

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2014-115911

(P2014-115911A)

(43) 公開日 平成26年6月26日(2014.6.26)

(51) Int.Cl.	F I	テーマコード (参考)
G06F 17/30 (2006.01)	G06F 17/30 340A	
	G06F 17/30 340B	
	G06F 17/30 210D	

審査請求 未請求 請求項の数 6 O L (全 10 頁)

(21) 出願番号	特願2012-270710 (P2012-270710)	(71) 出願人	000004226
(22) 出願日	平成24年12月11日 (2012.12.11)		日本電信電話株式会社
			東京都千代田区大手町一丁目5番1号
		(74) 代理人	100107766
			弁理士 伊東 忠重
		(74) 代理人	100070150
			弁理士 伊東 忠彦
		(74) 代理人	100124844
			弁理士 石原 隆治
		(72) 発明者	堤田 恭太
			東京都千代田区大手町二丁目3番1号 日
			本電信電話株式会社内
		(72) 発明者	戸田 浩之
			東京都千代田区大手町二丁目3番1号 日
			本電信電話株式会社内

最終頁に続く

(54) 【発明の名称】 情報推薦装置、情報推薦方法、及び情報推薦プログラム

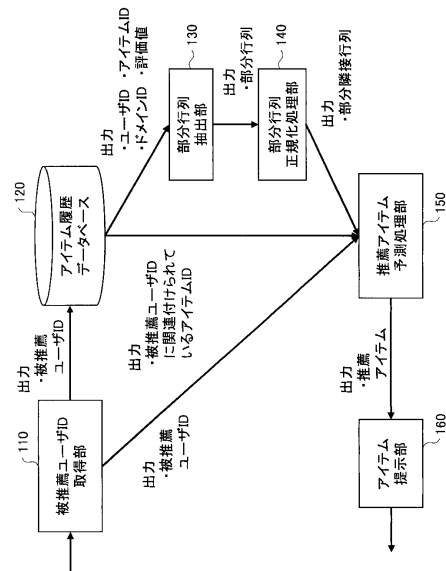
(57) 【要約】

【課題】複数ドメインをまたがったアイテムの推薦を高精度に行うことを可能とする情報推薦技術を提供する。

【解決手段】被推薦ユーザに対してアイテムを推薦する情報推薦装置において、ユーザID、アイテムID、及びドメインIDを含むアイテム履歴データベースを参照し、前記ドメインIDの情報を制約として用いることにより、アイテムに関する制約付きクラスタリングを行い、前記被推薦ユーザと関連の強いクラスタに属するアイテム集合を取得し、当該アイテム集合に基づき部分行列を抽出する部分行列抽出手段と、前記部分行列から部分隣接行列を作成し、当該部分隣接行列を用いて前記被推薦ユーザと各アイテムとの関連度を算出し、算出された関連度に基づき推薦アイテムを決定する推薦アイテム決定手段とを備える。

【選択図】 図1

本発明の一実施の形態における情報推薦装置の構成図



【特許請求の範囲】**【請求項 1】**

被推薦ユーザに対してアイテムを推薦する情報推薦装置であって、

ユーザID、アイテムID、及びドメインIDを含むアイテム履歴データベースを参照し、前記ドメインIDの情報を制約として用いることにより、アイテムに関する制約付きクラスタリングを行い、前記被推薦ユーザと関連の強いクラスタに属するアイテム集合を取得し、当該アイテム集合に基づき部分行列を抽出する部分行列抽出手段と、

前記部分行列から部分隣接行列を作成し、当該部分隣接行列を用いて前記被推薦ユーザと各アイテムとの関連度を算出し、算出された関連度に基づき推薦アイテムを決定する推薦アイテム決定手段と

を備えることを特徴とする情報推薦装置。

10

【請求項 2】

前記部分行列抽出手段は、互いに異なるドメインに属するアイテムの組の類似度を、制約なしの場合と比較して大きく評価することにより前記制約付きクラスタリングを行う

ことを特徴とする請求項 1 に記載の情報推薦装置。

【請求項 3】

前記部分行列抽出手段は、前記アイテム履歴データベースにおけるアイテムの組について、ドメインIDが異なる組を 1 とし、ドメインIDが同じ組を 0 とした制約ベクトルを求め、当該制約ベクトルを用いて前記制約付きクラスタリングを行う

ことを特徴とする請求項 2 に記載の情報推薦装置。

20

【請求項 4】

前記推薦アイテム決定手段は、RWR (Random Walk with Restart) のアルゴリズムを用いて前記被推薦ユーザと各アイテムとの関連度を算出する

ことを特徴とする請求項 1 ないし 3 のうちいずれか 1 項に記載の情報推薦装置。

【請求項 5】

被推薦ユーザに対してアイテムを推薦する情報推薦装置が実行する情報推薦方法であって、

ユーザID、アイテムID、及びドメインIDを含むアイテム履歴データベースを参照し、前記ドメインIDの情報を制約として用いることにより、アイテムに関する制約付きクラスタリングを行い、前記被推薦ユーザと関連の強いクラスタに属するアイテム集合を取得し、当該アイテム集合に基づき部分行列を抽出する部分行列抽出ステップと、

前記部分行列から部分隣接行列を作成し、当該部分隣接行列を用いて前記被推薦ユーザと各アイテムとの関連度を算出し、算出された関連度に基づき推薦アイテムを決定する推薦アイテム決定ステップと

を備えることを特徴とする情報推薦方法。

30

【請求項 6】

コンピュータを、請求項 1 ないし 4 のうちいずれか 1 項に記載の情報推薦装置の各手段として機能させるための情報推薦プログラム。

【発明の詳細な説明】**【技術分野】**

40

【0001】

本発明は、情報推薦において複数のサービスやドメイン（これらを総称して、以下、ドメインと呼ぶ）のコンテンツやアイテム（これらを総称して、以下、アイテムと呼ぶ）を横断的に扱ってアイテムを推薦する情報推薦の技術に関するものである。

【背景技術】**【0002】**

従来技術として、複数のドメインを対象としてアイテムの推薦を可能にするために、複数のドメインのアイテムに付与されたメタデータ間を関連付けることで、より高精度な情報推薦を実現する方法がある（例えば、特許文献1参照）。

【0003】

50

また、単一のドメインを対象として、ユーザ、アイテムの購買や利用履歴の関係を表すユーザ-アイテム行列について、その行列内から傾向の近いユーザ集合及びアイテム集合からなる複数の部分行列を発見するクラスタリングを施して、被推薦ユーザが属する部分行列のみを用いて一般的な協調フィルタリングの手法を適用することで推薦精度を改善する研究がある。(例えば、非特許文献1参照)。

【先行技術文献】

【特許文献】

【0004】

【特許文献1】特開2012-150561号公報

【非特許文献】

【0005】

【非特許文献1】Bin Xu, Jiajun Bu, Chun Chen, Deng Cai: An Exploration of Improving Collaborative Recommender Systems via User-Item Subgroups, Proceedings of the WWW2012, 2012.

【発明の概要】

【発明が解決しようとする課題】

【0006】

上記の特許文献1の方法を用いることで、複数のドメインを扱った情報推薦を行えるようになる。しかしながら、推薦システムが構築するグラフを表す隣接行列中に推薦精度の向上に寄与しないデータが存在する場合には推薦精度が低下するという問題がある。

【0007】

また、上記の非特許文献1の方法を用いることで、単一のドメインを扱った推薦システムの推薦精度向上が期待できるが、複数のドメインを対象としてクラスタリングを施す場合には、それぞれのドメインにクラスタが構築されやすいため、ドメインをまたがって推薦を行う場合には推薦できないアイテムが多くなり、結果として推薦精度が低下するという問題がある。

【0008】

特許文献1における上記課題を解決するために、グラフを表す隣接行列に非特許文献1のようにクラスタリングを施して、被推薦ユーザと関連の強いクラスタに属するアイテム集合のみからなる部分行列を用いることが考えられる。そうすることにより、異なるクラスタに属しやすい推薦精度に寄与しないアイテムを取り除くことができる。しかしながら、前述の通り、クラスタリングの結果としてドメイン間の情報が抜け落ちてしまうことによる推薦精度低下の懸念がある。

【0009】

本発明は上記の点に鑑みてなされたものであり、複数ドメインをまたがったアイテムの推薦を高精度に行うことを可能とする情報推薦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】

【0010】

上記の課題を解決するために、本発明は、被推薦ユーザに対してアイテムを推薦する情報推薦装置であって、

ユーザID、アイテムID、及びドメインIDを含むアイテム履歴データベースを参照し、前記ドメインIDの情報を制約として用いることにより、アイテムに関する制約付きクラスタリングを行い、前記被推薦ユーザと関連の強いクラスタに属するアイテム集合を取得し、当該アイテム集合に基づき部分行列を抽出する部分行列抽出手段と、

前記部分行列から部分隣接行列を作成し、当該部分隣接行列を用いて前記被推薦ユーザと各アイテムとの関連度を算出し、算出された関連度に基づき推薦アイテムを決定する推薦アイテム決定手段とを備えることを特徴とする情報推薦装置として構成される。

【0011】

前記部分行列抽出手段は、互いに異なるドメインに属するアイテムの組の類似度を、制約なしの場合と比較して大きく評価することにより前記制約付きクラスタリングを行うよ

10

20

30

40

50

うに構成することができる。

【0012】

例えば、前記部分行列抽出手段は、前記アイテム履歴データベースにおけるアイテムの組について、ドメインIDが異なる組を1とし、ドメインIDが同じ組を0とした制約ベクトルを求め、当該制約ベクトルを用いて前記制約付きクラスタリングを行う。

【0013】

また、前記推薦アイテム決定手段は、RWR (Random Walk with Restart) のアルゴリズムを用いて前記被推薦ユーザと各アイテムとの関連度を算出するように構成してもよい。

【0014】

本発明は、被推薦ユーザに対してアイテムを推薦する情報推薦装置が実行する情報推薦方法、及び、コンピュータを、情報推薦装置の各手段として機能させるための情報推薦プログラムとして構成することもできる。

【発明の効果】

【0015】

本発明によれば、複数のドメインをまたがるために必要な情報を残すような制約付きクラスタリングを行うこととしたので、複数ドメインのアイテムを、推薦精度に寄与するデータのみを抽出して用いることが可能となる。よって、複数ドメインをまたがったアイテムの推薦をより高精度に行うことが可能となる。

【図面の簡単な説明】

【0016】

【図1】本発明の一実施の形態における情報推薦装置の構成図である。

【図2】本発明の一実施の形態におけるアイテム履歴データベースのテーブルのイメージである。

【図3】本発明の一実施の形態における情報推薦装置の処理のフローチャートである。

【図4】本発明の一実施の形態における部分行列のイメージである。

【図5】本発明の一実施の形態における部分行列抽出処理で用いる制約ベクトル計算のフローチャートである。

【図6】本発明の一実施の形態における部分行列抽出処理のフローチャートである。

【図7】本発明の一実施の形態における部分行列正規化処理のフローチャートである。

【発明を実施するための形態】

【0017】

以下、図面を参照して本発明の実施の形態を説明する。なお、以下で説明する実施の形態は一例に過ぎず、本発明が適用される実施の形態は、以下の実施の形態に限られるわけではない。

【0018】

(装置構成)

図1は、本発明の一実施の形態における情報推薦装置の構成図である。

【0019】

同図に示す情報推薦装置は、被推薦ユーザID取得部110、アイテム履歴データベース120、部分行列抽出部130、部分行列正規化処理部140、推薦アイテム予測処理部150、アイテム提示部160を有する。各機能部の概要は以下のとおりである。

【0020】

被推薦ユーザID取得部110は、被推薦ユーザIDを取得し、アイテム履歴データベース120及び推薦アイテム予測処理部150に出力する。

【0021】

アイテム履歴データベース120は、アイテム履歴のテーブルを格納する格納手段を構成している。図2に、本発明の一実施の形態におけるアイテム履歴データベースのテーブルのイメージを示す。図2に示すように、アイテム履歴データベース120は、ユーザID、アイテムID、ドメインID、アイテムへの評価値がそれぞれ参照できるようなデー

10

20

30

40

50

データベースとなっている。

【0022】

図1の部分行列抽出部130は、アイテム履歴データベース120のデータを用いて後述する部分行列Aを計算する。部分行列正規化処理部140は、部分行列Aを正規化し、部分隣接行列を作成する。推薦アイテム予測処理部150は、被推薦ユーザID、被推薦ユーザIDに関連付けられているアイテムIDリスト、及び部分隣接行列に基づいて、推薦するアイテムを決定する。アイテム提示部160は、推薦アイテム予測処理部150の出力である推薦アイテムを提示する。なお、部分行列正規化処理部140と推薦アイテム予測処理部150とをまとめて推薦アイテム決定手段と呼んでもよい。

【0023】

本実施の形態における情報推薦装置は、例えば、1つ又は複数のコンピュータに、本実施の形態で説明する処理内容を記述したプログラムを実行させることにより実現可能である。すなわち、情報推薦装置の各部（被推薦ユーザID取得部、アイテム履歴データベース、部分行列抽出部、部分行列正規化処理部、推薦アイテム予測処理部、アイテム提示部）が有する機能は、当該情報推薦装置を構成するコンピュータに内蔵されるCPUやメモリ、ハードディスクなどのハードウェア資源を用いて、各部で実施される処理に対応するプログラムを実行することによって実現することが可能である。より具体的には、プログラムに従って、計算対象のデータをメモリから読み出し、CPUにより演算を行って、メモリに格納する動作を繰り返しながら処理が実行される。上記プログラムは、コンピュータが読み取り可能な記録媒体（可搬メモリ等）に記録して、保存したり、配布したりすることが可能である。また、上記プログラムをインターネットや電子メールなど、ネットワークを通して提供することも可能である。

【0024】

なお、アイテム履歴データベース120は、情報推薦装置内に存在しなくてもよい。例えば、アイテム履歴データベース120を外部のデータベースサーバ内に備え、情報推薦装置が当該データベースサーバにアクセスすることにより、必要なデータを取得することとしてもよい。

【0025】

（装置の動作）

図3は、本発明の一実施の形態における情報推薦装置の処理のフローチャートである。図3のフローチャートに沿って、図1に示した情報推薦装置により実行される被推薦ユーザに対して推薦アイテムを提示する処理について説明する。図3に示す各ステップの処理において、更に図を参照して詳細な説明を適宜行う。

【0026】

ステップ210)まず、被推薦ユーザID取得処理として、被推薦ユーザID取得部110は、被推薦ユーザのIDを取得し、アイテム履歴データベース120及び推薦アイテム予測処理部150に出力する。この処理は、例えば、ユーザが利用するサービスやシステムにログイン処理と同時にを行うものとする。

【0027】

ステップ220)部分行列抽出部130の処理として、アイテム履歴データベース120からユーザID、アイテムID、ドメインID、評価値のデータを受け取って、部分行列Aを生成し、部分行列正規化処理部140に部分行列Aを出力する。図4に、本発明の一実施の形態における部分行列Aのイメージを示す。後述するように、部分行列Aは、制約付きクラスタリングで得られたアイテムに関する複数のクラスタのうち、被推薦ユーザに関するベクトルとの間の距離の小さいクラスタに含まれるアイテムID集合を有効な行列のインデックスとする行列である。

【0028】

以下、図5及び図6のフローチャートを用いてより詳細な一実施形態における部分行列Aの構築手順例について説明する。

【0029】

10

20

30

40

50

図5に示すように、まず、アイテム履歴データベース120より出力される全ユーザ集合 $U=\{u_1, u_2, \dots, u_{|U|}\}$ 、全アイテム集合 $I=\{i_1, i_2, \dots, i_{|I|}\}$ を収集する(ステップ510)。

【0030】

次に、制約付きクラスタリングで用いる制約ベクトル $W=(w_{i,j})$ を求める。具体的には、前記全アイテムの全てのアイテムの組み $i,j(i \neq j)$ について、アイテム履歴データベース120上でドメインIDが異なるものについて $w_{i,j}$ に1を代入し、そうでない組合せについて0を代入する(ステップ530~ステップ537)。

【0031】

次に、上記のようにして作成された制約ベクトル $W=(w_{i,j})$ を用いて、制約付きクラスタリングを行い、部分行列Aを求める。この処理を図6を参照して説明する。

10

【0032】

まず、全てのアイテムについて、アイテム履歴データベース120での各ユーザの評点の平均を要素とするアイテムベクトル x_i を用意する(ステップ610)。全てのアイテムベクトルについて、K個のクラスタCを用意する(ステップ620)。クラスタの初期値は、 x_i をランダムにK個選択する。クラスタリング結果のクラスタ数を決定するパラメータKの値はK=10や50などを用いるが、データの規模や用途によって指定し規模が大きい時やアイテムのより詳細な違いを気にする場合は大きい値を用いる。

【0033】

x_i を最も近傍のクラスタ C_k にマージしてクラスタリングを行うための、近傍の計算は、前記制約ベクトル $w_{i,j}$ を用いて次式

20

$$\sum_{j=1}^K w_{i,j} \cdot w_{i,j} \cdot x_j \cdot c_k \text{Sim}(x_i, x_j) + \sum_{j=1}^K x_j \cdot c_k \text{Sim}(x_i, x_j)$$

の値が最大になるものを選ぶ。ここで、kは1~KのクラスタCを識別する添字、 $\text{Sim}(x_i, x_j)$ は、ベクトル x_i と x_j の類似度を表し、ユークリッド距離やJaccard係数などの任意の尺度を用いることができる。上記の式により、互いに異なるドメインに属するアイテムの組の類似度を、制約なしの場合と比較して大きく評価してクラスタリングを行うことができる。

【0034】

α_1 、 α_2 は制約付きの類似度と制約なしの類似度の影響度合いを制御するパラメータであり、 $\alpha_1=0.85$ 、 $\alpha_2=0.15$ などの実数を用いる。 α_1 と α_2 の和は必ずしも1でなくてもよい。例えば、 $\alpha_1=0.85$ 、 $\alpha_2=0.15$ のように、 α_1 が大きな値であれば、制約付きの類似度の影響が大きくなり、異なるドメインのアイテムを含むクラスタが構築され易くなる。このようにして算出されたクラスタに基づき生成された部分行列を用いることで、複数ドメインをまたがった精度の高い推薦を行うことが可能となる。

30

【0035】

全てのアイテムベクトルについて上記の制約付きクラスタリング処理を行い、制約付きクラスタリングの結果としてK個のクラスタCを得る(ステップ620~ステップ622)。

【0036】

更に、被推薦ユーザaにアイテム履歴データベース120に関連付けられているアイテムID集合 I_a を用いて、被推薦ユーザのベクトルを q_a 、その要素 $q_{a,j}$ を次式

40

$$q_{a,j} = (1/|I_a|) \cdot \sum_{i \in I_a} (x_{i,j} / (\sum_{j=1}^K x_{i,j}))$$

を用いて求める(ステップ630)。上記の式において、 $x_{i,j}$ はアイテムベクトル x_i の要素であり、jは x_i の添字である。 $x_{i,j} / (\sum_{j=1}^K x_{i,j})$ は、 x_i の要素の和が1になるように正規化することを意味する。

【0037】

最後に、K個のクラスタのうち、 q_a と最もベクトル間の距離の小さいクラスタ C_k に含まれるアイテムID集合を有効な行列のインデックスとする行列を部分行列 $A=(a_{i,j})$ として出力する(ステップ640)。すなわち、被推薦ユーザと関連の強いクラスタに属するアイテムID集合を取得し、当該アイテムID集合に基づき部分行列を抽出する。ベクトル

50

間の距離の基準は、ユークリッド距離やJaccard係数などの任意の基準を用いることができる。行列の要素 $a_{i,j}$ は、前記 $\text{Sim}(x_i, x_j)$ とする。

【0038】

ステップ230)部分行列正規化処理部140で部分行列の正規化を行う。この正規化処理を図7のフローチャートを参照して説明する。

【0039】

正規化処理では、まず、部分行列抽出部130より出力される部分行列 $A=(a_{j,k})$ を受け取る(ステップ710)。

【0040】

次に、部分行列 A の各要素 $a_{j,k}$ を列の値の和で除した値に次式

$$a_{j,k} = a_{j,k} / \sum_{g=1}^n a_{g,k}$$

を用いて更新し、全ての要素について行ったものを、部分隣接行列 A として出力する(ステップ720~ステップ740)。つまり、部分行列 A を列の値の和が1になるように正規化することで、部分隣接行列 A を求める。

【0041】

ステップ240)推薦アイテム予測処理部150が、被推薦ユーザID取得部110から被推薦ユーザID、アイテム履歴データベース120から被推薦ユーザIDに関連付けられているアイテムIDリスト、部分行列正規化処理部140から部分隣接行列をそれぞれ受け取り、推薦するアイテムを決定する。

【0042】

その計算には、取得した被推薦ユーザIDと各アイテムIDとの関連度を求め、その値の高いアイテムを推薦するアイテムとするRandom Walk with Restart (RWR)を用いる。具体的な計算式は、 $p^{(t)}$ を計算のステップ t におけるアイテムの関連度の列ベクトル、 α は $0 < \alpha < 1$ を満たす定数、 A を前記部分隣接行列、 q を被推薦ユーザがアイテム履歴データベース120上で関連付けられている n 個のアイテムを表す値が $1/n$ である列ベクトルとすると、次式のように表される。

【0043】

$$p^{(t+1)} = (1 - \alpha) A p^{(t)} + \alpha q$$

計算は部分隣接行列 A 上で t 回繰り返して $p^{(t)}$ が更新され、 $p^{(t)}$ の列ベクトルからアイテムを表すノードの関連度の高いものを推薦アイテムとして出力する。ここでの t は例えば30回とする。なお、この実施回数は30回に限定されることなく、十分に大きな値であればよい。

【0044】

ステップ250)推薦アイテム提示処理としてアイテム提示部160が、推薦アイテム予測処理部150の出力である推薦アイテムを提示し、推薦が完了する。

【0045】

本発明は、上記の実施の形態に限定されることなく、特許請求の範囲内において、種々変更・応用が可能である。

【符号の説明】

【0046】

- 110 被推薦ユーザID取得部
- 120 アイテム履歴データベース
- 130 部分行列抽出部
- 140 部分行列正規化処理部
- 150 推薦アイテム予測処理部
- 160 アイテム提示部

10

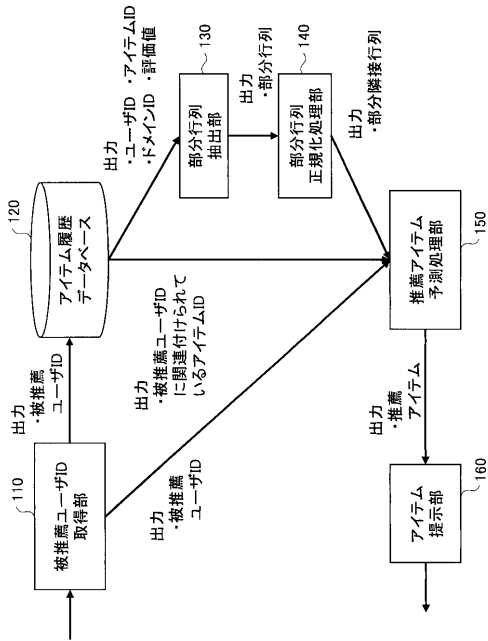
20

30

40

【 図 1 】

本発明の一実施の形態における情報推薦装置の構成図



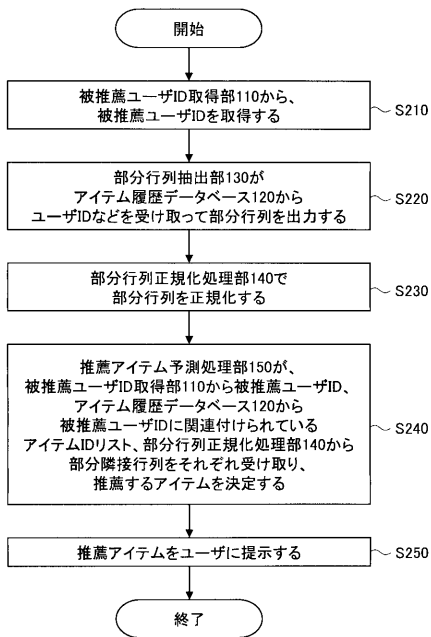
【 図 2 】

本発明の一実施の形態におけるアイテム履歴データベースのテーブルのイメージ

ユーザID	アイテムID	ドメインID	評価値
u_1	i_1	s_1	5
u_1	i_2	s_2	4
u_1	i_3	s_1	5
...
u_2	i_1	s_1	3
u_2	i_2	s_2	4
...

【 図 3 】

本発明の一実施の形態における情報推薦装置の処理のフローチャート



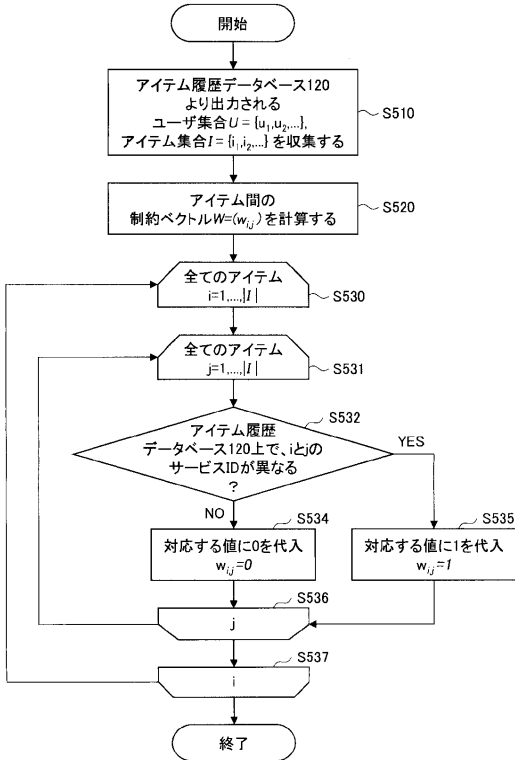
【 図 4 】

本発明の一実施の形態における部分行列のイメージ

	i_1	i_2	...	$i_{ I }$
i_1	0	1/3	...	1/3
i_2	1/2	0	...	0
...
$i_{ I }$	1/4	0	...	0

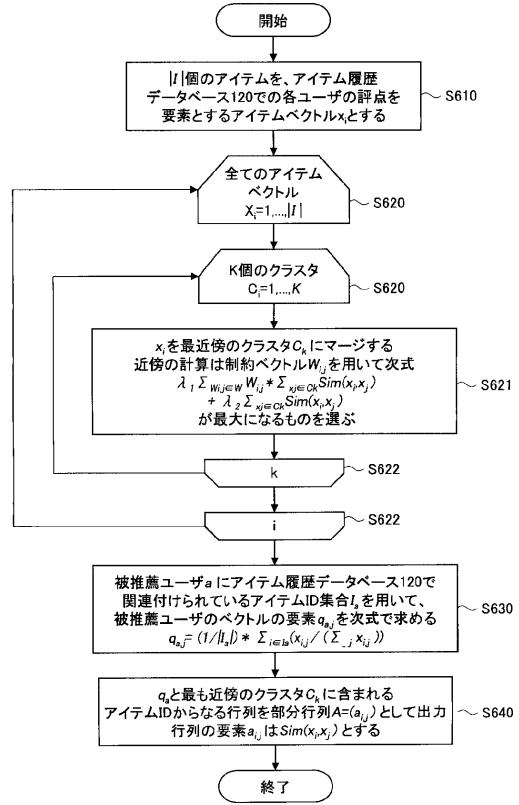
【 図 5 】

本発明の一実施の形態における部分行列抽出処理で用いる制約ベクトル計算のフローチャート



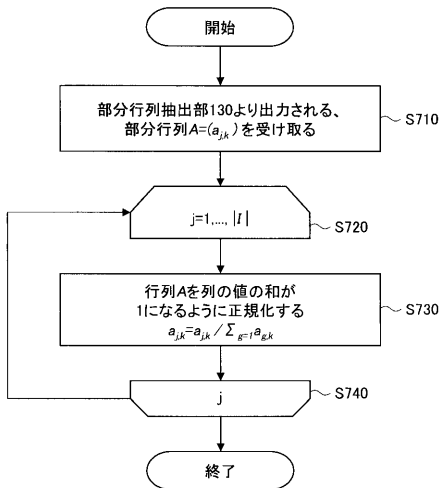
【 図 6 】

本発明の一実施の形態における部分行列抽出処理のフローチャート



【 図 7 】

本発明の一実施の形態における部分行列正規化処理のフローチャート



フロントページの続き

(72)発明者 内山 匡

東京都千代田区大手町二丁目3番1号 日本電信電話株式会社内