(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2017-27480
(P2017-27480A)
(43)公開日　平成29年2月2日(2017.2.2)

(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	350C	

審査請求 未請求　請求項の数 10　ＯＬ　（全 24 頁）

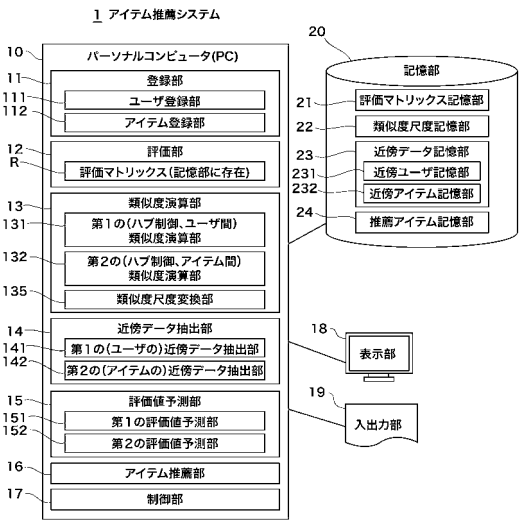
(21)出願番号　特願2015-147269 (P2015-147269)
(22)出願日　平成27年7月24日 (2015.7.24)

(71)出願人　504202472
大学共同利用機関法人情報・システム研究機構
東京都立川市緑町１０番３号
(74)代理人　100097320
弁理士　宮川　貞二
(74)代理人　100100398
弁理士　柴田　茂夫
(74)代理人　100131820
弁理士　金井　俊幸
(74)代理人　100155192
弁理士　金子　美代子

最終頁に続く

(54)【発明の名称】アイテム推薦システム及びアイテム推薦方法

(57)【要約】
【課題】偽ユーザを投入する攻撃に対して頑健なアイテム推薦システムを提供する。
【解決手段】ユーザのアイテムに係る評価値を記入する評価マトリックスＲを記憶する評価マトリックス記憶部２１と、ハブの出現を抑止する類似度尺度を用いてユーザ間の類似度を演算する類似度演算部１３と、類似度演算部１３にて演算された類似度を用いて、対象ユーザとの類似度の高い方からｋ人のユーザを抽出する近傍データ抽出部１４と、近傍データ抽出部１４にて抽出されたｋ人のユーザのアイテムに係る評価値を用いて、対象ユーザに係る未記入のセルに記入すべき評価値を予測する評価値予測部１５と、評価値予測部１５にて予測された評価値の高いアイテムから対象ユーザに推薦すべきアイテムを抽出して対象ユーザに推薦するアイテム推薦部１６とを備える。
【選択図】図５



【特許請求の範囲】
【請求項１】
ユーザのアイテムに係る評価値を記入する評価マトリックスを記憶する評価マトリックス記憶部と；
ハブの出現を抑制する類似度尺度を用いてユーザ間の類似度を演算する第１の類似度演算部と；
前記第１の類似度演算部にて演算された類似度を用いて、前記対象ユーザとの類似度の高い方からｋ人のユーザを抽出する第１の近傍データ抽出部と；
前記第１の近傍データ抽出部にて抽出されたｋ人のユーザのアイテムに係る評価値を用いて、前記対象ユーザに係る未記入のセルに記入すべき評価値を予測する第１の評価値予測部と；
前記第１の評価値予測部にて予測された評価値の高いアイテムから前記対象ユーザに推薦すべきアイテムを抽出する推薦アイテム抽出して、前記対象ユーザに推薦するアイテム推薦部とを備える；
アイテム推薦システム。
【請求項２】
ユーザのアイテムに係る評価値を記入する評価マトリックスを記憶する評価マトリックス記憶部と；
ハブの出現を抑制する類似度尺度を用いてアイテム間の類似度を演算する第２の類似度演算部と；
前記第２の類似度演算部にて演算された類似度を用いて、前記対象アイテムとの類似度の高い方からｋ個のアイテムを抽出する第２の近傍データ抽出部と；
前記第２の近傍データ抽出部にて抽出されたｋ個のアイテムに係るユーザの評価値を用いて、前記対象ユーザに係る未記入のセルに記入すべき評価値を予測する第２の評価値予測部と；
前記第２の評価値予測部にて予測された評価値の高いアイテムから前記対象ユーザに推薦すべきアイテムを抽出して、前記対象ユーザに推薦するアイテム推薦部とを備える；
アイテム推薦システム。
【請求項３】
前記ハブの出現を抑制する類似度尺度を記憶する類似度尺度記憶部を備える；
請求項１又は請求項２に記載のアイテム推薦システム。
【請求項４】
一般的な類似度尺度に基づく類似度を前記ハブの出現を抑制する類似度尺度に基づく類似度に変換する類似度尺度変換部を備える；
請求項１ないし請求項３のいずれか１項に記載のアイテム推薦システム。
【請求項５】
前記対象ユーザに係る未記入のセルに記入すべき評価値を予測するに際し、前記記入すべき評価値として、重み付けをした平均値を用いる；
請求項１ないし請求項４のいずれか１項に記載のアイテム推薦システム。
【請求項６】
ユーザのアイテムに係る評価値を記入する評価マトリックスを記憶する評価マトリックス記憶工程と；
ハブの出現を抑制する類似度尺度を用いてユーザ間の類似度を演算する第１の類似度演算工程と；
前記第１の類似度演算工程にて演算された類似度を用いて、前記対象ユーザとの類似度の高い方からｋ人のユーザを抽出する第１の近傍データ抽出工程と；
前記第１の近傍データ抽出工程にて抽出されたｋ人のユーザのアイテムに係る評価値を用いて、前記対象ユーザに係る未記入のセルに記入すべき評価値を予測する第１の評価値予測工程と；
前記第１の評価値予測工程にて予測された評価値の高いアイテムから前記対象ユーザに

推薦すべきアイテムを抽出して、前記対象ユーザに推薦するアイテム推薦工程とを備える；
アイテム推薦方法。
【請求項７】
ユーザのアイテムに係る評価値を記入する評価マトリックスを記憶する評価マトリックス記憶工程と；
ハブの出現を抑制する類似度尺度を用いてアイテム間の類似度を演算する第２の類似度演算工程と；
前記第２の類似度演算工程にて演算された類似度を用いて、前記対象アイテムとの類似度の高い方からｋ個のアイテムを抽出する第２の近傍データ抽出工程と；
前記第２の近傍データ抽出工程にて抽出されたｋ個のアイテムに係るユーザの評価値を用いて、前記対象ユーザに係る未記入のセルに記入すべき評価値を予測する第２の評価値予測工程と；
前記第２の評価値予測工程にて予測された評価値の高いアイテムから前記対象ユーザに推薦すべきアイテムを抽出して、前記対象ユーザに推薦するアイテム推薦工程とを備える；
アイテム推薦方法。
【請求項８】
一般的な類似度尺度に基づく類似度を前記ハブの出現を抑制する類似度尺度に基づく類似度に変換する類似度尺度変換工程を備える；
請求項６又は請求項７に記載のアイテム推薦方法。
【請求項９】
請求項６ないし請求項８のいずれか１項に記載のアイテム推薦方法をコンピュータに実行させるためのプログラム。
【請求項１０】
請求項９に記載のプログラムを記録したコンピュータ読み取り可能な記録媒体。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明はアイテム推薦システム及びアイテム推薦方法に関する。詳しくは、ユーザベースあるいはアイテムベースに代表される協調フィルタリング（ＣＦ）において、シリングアタック、すなわち、システムがユーザに推薦するアイテムを決定する工程に介入するために偽ユーザを不正投入する攻撃に対して、頑健なアイテム推薦システム及びアイテム推薦方法に関する。
【背景技術】
【０００２】
ユーザベースのＣＦは、類似度演算に例えばｋ近傍法を用い、アイテムに対する嗜好が類似する他のユーザの過去の評価値を参照してアイテムをユーザに推薦するシステムである。すなわち、アイテムに対する評価値の与え方が類似する他のユーザｋ人を選んで、アイテムに係る評価値を予測し、高い評価値が得られたアイテムをユーザに推薦する。
しかしながら、例えばアイテムが商品で、評価値が嗜好度の場合、平均的な嗜好度を有するように設計された偽ユーザがユーザベースのＣＦシステムに投入される（アベレジアタックと呼ばれるシリングアタック）と、偽ユーザはどのユーザとも高い類似度を示すハブユーザ、すなわち、インフルエンサとなるため、偽ユーザの嗜好する商品が何時も推薦されるようになるおそれがある。
アイテムベースのＣＦ、すなわち、類似する他のアイテムに対するユーザの過去の評価を参照してユーザに推薦するアイテムを決める推薦システム及び推薦方法に対しては、セグメントアタックあるいはポピュラーアタックと呼ばれるシリングアタックが効果を持つ。
【０００３】

他方、発明者達は、ｋ近傍法でハブを軽減する方法を提案した。すなわち、大規模高次元データセットに対して類似度尺度にラプラシアンベースのカーネルを適用する方法（非特許文献１参照）、センタリングを適用する方法（非特許文献２参照）、及び、局在的センタリングを適用する方法（非特許文献３参照）を提案した。

これらのハブを軽減する方法をユーザベースのＣＦあるいはアイテムベースのＣＦに適用することにより、ターゲットアイテムの評価を不正に高めるために攻撃者により偽ユーザが投入されたとしても、ターゲットアイテムの評価が変動しないようにするアイテム推薦システム及びアイテム推薦方法を提供できると期待される。

【先行技術文献】

【非特許文献】

【０００４】

【非特許文献１】Ｉｋｕｍｉ　Ｓｕｚｕｋｉ，Ｋａｚｕｏ　Ｈａｒａ，Ｍａｓａｓｈｉ　Ｓｈｉｍｂｏ，Ｙｕｊｉ　Ｍａｔｓｕｍｏｔｏ，Ｍａｒｃｏ　Ｓａｅｒｅｎｓ，「Ｉｎｖｅｓｔｉｇａｔｉｎｇ　ｔｈｅ　Ｅｆｆｅｃｔｉｖｅｎｅｓｓ　ｏｆ　Ｌａｐｌａｃｉａｎ－ｂａｓｅｄ　Ｋｅｒｎｅｌｓ　ｉｎ　Ｈｕｂ　Ｒｅｄｕｃｔｉｏｎ」、Ｉｎ　Ｐｒｏｃ．２６ｔｈ　ＡＡＡＩ　Ｃｏｎｆｅｒｅｎｃｅ　ｏｎ　Ａｒｔｉｆｉｃｉａｌ　Ｉｎｔｅｌｌｉｇｅｎｃｅ、ｐｐ．１１１２－１１１８、２０１２年

【非特許文献２】鈴木郁美、原一夫、新保仁「ｋ近傍法でハブを軽減する類似度尺度」、情報処理学会研究報告、自然言語処理研究会、２０１２－ＮＬ－２０９、Ｎｏ．１１、ｐｐ．１－８、２０１２年

【非特許文献３】Ｋａｚｕｏ　Ｈａｒａ，Ｉｋｕｍｉ　Ｓｕｚｕｋｉ，Ｍａｓａｓｈｉ　Ｓｈｉｍｂｏ，Ｋｅｉ　Ｋｏｂａｙａｓｈｉ，Ｋｅｎｊｉ　Ｆｕｋｕｍｉｚｕ，Ｍｉｌｏｓ　Ｒａｄｏｖａｎｏｖｉｃ，「Ｌｏｃａｌｉｚｅｄ　Ｃｅｎｔｅｒｉｎｇ：Ｒｅｄｕｃｉｎｇ　Ｈｕｂｎｅｓｓ　ｉｎ　Ｌａｒｇｅ－Ｓａｍｐｌｅ　Ｄａｔａ」、Ｉｎ　Ｐｒｏｃ．２９ｔｈ　ＡＡＡＩ　Ｃｏｎｆｅｒｅｎｃｅ　ｏｎ　Ａｒｔｉｆｉｃｉａｌ　Ｉｎｔｅｌｌｉｇｅｎｃｅ、ｐｐ．２６４５－２６５１、２０１５年

【発明の概要】

【発明が解決しようとする課題】

【０００５】

ユーザベースのＣＦは、アベレジアタックと呼ばれる攻撃、すなわち、どのユーザとも高い類似度を示す多数の偽ユーザを投入する攻撃を受けると、偽ユーザの嗜好する商品が何時も推薦されるようになるおそれがある。

また、アイテムベースのＣＦは、セグメントアタックあるいはポピュラーアタックと呼ばれる攻撃、すなわち、ある特定のトピック（例えば、アクション映画、ホラー映画などのトピック）において、ポピュラーなアイテムと高い類似度をターゲットアイテムに持たせるために、ターゲットアイテムとポピュラーアイテムの両方に高い評価値を与える偽ユーザを多数投入する攻撃を受けると、当該トピックに属するアイテムを好むユーザに対して、ターゲットアイテムが推薦され易くなる。

上記のような推薦は不自然であり、推薦システムの本来の機能を阻害するという問題があった。

【０００６】

本発明は、ハブの出現が抑制された類似度尺度を用いる、あるいは、与えられた類似度尺度をハブが出現しにくくなるように変換して用いることにより、インフルエンサとなるユーザ、あるいは、インフルエンサとなるアイテムの出現を抑制し、これらの影響力を低減することによって、結果的に攻撃者の意図通りにターゲットアイテムの評価値を変更されることがないようにする。

本発明は、偽ユーザを投入する攻撃を受けても、結果として攻撃の影響を受けることが少ない、アイテム推薦システム及びアイテム推薦方法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、本発明の第１の態様に係るアイテム推薦システム１は、例えば図５に示すように、ユーザｕのアイテムｉに係る評価値Ｒ（ｕ，ｉ）を記入する評価マトリックスＲを記憶する評価マトリックス記憶部２１と、ハブの出現を抑制する類似度尺度を用いてユーザ間の類似度を演算する第１の類似度演算部１３１と、第１の類似度演算部１３１にて演算された類似度を用いて、対象ユーザとの類似度の高い方からｋ人のユーザを抽出する第１の近傍データ抽出部１４１と、第１の近傍データ抽出部１４１にて抽出されたｋ人のユーザのアイテムに係る評価値を用いて、対象ユーザに係る未記入のセルに記入すべき評価値を予測する第１の評価値予測部１５１と、第１の評価値予測部１５１にて予測された評価値の高いアイテムから対象ユーザに推薦すべきアイテムを抽出して、対象ユーザに推薦するアイテム推薦部１６とを備える。

【０００８】

ここにおいて、アイテムは典型的には商品又はサービスである。さらに、商品又はサービスの種類、提供時期、提供地方、価格帯を限定する（夏季果物、Ｘ月公開映画等）等の条件を定めても良い。ただし、商品又はサービスに限定されず、評価可能であれば動植物、山河、都市、建築、絵画、音楽、演劇、武道、学問、生産性、効果でも良い。

また、マトリックスＲは典型的にはユーザ数×アイテム数の評価マトリックスである。評価値Ｒ（ｕ，ｉ）として、典型的にはユーザｕのアイテムｉに係る嗜好度が使用される。ただし、嗜好度に限られず、定量的に評価可能であれば良い。例えば、健康への寄与度でも、不動産の価値でも、目的地への所要時間でも良い。また、定量的な評価はランク、レベルで表現するものでも良い。

【０００９】

また、類似度尺度とは、２つのデータの類似性を測る尺度として使用できるものすべてを含む。典型的には、内積、コサイン、ピアソン相関、距離が使用される。内積は２つのベクトルデータのスカラー積であり、コサインは長さ１に規格化されたベクトルデータの内積である。ピアソン相関は要素和がゼロになるように各要素値から要素和を差し引いた後に長さ１に規格化されたベクトルデータの内積である。さらに、内積の一般化とみなせる（機械学習分野で主に呼ばれるところの各種の）カーネルも含む。距離の典型は、２つのベクトルデータ間のユークリッド距離（Ｌ２ノルム）であるが、ユークリッド距離を一般化した距離（マンハッタン距離やＬｐノルムなど）も含む。さらに、ドメインの知識を持つ人間が各タスクの目的に応じて適宜定めた類似度スコア計算方法（ＢＬＡＳＴなど）が出力する類似度も、ここでの類似度尺度に含まれる。これらを一般的な類似度尺度ということとする。

【００１０】

また、ハブの出現を抑制する類似度尺度として、例えば全てのデータ対象がデータ中心に同等に類似になるように変換された類似度尺度、すなわちＳｐａｔｉａｌ Ｃｅｎｔｒａｌｉｔｙのない類似度尺度が該当する。例えば、上記一般的な類似度尺度に対して、原点をデータセットの平均（グローバルセントロイド）に移動する「（グローバル）センタリング」を適用して変換したもの、原点をローカルな部分集合の中心としてのローカルセントロイドに移動する「局在的センタリング」を適用して変換したものが挙げられる。さらに、ラプラシアンベースのカーネル、たとえば「コミュートタイムカーネル」（Ｍａｒｃｏ Ｓａｅｒｅｎｓ，Ｆｒａｎｃｏｉｓ Ｆｏｕｓｓ，Ｌｕｈ Ｙｅｎ，ａｎｄ Ｐｉｅｒｒｅ Ｄｕｐｏｎｔ．「Ｔｈｅ ｐｒｉｎｃｉｐａｌ ｃｏｍｐｏｎｅｎｔｓ ａｎａｌｙｓｉｓ ｏｆ ｇｒａｐｈ，ａｎｄ ｉｔｓ ｒｅｌａｔｉｏｎｓｈｉｐｓ ｔｏ ｓｐｅｃｔｒａｌ ｃｌｕｓｔｅｒｉｎｇ」、Ｉｎ Ｐｒｏｃ．１５ｔｈ Ｅｕｒｏｐｅａｎ Ｃｏｎｆｅｒｅｎｃｅ ｏｎ Ｍａｃｈｉｎｅ Ｌｅａｒｎｉｎｇ（ＥＣＭＬ），ｐｐ．３７１－３８３、２００４年）を適用して変換したものが該当する。

また、ハブの出現を抑制する類似度尺度として、全てのデータ対象がデータ中心に同等に類似になるように変換された類似度尺度以外にも、ミューチュアルプロキシミティ、ローカルスケーリング等が挙げられる。

【００１１】

また、「ユーザとの類似度の高い方からｋ人のユーザを抽出する」とは、典型的にはｋ近傍法を使用して抽出することをいう。ｋとして任意の数値が可能であるが、たとえば、ムービーレンズデータセットでは、教師あり学習の結果、予測精度を高くするには、３０＜＝ｋ＜＝１００が好ましく、４０＜＝ｋ＜＝７０がより好ましく、ｋ＝５０が最も好ましい（図７参照）。

また、評価値を予測する際に、ｋ人の平均値を使用できる。さらに、平均値を用いる際に、後述のように重み付けした平均値を用いると好ましい。重み付けには例えばユーザ間の類似度、季節による係数（果物の品質は季節に影響を受ける）等を使用できる。

【００１２】

ユーザベースのＣＦは、最近傍のｋ人（ｋ近傍法（ｋＮＮ）により抽出された最も近い方、及び類似度が高い方からｋ個のデータ）のユーザの過去の評価値を参照して未評価アイテムに対するユーザの評価値を予測する形態の推薦システムである。ユーザベースのＣＦの想定可能な欠点は、シリングアタックへの脆弱性である。シリングアタックは、システムに攻撃者によるバイアスのかかった（恣意的な）推薦を強制させるために、偽ユーザを推薦システムに投入する。ユーザベースのＣＦは、アイテムの特徴に基づく推薦を行わず、アイテムに対する他のユーザの過去の評価値に基づいて推薦を行う。このため、ユーザベースのＣＦは、どのユーザとも似るように偽ユーザを設計し、これを投入してシステムによる推薦アイテムの決定を変えさせようとする攻撃に対して、脆弱性を持つ。

【００１３】

他方、高次元データセットには、いわゆる「次元の呪い」の結果として、ハブデータが出現し易いことが見出された（Ｍｉｌｏｓ Ｒａｄｏｖａｎｏｖｉｃ, Ａｌｅｘａｎｄｒｏｓ Ｎａｎｏｐｏｕｌｏｓ, Ｎｉｒｊａｎａ Ｉｖａｎｏｖｉｃ.「Ｈｕｂｓ ｉｎ Ｓｐａｃｅ：Ｐｏｐｕｌａｒ Ｎｅａｒｅｓｔ Ｎｅｉｇｈｂｏｒｓ ｉｎ Ｈｉｇｈ-Ｄｉｍｅｎｓｉｏｎａｌ Ｄａｔａ」、Ｊｏｕｒｎａｌ ｏｆ Ｍａｃｈｉｎｅ Ｌｅａｒｎｉｎｇ Ｒｅｓｅａｒｃｈ, ｐｐ. ２４８７－２５３１, ２０１０年）。すなわち、高次元ではハブと呼ばれる少数のデータが他のデータのｋＮＮに頻繁に現れる。ユーザベースのＣＦシステムにおいて、ｋＮＮが計算される時、各々のユーザはアイテム数の次元を持つベクトルとして表されるが、アイテムは一般に数多く存在するため、ベクトルは高次元ベクトルとなる。したがって、ハブとなるデータ（ハブデータ）が出現する。ハブユーザは推薦工程にインフルエンサとして寄与するので、推薦システムによる推薦アイテムの決定に大きな影響を与える。

シリングアタックは、ハブを利用する攻撃と見て取れる。ユーザベースのＣＦに対する攻撃では、システムによる推薦アイテムの決定を意図的にコントロールすることを目的とし、インフルエンサ、すなわち、ハブとなる偽ユーザを投入する。具体的には、偽ユーザをユーザに関するデータ中心に類似するように設計し、投入する。

【００１４】

そこで、攻撃の影響を回避するために、ｋＮＮを求めるために使用される類似度尺度を全てのユーザがデータ中心に同等に類似するように変換することによって、ハブユーザ、すなわち、インフルエンサの出現自体を抑制し、偽ユーザをインフルエンサとしてシステムに送り込む攻撃者の企てを無効化することを提案する。ハブの出現を抑制する方法はいくつか提案されているが、たとえば、与えられた類似度マトリックスからコミュートタイムカーネルを計算することによって、又はより簡易に類似度マトリックスをセンタリングすることによって達成できる。ムービーレンズデータセットを用いて、かかる方法適用後に、偽ユーザがハブユーザと成りにくくなる傾向の存在を確認した（図８及び図９参照）。結果として、かかる類似度尺度の変換は、アイテム推薦の精度を劣化させることなく（図７参照）、攻撃に対して耐性の有るシステムを提供する。

【００１５】

本態様のように構成すると、ハブの出現を抑制する類似度尺度を使用して類似度を演算するのでハブの出現を抑制でき、偽ユーザを投入する攻撃を受けても、結果として攻撃の影響を受けることが少ないアイテム推薦システムを提供することができる。

【００１６】
上記課題を解決するために、本発明の第２の態様に係るアイテム推薦システム１は、例えば図５に示すように、ユーザｕのアイテムｉに係る評価値Ｒ（ｕ，ｉ）を記入する評価マトリックスＲを記憶する評価マトリックス記憶部２１と、ハブの出現を抑制する類似度尺度を用いてアイテム間の類似度を演算する第２の類似度演算部１３２と、第２の類似度演算部１３２にて演算された類似度を用いて、対象アイテムとの類似度の高い方からｋ個のアイテムを抽出する第２の近傍データ抽出部１４２と、第２の近傍データ抽出部１４２にて抽出されたｋ個のアイテムに係る対象ユーザの評価値を用いて、対象ユーザに係る未記入のセルに記入すべき評価値を予測する第２の評価値予測部１５２と、第２の評価値予測部１５２にて予測された評価値の高いアイテムから対象ユーザに推薦すべきアイテムを抽出して、対象ユーザに推薦するアイテム推薦部１６を備える。
【００１７】
第１の態様では、ユーザ間の類似度に基づいて評価値を求めたが、本態様ではアイテム間の類似度に基づいて評価値を求める。第１の態様では、ｋ人の平均値を使用したが、本態様ではｋ個のアイテムの平均値を用いる。しかし、その他のシステム構成は第１の態様と同様であり、第１の態様と同様に、ハブの出現を低減できるので、偽ユーザが投入されても、結果として推薦アイテムの決定が偽ユーザの投入に影響されにくい、すなわち、攻撃に対して頑健なアイテム推薦システムを提供することができる。
このように構成すると、ハブの出現を抑制する類似度尺度を使用するので、偽ユーザを投入する攻撃を受けても、結果として攻撃の影響を受けることが少ないアイテム推薦システムを提供することができる。
【００１８】
また、本発明の第３の態様に係るアイテム推薦システム１は、第１又は第２の態様において、ハブの出現を抑制する類似度尺度を記憶する類似度尺度記憶部２２を備える。
このように構成すると、システムに記憶されたハブの出現を抑制する類似度尺度を使用して類似度を演算するのでハブの出現を抑制でき、アイテム推薦時に偽ユーザによる影響を少なくすることができる。
【００１９】
また、本発明の第４の態様に係るアイテム推薦システムは、第３の態様において、一般的な類似度尺度に基づく類似度を前記ハブの出現を抑制する類似度尺度に基づく類似度に変換する類似度尺度変換部１３５を備える。
ここにおいて、類似度の変換には、例えば一般的な類似度尺度の式をハブの出現を抑制する類似度尺度の式に変換して類似度を求める方法、一般的な類似度尺度で求めた類似度を行列によりハブの出現を抑制する類似度尺度に基づく類似度に変換する方法等が使用される。
このように構成すると、一般的な類似度尺度をハブの出現を抑制する類似度尺度に変換して使用することによりハブの出現を抑制でき、アイテム推薦時に偽ユーザによる影響を少なくすることができる。
【００２０】
また、本発明の第５の態様に係るアイテム推薦システムは、第1ないし第４のいずれかの態様において、対象ユーザに係る未記入のセルに記入すべき評価値を予測するに際し、記入すべき評価値として、重み付けをした平均値を用いる。
ここにおいて、重み付けには例えばユーザ間あるいはアイテム間の類似度、季節による係数（果物の品質は季節に影響を受ける）等を使用できる。本態様のように構成すると予測精度を向上できる。
【００２１】
上記課題を解決するために、本発明の第６の態様に係るアイテム推薦方法は、例えば図６（ａ）に示すように、ユーザｕのアイテムｉに係る評価値Ｒ（ｕ，ｉ）を記入する評価マトリックスＲを記憶する評価マトリックス記憶工程（Ｓ１０４）と，ハブの出現を抑制する類似度尺度を用いてユーザ間の類似度を演算する第１の類似度演算工程（Ｓ１０７）

と，第１の類似度演算工程（Ｓ１０７）にて演算された類似度を用いて、対象ユーザとの類似度の高い方からｋ人のユーザを抽出する第１の近傍データ抽出工程（Ｓ１０８）と、第１の近傍データ抽出工程（Ｓ１０８）にて抽出されたｋ人のユーザのアイテムに係る評価値を用いて、対象ユーザに係る未記入のセルに記入すべき評価値を予測する第１の評価値予測工程（Ｓ１０９）と、第１の評価値予測工程（Ｓ１０９）にて予測された評価値の高いアイテムから対象ユーザに推薦すべきアイテムを抽出して、対象ユーザに推薦するアイテム推薦工程（Ｓ１１０）とを備える。

【００２２】

本態様は第１の態様に係るアイテム推薦システムに対応するアイテム推薦方法である。

本態様のように構成すると、ハブユーザの出現を低減できるので、偽ユーザを投入する攻撃を受けても、結果として攻撃の影響を受けることが少ないアイテム推薦方法を提供することができる。

【００２３】

上記課題を解決するために、本発明の第７の態様に係るアイテム推薦方法は、例えば図６（ｂ）に示すように、ユーザｕのアイテムｉに係る評価値Ｒ（ｕ，ｉ）を記入する評価マトリックスＲを記憶する評価マトリックス記憶工程（Ｓ１０４）と，ハブの出現を抑制する類似度尺度を用いてアイテム間の類似度を演算する第２の類似度演算工程（Ｓ２０７）と，第２の類似度演算工程（Ｓ２０７）にて演算された類似度を用いて、対象アイテムとの類似度の高い方からｋ個のアイテムを抽出する第２の近傍データ抽出工程（Ｓ２０８）と、第２の近傍データ抽出工程（Ｓ２０８）にて抽出されたｋ個のアイテムに係る対象ユーザの評価値を用いて、対象ユーザに係る未記入のセルに記入すべき評価値を予測する第２の評価値予測工程（Ｓ２０９）と、第２の評価値予測工程（Ｓ２０９）にて予測された評価値の高いアイテムから対象ユーザに推薦すべきアイテムを抽出して、対象ユーザに推薦するアイテム推薦工程（Ｓ１１０）とを備える。

【００２４】

本態様は第２の態様に係るアイテム推薦システムに対応するアイテム推薦方法である。

本態様のように構成すると、ハブアイテムの出現を低減できるので、偽ユーザを投入する攻撃を受けても、結果として攻撃の影響を受けることが少ないアイテム推薦方法を提供することができる。

【００２５】

また、本発明の第８の態様に係るアイテム推薦システムは、第６又は第７の態様において、一般的な類似度尺度に基づく類似度をハブの出現を抑制する類似度尺度に基づく類似度に変換する類似度尺度変換工程（Ｓ１０６）を備える。

このように構成すると、一般的な類似度尺度に基づく類似度をハブの出現を抑制する類似度尺度に基づく類似度に変換して使用することによりハブの出現を抑制でき、アイテム推薦時に偽ユーザによる影響を少なくすることができる。

【００２６】

また、本発明の第９の態様に係るプログラムは、第６ないし第８のいずれかの態様のアイテム推薦方法をコンピュータに実行させるためのプログラムである。

【００２７】

また、本発明の第１０の態様に係る記録媒体は、第９の態様に係るプログラムを記録したコンピュータ読み取り可能な記録媒体である。

【発明の効果】

【００２８】

本発明によれば、偽ユーザを投入する攻撃を受けても、結果として攻撃の影響を受けることが少ないアイテム推薦システム及びアイテム推薦方法を提供できる。

【図面の簡単な説明】

【００２９】

【図１】評価マトリックスＲの例を示す図である。

【図２】ユーザ間の相関を示す図である。図２（ａ）はユーザｕとユーザｖとの相関を説

明するための図、図２（ｂ）は相関が強い場合の散布図、図２（ｃ）は相関が弱い場合の散布図である。
【図３】低次元と高次元におけるN_{10}分布を示す図である。図３（ａ）は低次元、図３（ｂ）は高次元におけるN_{10}のヒストグラムである。
【図４】N_{10}値とデータ中心への類似度の関係を示す散布図である。図４（ａ）は低次元、図４（ｂ）は高次元における図である。
【図５】実施例１及び実施例２におけるアイテム推薦システム１の構成例を示す図である。
【図６】実施例１及び実施例２におけるアイテム推薦方法の処理フロー例を示す図である。図６（ａ）は実施例１における処理フロー例を示す図、図６（ｂ）は実施例２における処理フロー例を示す図である。
【図７】最近傍パラメータｋを横軸、平均絶対誤差（ＭＡＥ）を縦軸とし、異なる類似度尺度を用いたアイテム推薦システムを比較する図である。
【図８】ユーザ間類似度尺度として一般的なピアソン相関を用いた場合、投入された偽ユーザがハブになっていることを示す図（その１）で、誠実なユーザ（偽ユーザ以外のユーザ）を含む全ユーザに関するN_{50}とデータ中心への類似度に係る散布図である。
【図９】ユーザ間類似度尺度として一般的なピアソン相関を用いた場合、投入された偽ユーザが（ａ）ハブとなること、（ｂ），（ｃ）類似度尺度の変換によってハブとなりにくくなることを示す図（その２）である。図９（ａ）はユーザ間類似度尺度として一般的なピアソン相関を用いた場合のN_{50}に関するヒストグラムである。図９（ｂｃ）はセンタリングによる変換後のN_{50}に関するヒストグラム、図９（ｃ）はコミュートタイムカーネルによる変換後のN_{50}に関するヒストグラムである。
【発明を実施するための形態】
【００３０】
図面を参照して以下に本発明の実施の形態について説明する。なお、各図において、互いに同一又は相当する部分には同一符号を付し、重複した説明は省略する。
【００３１】
〔ユーザベースのＣＦ〕
ユーザ数N_{user}×アイテム数N_{item}のマトリックスＲを、アイテムに対するユーザの過去の反応（評価値）からなるデータセットとする。Ｒ（ｕ，ｉ）はｕ番目のユーザのｉ番目のアイテムへの評価値を示す。マトリックスＲはｎｉｌと称する値の無い空欄を含んでいる。このｎｉｌの値は、ユーザのアイテムに対する評価がまだ与えられていないことを意味する。一般に、マトリックスＲは空欄が多く、大部分がｎｉｌである。ユーザベースのＣＦは（後述するアイテムベースのＣＦも）、ｋ近傍法を利用してこれらの値を予測するものである。
【００３２】
【数１】

$$Pred(u,i)=\frac{\sum_{n\in U} Sim(u,n)\{R(n,i)-\bar{R}(n)\}}{\sum_{n\in U} Sim(u,n)}+\bar{R}(u) \quad (1)$$

式（１）はｕ番目のユーザのｉ番目のアイテムへの評価値Ｒ（ｕ，ｉ）を予測する予測関数Ｐｒｅｄ（ｕ，ｉ）を示す。ユーザｕとユーザｖ間の類似度をＳｉｍ（ｕ，ｖ）、類似度Ｓｉｍのもとでユーザｕと最近傍となるｋ人のユーザの集合をＵとし、Ｕに属するユーザｎについて使用する評価値は、Ｒ（ｎ，ｉ）≠ｎｉｌを満たす。
【００３３】
さらに

【数２】

$$\bar{R}(u) = \frac{\sum_{i=1}^{N_{\mathrm{item}}} R(u,i)\delta[R(u,i) \neq \mathrm{nil}]}{\sum_{i=1}^{N_{\mathrm{item}}} \delta[R(u,i) \neq \mathrm{nil}]}$$

である。

【数３】

$$\bar{R}(u)$$

はユーザｕが評価したアイテムに対する平均の評価値である。δ〔．〕は〔〕内の条件が満たされれば１、それ以外は０となる指示関数である。

【００３４】

図１にマトリックスＲの例を示す。図１（ａ）は偽ユーザ投入前、図１（ｂ）は偽ユーザ投入後の例を示す。列方向にアイテムｉを、行方向にユーザｕを配置し、その交点となるセル（欄）に評価値Ｒ（ｕ，ｉ）が記入されている。ここでは評価値Ｒ（ｕ，ｉ）は１から５の５段階の整数で評価されている。ターゲットアイテムの評価値の引き上げを目的とするアベレジアタックでは、図１（ｂ）の下側のように、ターゲットアイテムであるアイテム１に高い評価を、その他のアイテムには過去にそのアイテムに対して評価を与えたユーザが付与した評価の平均に近い値を与える偽ユーザが投入される。

【００３５】

図２を用いて、ユーザ間の類似度を、アイテムに与える評価値のピアソン相関により測る方法について説明する。図２（ａ）はユーザｕとユーザｖの相関を説明するための図である。図１（ａ）に示したアイテム２（Ｒ（ｕ，２）＝１，Ｒ（ｖ，２）＝１）、および、アイテム３（Ｒ（ｕ，３）＝５，Ｒ（ｖ，３）＝４）が図２（ａ）にプロットされている。図２（ｂ），（ｃ）は全アイテムｉ（Ｒ（ｕ，ｉ）、Ｒ（ｖ，ｉ））をプロットした散布図であり、図２（ｂ）はユーザｕとユーザｖの（正の）相関が強い場合、図２（ｃ）はユーザｕとユーザｖの相関が弱い場合の例である。相関が強い場合は、全アイテムのプロットは直線に乗り、相関が弱い場合はアトランダムとなる。

【００３６】

〔ユーザベースのＣＦに一般的に使用されるユーザ間類似度尺度〕

ユーザベースのＣＦにおいて、ユーザ間類似度を与える関数Ｓｉｍ（．，．）を適切に選定することはが重要である。なぜなら、類似度関数はｋＮＮに入るユーザ、及び式（１）に係るｋＮＮに入るユーザの重みを決定するからである。

一般的な類似度尺度関数として、マトリックスＲのｎｉｌを０に置換した後に、行ベクトル（各ユーザが与えた評価値のベクトル）がなす角度のコサイン（ｃｏｓ）を計算するコサイン類似度がある。式（２）にこれを示す。

【００３７】

【数４】

$$Sim(u,v) = Cos(u,v) = \frac{\langle x_u, x_v \rangle}{\sqrt{\langle x_u, x_u \rangle}\sqrt{\langle x_v, x_v \rangle}} \qquad (2)$$

ここに、x_uはN_{item}次元ベクトルで、その成分はＲ（ｕ，ｉ）≠ｎｉｌならばx_u（ｉ）＝Ｒ（ｕ，ｉ）、それ以外はx_u（ｉ）＝０となる。上記関数使用の１つの欠点は、

各々のユーザｕがアイテムに与える平均的評価
【数５】

$$\bar{R}(u)$$

の違いに基づくバイアスが無視されるという点である。それ故、上記欠点の修正方法として、各ベクトル成分から
【数６】

$$\bar{R}(u)$$

を差し引く方法が一般的に使用される。
【００３８】
ユーザのバイアス
【数７】

$$\bar{R}(u)$$

を差し引いたベクトルを用いた類似度は、式（３）のように計算され、これはユーザ間のピアソン（Ｐｅａｒｓｏｎ）相関と呼ばれる。
【数８】

$$Sim(u,v) = Pearson(u,v) = \frac{\langle x'_u, x'_v \rangle}{\sqrt{\langle x'_u, x'_u \rangle}\sqrt{\langle x'_v, x'_v \rangle}} \quad (3)$$

ここに、もし、Ｒ（ｕ，ｉ）≠ｎｉｌかつＲ（ｖ，ｉ）≠ｎｉｌならば、
【数９】

$$x'_u(i) = R(u,i) - \bar{R}(u), x'_v(i) = R(u,i) - \bar{R}(v)$$

であり、そうでなければｘ’$_{u}$（ｉ）＝０、ｘ’$_{v}$（ｉ）＝０となる。
【００３９】
〔ＣＦへのアタック〕
ユーザベースのＣＦ、すなわち、ユーザと類似する他のユーザの過去の評価を参照してユーザに推薦するアイテムを決める推薦システム及び推薦方法に対しては、アベレジアタックと呼ばれるシリングアタックが効果を持つ。システムが持つ評価値マトリックスＲを、不正な評価値を加えることによって改ざんすれば、推薦されるアイテムは変更される。この目的で偽ユーザを投入する攻撃をシリングアタックと呼び、アベレジアタックはその一つである。この攻撃を受けると、どのユーザも偽ユーザとの類似度が高くなる。つまり、偽ユーザは、推薦アイテムの決定に影響力を持つインフルエンサ、すなわち、ハブユーザとなる。
アベレジアタックにおいて投入される偽ユーザは、ターゲットアイテム（攻撃対象アイテム）を好む振る舞いをし、他のアイテムに対しては誠実なユーザ（偽ユーザ以外のユーザ）の平均的な振る舞いをする。すなわち、偽ユーザはターゲットアイテムには高い評価値点を与え、残りのアイテムには平均的な評価値を与える。結果として、偽ユーザはターゲットアイテムを好み、かつ、任意の誠実なユーザとの類似が高くなる。それ故に、ユー

ザベースのＣＦは、アベレジアタックを受けると、偽ユーザが高い評価を与えるターゲットアイテムを全ての誠実なユーザに推薦しやすくなる。

【００４０】

アイテムベースのＣＦ、すなわち、類似する他のアイテムに対するユーザの過去の評価を参照してユーザに推薦するアイテムを決める推薦システム及び推薦方法に対しては、セグメントアタックあるいはポピュラーアタックと呼ばれるシリングアタックが効果を持つ。この攻撃を受けると、攻撃対象となるターゲットアイテムは、どのユーザも高い評価を与えるポピュラーアイテムとの類似度が高くなる。攻撃者は、ポピュラーアイテムが推薦アイテムの決定に影響力を持つインフルエンサ、すなわち、ハブアイテムであることを悪用し、システムによるターゲットアイテムの評価値を不当に高く変更しようとする。

【００４１】

〔ハブ現象〕

ハブ現象は、「次元の呪い」の結果として起こる現象の一つである。Ｄをｄ次元データの集合とし、N_k（ｘ）は、Ｄ内のデータｘ∈ＤがＤ内の他のデータのｋＮＮ内に入る回数を示す。次元ｄが増加すると、N_kの分布形状は右に長い尾を引くように変わる（図３参照）。又は、少数のデータが大きなN_k値をとるようになる。かかる大きなN_k値を示すデータをハブといい、かかる現象をハブネス（ハブ現象）という。

【００４２】

ここでは、人工データセットを用いてハブ現象について説明する。推薦システムでは一般に各ユーザは数個のアイテムに対してのみ評価値を与えるため、評価マトリックスＲは空欄の多いスパースなマトリックスとなるが、この情況を模してスパースなデータセットを人工的に生成した。データセットは２０００個のデータからなり、それぞれｄ次元ベクトルである。各データの生成方法は次の通りである：まず、各次元ｉ＝１，．．．，ｄに対して、Ｌｏｇｎｏｒｍａｌ（５；１）分布にしたがって発生させた正の実数を丸め、整数ｎｉを得る。そして、２０００個のデータからランダムにn_i個を選択し、その各々に対して、範囲〔０，１〕から一様に乱数を発生させ、それを各々のベクトルのｉ番目の要素（ｉ次元成分）とする。

【００４３】

図３は、データ間の類似度をベクトル間の角度、すなわち、ｃｏｓ（コサイン類似度）を用いて測ったときの、N_{10}分布を示すヒストグラムである。図３（ａ）は低次元、図３（ｂ）は高次元の場合のヒストグラムである。ハブ現象の出現を説明するために、次元が低い場合（ｄ＝５０）と高い場合（ｄ＝１０００）の２ケースにおいてN_{10}分布を比較した。図３は、高次元では大きなN_{10}値を持つデータが出現し、結果としてN_{10}の分布が歪む（対称でなくなる）ことを示す。最大となるN_{10}は図３（ａ）で３８、図３（ｂ）で１３３である。

【００４４】

図４はN_{10}値とデータ中心への類似度の関係を示す散布図である。図４（ａ）は低次元、図４（ｂ）は高次元における図である。N_{10}値とデータ中心への類似度との間には、高次元で強い相関がみられることから、ハブ現象の起源は、高次元で発生するデータ中心へのバイアス、すなわち、Ｓｐａｔｉａｌ　Ｃｅｎｔｒａｌｉｔｙであることが分かる。

【００４５】

〔攻撃シリングアタックとハブ現象との関係〕

ナノポウラス達（Ａ．Ｎａｎｏｐｏｕｌｏｓ，ａｎｄ Ｍ．Ｒａｄｏｖａｎｏｖｉｃ，Ｍ．Ｉｖａｎｏｖｉｃ．Ｈｏｗ ｄｏｅｓ ｈｉｇｈ ｄｉｍｅｎｓｉｏｎａｌｉｔｙ ａｆｆｅｃｔ ｃｏｌｌａｂｏｒａｔｉｖｅ ｆｉｌｔｅｒｉｎｇ？ Ｉｎ Ｐｒｏｃ．３ｒｄ ＡＣＭ Ｃｏｎｆ．ｏｎ Ｒｅｃｏｍｍｅｎｄｅｒ Ｓｙｓｔｅｍｓ（ＲｅｃＳｙｓ），ｐａｇｅｓ ２９３－２９６，２００９．）及びニース達（Ｐ．Ｋｎｅｅｓ，Ｄ．Ｓｃｈｎｉｔｚｅｒ，ａｎｄ Ａ．Ｆｌｅｘｅｒ．「Ｉｍｐｒｏｖｉｎｇ ｎｅｉｇｈｂｏｒｈｏｏｄ－ｂａｓｅｄ ｃｏｌｌａｂｏｒａｔｉｖｅ ｆｉｌ

ｔｅｒｉｎｇ ｂｙ ｒｅｄｕｃｉｎｇ ｈｕｂｎｅｓｓ」、Ｉｎ Ｐｒｏｃ．ＩＣＭＲ’１４，ｐａｇｅｓ １６１－１６８，２０１４年）は、ユーザベースあるいはアイテムベースのＣＦにおいては、ｋＮＮは高次元で計算されるので、ハブ現象が出現すると報告した。通常、ユーザ数及びアイテム数は大きいので、コサイン類似度やピアソン相関のような類似度の計算に使用されるベクトルは高次元となり、ハブ現象が生じる。そして、他のデータのｋＮＮ内に頻繁に現れるハブデータは、多くの推薦を決定するのに影響する。しかし、ハブデータは多くのデータにとってあまり意味を持たないデータである。なぜなら、ハブデータは高次元でデータ中心に類似するという理由によってのみｋＮＮの中に頻繁に生じるのであり、個々のデータを特徴付けるための役には立たないからである。事実、ニース達の文献によれば、推薦システムのパフォーマンスはハブデータの存在により悪化する。さらに、ハブデータはシステムによる推薦アイテムの決定に強い影響を持つデータなので、もしもシステム外からハブデータを操ることができれば、システムを効果的に攻撃することが可能となる。

実際、ハブ現象は推薦システムを攻撃に対して危うくする。例えば、アベレジアタックによりシステムに投入された偽ユーザは、ハブデータとなることで、システムに大きな影響を与える。よって、ハブ現象の発生を抑えることは攻撃回避につながると考えられる。

【００４６】

〔データ中心へのバイアス削減によるハブの抑制〕

データ中心との類似度が高い少数のデータがハブになるというのであれば、類似度尺度を全てのデータ対象がデータ中心に同等に類似になる類似度尺度に変換することにより、ハブ現象を抑制できると考えられる。かかる類似度（尺度）の変換は、与えられた類似度からコミュートタイムカーネルを計算することによって得られ、より簡易には、与えられた類似度をセンタリングすることによって得られる。

Ｎをデータ数とし、Ｋをサイズ Ｎの類似度行列とする。Ｋに対するコミュートタイムカーネルK^{CT}は、式（４）で与えられる。

$K^{CT} = L^{+}$（Ｌの一般化逆行列）・・・（４）

ここに、Ｌ＝Ｄ－Ｋはグラフラプラシアンと呼ばれる。Ｄは$D_{ii} = \Sigma_j K_{ij}$となる対角行列である。

【００４７】

次に、Ｉを単位行列、

【数１０】

$$\vec{1}$$

を全要素が１であるＮ次元ベクトルとする。Ｋをセンタリングした類似度行列K^{CENT}は式（５）のように計算される。

【数１１】

$$\mathbf{K}^{\mathrm{CENT}} = \left(\mathbf{I} - \frac{1}{N}\vec{1}\vec{1}^{\mathrm{T}}\right)\mathbf{K}\left(\mathbf{I} - \frac{1}{N}\vec{1}\vec{1}^{\mathrm{T}}\right) \qquad (5)$$

【実施例１】

【００４８】

図５に実施例１におけるアイテム推薦システム１の構成例を示す。

本実施例では、アイテム推薦システム１としてユーザベースのＣＦを説明する。すなわち、類似度演算に例えばｋ近傍法を用い、対象ユーザと評価が似ている他のユーザの過去

の評価値を参照してアイテムを推薦するシステムである。アイテムが商品で、評価値が嗜好度の場合は、対象ユーザと嗜好が似ている他のユーザの過去の嗜好度を参照して商品を推薦するシステムである。

なお、図５の構成は実施例１（ユーザベースのＣＦ）及び実施例２（アイテムベースのＣＦ）の両者に適用可能な構成である。このため、ユーザベースのＣＦ及びアイテムベースのＣＦに共通する説明は本実施例で行うこととし、実施例２では差異を説明する程度とする。

【００４９】

アイテム推薦システム１は、データ及びコマンドを処理するパーソナルコンピュータ（ＰＣ）１０、各部で処理された又は入出力されたデータ・コマンド等を表示する表示部１８、データ及びコマンドを入出力するための入出力部１９、及び各部で処理された又は入出力されたデータ・コマンド等を記憶する記憶部２０を含んで構成される。

パーソナルコンピュータ（ＰＣ）１０は、ユーザ及びアイテムを登録する登録部１１、ユーザのアイテムに係る評価の程度を表す評価値を評価マトリックスに記入する評価部１２、類似度尺度を用いてユーザ間の類似度及び／又はアイテム間の類似度を演算する類似度演算部１３、類似度の高い方から例えばｋ個の対象データ（ユーザ及び／又はアイテム）を抽出する近傍データ抽出部１４、評価マトリックスＲのセルに評価値が記入されていない時に、近傍データ抽出部１４で抽出された評価値に基づいて、セルに記入されるであろうと予測される評価値を予測する評価値予測部１５、評価値予測部１５にて高い評価値を予測されたアイテムを推薦するアイテム推薦部１６、アイテム推薦システム１の各部を制御して、アイテム推薦システムとして機能させる制御部１７を備える。

【００５０】

ここにおいて、登録部１１はユーザを登録するユーザ登録部１１１とアイテムを登録するアイテム登録部１１２を有する。本実施例では、類似度演算部１３は、類似度尺度として、ハブを抑制する類似度尺度を用いて、ユーザ間の類似度及び／又はアイテム間の類似度を演算する。詳しくは、類似度演算部１３はハブを抑制する類似度尺度を用いて、評価マトリックスＲの各行のユーザの評価値に着目してユーザ間の類似度を演算する第１の類似度演算部１３１と、各列のアイテムの評価値に着目してアイテム間の類似度を演算する第２の類似度演算部１３２を有する。また、一般的な類似度尺度に基づく類似度からハブを抑制する類似度尺度に基づく類似度への変換を行う類似度尺度変換部１３５を有する。近傍データ抽出部１４は、類似度の高い方から例えばｋ個のユーザを抽出する第１の近傍データ抽出部１４１と、類似度の高い方から例えばｋ個のアイテムを抽出する第２の近傍データ抽出部１４２を有する。評価値予測部１５は、第１の近傍データ抽出部１４１で抽出された評価値に基づいて、対象ユーザの評価値を予測する第１の評価値予測部１５１と、第２の近傍データ抽出部１４２で抽出された評価値に基づいて、対象ユーザの評価値を予測する第２の評価値予測部１５２を有する。第１の評価値予測部１５１及び第２の評価値予測部は、対象ユーザに係る未記入のセルに記入すべき評価値を予測するに際し、記入すべき評価値として例えば平均値を用いることができる。また、重み付けをした平均値を用いるのが、推薦精度を高くできるので好ましい。

【００５１】

また、記憶部２０は評価マトリックスＲを記憶する評価マトリックス記憶部２１、類似度尺度及び第１の類似度演算部１３１で演算されたユーザに関する類似度データ、及び／又は、第２の類似度演算部１３２で演算されたアイテムに関する類似度データを記憶する類似度尺度記憶部２２を有する。第１の類似度演算部１３１及び第２の類似度演算部１３２の演算データはハブデータが出現しにくい類似度尺度を用いて演算したものである。類似度尺度記憶部２２は、一般的な類似度尺度を記憶してもよい。また、記憶部２０は近傍データ抽出部１４にて抽出されたデータ（ユーザ及び／又はアイテム）を記憶する近傍データ記憶部２３、評価値推定部１５で推定されたアイテムを記憶する推薦アイテム記憶部２４を有する。類似度尺度記憶部２２は、類似度データの他に一般的な類似度尺度及び／又はハブの出現を抑制する類似度尺度を記憶する。ハブの出現を抑制する類似度尺度を記

憶せず、一般的な類似度尺度を記憶している場合には、類似度尺度変換部１３５にて一般的な類似度尺度をハブの出現を抑制する類似度尺度へ変換し、類似度尺度記憶部２２には得られたハブの出現を抑制する類似度尺度が記憶し直される。類似度の変換には、例えば一般的な類似度尺度の式をハブの出現を抑制する類似度尺度の式に変換して類似度を求める方法、一般的な類似度尺度で求めた類似度を行列によりハブの出現を抑制する類似度尺度に基づく類似度に変換する方法等が使用される。この場合、変換前の一般的な類似度尺度を消去せずに残しておいても良い。一般的な類似度尺度とハブの出現を抑制する類似度尺度を共に記憶しておくと、類似度演算結果及びアイテム推薦に係る評価値予測結果を組み合わせて用いたり、比較したりすることができる。近傍データ記憶部２３は、第１の近傍データ抽出部１４１にて抽出されたユーザを記憶する第１の近傍データ記憶部２３１と、第２の近傍データ抽出部１４２にて抽出されたアイテムを記憶する第２の近傍データ記憶部２３２を有する。推薦アイテム記憶部２４には、ユーザに対してアイテム推薦時に表示したい内容が記憶される。例えば、アイテム名の他に、アイテムについての説明、アイテムを使用するための説明、アイテムの画像等を記憶する。これらの内容は推薦時に表示部１８に表示される。

【００５２】

なお、本実施例では、ユーザと類似度の高い方からｋ人を抽出し、ｋ人の評価値の平均としてユーザの評価値を予測するので、類似度演算部１３として第１の類似度演算部１３１、近傍データ抽出部１４として第１の近傍データ抽出部１４１、評価値予測部１５として第１の評価値予測部１５１、近傍データ記憶部２３としての近傍ユーザ記憶部２３１を使用できれば良く、第２の類似度演算部１３２、第２の近傍データ抽出部１４２、第２の評価値予測部１５２、近傍アイテム記憶部２３２は無くても良い。これらは実施例２で使用される。

【００５３】

図６に実施例１におけるアイテム推薦方法の処理フロー例を示す。図６（ａ）は実施例１における処理フロー例を示す図、図６（ｂ）は後述する実施例２における処理フロー例を示す図である。

まず、評価マトリックスＲにアイテムｉ（Ｓ１０１：アイテム登録工程）、及びユーザｕを登録する（Ｓ１０２：ユーザ登録工程）。アイテムｉの登録とユーザｕの登録はどちらが先でも良く、並行して行っても良い。次に、評価マトリックスＲにユーザｕのアイテムｉに係る評価の程度を表す評価値Ｒ（ｕ，ｉ）を登録する（Ｓ１０３：評価値登録工程）。本実施例ではアイテムを商品とし、評価値を嗜好度とし、評価マトリックスＲを嗜好度マトリックスとする。評価は例えば５段階評価（１～５の整数)）で行う。ユーザ自身が登録しても良く、システム側で過去のユーザの当該アイテムに係る振る舞いを参照して登録しても良い。必ずしもマトリックスＲ全体を記入する必要はなく、空欄のセルがあっても良く、通常は大部分が空欄になっている。評価マトリックスＲは評価マトリックス記憶部２１に記憶される（Ｓ１０４：評価マトリックス記憶工程）。

【００５４】

次に、評価マトリックスＲに基づいて各ユーザに対して推薦すべきアイテムを定める。まず、ユーザ本人に似た他のユーザを求めるための類似度演算を行う。類似度演算を行うに際し、類似度尺度記憶部２２には類似度尺度として予め一般的な類似度尺度又はハブの出現を抑制する類似度尺度が記憶されているものとする。まず、類似度尺度記憶部２２にハブの出現を抑制する類似度尺度が有るか無いかを判定する（Ｓ１０５：ハブ抑制類似度尺度の有無判定工程）。類似度尺度記憶部２２に、ハブの出現を抑制する類似度尺度が記憶されておらず、一般的な類似度尺度が記憶されている場合には（Ｓ１０５でＮｏの場合）、一般的な類似度尺度に基づく類似度をハブの出現を抑制する類似度尺度に基づく類似度に変換する（Ｓ１０６：類似度尺度変換工程）。ハブの出現を抑制する類似度尺度として、例えば全てのデータ対象がデータ中心に同等に類似になる類似度尺度、すなわちＳｐａｔｉａｌ　Ｃｅｎｔｒａｌｉｔｙのない類似度尺度を使用できる。具体的には、例えば、センタリングを行う又はコミュートタイムカーネルへの変換を行う。変換されたハブを

抑制する類似度尺度は類似度尺度記憶部２２に記憶される。次に、ハブの出現を抑制する類似度尺度を用いてユーザ間の類似度を演算する（Ｓ１０７：第１の類似度演算工程）。なお、類似度の変換を行列で行う場合には類似度尺度変換工程（Ｓ１０６）と第１の類似度演算工程（Ｓ１０７）とが一括して行われる。この場合類似度尺度は必ずしも式として残るとは限らないが、演算結果においてハブを抑制する類似度尺度に基づく類似度データに内在して残ることになる。類似度尺度記憶部２２に、ハブの出現を抑制する類似度尺度がすでに記憶されている場合には（Ｓ１０５でＹｅｓの場合）、類似度尺度変換工程（Ｓ１０６）を省略し、ハブの出現を抑制する類似度尺度を用いてユーザ間の類似度を演算する（Ｓ１０７：第１の類似度演算工程）。ハブの出現を抑制する類似度尺度を用いて演算された結果は、類似度尺度記憶部２２に記憶される。

【００５５】

次に、類似度尺度記憶部２２に記憶された類似度が高い方から例えばｋ人のユーザを抽出する（Ｓ１０８：第１の近傍データ抽出工程）。そして、抽出されたｋ人のユーザの平均評価値等に基づき、対象ユーザの空欄になっている評価値を予測する（Ｓ１０９：第１の評価値予測工程）。第１の評価値予測工程（Ｓ１０９）では、対象ユーザに係る未記入のセルに記入すべき評価値を予測するに際し、例えば平均値を用いて予測することができる。また、重み付けをした平均値を用いるのが好ましい。最後に、予測値の高いアイテムを推薦する（Ｓ１１０：アイテム推薦工程）。例えば、アイテムを提供するインターネットのサイトにユーザが訪れた時に、当該ユーザに関して予測値の高い順にアイテムを提示する。また、電子メールで当該ユーザ宛に配信しても良い。

【００５６】

〔実験〕

ハブデータの出現を抑制することが、ユーザベースのＣＦを、アベレジアタックに対して頑健にすることを確かめる実験を行った。実験には、推薦タスクのベンチマークデータとして使用されるムービーレンズ１Ｍデータセット（ｍ１－１ｍ）を用いた。このデータセットは、６，０４０ユーザ、３，７０６アイテムに対する１，０００，２０９個の評価値（整数１～５の５段階評価）から成る。どのユーザも少なくとも２０アイテムを評価している。ベースラインとなるユーザ間の類似度尺度として、一般的に使われるコサイン類似度（Ｃｏｓ）、及び、ピアソン相関とｓｈｒｕｎｋｅｎピアソン相関（Ｐｅａｒｓｏｎ）を用い、式（１）を用いて評価値を予測した。ｓｈｒｕｎｋｅｎピアソン相関がピアソン相関より良い精度が出ることが知られているので、今回はピアソン相関として、ｓｈｒｕｎｋｅｎピアソン相関を使用する。今後は、ｓｈｒｕｎｋｅｎピアソン相関をピアソン相関（Ｐｅａｒｓｏｎ）と表記する。ピアソン相関（Ｐｅａｒｓｏｎ）のパラメータβは、過去の研究報告に倣い、β＝１００に設定した。ハブデータの出現を抑制するための方法として、ベースラインとなる類似度を式（４）によりコミュートタイムカーネルに変換する方法（ＣＴ）、又は、式（５）によりセンタリング変換する方法（C_{ENT}）を用いた。この実験の主な目的は変換の前後における攻撃に対するシステムのロバスト性（耐性）を比較することである。

【００５７】

〔攻撃が無いときの予測精度〕

攻撃に対するロバスト性を調べる前に、攻撃が無いときに、ＣＴ変換及びC_{ENT}変換が評価値の予測精度を劣化させることがないか否かを検証した。推薦業務をシミュレートするために、データセット中の１，０００，２０９個の評価値を、９３９，８０９個の訓練データ（テストデータの予測に使用するデータ）と６０，４００個のテストデータ（予測の対象となるデータ）に分割した。ＣＦアルゴリズムの評価に一般的に使用される平均絶対誤差（ＭＡＥ）を用いて、変換前後の類似度尺度の良し悪しを比較した。

$$MAE = 1/|T| \Sigma_{(u,i)\in T} |Pred(u,i) - R(u,i)|$$

として計算した。ここにＴはテストデータ（|Ｔ|＝６０４００）として与えられたユー

ザーアイテムのペアの組である。

【００５８】

図７に最近傍パラメータｋを１０から１００の間で変動させ、ベースラインとなる類似度尺度であるコサイン類似度（Ｃｏｓ）及びピアソン相関（Ｐｅａｒｓｏｎ）を、コミュートタイムカーネル変換（ＣＴ）あるいはセンタリング変換（C_{ENT}）した場合の、平均絶対誤差（ＭＡＥ）を比較して示す。図７より、C_{ENT}は殆どの場合にＭＡＥを減少（予測精度を増加し）させ、ＣＴはピアソン相関の場合はＭＡＥを減少する。このことから、C_{ENT}変換及びピアソン相関に対するＣＴ変換は、攻撃が無い時の予測精度を悪化させるどころか、改良することが分かる。

以下でアベレジアタックに対するロバスト性を評価するに際し、上記実験で概ねベストとなるＭＡＥを達成するｋ＝５０と設定する。

【００５９】

〔攻撃に対するロバスト性〕

アベレジアタックのターゲットアイテムとして２１個の映画アイテムを選択した。これらのアイテムは、アベレジアタックに関する最初の研究を行ったラム達（Ｓ．Ｋ．Ｌａｍ ａｎｄ Ｊ．Ｒｉｅｄｌ．Ｓｈｉｌｌｉｎｇ．Ｒｅｃｏｍｍｅｎｄｅｒ Ｓｙｓｔｅｍｓ ｆｏｒ Ｆｕｎ ａｎｄ Ｐｒｏｆｉｔ．Ｉｎ Ｐｒｏｃ．ＷＷＷ ’０４，ｐａｇｅｓ ３９３－４０２，２００４年）が実験で用いたアイテムにできるだけ近いもの（評価ユーザ数、平均評価値の観点から）になるように選んだ。アベレジアタックとして、１００の偽ユーザを投入し、偽ユーザはターゲットアイテムには高い評価（すなわち５）を付与し、残り他のアイテムにはノイズを加えた平均的評価を付与するよう作成した。すなわち、残りの各アイテムの各々に対して、μ＝平均評価値、σ＝１．０となる、正規分布（μ；σ）に従う乱数を生成し、もっとも近い整数値１～５に変換して付与した。予測シフトと呼ばれる値、すなわち、攻撃前後の予測評価値の差となる値を用い、変換前後の類似度尺度の良し悪しを比較した。より正確には、訓練用データを除き、誠実なユーザ（偽ユーザを除く全ユーザ）とターゲットアイテムの各ペアに対する予測シフトを計算し、その平均値を比較に用いた。

【００６０】

【表１】

アベレジアタックにより生じた予測シフトと *Nk* 分布の歪度(*k* =50)

類似度尺度	変換	歪度	予測シフト
コサイン類似度	(オリジナル)	3.422	0.902
	CENT	1.629	0.319
	CT	1.526	0.205
ピアソン相関	(オリジナル)	6.389	1.542
	CENT	2.648	1.084
	CT	1.173	0.981

表１はアベレジアタックにより生じた予測シフトと、N_k分布の歪度を示す。大きいN_kを持つデータ、すなわちハブデータが存在するほど、N_k分布の歪度は大きい値となる。N_k分布の歪度は、$S_{Nk}=E〔(N_k-\mu_{Nk})^3/\sigma_{Nk}^3〕$（E〔 〕は期待オペレータ、$\mu_{Nk}$と$\sigma_{Nk}$はそれぞれＮｋ分布の平均と標準分散である）で表される。N_k分布の歪度、予測シフトのどちらも、C_{ENT}又はＣＴ変換後に減少した。このことは、変換された類似度尺度の使用は、ハブデータの出現を抑制し、その結果、推薦システムを攻撃に対してロバストにすることを示している。

【００６１】
図８及び図９は、ユーザ間類似度尺度として一般的なピアソン相関を用いた場合、投入された偽ユーザがハブとなること、及び、類似度尺度の変換によって偽ユーザがハブとなりにくくなることを示す図（その１及びその２）である。図８は、ユーザに関するN_{50}値とデータ中心への類似度との関係を見るために、各々のユーザをプロットした散布図である。横軸はN_{50}、縦軸はデータ中心への類似性を示す。図８より、投入された偽ユーザはデータ中心と高い類似度を持ち、ゆえにハブとなっていることが見て取れる。図９（ａ），（ｂ），（ｃ）は、N_{50}値に係るユーザのヒストグラムである。図９（ａ）はピアソン相関（オリジナル）を、図９（ｂ）はセンタリング変換後のピアソン相関を、図９（ｃ）はコミュートタイムカーネル変換後のピアソン相関を、それぞれ類似度尺度として使用した場合のヒストグラムである。
次に、ハブ現象において、何故、C_{ENT}又はＣＴがロバスト性を提供するかを解析する。

【００６２】
図８の散布図において、最大９６１のN_{50}値を持つハブユーザが存在すること、及び、N_{50}値とデータ中心への類似度との間に強い相関が生じていることが見て取れる。誠実なユーザは○、投入された偽ユーザは×で示されるが、投入された偽ユーザは平均的な誠実なユーザを模倣して作られているため、ユーザに関するデータ中心と高い類似度を持つ。それ故、図９（ａ）のN_{50}分布から分かるように、多くの誠実なユーザと比較して、投入された偽ユーザは大きいN_{50}値（最小４６５、最大９６１）を有するハブユーザ（インフルエンサ）となる。これに対して、C_{ENT}又はＣＴ変換は、ハブ現象の発生を抑制する。結果的に、図９（ｂ）に示すように、C_{ENT}では、偽ユーザのN_{50}値は最小１０１、最大１５６に減少した。また、図９（ｃ）に示すように、ＣＴを用いた場合は、最小０、最大４に減少した。このことは、オリジナルのピアソン相関をユーザ間類似度尺度として使用する場合と比較して、C_{ENT}又はＣＴ変換後の類似度尺度を使用することにより、投入された偽ユーザは、他のユーザのｋＮＮにさほど頻繁に表れなくなったことを明確に示している。つまり、偽ユーザは推薦アイテムの決定にさほど影響しないようになった（もはや投入された偽ユーザはインフルエンサではない）。

【００６３】
以上から、ハブ現象の発生を抑制するように類似度尺度を変換することにより、攻撃に対してロバスト、かつ、オリジナルな類似度尺度と同等又は良い予測精度を示す推薦システムを得られることが分かった。

【００６４】
〔結論〕
外部から偽ユーザを投入することによって推薦されるアイテムの変更を狙う攻撃に対し、ハブ現象を抑制することによって協調フィルタリング（ＣＦ）をロバストにする方法を提案した。
我々のアプローチは、ハブ現象はシリングアタックにより利用される主要因子の１つであるという基盤に立つものである。我々は、ハブデータの出現を抑制する２つの変換（センタリング及びコミュートタイムカーネルへの変換）を、一般的に使用される類似度尺度（コサイン類似度及びピアソン相関）に適用した。ムービーレンズデータセットを用いて、これらの変換が推薦システムを、推薦精度を劣化させることなく、シリングアタックに対してロバストにすることを示した。

【００６５】
以上により、本実施例によれば、偽ユーザを投入する攻撃を受けても、結果として攻撃の影響を受けることが少ない、アイテム推薦システム及びアイテム推薦方法を提供できる。

【実施例２】
【００６６】
実施例１（ユーザベースＣＦ）では、ユーザ間の類似度に基づいて評価値を予測したが

、本態様ではアイテム間の類似度に基づいて評価値を予測する例について説明する。すなわち、実施例２（アイテムベースＣＦ）では、対象アイテムと類似度の高い方からｋ個のアイテムを抽出し、そのｋ個のアイテムに対して対象ユーザが過去に与えた評価値の平均として、対象ユーザの対象アイテムに対する評価値を予測する例について説明する。

【００６７】

実施例１に比して、類似度演算部１３では、ユーザ間の類似度を演算する第１の類似度演算部１３１に代えて、アイテム間の類似度を演算する第２の類似度演算部１３２を使用する。アイテム間の類似度尺度として例えば全てのアイテムがアイテムに関するデータ中心に同等に類似になるものを使用する。近傍データ抽出部１４では、第１の近傍データ抽出部１４１に代えて、第２の類似度演算部１３２にて演算されたアイテム間の類似度を用いて、ターゲットアイテムとの類似度の高い方からｋ個のアイテムを抽出する第２の近傍データ抽出部１４２を使用する。評価値予測部１５では、第１の評価値予測部１５１に代えて、第２の近傍データ抽出部１４２にて抽出されたｋ個のアイテムに係る評価値を用いて、対象ユーザのターゲットアイテム対する未記入のセルに記入すべき評価値を予測する第２の評価値予測部１５２を使用する。その他の構成は実施例１と同様である。

【００６８】

実施例１に比して、類似度演算工程では、ユーザ間の類似度を演算する第１の類似度演算工程（Ｓ１０７）に代えて、アイテム間の類似度を演算する第２の類似度演算工程（Ｓ２０７）を行う。アイテム間の類似度尺度として例えば全てのアイテムがアイテムに関するデータ中心に同等に類似になるものを使用する。近傍データ抽出工程では、第１の近傍データ抽出工程（Ｓ１０８）に代えて、第２の類似度演算工程（Ｓ２０７）にて演算されたアイテム間の類似度を用いて、ターゲットアイテムとの類似度の高い方からｋ個のアイテムを抽出する第２の近傍データ抽出工程（Ｓ２０８）を行う。評価値予測工程では、第１の評価値予測工程（Ｓ１０９）に代えて、第２の近傍データ抽出工程（Ｓ２０７）にて抽出されたｋ個のアイテムに係る評価値を用いて、対象ユーザのターゲットアイテムに対する未記入のセルに記入すべき評価値を予測する第２の評価値予測工程（Ｓ２０９）を行う。その他の処理フローは実施例１と同様である。

このように、対象アイテムと類似度の高い方からｋ個のアイテムを抽出し、ｋ個のアイテムの評価値の（重み付き）平均としてユーザの評価値を予測する。

【００６９】

本実施例によれば、実施例１と同様に、偽ユーザを投入する攻撃を受けても、結果として攻撃の影響を受けることが少ない、アイテム推薦システム及びアイテム推薦方法を提供できる。

【実施例３】

【００７０】

以上の実施例では、偽ユーザを投入する攻撃について説明したが、攻撃が無いときでも、高次元又は大規模データセットには元来ハブデータが存在し易い。この場合でも、ハブの出現を抑制するように類似度尺度を変換すれば、インフルエンサとなるデータの出現を抑制できる。これにより、本来推薦したいアイテムを推薦することが可能になる（ニース達による研究報告、前述の〔攻撃シリングアタックとハブ現象との関係〕に記載、を参照）。

本実施例においては、偽ユーザ投入による攻撃が無いときでも、インフルエンサによる推薦のバイアスを受けない、アイテム推薦システム及びアイテム推薦方法を提供できる。

【実施例４】

【００７１】

以上の実施例では、ユーザ毎に嗜好に合うアイテムを推薦する例について説明したが、大勢の人、大衆に広告を出す場合を想定してみる。もし、平均的なユーザ向けに広告するのが良いと仮定する。この場合、各ユーザ間の類似度を演算する代わりに、平均的な評価値を有する仮のユーザを生成し、当該仮のユーザについて、他のユーザとの類似度を演算して、ｋ人のユーザを抽出し、評価値を予測すれば、大衆を対象としてアイテムを推薦す

ると好適である。
本実施例においても、偽ユーザを投入する攻撃を受けても、結果として攻撃の影響を受けることが少ない、アイテム推薦システム及びアイテム推薦方法を提供できる。
【００７２】
また、本発明は、以上の実施例のフローチャート等に記載のアイテム推薦方法をコンピュータに実行させるためのプログラムとしても実現可能である。プログラムはアイテム推薦システムの記憶部に蓄積して使用してもよく、外付けの記憶装置に蓄積して使用してもよく、インターネットからダウンロードして使用しても良い。また、当該プログラムを記録した記録媒体としても実現可能である。
【００７３】
以上、本発明の実施の形態について説明したが、実施の形態は以上の例に限られるものではなく、本発明の趣旨を逸脱しない範囲で、種々の変更を加え得ることは明白である。
【００７４】
例えば、アイテム及び評価値については、本明細書中に列挙しなかったアイテム及び評価値についても定量的に評価可能であれば本発明を適用できる。また、類似度尺度については、パラメータを用いて調整可能としても良い。アイテム推薦については、アイテム名に添えて画像、説明文を追記可能である。また、推薦は、各ユーザがウェブページにアクセスした時のほか、各ユーザへのメールからアクセス可能にしても良く、メールで配信することも可能である。また、評価マトリックスの寸法、ｋ近傍法のパラメータｋは目的、状況に応じて適宜定めることができる。
【産業上の利用可能性】
【００７５】
本発明はユーザベースあるいはアイテムベースに代表される協調フィルタリングに基づく推薦システムに利用される。
【符号の説明】
【００７６】
１　　アイテム推薦システム
１０　パーソナルコンピュータ（ＰＣ）
１１　登録部
１２　評価部
１３　類似度演算部
１４　近傍データ抽出部
１５　評価値予測部
１６　アイテム推薦部
１７　制御部
１８　表示部
１９　入出力部
２０　記憶部
２１　評価マトリックス記憶部
２２　ハブデータを抑制した類似度尺度記憶部
２３　近傍データ記憶部
２４　推薦アイテム記憶部
１１１　ユーザ登録部
１１２　アイテム登録部
１３１　第１の類似度演算部
１３２　第２の類似度演算部
１３５　類似度尺度変換部
１４１　第１の近傍データ記憶部
１４２　第２の近傍データ記憶部
１５１　第１の評価値予測部

１５２　第２の評価値予測部
２３１　近傍ユーザ記憶部
２３２　近傍アイテム記憶部
ｉ　　アイテム
Ｒ　　評価マトリクス
Ｒ（ｕ，ｉ）　評価値
ｕ　　ユーザ

【図１】

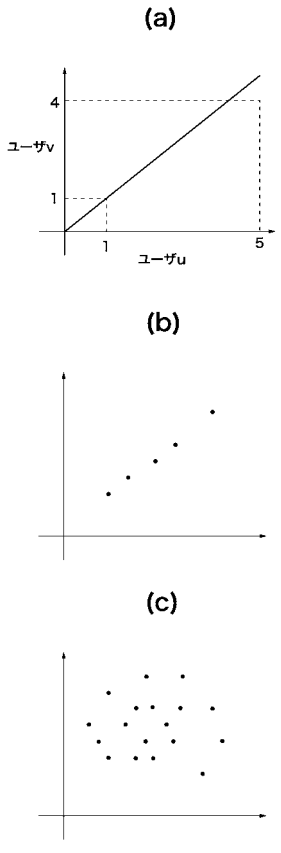
(a)

	アイテム1	アイテム2	アイテム3	アイテム4	……	アイテムm
ユーザ1	5	4		4		
ユーザ2	4			1	1	
ユーザ3						
ユーザ4						
ユーザu	5	1	5	4		
ユーザv	5	1	4	4	1	
ユーザn						

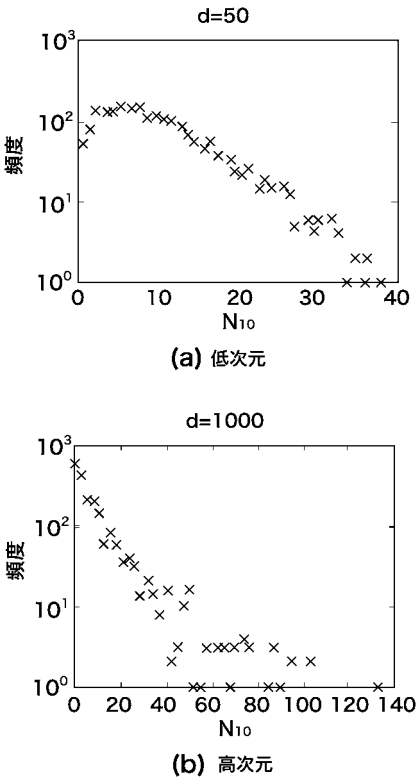
(b)

	アイテム1	アイテム2	アイテム3	アイテム4	……	アイテムm
ユーザ1	5	4		4		
ユーザ2	4			1	1	
ユーザ3						
ユーザ4						
ユーザu	5	1	5	4		
ユーザv	5	1	4	4	1	
ユーザn						
ニセ1	5	1	5	4		
ニセ2	5	1	4	4		
……	……					
ニセ100	5	1	4	4		

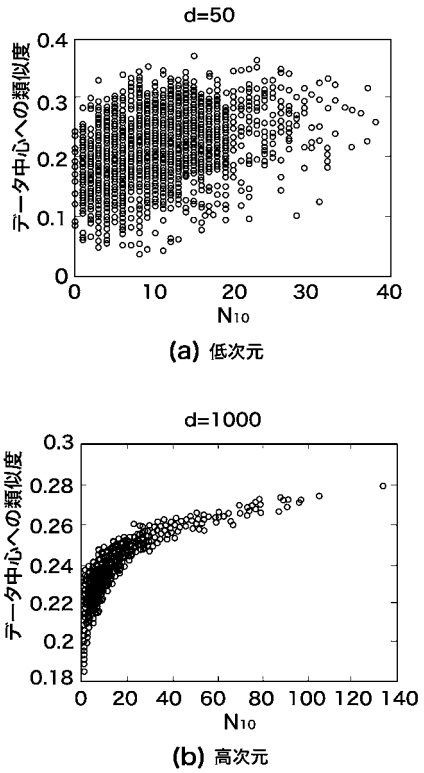
【図２】



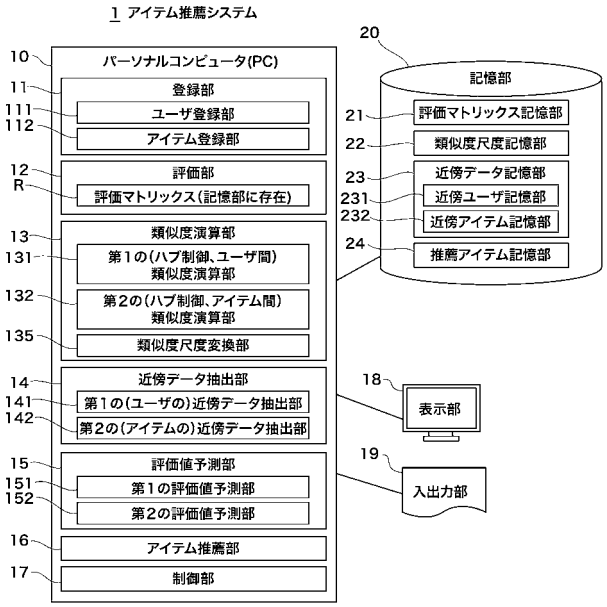
【図３】


d=50
頻度
N10
(a) 低次元
d=1000
頻度
N10
(b) 高次元

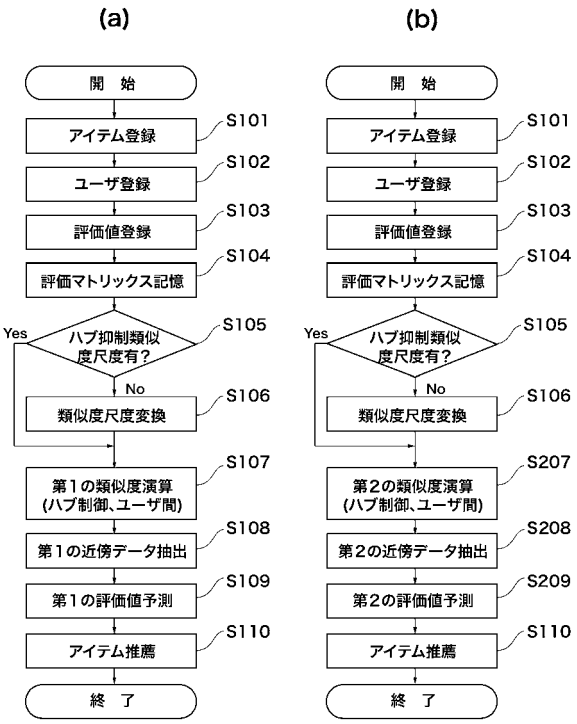
【図４】


d=50
データ中心への類似度
N10
(a) 低次元
d=1000
データ中心への類似度
N10
(b) 高次元

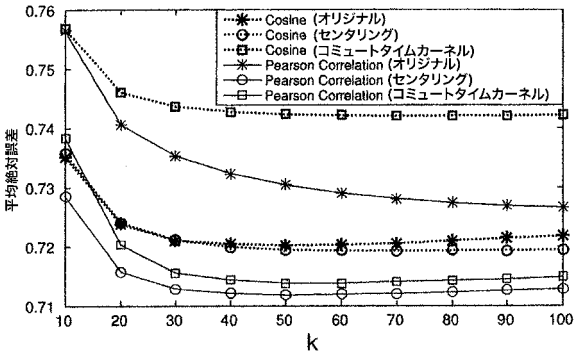
【図５】


1 アイテム推薦システム
10 パーソナルコンピュータ(PC)
11 登録部
111 ユーザ登録部
112 アイテム登録部
12 評価部
R 評価マトリックス(記憶部に存在)
13 類似度演算部
131 第１の(ハブ制御、ユーザ間)類似度演算部
132 第２の(ハブ制御、アイテム間)類似度演算部
135 類似度尺度変換部
14 近傍データ抽出部
141 第１の(ユーザの)近傍データ抽出部
142 第２の(アイテムの)近傍データ抽出部
15 評価値予測部
151 第１の評価値予測部
152 第２の評価値予測部
16 アイテム推薦部
17 制御部
20 記憶部
21 評価マトリックス記憶部
22 類似度尺度記憶部
23 近傍データ記憶部
231 近傍ユーザ記憶部
232 近傍アイテム記憶部
24 推薦アイテム記憶部
18 表示部
19 入出力部

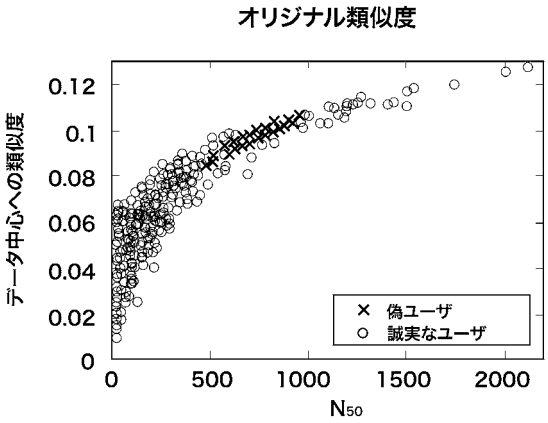
【図６】


(a)
開 始
アイテム登録 S101
ユーザ登録 S102
評価値登録 S103
評価マトリックス記憶 S104
Yes
ハブ抑制類似度尺度有？ S105
No
類似度尺度変換 S106
第１の類似度演算(ハブ制御、ユーザ間) S107
第１の近傍データ抽出 S108
第１の評価値予測 S109
アイテム推薦 S110
終 了
(b)
開 始
アイテム登録 S101
ユーザ登録 S102
評価値登録 S103
評価マトリックス記憶 S104
Yes
ハブ抑制類似度尺度有？ S105
No
類似度尺度変換 S106
第２の類似度演算(ハブ制御、ユーザ間) S207
第２の近傍データ抽出 S208
第２の評価値予測 S209
アイテム推薦 S110
終 了

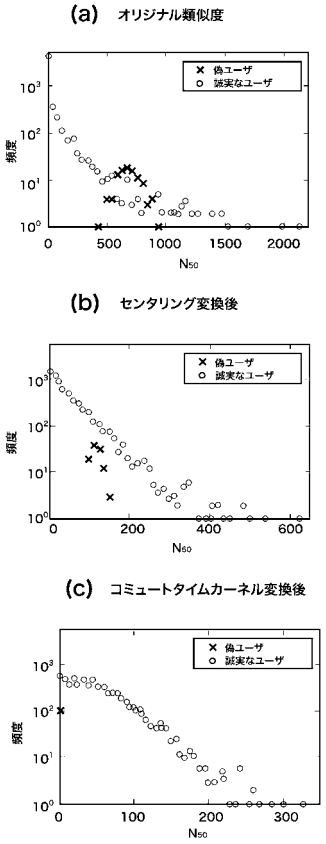
【図７】


Cosine (オリジナル)
Cosine (センタリング)
Cosine (コミュートタイムカーネル)
Pearson Correlation (オリジナル)
Pearson Correlation (センタリング)
Pearson Correlation (コミュートタイムカーネル)
平均絶対誤差
k

【図８】


オリジナル類似度
データ中心への類似度
偽ユーザ
誠実なユーザ
N50

【図９】


(a) オリジナル類似度
(b) センタリング変換後
(c) コミュートタイムカーネル変換後
偽ユーザ
誠実なユーザ
頻度
N50

フロントページの続き

(72)発明者　原　一夫
静岡県三島市谷田１１１１　大学共同利用機関法人情報・システム研究機構　国立遺伝学研究所内
(72)発明者　鈴木　郁美
東京都立川市緑町１０番３号　大学共同利用機関法人情報・システム研究機構　統計数理研究所内