユーザが選択した新規コンテンツの情報で更新された嗜好ベクトルが得られる。なお、ここで説明したユーザ嗜好データベース 1 0 2 の更新処理は、嗜好抽出エンジン 1 0 1 により実行されてもよい。

【 0 0 9 9 】

(推薦リストの生成処理)

次に、図 1 5 ~ 図 1 8 を参照しながら、推薦リストの生成処理について説明する。

【 0 1 0 0 】

図 1 5 に示すように、まず、推薦エンジン 1 0 7 は、ユーザ嗜好データベース 1 0 2 から対象ユーザの嗜好ベクトルを取得し、状況 C のワードベクトルに設定する (S 1 2 1)。なお、新規コンテンツの選択に応じた嗜好ベクトルの更新方法については既に説明した通りであり、嗜好ベクトル (状況 C のワードベクトル) の生成方法については図 1 6 に模式的に示されている。次いで、推薦エンジン 1 0 7 は、事例データベース 1 0 6 から 1 つの事例を選択し、その事例に対応する関係 R のワードベクトルを取得する (S 1 2 2)。

【 0 1 0 1 】

次いで、推薦エンジン 1 0 7 は、図 1 8 (1) に示すように、状況 C のワードベクトルから単語 (この例では 2 単語) を抽出し、関係 R のワードベクトルから単語 (この例では 1 単語) を抽出して、単語の組み合わせを生成する (S 1 2 3)。次いで、推薦エンジン 1 0 7 は、単語の各組み合わせについて、それぞれ共起するコンテンツ群を検索し、各コンテンツに対するスコアを算出する (S 1 2 4)。このスコアは、例えば、下記の式 (9) に基づいて算出される。但し、下記の式 (9) に含まれる s_1 、 s_2 、 s_3 は、下記の式 (1 0) ~ (1 2) により定義される。また、 c_1 、 c_2 は、状況 C のワードベクトルから抽出された単語を表す。

【 0 1 0 2 】

【 数 3 】

$$score(ri) = s1(i) \cdot s2(i) \cdot s3(i)$$

... (9)

$$s1(i) = \log \left\{ 1 + \frac{ht(c1, ri) \cdot N}{ht(c1) \cdot ht(ri)} \right\}$$

... (1 0)

10

20

30

40

$$s2(i) = \log \left\{ 1 + \frac{ht(c2, ri) \cdot N}{ht(c2) \cdot ht(ri)} \right\}$$

... (1 1)

10

$$s3(i) = \log \left\{ 1 + \frac{ht(c1, c2, ri) \cdot N}{ht(c1, c2) \cdot ht(ri)} \right\}$$

... (1 2)

【 0 1 0 3 】

状況 C のワードベクトルから抽出される単語と、関係 R のワードベクトルから抽出される単語との組み合わせは多数存在する。そこで、推薦エンジン 1 0 7 は、全ての組み合わせについて上記のスコアを算出し、コンテンツ毎にスコアを全て足し合わせて、各コンテンツのスコアをリストにした推薦リストを生成する (S 1 2 5)。次いで、推薦エンジン 1 0 7 は、全ての事例について推薦リストを生成したか否かを判定する (S 1 2 6)。

20

【 0 1 0 4 】

全ての事例について推薦リストを生成した場合、推薦エンジン 1 0 7 は、処理をステップ S 1 2 7 に進める。一方、全ての事例について推薦リストを生成していない場合、推薦エンジン 1 0 7 は、処理をステップ S 1 2 2 に戻し、推薦リストを生成していない事例に関する推薦リストを生成する。処理をステップ S 1 2 7 に進めた場合、推薦エンジン 1 0 7 は、全ての事例に対応する推薦リストを組み合わせせて対象ユーザに提示し (S 1 2 7)、一連の処理を終了する。

30

【 0 1 0 5 】

(変形例 : 計算量の削減方法)

図 1 5 に示す処理手順の中では 1 つ 1 つの事例について推薦リストを生成していた。しかし、次のように変形することで処理を簡略化することができる。

【 0 1 0 6 】

まず、推薦エンジン 1 0 7 は、図 1 7 に示すように、状況 C のワードベクトルを取得する。次いで、推薦エンジン 1 0 7 は、状況 C のワードベクトルから単語 (この例では 2 単語) を抽出する (1)。次いで、推薦エンジン 1 0 7 は、事例データベース 1 0 6 を参照し、状況 C のワードベクトルから抽出した単語 (この例では 2 単語) が状況 A のワードベクトル (word A 1 , word A 2) に一致する状況 A、結果 B、関係 R の組を抽出する (2)。次いで、推薦エンジン 1 0 7 は、抽出した組から関係 R のワードベクトルを抽出する (3)。そして、推薦エンジン 1 0 7 は、抽出した関係 R のワードベクトルをマージしてマージベクトル R m e r g e を生成する (4)。

40

【 0 1 0 7 】

次いで、図 1 8 に示すように、推薦エンジン 1 0 7 は、状況 C のワードベクトルから単語 (この例では 2 単語) を抽出し、マージベクトル R m e r g e から単語 (この例では 1 単語) を抽出する (1)。次いで、推薦エンジン 1 0 7 は、抽出した単語の組が共起するコンテンツを検索する (2)。次いで、推薦エンジン 1 0 7 は、検索処理により抽出されたコンテンツの推薦リストを生成する (3)。但し、各コンテンツに付与されるスコアの

50

算出方法は、上記の方法と同じである。次いで、推薦エンジン 107 は、上記 (3) で生成した推薦リストをユーザに提示する。このような処理手順にすることで、計算量を削減することが可能になる。

【0108】

以上、本技術の第 1 実施形態について説明した。

【0109】

< 3 : 第 2 実施形態 >

次に、本技術の第 2 実施形態について説明する。

【0110】

[3 - 1 : システム構成]

まず、図 19 を参照しながら、本実施形態に係る推薦システム 200 のシステム構成について説明する。図 19 は、本実施形態に係る推薦システム 200 のシステム構成について説明するための説明図である。

【0111】

図 19 に示すように、推薦システム 200 は、主に、嗜好抽出エンジン 201 と、ユーザ嗜好データベース 202 と、コンテンツ特徴量抽出エンジン 203 と、コンテンツ特徴量データベース 204 と、事例関係抽出エンジン 205 と、事例データベース 206 と、推薦エンジン 207 と、R パターン抽出エンジン 208 と、R パターンデータベース 209 (重心データベース) とにより構成される。

【0112】

なお、嗜好抽出エンジン 201、コンテンツ特徴量抽出エンジン 203、事例関係抽出エンジン 205、推薦エンジン 207、R パターン抽出エンジン 208 が有する機能は、図 38 に示すハードウェア構成のうち、CPU 902 などの機能により実現される。また、ユーザ嗜好データベース 202、コンテンツ特徴量データベース 204、事例データベース 206、R パターンデータベース 209 は、図 38 に示すハードウェア構成のうち、ROM 904、RAM 906、記憶部 920、リムーバブル記録媒体 928 などの機能により実現される。また、推薦システム 200 の機能は、単一のハードウェアを利用して実現されてもよいし、ネットワーク又は専用回線により接続された複数のハードウェアを利用して実現されてもよい。

【0113】

(コンテンツ特徴量抽出エンジン 203、コンテンツ特徴量データベース 204)

まず、コンテンツ特徴量抽出エンジン 203 及びコンテンツ特徴量データベース 204 について説明する。

【0114】

コンテンツ特徴量抽出エンジン 203 は、コンテンツ特徴量データベース 204 を構築する手段である。コンテンツ特徴量抽出エンジン 203 は、まず、コンテンツのメタデータを取得する。そして、コンテンツ特徴量抽出エンジン 203 は、取得したメタデータの構造を参照してコンテンツを構成する各領域を識別し、TF-IDF 値などに基づいて各領域を特徴付ける一又は複数の単語を抽出する。さらに、コンテンツ特徴量抽出エンジン 203 は、コンテンツの情報、領域の情報、抽出した単語の情報などをコンテンツ特徴量データベース 204 に格納する。なお、コンテンツ特徴量データベース 204 の構造は、図 6 に示したコンテンツ特徴量データベース 104 の構造と実質的に同じである。このコンテンツ特徴量データベース 204 は、嗜好抽出エンジン 201、事例関係抽出エンジン 205、推薦エンジン 207 により利用される。

【0115】

(嗜好抽出エンジン 201、ユーザ嗜好データベース 202)

次に、嗜好抽出エンジン 201 及びユーザ嗜好データベース 202 について説明する。

【0116】

ユーザが機器 10 を介して情報を入力すると、入力された情報は、嗜好抽出エンジン 201 に入力される。例えば、ユーザの操作ログは、嗜好抽出エンジン 201 に入力される

10

20

30

40

50

。ユーザの操作ログが入力されると、嗜好抽出エンジン 201 は、入力された操作ログに基づいてユーザの嗜好を抽出する。嗜好抽出エンジン 201 により抽出されたユーザの嗜好を示す情報は、ユーザ嗜好データベース 202 に格納される。なお、ユーザ嗜好データベース 202 の構造は、図 7 に示したユーザ嗜好データベース 102 の構造と実質的に同じである。このユーザ嗜好データベース 202 は、推薦エンジン 207 で利用される。

【0117】

(事例関係抽出エンジン 205、事例データベース 206)

次に、事例関係抽出エンジン 205 及び事例データベース 206 について説明する。

【0118】

事例関係抽出エンジン 205 は、コンテンツ特徴量データベース 204 に格納された情報に基づいて事例関係を抽出する。事例関係抽出エンジン 205 により抽出された事例関係を示す情報は、事例データベース 206 に格納される。なお、この事例データベース 206 の構造は、図 8 に示した事例データベース 106 の構造と実質的に同じである。この構築された事例データベース 206 は、推薦エンジン 207 及び R パターン抽出エンジン 208 により利用される。

【0119】

(R パターン抽出エンジン 208、R パターンデータベース 209)

次に、R パターン抽出エンジン 208 及び R パターンデータベース 209 について説明する。

【0120】

R パターン抽出エンジン 208 は、事例データベース 206 に格納された関係 R のワードベクトルを用いて各事例をクラスタリングする。次いで、R パターン抽出エンジン 208 は、クラスタ毎に関係 R のワードベクトルに含まれる各単語の出現回数を算出する。次いで、R パターン抽出エンジン 208 は、出現回数が最大の単語又は出現回数が多い単語の組をクラスタの重心を示す単語(以下、重心ワード)として選択し、R パターンデータベース 209 に登録する。さらに、R パターン抽出エンジン 208 は、各クラスタと各事例との対応関係を R パターンデータベース 209 に登録する。

【0121】

R パターンデータベース 209 は、図 20 及び図 21 に示すような構造を有する。なお、ここでは重心データベースを R パターンデータベース 209 の一部として捉えているが、重心データベースを R パターンデータベース 209 とは別に設けてもよい。さて、重心データベースは、図 20 に示すような構造を有する。図 20 に示すように、重心データベースは、クラスタ ID と、重心ワード ID とが対応付けて格納されている。クラスタ ID は、クラスタを識別するための識別情報である。また、重心ワード ID は、重心ワードを識別するための識別情報である。

【0122】

また、クラスタと事例との関係について、R パターンデータベース 209 は、図 21 に示すような構造を有する。図 21 に示すように、R パターンデータベース 209 には、クラスタ ID、状況 A のワードベクトル、結果 B のワードベクトル、及び関係 R のワードベクトルが対応付けて格納されている。この R パターンデータベース 209 は、推薦エンジン 207 により利用される。

【0123】

(推薦エンジン 207)

次に、推薦エンジン 207 について説明する。

【0124】

推薦エンジン 207 は、機器 10 を介してユーザにより推薦要求を受けると、ユーザ嗜好データベース 202、コンテンツ特徴量データベース 204、事例データベース 206、R パターンデータベース 209 に格納された情報に基づいて推薦コンテンツを抽出する。そして、推薦エンジン 207 は、抽出した推薦コンテンツのリストを推薦結果としてユーザに提示する。例えば、あるコンテンツ(以下、新規コンテンツ)をユーザが選択した

10

20

30

40

50

場合、新規コンテンツの情報が推薦エンジン 207 に入力される。新規コンテンツの情報が入力されると、推薦エンジン 207 は、入力された新規コンテンツを含むコンテンツ群から、状況 C のワードベクトルを抽出する。

【0125】

状況 C のワードベクトルを抽出すると、推薦エンジン 207 は、R パターンデータベース 209 からクラスタの情報（以下、R パターン）を読み出してユーザに提示する。例えば、推薦エンジン 207 は、重心ワードをユーザに提示する。提示された R パターンの中からユーザが R パターンを選択すると、推薦エンジン 207 は、選択された R パターンに対応する関係 R のワードベクトルを抽出する。さらに、推薦エンジン 207 は、抽出した状況 C のワードベクトル及び関係 R のワードベクトルを用いて推薦コンテンツを抽出する。そして、推薦エンジン 207 は、各推薦コンテンツに対するスコアを算出し、スコア付けされた推薦コンテンツのリストをユーザに提示する。

10

【0126】

以上、推薦システム 200 のシステム構成について説明した。次に、推薦システム 200 により実行される処理の流れについて詳細に説明する。

【0127】

[3-2: オフライン処理の流れ]

本実施形態のオフライン処理は、事例データベース 206 を生成する段階までは第 1 実施形態のオフライン処理と実質的に同じである。

【0128】

20

まず、図 24（この例は $j = 1$ の場合）に示すように、事例関係抽出エンジン 205 は、状況 A_j のワードベクトルから単語（この例では 2 単語）を抽出すると共に、結果 B_j のワードベクトルから単語（この例では 2 単語）を抽出する。次いで、事例関係抽出エンジン 205 は、抽出した単語が共起するコンテンツを検索する。そして、事例関係抽出エンジン 205 は、検索結果として抽出されたコンテンツの領域 R からスコアの高い単語を抽出して関係 R_j のワードベクトルに設定する。このようにして設定された関係 R_j のワードベクトルは、事例データベース 206 に格納される。

【0129】

本実施形態のオフライン処理と第 1 実施形態のオフライン処理との違いは、R パターンデータベース 209 を構築する処理の有無にある。そこで、R パターン抽出エンジン 208 により実行される R パターンデータベース 209 の構築処理について、図 25 及び図 26 を参照しながら説明する。

30

【0130】

図 26 に示すように、R パターン抽出エンジン 208 は、まず、関係 R のワードベクトルを用いて各事例をクラスタリングする（S211；図 25 を参照）。次いで、R パターン抽出エンジン 208 は、各クラスタに属する関係 R のワードベクトルに含まれる各単語の出現回数を算出する（S212）。次いで、R パターン抽出エンジン 208 は、ステップ S212 で算出した出現回数が最も多い単語又は出現回数の多い単語の組を重心ワードとして選択し、重心データベース（R パターンデータベース 209）に登録する（S213）。次いで、R パターン抽出エンジン 208 は、クラスタと事例との対応関係を R パターンデータベース 209 に登録し（S214）、R パターンデータベース 209 の構築に関する一連の処理を終了する。

40

【0131】

以上、オフライン処理の流れについて説明した。

【0132】

[3-3: オンライン処理の流れ]

次に、図 22、図 23、図 27～図 29 を参照しながら、推薦システム 200 により実行される処理のうち、オンライン処理の流れについて説明する。図 22、図 23、図 27～図 29 は、推薦システム 200 により実行される処理のうち、オンライン処理の流れについて説明するための説明図である。

50

【 0 1 3 3 】

(図 2 2 : オンライン処理の概要)

図 2 2 に示すように、オンライン処理として実行される主な処理内容は、主に推薦エンジン 2 0 7 による R パターンの提示、推薦コンテンツの検索及び検索結果の提示である。

【 0 1 3 4 】

上記の通り、オンライン処理は、推薦要求を受けた際に実行される。つまり、ユーザが新規コンテンツを選択した場合にオンライン処理が実行される。新規コンテンツが選択されると ((1) ユーザ入力)、推薦エンジン 2 0 7 は、状況 C のワードベクトルを抽出する (2)。このとき、推薦エンジン 2 0 7 は、まず、ユーザ嗜好データベース 2 0 2 から嗜好ベクトルを抽出し、新規コンテンツの領域 C を特徴付ける単語を用いて嗜好ベクトルを更新する。次いで、推薦エンジン 2 0 7 は、更新後の嗜好ベクトルを状況 C のワードベクトルに設定する。

10

【 0 1 3 5 】

次いで、推薦エンジン 2 0 7 は、状況 C のワードベクトルから単語 (この例では 2 単語) を抽出する (3)。次いで、推薦エンジン 2 0 7 は、R パターンデータベース 2 0 9 から R パターンを読み出し、読み出した R パターンをユーザに提示する。提示した R パターンの中からユーザが R パターンを選択すると、推薦エンジン 2 0 7 は、ユーザが選択した R パターンに対応する関係 R のワードベクトルから単語 (この例では 1 単語) を抽出する (4)。次いで、推薦エンジン 2 0 7 は、状況 C のワードベクトルから抽出した単語が領域 C に出現し、かつ、関係 R のワードベクトルから抽出した単語が領域 R に出現するコンテンツを検索する (5)。次いで、推薦エンジン 2 0 7 は、検索処理により抽出されたコンテンツを示すアイテム ID のリスト (推薦リスト) を生成する (6)。

20

【 0 1 3 6 】

状況 C のワードリストから抽出される単語、及び関係 R のワードリストから抽出される単語の組み合わせは多数存在する。そのため、推薦リストの生成処理は異なる組み合わせについて繰り返し実行され、複数の推薦リストが生成される。推薦エンジン 2 0 7 は、複数の推薦リストを統合すると共に、各推薦コンテンツに対してスコアを付与する。そして、推薦エンジン 2 0 7 は、付与したスコアに基づいて推薦対象となる推薦コンテンツの組み合わせを選択し、選択された推薦コンテンツを含む推薦リストを生成する (7)。

30

【 0 1 3 7 】

(図 2 3、図 2 7 ~ 図 2 9 : オンライン処理の詳細)

上記のように、本実施形態に係るオンライン処理には、R パターンをユーザに選択させる処理が含まれる。ここで、オンライン処理の流れについて、より詳細に説明する。

【 0 1 3 8 】

図 2 3 に示すように、まず、推薦エンジン 2 0 7 は、ユーザ嗜好データベース 2 0 2 から対象ユーザの嗜好ベクトルを取得し、状況 C のワードベクトルに設定する (S 2 0 1)。なお、新規コンテンツの選択に応じた嗜好ベクトルの更新方法、及び嗜好ベクトル (状況 C のワードベクトル) の生成方法は第 1 実施形態の中で説明した方法と実質的に同じである。次いで、推薦エンジン 2 0 7 は、図 2 7 に示すように、複数の R パターンをユーザに提示し、1つの R パターンをユーザに選択させる (S 2 0 2)。次いで、推薦エンジン 2 0 7 は、選択された R パターンに対応する関係 R のワードベクトルを取得する (S 2 0 3)。

40

【 0 1 3 9 】

次いで、推薦エンジン 2 0 7 は、図 2 9 に示すように、状況 C のワードベクトルから単語 (この例では 2 単語) を抽出し、関係 R のワードベクトルから単語 (この例では 1 単語) を抽出して、単語の組み合わせを生成する (S 2 0 4)。次いで、推薦エンジン 2 0 7 は、単語の各組み合わせについて、それぞれ共起するコンテンツ群を検索し、各コンテンツに対するスコアを算出する (S 2 0 5)。なお、スコアの算出方法は、第 1 実施形態の中で説明した方法と実質的に同じである。

【 0 1 4 0 】

50

状況Cのワードベクトルから抽出される単語と、関係Rのワードベクトルから抽出される単語との組み合わせは多数存在する。そこで、推薦エンジン207は、全ての組み合わせについて上記のスコアを算出し、コンテンツ毎にスコアを全て足し合わせて、各コンテンツのスコアをリストにした推薦リストを生成する(S206)。次いで、推薦エンジン207は、全ての事例について推薦リストを生成したか否かを判定する(S207)。

【0141】

全ての事例について推薦リストを生成した場合、推薦エンジン207は、処理をステップS208に進める。一方、全ての事例について推薦リストを生成していない場合、推薦エンジン207は、処理をステップS203に戻し、推薦リストを生成していない事例に関する推薦リストを生成する。処理をステップS208に進めた場合、推薦エンジン207は、全ての事例に対応する推薦リストを組み合わせ対象ユーザに提示し(S208)、一連の処理を終了する。

10

【0142】

(変形例：計算量の削減方法)

図23に示す処理手順の中では1つ1つの事例について推薦リストを生成していた。しかし、第1実施形態の場合と同様、次のように変形することで処理を簡略化することができる。

【0143】

まず、推薦エンジン207は、図28に示すように、状況Cのワードベクトルを取得する。次いで、推薦エンジン207は、状況Cのワードベクトルから単語(この例では2単語)を抽出する(1)。次いで、推薦エンジン207は、事例データベース206を参照し、状況Cのワードベクトルから抽出した単語(この例では2単語)が状況Aのワードベクトル(word A1, word A2)に一致する状況A、結果B、関係Rの組を抽出する(2)。但し、ユーザにより選択されたRパターンに対応する事例の中から、状況A、結果B、関係Rの組が抽出される。次いで、推薦エンジン207は、抽出した組から関係Rのワードベクトルを抽出する(3)。そして、推薦エンジン207は、抽出した関係RのワードベクトルをマージしてマージベクトルRmergeを生成する(4)。

20

【0144】

次いで、図29に示すように、推薦エンジン207は、状況Cのワードベクトルから単語(この例では2単語)を抽出し、マージベクトルRmergeから単語(この例では1単語)を抽出する(1)。次いで、推薦エンジン207は、抽出した単語の組が共起するコンテンツを検索する(2)。次いで、推薦エンジン207は、検索処理により抽出されたコンテンツの推薦リストを生成する(3)。但し、各コンテンツに付与されるスコアの算出方法は、第1実施形態の中で説明した方法と実質的に同じである。次いで、推薦エンジン207は、上記(3)で生成した推薦リストをユーザに提示する。このような処理手順にすることで、計算量を削減することが可能になる。

30

【0145】

以上、本技術の第2実施形態について説明した。

【0146】

<4：第3実施形態>

40

次に、本技術の第3実施形態について説明する。

【0147】

[4-1：システム構成]

まず、図30を参照しながら、本実施形態に係る推薦システム300のシステム構成について説明する。図30は、本実施形態に係る推薦システム300のシステム構成について説明するための説明図である。

【0148】

図30に示すように、推薦システム300は、主に、嗜好抽出エンジン301と、ユーザ嗜好データベース302と、コンテンツ特徴量抽出エンジン303と、コンテンツ特徴量データベース304と、事例関係抽出エンジン305と、事例データベース306と、

50

推薦エンジン 307 と、推薦リスト計算エンジン 308 と、推薦リストデータベース 309 とにより構成される。

【0149】

なお、嗜好抽出エンジン 301、コンテンツ特徴量抽出エンジン 303、事例関係抽出エンジン 305、推薦エンジン 307、推薦リスト計算エンジン 308 が有する機能は、図 38 に示すハードウェア構成のうち、CPU 902 などの機能により実現される。また、ユーザ嗜好データベース 302、コンテンツ特徴量データベース 304、事例データベース 306、推薦リストデータベース 309 は、図 38 に示すハードウェア構成のうち、ROM 904、RAM 906、記憶部 920、リムーバブル記録媒体 928 などの機能により実現される。また、推薦システム 300 の機能は、単一のハードウェアを利用して実現されてもよいし、ネットワーク又は専用回線により接続された複数のハードウェアを利用して実現されてもよい。

10

【0150】

(コンテンツ特徴量抽出エンジン 303、コンテンツ特徴量データベース 304)

まず、コンテンツ特徴量抽出エンジン 303 及びコンテンツ特徴量データベース 304 について説明する。

【0151】

コンテンツ特徴量抽出エンジン 303 は、コンテンツ特徴量データベース 304 を構築する手段である。コンテンツ特徴量抽出エンジン 303 は、まず、コンテンツのメタデータを取得する。そして、コンテンツ特徴量抽出エンジン 303 は、取得したメタデータの構造を参照してコンテンツを構成する各領域を識別し、TF-IDF 値などに基づいて各領域を特徴付ける一又は複数の単語を抽出する。さらに、コンテンツ特徴量抽出エンジン 303 は、コンテンツの情報、領域の情報、抽出した単語の情報などをコンテンツ特徴量データベース 304 に格納する。なお、コンテンツ特徴量データベース 304 の構造は、図 6 に示したコンテンツ特徴量データベース 104 の構造と実質的に同じである。このコンテンツ特徴量データベース 304 は、嗜好抽出エンジン 301、事例関係抽出エンジン 305、推薦エンジン 307 により利用される。

20

【0152】

(嗜好抽出エンジン 301、ユーザ嗜好データベース 302)

次に、嗜好抽出エンジン 301 及びユーザ嗜好データベース 302 について説明する。

30

【0153】

ユーザが機器 10 を介して情報を入力すると、入力された情報は、嗜好抽出エンジン 301 に入力される。例えば、ユーザの操作ログは、嗜好抽出エンジン 301 に入力される。ユーザの操作ログが入力されると、嗜好抽出エンジン 301 は、入力された操作ログに基づいてユーザの嗜好を抽出する。嗜好抽出エンジン 301 により抽出されたユーザの嗜好を示す情報は、ユーザ嗜好データベース 302 に格納される。なお、ユーザ嗜好データベース 302 の構造は、図 7 に示したユーザ嗜好データベース 102 の構造と実質的に同じである。このユーザ嗜好データベース 302 は、推薦エンジン 307 で利用される。

【0154】

(事例関係抽出エンジン 305、事例データベース 306)

次に、事例関係抽出エンジン 305 及び事例データベース 306 について説明する。

40

【0155】

事例関係抽出エンジン 305 は、コンテンツ特徴量データベース 304 に格納された情報に基づいて事例関係を抽出する。事例関係抽出エンジン 305 により抽出された事例関係を示す情報は、事例データベース 306 に格納される。なお、この事例データベース 306 の構造は、図 8 に示した事例データベース 106 の構造と実質的に同じである。この構築された事例データベース 306 は、推薦エンジン 307 及び推薦リスト計算エンジン 308 により利用される。

【0156】

(推薦リスト計算エンジン 308、推薦リストデータベース 309)

50

次に、推薦リスト計算エンジン 308 及び推薦リストデータベース 309 について説明する。

【0157】

推薦リスト計算エンジン 308 は、事例データベース 306 に格納された情報に基づいて推薦リストを生成する。なお、推薦リスト計算エンジン 308 による推薦リストの生成方法は、上記の第 1 実施形態に係る推薦エンジン 107 又は第 2 実施形態に係る推薦エンジン 207 による推薦リストの生成方法と実質的に同じである。但し、推薦リスト計算エンジン 308 は、上記の推薦エンジン 107、207 とは異なり、オフライン処理で推薦リストを生成する。そのため、推薦リスト計算エンジン 308 は、状況 C のワードベクトルとして、ユーザ嗜好データベース 302 に格納されている嗜好ベクトル又は所定の単語群を用いる。推薦リスト計算エンジン 308 により生成された推薦リストは、推薦リストデータベース 309 に格納される。

10

【0158】

推薦リストデータベース 309 は、図 31 に示すような構造を有する。図 31 に示すように、推薦リストデータベース 309 には、状況 C のワードベクトルと、推薦リストに含まれるコンテンツのアイテム ID 及びスコアとが含まれる。この推薦リストデータベース 309 は、推薦エンジン 307 により利用される。

【0159】

(推薦エンジン 307)

次に、推薦エンジン 307 について説明する。

20

【0160】

推薦エンジン 307 は、機器 10 を介してユーザにより推薦要求を受けると、ユーザ嗜好データベース 302、コンテンツ特徴量データベース 304、事例データベース 306、推薦リストデータベース 309 に格納された情報に基づいて推薦リストを抽出し、推薦結果としてユーザに提示する。例えば、あるコンテンツ（以下、新規コンテンツ）をユーザが選択した場合、新規コンテンツの情報が推薦エンジン 307 に入力される。新規コンテンツの情報が入力されると、推薦エンジン 307 は、入力された新規コンテンツを含むコンテンツ群から、状況 C のワードベクトルを抽出する。次いで、推薦エンジン 307 は、抽出した状況 C のワードベクトルから単語の組を抽出し、抽出した単語の組に対応する推薦リストを推薦リストデータベース 309 から抽出する。そして、推薦エンジン 307 は、推薦リストデータベース 309 から抽出した推薦リストをユーザに提示する。

30

【0161】

以上、推薦システム 300 のシステム構成について説明した。次に、推薦システム 300 により実行される処理の流れについて詳細に説明する。

【0162】

[4-2: オフライン処理の流れ]

本実施形態のオフライン処理は、事例データベース 306 を生成する段階までは第 1 実施形態のオフライン処理と実質的に同じである。

【0163】

まず、図 34（この例は $j = 1$ の場合）に示すように、事例関係抽出エンジン 305 は、状況 A $_j$ のワードベクトルから単語（この例では 2 単語）を抽出すると共に、結果 B $_j$ のワードベクトルから単語（この例では 2 単語）を抽出する。次いで、事例関係抽出エンジン 305 は、抽出した単語が共起するコンテンツを検索する。そして、事例関係抽出エンジン 305 は、検索結果として抽出されたコンテンツの領域 R からスコアの高い単語を抽出して関係 R $_j$ のワードベクトルに設定する。このようにして設定された関係 R $_j$ のワードベクトルは、事例データベース 306 に格納される。

40

【0164】

本実施形態のオフライン処理と第 1 実施形態のオフライン処理との違いは、オフライン処理で予め推薦リストを生成し、推薦リストデータベース 309 に保持しておく点にある。図 32 に示すように、まず、推薦リスト計算エンジン 308 は、ユーザ嗜好データベー

50

ス 3 0 2 に格納された嗜好ベクトル又は所定の単語群（以下、全単語リスト C'）から単語（この例では 2 単語）を抽出する（1）。次いで、推薦リスト計算エンジン 3 0 8 は、抽出した単語の組と状況 A のワードベクトルとが一致する事例を抽出する（2）。次いで、推薦リスト計算エンジン 3 0 8 は、抽出した事例から関係 R のワードベクトルを抽出する（3）。次いで、推薦リスト計算エンジン 3 0 8 は、抽出した関係 R のワードベクトルをマージしてマージベクトル R m e r g e を生成する（4）。

【 0 1 6 5 】

次いで、図 3 3 に示すように、推薦リスト計算エンジン 3 0 8 は、全単語リスト C' から単語（この例では 2 単語）を抽出すると共に、マージベクトル R m e r g e から単語（この例では 1 単語）を抽出する（1）。次いで、推薦リスト計算エンジン 3 0 8 は、全単語リスト C' から抽出した単語を領域 C に含み、かつ、マージベクトル R m e r g e から抽出した単語を領域 R に含むコンテンツを検索する（2）。次いで、推薦リスト計算エンジン 3 0 8 は、検索処理により抽出された各コンテンツのスコアを算出し、推薦リストを生成する（3）。

【 0 1 6 6 】

なお、上記（1）で抽出される単語の組み合わせは多数存在する。そのため、推薦リスト計算エンジン 3 0 8 は、これら多数の組み合わせについて、上記（2）の検索処理及び上記（3）の推薦リストの生成処理を繰り返し実行する。このようにして生成された推薦リストは、推薦リストデータベース 3 0 9 に格納される。

【 0 1 6 7 】

以上、オフライン処理の流れについて説明した。

【 0 1 6 8 】

[4 - 3 : オンライン処理の流れ]

次に、図 3 5 及び図 3 6 を参照しながら、推薦システム 3 0 0 により実行される処理のうち、オンライン処理の流れについて説明する。図 3 5 及び図 3 6 は、推薦システム 3 0 0 により実行される処理のうち、オンライン処理の流れについて説明するための説明図である。

【 0 1 6 9 】

（オンライン処理の概要）

まず、図 3 5 を参照する。オンライン処理は、推薦要求を受けた際に実行される。ユーザにより新規コンテンツが選択されると（（1）ユーザ入力）、推薦エンジン 3 0 7 は、状況 C のワードベクトルを抽出する（2）。このとき、推薦エンジン 3 0 7 は、まず、ユーザ嗜好データベース 3 0 2 から嗜好ベクトルを抽出し、新規コンテンツの領域 C を特徴付ける単語を用いて嗜好ベクトルを更新する。次いで、推薦エンジン 3 0 7 は、更新後の嗜好ベクトルを状況 C のワードベクトルに設定する。

【 0 1 7 0 】

次いで、推薦エンジン 3 0 7 は、状況 C のワードベクトルから単語（この例では 2 単語）を抽出する（3）。次いで、推薦エンジン 3 0 7 は、推薦リストデータベース 3 0 9 を参照し、抽出した単語の組と同じワードベクトルに対応するコンテンツとスコアとを抽出し、これらコンテンツとスコアとを組み合わせる推薦リストを生成する（4）。状況 C のワードリストから抽出される単語、及び関係 R のワードリストから抽出される単語の組み合わせは多数存在する。そのため、推薦リストの生成処理は異なる組み合わせについて繰り返し実行され、複数の推薦リストが生成される。推薦エンジン 3 0 7 は、複数の推薦リストを統合し、推薦結果としてユーザに提示する。

【 0 1 7 1 】

（オンライン処理の流れ）

図 3 5 を参照しながら説明したオンライン処理の流れは、図 3 6 のようなフローチャートに纏めることができる。図 3 6 に示すように、推薦エンジン 3 0 7 は、状況 C のワードベクトルから単語（この例では 2 単語）の組み合わせを抽出する（S 3 1 1）。次いで、推薦エンジン 3 0 7 は、ステップ S 3 1 1 で抽出した各組み合わせに対応する推薦スコア

を推薦リストデータベース309から取得する(S312)。次いで、推薦エンジン307は、ステップS313で取得した推薦スコアとコンテンツとを組み合わせた推薦リストを生成してユーザに提示し(S313)、一連の処理を終了する。このように、推薦リストをオフライン処理で予め生成しておくことにより、オンライン処理が大幅に簡略化される。

【0172】

以上、本技術の第3実施形態について説明した。

【0173】

<5：応用例(クロスカテゴリ推薦)>

ここで、上記の第1～第3実施形態に係る応用例として、クロスカテゴリ推薦への応用について説明する。クロスカテゴリ推薦とは、カテゴリを跨いだ構造写像に基づいてコンテンツの推薦を実行することを意味する。クロスカテゴリ推薦の例としては、図38に示すように、状況Aと結果Bとの関係Rをレシピの紹介ページから抽出し、その関係Rと状況Cとを用いて、ユーザの嗜好に合った本の紹介ページを推薦する例が挙げられる。レシピの紹介ページと本の紹介ページとは互いに異なるカテゴリに属するコンテンツである。しかし、上記の第1～第3実施形態に係る技術を適用することにより、このようなカテゴリを跨いだ推薦を実現することが可能になる。

【0174】

<6：ハードウェア構成例>

上記の推薦システム100、200、300が有する各構成要素の機能は、例えば、図38に示す情報処理装置のハードウェア構成を用いて実現することが可能である。つまり、当該各構成要素の機能は、コンピュータプログラムを用いて図38に示すハードウェアを制御することにより実現される。なお、このハードウェアの形態は任意であり、例えば、パーソナルコンピュータ、携帯電話、PHS、PDA等の携帯情報端末、ゲーム機、又は種々の情報家電がこれに含まれる。但し、上記のPHSは、Personal Handy-phone Systemの略である。また、上記のPDAは、Personal Digital Assistantの略である。

【0175】

図38に示すように、このハードウェアは、主に、CPU902と、ROM904と、RAM906と、ホストバス908と、ブリッジ910と、を有する。さらに、このハードウェアは、外部バス912と、インターフェース914と、入力部916と、出力部918と、記憶部920と、ドライブ922と、接続ポート924と、通信部926と、を有する。但し、上記のCPUは、Central Processing Unitの略である。また、上記のROMは、Read Only Memoryの略である。そして、上記のRAMは、Random Access Memoryの略である。

【0176】

CPU902は、例えば、演算処理装置又は制御装置として機能し、ROM904、RAM906、記憶部920、又はリムーバブル記録媒体928に記録された各種プログラムに基づいて各構成要素の動作全般又はその一部を制御する。ROM904は、CPU902に読み込まれるプログラムや演算に用いるデータ等を格納する手段である。RAM906には、例えば、CPU902に読み込まれるプログラムや、そのプログラムを実行する際に適宜変化する各種パラメータ等が一時的又は永続的に格納される。

【0177】

これらの構成要素は、例えば、高速なデータ伝送が可能なホストバス908を介して相互に接続される。一方、ホストバス908は、例えば、ブリッジ910を介して比較的データ伝送速度が低速な外部バス912に接続される。また、入力部916としては、例えば、マウス、キーボード、タッチパネル、ボタン、スイッチ、及びレバー等が用いられる。さらに、入力部916としては、赤外線やその他の電波を利用して制御信号を送信することが可能なりモートコントローラ(以下、リモコン)が用いられることもある。

【0178】

10

20

30

40

50

出力部 918 としては、例えば、CRT、LCD、PDP、又は ELD 等のディスプレイ装置、スピーカ、ヘッドホン等のオーディオ出力装置、プリンタ、携帯電話、又はファクシミリ等、取得した情報を利用者に対して視覚的又は聴覚的に通知することが可能な装置である。但し、上記の CRT は、Cathode Ray Tube の略である。また、上記の LCD は、Liquid Crystal Display の略である。そして、上記の PDP は、Plasma Display Panel の略である。さらに、上記の ELD は、Electro-Luminescence Display の略である。

【0179】

記憶部 920 は、各種のデータを格納するための装置である。記憶部 920 としては、例えば、ハードディスクドライブ (HDD) 等の磁気記憶デバイス、半導体記憶デバイス、光記憶デバイス、又は光磁気記憶デバイス等が用いられる。但し、上記の HDD は、Hard Disk Drive の略である。

【0180】

ドライブ 922 は、例えば、磁気ディスク、光ディスク、光磁気ディスク、又は半導体メモリ等のリムーバブル記録媒体 928 に記録された情報を読み出し、又はリムーバブル記録媒体 928 に情報を書き込む装置である。リムーバブル記録媒体 928 は、例えば、DVD メディア、Blu-ray メディア、HD DVD メディア、各種の半導体記憶メディア等である。もちろん、リムーバブル記録媒体 928 は、例えば、非接触型 IC チップを搭載した IC カード、又は電子機器等であってもよい。但し、上記の IC は、Integrated Circuit の略である。

【0181】

接続ポート 924 は、例えば、USB ポート、IEEE 1394 ポート、SCSI、RS-232C ポート、又は光オーディオ端子等のような外部接続機器 930 を接続するためのポートである。外部接続機器 930 は、例えば、プリンタ、携帯音楽プレーヤ、デジタルカメラ、デジタルビデオカメラ、又は IC レコーダ等である。但し、上記の USB は、Universal Serial Bus の略である。また、上記の SCSI は、Small Computer System Interface の略である。

【0182】

通信部 926 は、ネットワーク 932 に接続するための通信デバイスであり、例えば、有線又は無線 LAN、Bluetooth (登録商標)、又は WUSB 用の通信カード、光通信用のルータ、ADSL 用のルータ、又は各種通信用のモデム等である。また、通信部 926 に接続されるネットワーク 932 は、有線又は無線により接続されたネットワークにより構成され、例えば、インターネット、家庭内 LAN、赤外線通信、可視光通信、放送、又は衛星通信等である。但し、上記の LAN は、Local Area Network の略である。また、上記の WUSB は、Wireless USB の略である。そして、上記の ADSL は、Asymmetric Digital Subscriber Line の略である。

【0183】

<7: まとめ>

最後に、本実施形態の技術的思想について簡単に纏める。以下に記載する技術的思想は、例えば、PC、携帯電話、携帯ゲーム機、携帯情報端末、情報家電、カーナビゲーションシステム等、種々の情報処理装置に対して適用することができる。また、以下に記載する情報処理装置の一部又は全部の機能をクラウドシステムなどの情報処理システムにより実現することも可能である。

【0184】

上記の情報処理装置の機能構成は次のように表現することができる。当該情報処理装置は、以下に示す第 1 特徴量生成部、第 2 特徴量生成部、関係特徴量生成部、及び推薦コンテンツ検索部を有する。上記の第 1 特徴量生成部は、対象ユーザが過去に選択した第 1 のコンテンツに含まれる第 1 の種類の情報に基づいて第 1 の特徴量を生成する手段である。また、上記の第 2 特徴量生成部は、前記第 1 のコンテンツを選択した後に前記対象ユーザ

が選択した第2のコンテンツに含まれる第2の種類の情報に基づいて第2の特徴量を生成する手段である。

【0185】

例えば、上記の第1特徴量生成部は、対象ユーザが過去に選択したコンテンツX1、...、Xnのメタデータに基づき、コンテンツX1、...、Xnのそれぞれについて、予め設定された第1の種類のメタデータが付与された一又は複数の情報領域を認識する。次いで、上記の第1特徴量生成部は、認識した情報領域から所定数の特徴的な情報単位（例えば、単語など）を抽出し、その情報単位を要素に持つ特徴ベクトル（以下、第1の特徴ベクトル）を生成する。なお、コンテンツX1、...、Xnは、X1、...、Xnの順に選択されたものとする。一方、上記の第2特徴量生成部は、対象ユーザにより最後に選択されたコンテンツXnの次に選択されたコンテンツYに注目する。

10

【0186】

まず、上記の第2特徴量生成部は、コンテンツYのメタデータに基づき、予め設定された第2の種類のメタデータが付与された一又は複数の情報領域を認識する。次いで、上記の第2特徴量生成部は、認識した情報領域から所定数の特徴的な情報単位（例えば、単語など）を抽出し、その情報単位を要素に持つ特徴ベクトル（以下、第2の特徴ベクトル）を生成する。この例におけるコンテンツYは、対象ユーザが過去にコンテンツX1、...、Xnを選択したという事前条件（以下、状況A）の下で選択されたものである。つまり、上記の第1の特徴ベクトルは対象ユーザの状況Aを特徴付けるものであり、上記の第2の特徴ベクトルは状況Aの結果（以下、結果B）を特徴付けるものである。

20

【0187】

さて、上記の関係特徴量生成部は、前記第1特徴量生成部により生成された第1の特徴量と、前記第2特徴量生成部により生成された第2の特徴量とに基づき、前記第1のコンテンツと前記第2のコンテンツとの間の関係を示す関係特徴量を生成する手段である。上記の例に表現を合わせると、上記の関係特徴量生成部は、第1の特徴ベクトルと、第2の特徴ベクトルとを用いて、状況Aと結果Bとの間の関係を示す関係特徴量を生成する手段である。

【0188】

例えば、上記の関係特徴量生成部は、第1の特徴ベクトルから抽出した所定数の情報要素と、第2の特徴ベクトルから抽出した所定数の情報要素とが共起するコンテンツを検索する。そして、上記の関係特徴量生成部は、検出されたコンテンツの中から所定数の特徴的な情報単位（例えば、単語など）を抽出し、その情報単位を要素に持つ特徴ベクトル（以下、関連特徴ベクトル）を生成する。このとき、上記の関係特徴量生成部は、所定の情報領域の中から情報単位を抽出するように構成されていてもよい。その場合、上記の関係特徴量生成部は、コンテンツに付与されたメタデータを参照して、情報単位の抽出に用いる情報領域を選択する。なお、上記の関連特徴ベクトルは、状況Aと結果Bとの関係（以下、関係R）を特徴付けるものである。

30

【0189】

さて、上記の推薦コンテンツ検索部は、前記対象ユーザが新たに選択したコンテンツに含まれる前記第1の種類の情報と、前記関係特徴量生成部により生成された関係特徴量とを用いて、前記対象ユーザに推薦するコンテンツを検索する手段である。上記の例に表現を合わせると、上記の推薦コンテンツ検索部は、上記の状況Aに対応する新たな状況Cが与えられた場合に、その状況Cと関係Rとを用いて、対象ユーザに推薦すべきコンテンツを検索する手段である。ここで、上記の推薦コンテンツ検索部は、上記の関係Rを用いて、上記の状況Cに対応する結果Xを得る手段ではない点に注意されたい。

40

【0190】

このように、対象ユーザの履歴に基づいて状況Aと結果Bとの関係Rを算出し、その関係Rを用いて推薦すべきコンテンツを検索することにより、対象ユーザの嗜好に適合した新鮮な情報を含むコンテンツを得ることが可能になる。

50

【 0 1 9 1 】

(備考)

上記の推薦システム 1 0 0、2 0 0、3 0 0 は、コンテンツ推薦装置の一例である。上記の事例関係抽出エンジン 1 0 5、2 0 5、3 0 5 は、第 1 特徴量生成部、第 2 特徴量生成部、関係特徴量生成部の一例である。上記の推薦エンジン 1 0 7、2 0 7、3 0 7 は、推薦コンテンツ検索部の一例である。上記の R パターン抽出エンジン 2 0 8 は、関係特徴量生成部の一例である。上記の推薦エンジン 2 0 7 は、関係選択要求部の一例である。上記の事例データベース 1 0 6、2 0 6、3 0 6 は、特徴量格納部の一例である。

【 0 1 9 2 】

以上、添付図面を参照しながら本技術に係る好適な実施形態について説明したが、本技術はここで開示した構成例に限定されないことは言うまでもない。当業者であれば、特許請求の範囲に記載された範疇内において、各種の変更例又は修正例に想到し得ることは明らかであり、それらについても当然に本技術の技術的範囲に属するものと了解される。

10

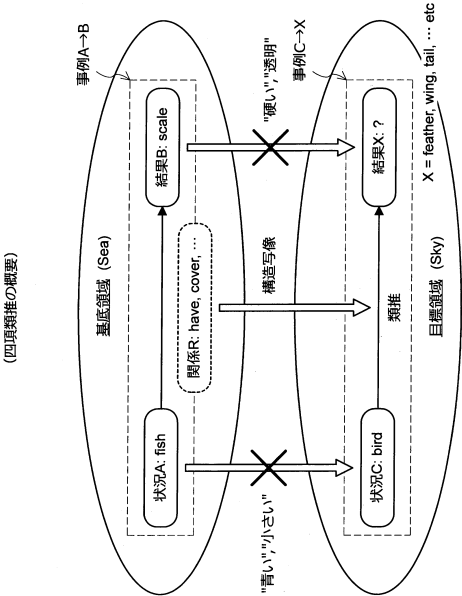
【符号の説明】

【 0 1 9 3 】

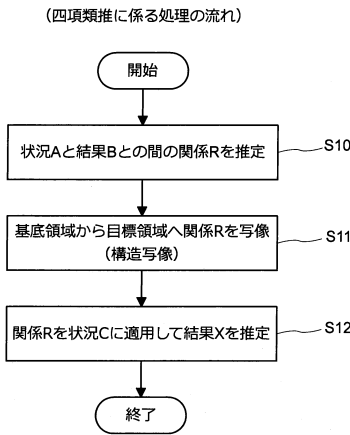
1 0	機器	
1 0 0、2 0 0、3 0 0	推薦システム	
1 0 1、2 0 1、3 0 1	嗜好抽出エンジン	
1 0 2、2 0 2、3 0 2	ユーザ嗜好データベース	
1 0 3、2 0 3、3 0 3	コンテンツ特徴量抽出エンジン	
1 0 4、2 0 4、3 0 4	コンテンツ特徴量データベース	
1 0 5、2 0 5、3 0 5	事例関係抽出エンジン	
1 0 6、2 0 6、3 0 6	事例データベース	
1 0 7、2 0 7、3 0 7	推薦エンジン	
2 0 8	R パターン抽出エンジン	
2 0 9	R パターンデータベース	
3 0 8	推薦リスト計算エンジン	
3 0 9	推薦リストデータベース	

20

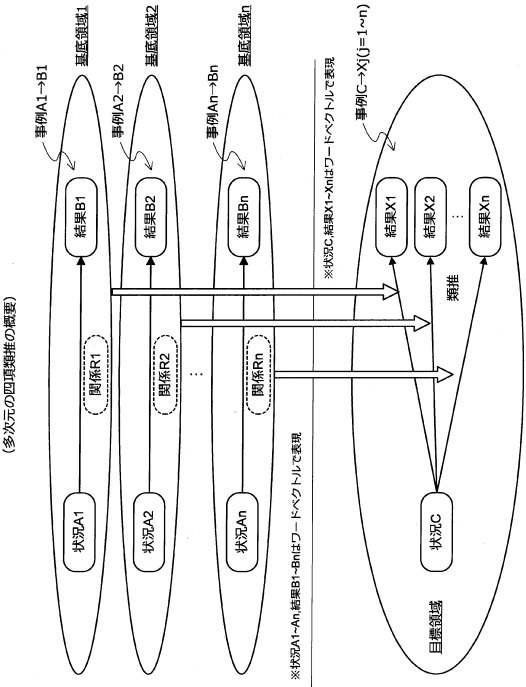
【図 1】



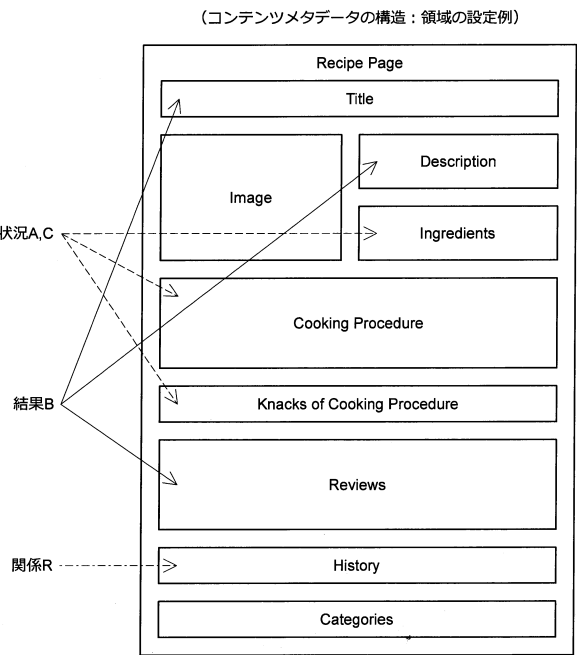
【図 2】



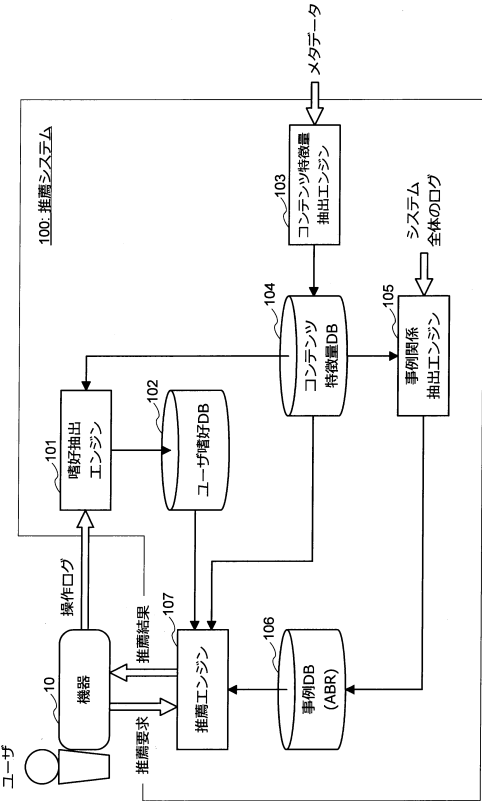
【図 3】



【図 4】



【図 5】



【図 6】

(コンテンツ特徴量DBの構造)

アイテムID	領域ID	特徴量ID	更新回数	重要度
0001	1	2	10	1.58
0002	1	30	21	1.12
0003	2	52	2	2.11
0004	11	4	8	1.01
⋮	⋮	⋮	⋮	⋮

【図 7】

(ユーザ嗜好DBの構造)

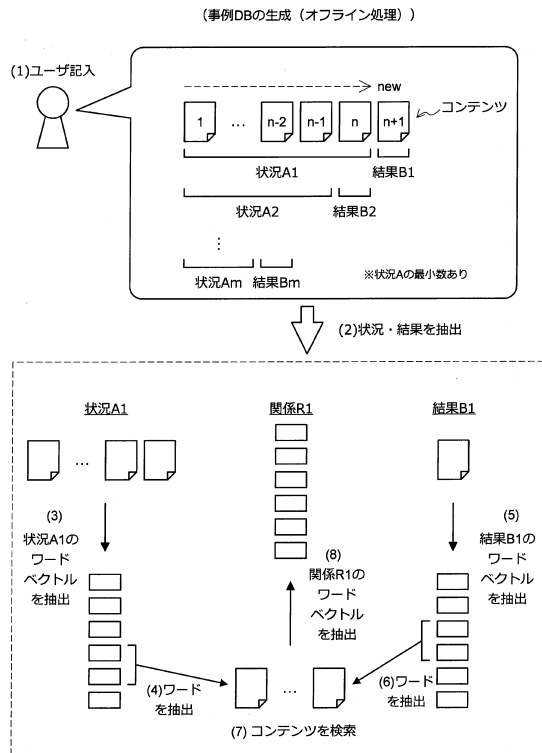
ユーザID	領域ID	特徴量ID	重要度
1	1	30	0.5
1	2	2	0.3
1	2	4	1.3
1	11	52	12.4
2	1	3	0.9
2	2	4	1.2
2	5	65	3.2
2	6	3	0.9
⋮	⋮	⋮	⋮

【図 8】

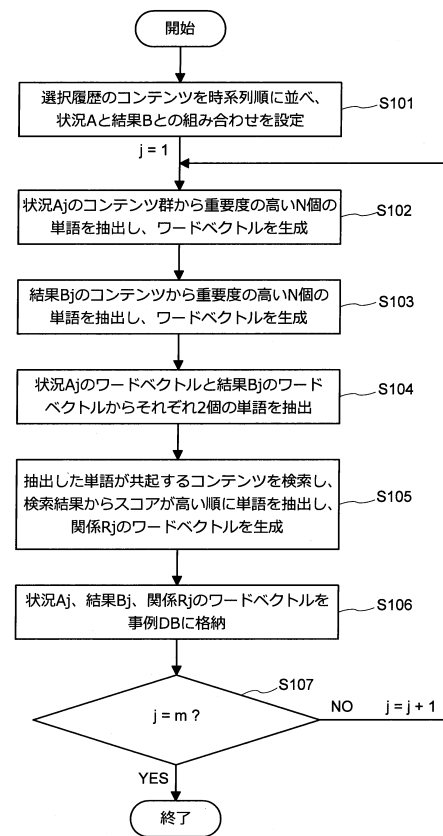
(事例DBの構造)

状況Aのワードベクトル		結果Bのワードベクトル		関係R
word A1	word A2	word B1	word B2	word R
1	5	2	1	8
1	5	2	1	3
1	5	2	1	10
1	5	2	1	15
1	5	2	7	4
1	5	2	7	23
1	5	2	7	4
1	6	1	5	2
1	6	1	5	1
1	6	1	5	7

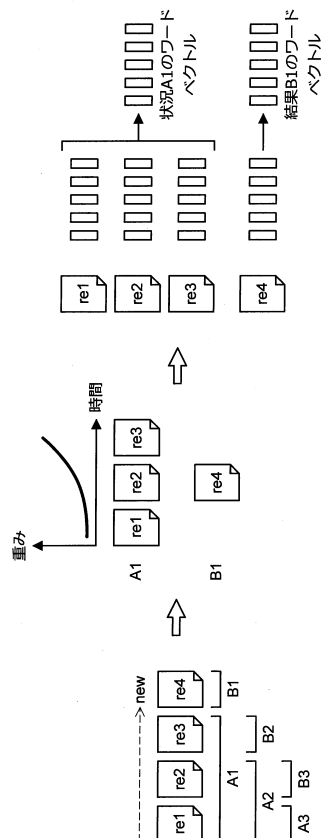
【図 9】



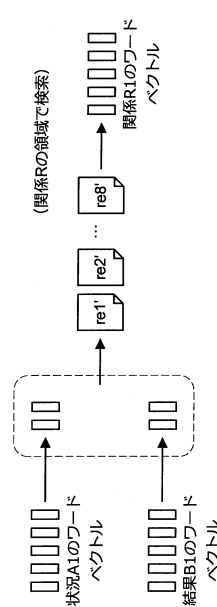
【図 10】



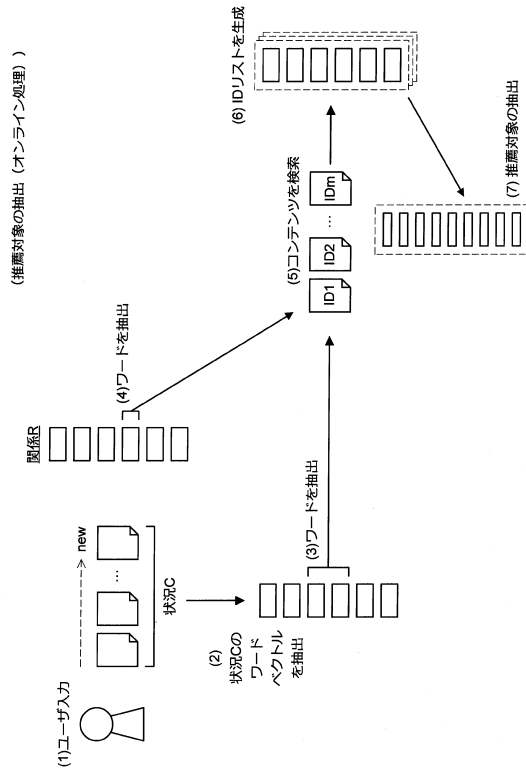
【図 11】



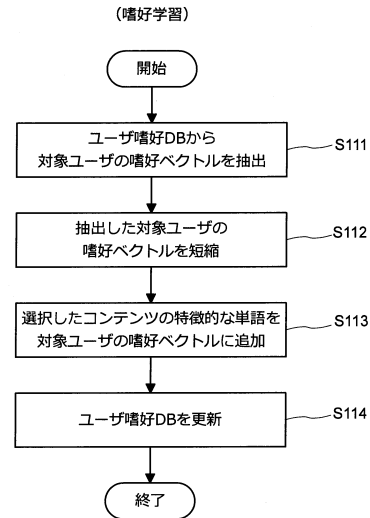
【図 12】



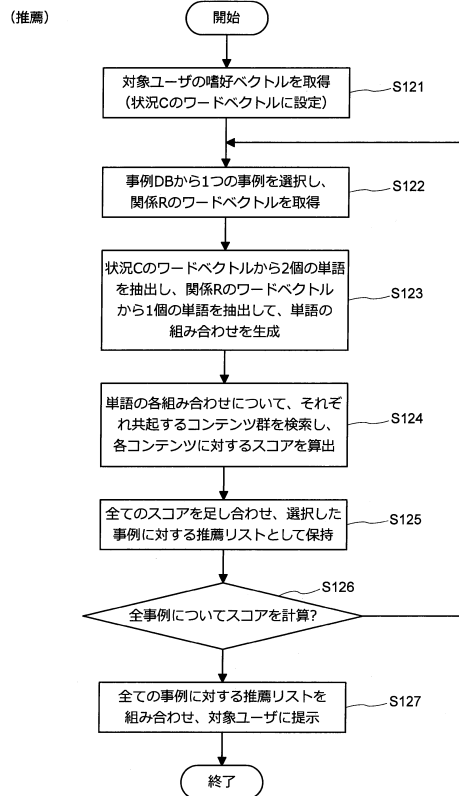
【図 13】



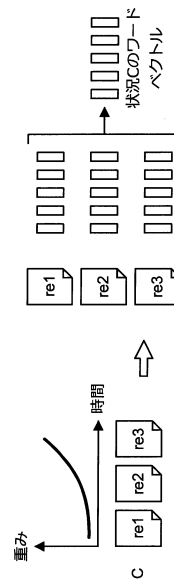
【図 14】



【図 15】



【図 16】

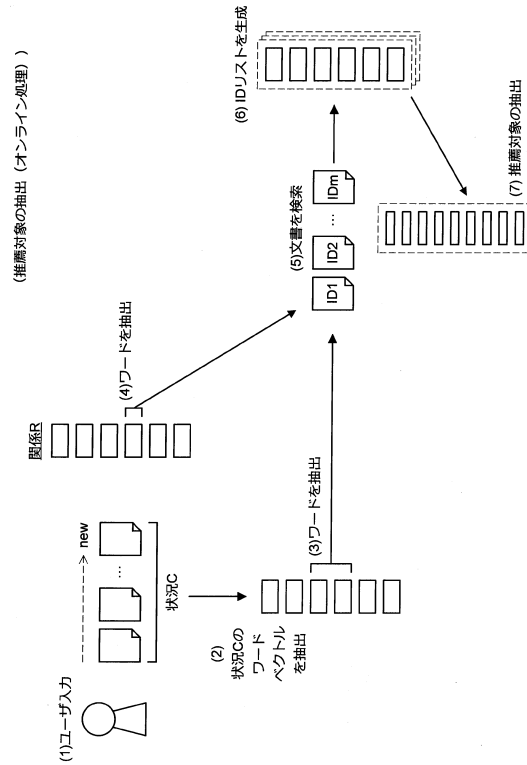


【図 2 1】

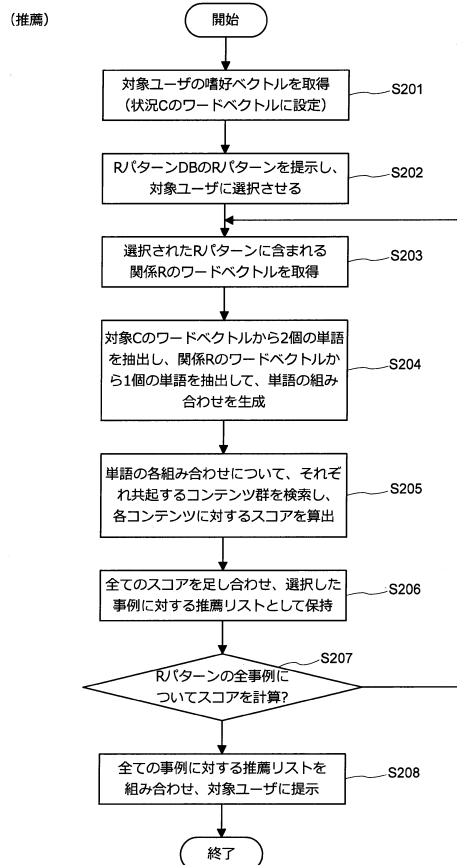
(RパターンDBの構造)

クラスID	状況Aのワードベクトル			結果Bのワードベクトル			関係R
	word A1	word A2		word B1	word B2	word R	
1	1	5		2	1	8	
1	1	5		2	1	3	
1	1	5		2	1	10	
1	1	5		2	1	15	
2	1	5		2	7	4	
2	1	5		2	7	23	
2	1	5		2	7	4	
2	1	6		1	5	2	
1	1	6		1	5	1	
1	1	6		1	5	7	

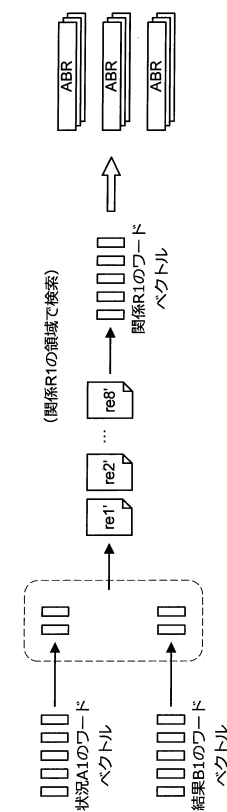
【図 2 2】



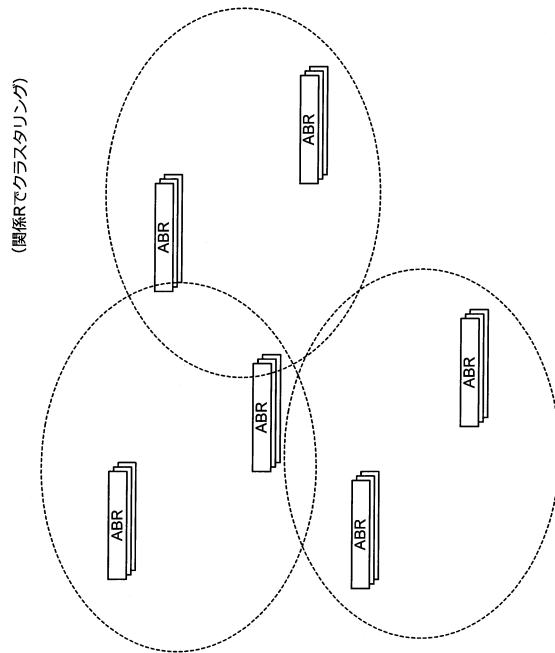
【図 2 3】



【図 2 4】

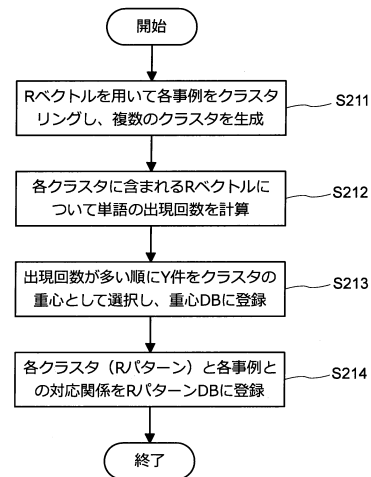


【図 25】

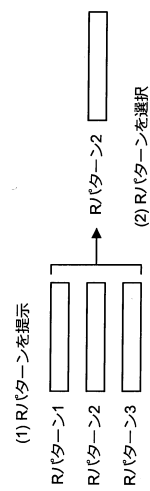


【図 26】

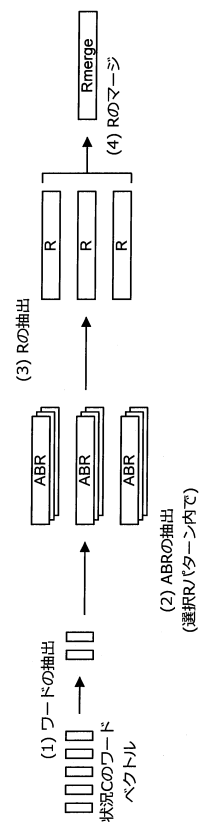
(クラスタリング)



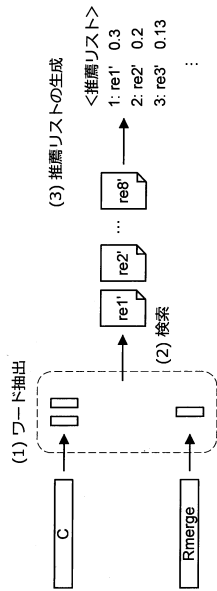
【図 27】



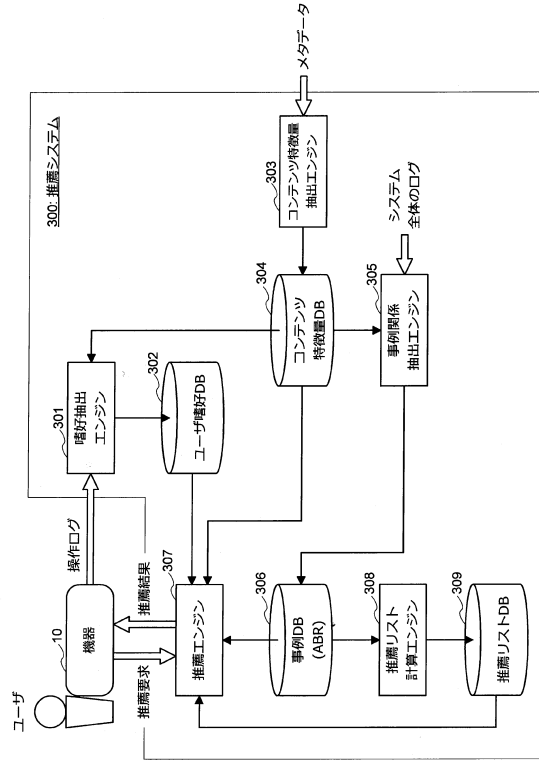
【図 28】



【図 29】



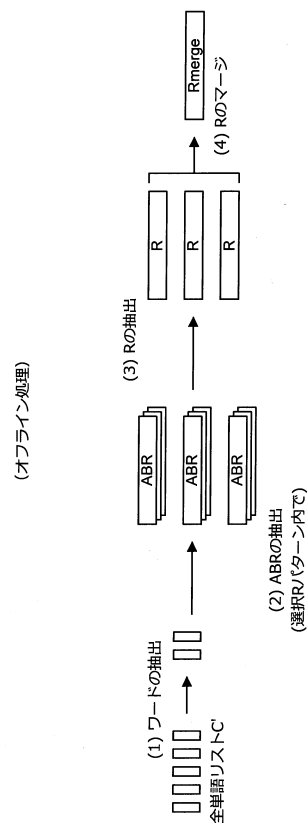
【図 30】



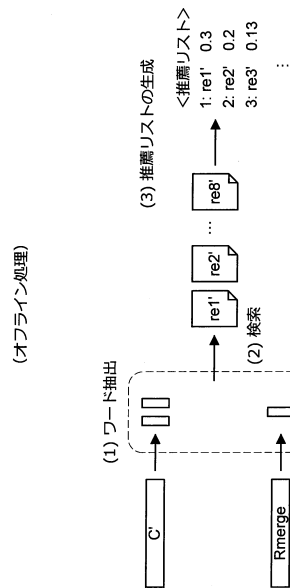
【図 31】

(推薦リストDBの構造)			
状況Cのワードベクトル		アイテムID	スコア
word C1	word C2		
704	706	81925	0.12
704	706	99623	0.20
704	706	11245	0.18
4315	4369	21907	0.83
4315	4660	28	0.62
4315	4660	1397	0.11
2794	2796	11287	0.01
4762	4806	45902	0.91
6110	6997	7239	1.12
⋮	⋮	⋮	⋮

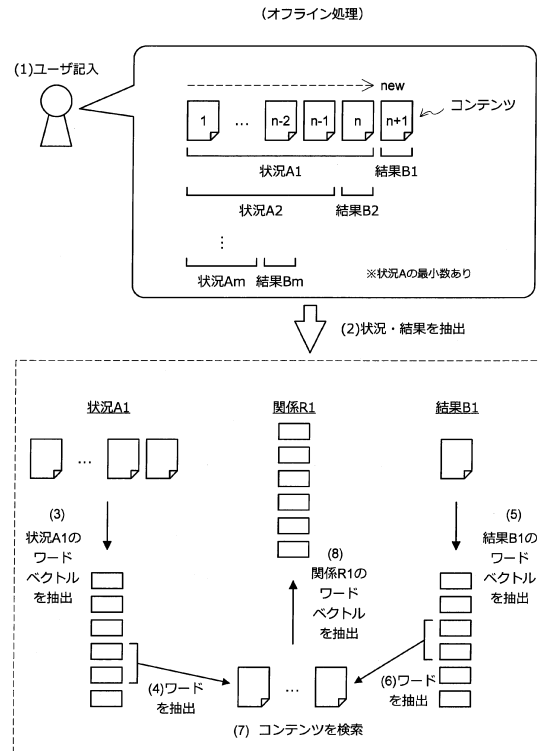
【図 32】



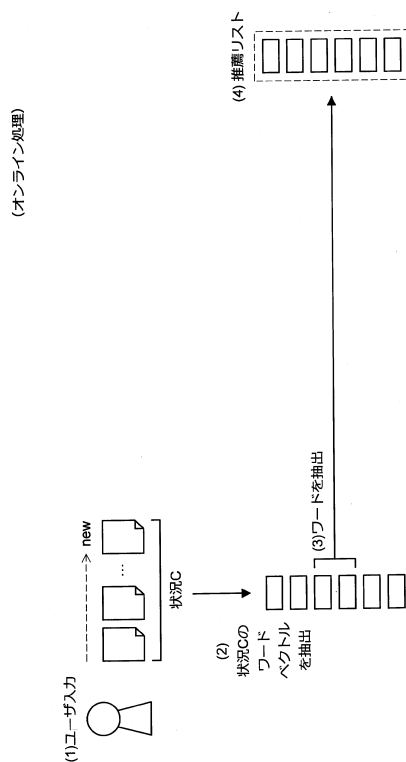
【図 3 3】



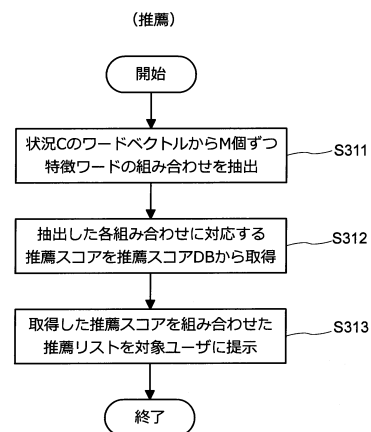
【図 3 4】



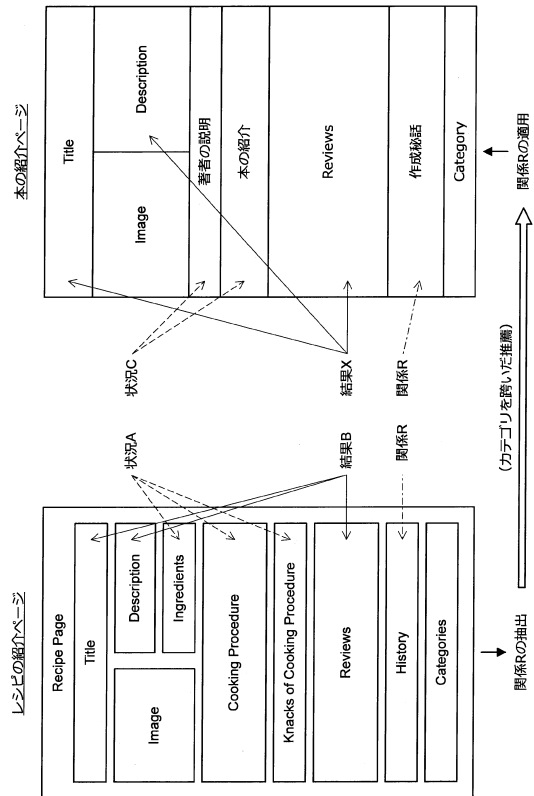
【図 3 5】



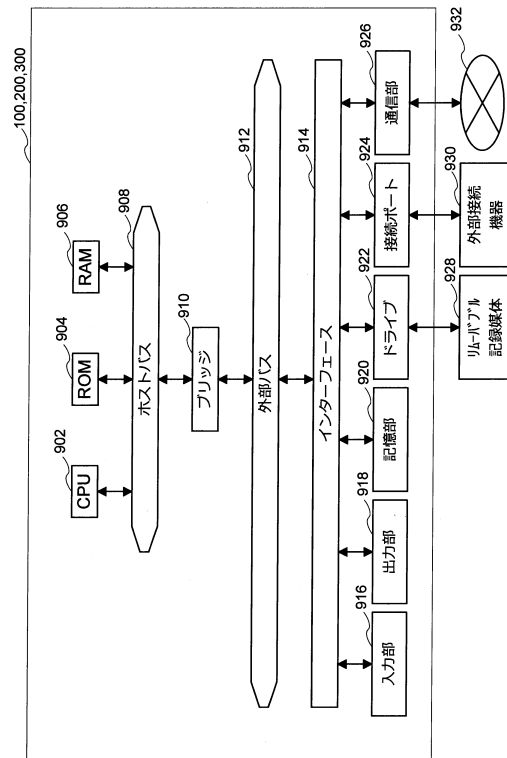
【図 3 6】



【図 37】



【図 38】



フロントページの続き

(72)発明者 宮原 正典

東京都港区港南1丁目7番1号 ソニー株式会社内

(72)発明者 高木 友博

神奈川県川崎市多摩区東三田1-1-1 学校法人明治大学 生田校舎内

審査官 野崎 大進

(56)参考文献 特開2010-225115(JP,A)

特開2004-220251(JP,A)

特開2002-251408(JP,A)

特開2010-170173(JP,A)

特開2010-204735(JP,A)

佐藤 大樹 他, ブラウザの閲覧履歴に基づく探索行動抽出手法, 電子情報通信学会技術研究報告 Vol.110 No.105 AI2010-1-AI2010-9 人工知能と知識処理, 日本, 社団法人電子情報通信学会, 2010年 6月18日, pp.31-35

(58)調査した分野(Int.Cl., DB名)

G06F 17/30

G06Q 30/02

JSTPlus(JDreamIII)