

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第6700146号
(P6700146)

(45) 発行日 令和2年5月27日 (2020.5.27)

(24) 登録日 令和2年5月7日 (2020.5.7)

(51) Int. Cl. F I
G 0 6 Q 30/02 (2012.01)
 G 0 6 Q 30/02 3 8 2
 G 0 6 Q 30/02 3 9 8

請求項の数 8 (全 23 頁)

(21) 出願番号	特願2016-180181 (P2016-180181)	(73) 特許権者	000005108
(22) 出願日	平成28年9月15日 (2016.9.15)		株式会社日立製作所
(65) 公開番号	特開2018-45476 (P2018-45476A)		東京都千代田区丸の内一丁目6番6号
(43) 公開日	平成30年3月22日 (2018.3.22)	(74) 代理人	110001678
審査請求日	平成30年11月27日 (2018.11.27)		特許業務法人藤央特許事務所
		(72) 発明者	グエン フォン
			東京都千代田区丸の内一丁目6番6号 株
			式会社日立製作所内
		(72) 発明者	秋山 高行
			東京都千代田区丸の内一丁目6番6号 株
			式会社日立製作所内
		審査官	梅岡 信幸

最終頁に続く

(54) 【発明の名称】 評価値に基づいて推奨するコンテンツを決定するシステム

(57) 【特許請求の範囲】

【請求項 1】

評価値に基づいて推奨するコンテンツを決定するシステムであって、
 記憶装置と、
 前記記憶装置に格納されているプログラムに従って動作するプロセッサと、を含み、
 前記記憶装置は、複数コンテンツの各コンテンツの複数の特徴量を管理する特徴量管理情報と、前記複数コンテンツの各コンテンツの評価値及び前記複数の特徴量の評価値を管理する評価値管理情報と、を格納し、
 前記プロセッサは、
 前記評価値管理情報が管理する前記複数コンテンツの評価値に基づいて、前記複数コン 10
テンツから、提示する第1コンテンツを選択し、
 提示された前記第1コンテンツに対する視聴者の反応の情報を取得し、
 前記反応の情報と予め定められた基準とに基づいて、前記第1コンテンツに対する報酬
 値を決定し、
 前記報酬値に基づいて、前記評価値管理情報が管理する前記第1コンテンツの評価値及
び前記複数の特徴量の評価値を更新し、
 前記第1コンテンツと前記類似するコンテンツとの間において、前記特徴量管理情報に
 基づいて、同一の値の特徴量を特定し、
 前記評価値管理情報が管理する前記類似するコンテンツの前記同一の値の特徴量の評価
 値を、前記第1コンテンツの前記同一の値の特徴量の評価値に基づいて更新し、前記類似 20

するコンテンツの複数の特徴量の評価値に基づき前記評価値管理情報が管理する前記類似するコンテンツの評価値を更新し、

前記類似するコンテンツの前記同一の値の特徴量の評価値の更新は、次の式に従う、

【数 1】

$$Q'_{feature}(s, a_{similar}) \leftarrow Q(s, a_{similar}) + \alpha [Q_{feature}(s, a) + \gamma \max_{a'} Q'_{feature}(s', a') - Q_{feature}(s, a_{similar})]$$

$Q'_{feature}$: 前記類似するコンテンツの一つの特徴量の評価値

s : 現在の状態

s' : 次の状態

$Q_{feature}$: 前記第 1 コンテンツの前記一つの特徴量の評価値

a : 前記第 1 コンテンツ

$a_{similar}$: 前記類似するコンテンツ

a' : 次の状態でのコンテンツ

α, γ : 予め定められた 0 より大きく 1 以下の数値

$\max_{a'} Q'_{feature}(s', a')$: 次の状態 s' における、前記一つの特徴量の値が同一であるコンテンツのうちの、前記一つの特徴量の最大評価値、システム。

【請求項 2】

評価値に基づいて推奨するコンテンツを決定するシステムであって、

記憶装置と、

前記記憶装置に格納されているプログラムに従って動作するプロセッサと、を含み、

前記記憶装置は、複数コンテンツの評価値を管理する評価値管理情報を格納し、

前記プロセッサは、

前記評価値管理情報が管理する前記複数コンテンツの評価値に基づいて、前記複数コンテンツから提示する第 1 コンテンツを選択し、

提示された前記第 1 コンテンツに対する視聴者の反応の情報を取得し、

前記反応の情報と予め定められた基準とに基づいて、前記第 1 コンテンツに対する報酬値を決定し、

前記報酬値に基づいて、前記評価値管理情報が管理する前記第 1 コンテンツの評価値を更新し、

前記第 1 コンテンツの評価値に基づいて、前記評価値管理情報が管理する前記第 1 コンテンツに類似するコンテンツの評価値を更新し、

予め定められた確率で、提示するコンテンツを、前記評価値に基づいて選択するか、前記評価値から独立して選択するか、決定し、

前記評価値から独立して前記提示するコンテンツを選択する場合に、これまでに提示されたことがないコンテンツであって、前記評価値管理情報において評価値の更新度合が最も少ないコンテンツから、前記提示するコンテンツを選択する、システム。

【請求項 3】

請求項 2 に記載のシステムであって、

前記記憶装置は、前記複数コンテンツの各コンテンツの、複数の特徴量を管理する、特徴量管理情報を格納し、

前記プロセッサは、

前記第 1 コンテンツと前記類似するコンテンツとの間において、前記特徴量管理情報に基づいて、同一の値の特徴量を特定し、

前記第 1 コンテンツの前記同一の値の特徴量の評価値に基づいて、前記類似するコンテンツの評価値を更新する、システム。

【請求項 4】

請求項 3 に記載のシステムであって、

前記評価値管理情報は、前記複数コンテンツの各コンテンツの、前記複数の特徴量それぞれの評価値、を管理し、

10

20

30

40

50

前記プロセッサは、

前記評価値管理情報が管理する、前記類似するコンテンツの、前記同一の値の特徴量の評価値を、前記第1コンテンツの前記同一の値の特徴量の評価値に基づいて更新し、前記類似するコンテンツの評価値を前記類似するコンテンツの前記複数の特徴量の評価値に基づき決定する、システム。

【請求項5】

請求項4に記載のシステムであって、

前記類似するコンテンツの、前記同一の値の特徴量の評価値の更新は、次の式に従う、

【数2】

$$Q'_{feature}(s, a_{similar}) \leftarrow Q'(s, a_{similar}) + \alpha [Q_{feature}(s, a) + \gamma \max_{a'} Q'_{feature}(s', a') - Q'_{feature}(s, a_{similar})]$$

10

$Q'_{feature}$: 前記類似するコンテンツの一つの特徴量の評価値

s : 現在の状態

s' : 次の状態

$Q_{feature}$: 前記第1コンテンツの前記一つの特徴量の評価値

a : 前記第1コンテンツ

$a_{similar}$: 前記類似するコンテンツ

a' : 次の状態でのコンテンツ

α, γ : 予め定められた0より大きく1以下の数値

$\max_{a'} Q'_{feature}(s', a')$: 次の状態 s' における、前記一つの特徴量の値が同一であるコンテンツのうちの、前記一つの特徴量の最大評価値、

20

システム。

【請求項6】

請求項1又は2に記載のシステムであって、

前記評価値管理情報は、視聴者の特徴量の値と前記複数コンテンツそれぞれの評価値との関係を管理し、

前記プロセッサは、

前記提示されたコンテンツの、前記視聴者の特徴量の値に関連付けられている評価値を更新し、

前記類似するコンテンツの、前記視聴者の特徴量の値に関連付けられている評価値を更新する、システム。

30

【請求項7】

計算機システムが、評価値に基づいて推奨するコンテンツを決定する方法であって、

前記計算機システムが、

複数コンテンツの評価値を管理する評価値管理情報が示す前記評価値に基づいて、前記複数コンテンツから、提示する第1コンテンツを選択し、

提示された前記第1コンテンツに対する視聴者の反応の情報を取得し、

前記反応の情報と予め定められた基準とに基づいて、前記第1コンテンツに対する報酬値を決定し、

前記報酬値に基づいて、前記評価値管理情報において、前記第1コンテンツの評価値を更新し、

40

前記第1コンテンツの評価値に基づいて、前記評価値管理情報において、前記第1コンテンツに類似するコンテンツの評価値を更新し、

前記第1コンテンツの評価値に基づいて、前記評価値管理情報において、前記第1コンテンツに類似するコンテンツの評価値を更新し、

予め定められた確率で、提示するコンテンツを、前記評価値に基づいて選択するか、前記評価値から独立して選択するか、決定し、

前記評価値から独立して前記提示するコンテンツを選択する場合に、これまでに提示されたことがないコンテンツであって、前記評価値管理情報において評価値の更新度合が最も少ないコンテンツから、前記提示するコンテンツを選択する、方法。

50

【請求項 8】

計算機システムが、評価値に基づいて推奨するコンテンツを決定する方法であって、
 前記計算機システムが、
 複数コンテンツの各コンテンツの評価値及び複数の特徴量の評価値を管理する評価値管理情報を参照して決定した前記複数コンテンツの評価値に基づいて、前記複数コンテンツから、提示する第 1 コンテンツを選択し、
 提示された前記第 1 コンテンツに対する視聴者の反応の情報を取得し、
 前記反応の情報と予め定められた基準とに基づいて、前記第 1 コンテンツに対する報酬値を決定し、

前記報酬値に基づいて、前記評価値管理情報が管理する前記第 1 コンテンツの評価値及び前記複数の特徴量の評価値を更新し、

前記複数コンテンツの各コンテンツの前記複数の特徴量を管理する特徴量管理情報を参照して、前記第 1 コンテンツと前記類似するコンテンツとの間において、同一の値の特徴量を特定し、

前記評価値管理情報が管理する前記類似するコンテンツの前記同一の値の特徴量の評価値を、前記第 1 コンテンツの前記同一の値の特徴量の評価値に基づいて更新し、前記類似するコンテンツの複数の特徴量の評価値に基づき前記評価値管理情報が管理する前記類似するコンテンツの評価値を更新し、

前記類似するコンテンツの前記同一の値の特徴量の評価値の更新は、次の式に従う、

【数 3】

$$Q'_{feature}(s, a_{similar}) \leftarrow Q'(s, a_{similar}) + \alpha [Q_{feature}(s, a) + \gamma \max_{a'} Q'_{feature}(s', a') - Q'_{feature}(s, a_{similar})]$$

$Q'_{feature}$: 前記類似するコンテンツの一つの特徴量の評価値

s : 現在の状態

s' : 次の状態

$Q_{feature}$: 前記第 1 コンテンツの前記一つの特徴量の評価値

a : 前記第 1 コンテンツ

$a_{similar}$: 前記類似するコンテンツ

a' : 次の状態でのコンテンツ

α : 予め定められた 0 より大きく 1 以下の数値

$\max_{a'} Q'_{feature}(s', a')$: 次の状態 s' における、前記一つの特徴量の値が同一であるコンテンツのうちの、前記一つの特徴量の最大評価値、

方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は評価値に基づいて推奨するコンテンツを決定するシステムに関する。

【背景技術】

【0002】

従来、バナー、サイネージ、又は何らかの情報の推奨コンテンツ（おすすめコンテンツ）は、静的な態様で示されており、視聴者の嗜好に対応するように変化することはない。従来のシステムは、誰が、又はどのようなタイプの人が、コンテンツを視聴しているかについて、情報を有していない。ユーザについての情報を、クッキー、ブラウジングの履歴、又は行動追跡によって集めることができるオンライン環境と異なり、オフライン環境は、そのような情報を有していない。異なる視聴者が異なる嗜好を有することを鑑みれば、静的にコンテンツを全てのタイプの視聴者に表示する従来の方法は、非効率である。

【0003】

繰り返し複数のコンテンツを表示する方法がある。しかし、これらコンテンツは、視聴者の興味に必ずしも沿ったものではなく、また、人手により適宜更新しないと、シーズンの傾向から外れたものになってしまう。

10

20

30

40

50

【 0 0 0 4 】

推奨コンテンツを視聴している視聴者のタイプを認識し、最も適切なコンテンツを選択して視聴者に表示し、さらに、以前に表示されたコンテンツに対する視聴者の反応に基づいて、より良い選択をするように調整できる、自動情報レコメンドシステムが望まれる。自動情報レコメンドシステムは、Q学習を使用して、この目的を達成できる。

【 0 0 0 5 】

Q学習は、モデルフリーの強化学習技術であり、エージェントが、その状態及び行動空間を探索し、所与の状態において実行する最善の行動を見出す。何が最善の行動であるかを判定するために、エージェントは、一般に、一つの行動を試し、取った行動から報酬又はペナルティを経験する。

10

【 0 0 0 6 】

Q値は、(状態、行動)ペアの評価値を示すものであって、多くの繰り返しによって、所与の状態における最善の行動を見つけるために、計算及び更新される。一般のQ学習方法は、「次元の呪い」と呼ばれる問題を有している。これは、状態と行動の数が増加すると、指数的に必要な計算量が増加する問題である。

【 0 0 0 7 】

Q学習を使用した情報レコメンドシステムの設定においては、行動空間は、視聴者に提供しようとしているコンテンツプールで表わすことができる。視聴者に提供するコンテンツの可能なオプションが増加するにつれて、より多くの状態と行動の組み合わせを探索することが必要となり、結果的に、長い学習時間が必要となる。

20

【 0 0 0 8 】

視聴者に表示するために入手可能なコンテンツの数は膨大である。視聴者に全ての可能なオプションを示すには、実行不能な量の学習時間がかかり得る。学習時、システムは、まだ適切な行動を行わない。従って、長い学習時間により、多くの利益を失い、学習のために多くのコストがかかる。

【 0 0 0 9 】

行動空間を階層的にクラスタリングする技術が知られている。各行動は、一意のコンテンツを選択し、同一クラスタ内の全ての行動を同様に扱って、行動空間における行動の数を低減する。このような技術は、例えば、特開 2 0 0 7 - 1 6 4 4 0 6 号公報に開示されている。

30

【 先行技術文献 】

【 特許文献 】

【 0 0 1 0 】

【 特許文献 1 】 特開 2 0 0 7 - 1 6 4 4 0 6 号公報

【 発明の概要 】

【 発明が解決しようとする課題 】

【 0 0 1 1 】

しかし、上記アプローチは、同一クラスタ内の行動が異なり、異なるクラスタの行動が類似している場合に、正確ではない。このように、従来の技術は、上記のような孤立した大きな行動空間の課題を解決する具体的な方法を有していない。従って、短い学習時間において、視聴者に対してより適切なコンテンツを選択して提示できる、システムが望まれる。

40

【 課題を解決するための手段 】

【 0 0 1 2 】

本発明の一態様は、評価値に基づいて推奨するコンテンツを決定するシステムであって、記憶装置と、前記記憶装置に格納されているプログラムに従って動作するプロセッサと、を含み、前記記憶装置は、複数コンテンツの評価値を管理する評価値管理情報を格納し、前記プロセッサは、前記評価値管理情報が示す前記評価値に基づいて、前記複数コンテンツから、提示する第 1 コンテンツを選択し、提示された前記第 1 コンテンツに対する視聴者の反応の情報を取得し、前記反応の情報と予め定められた基準とに基づいて、前記第

50

1 コンテンツに対する報酬値を決定し、前記報酬値に基づいて、前記評価値管理情報において、前記第 1 コンテンツの評価値を更新し、前記第 1 コンテンツの評価値に基づいて、前記評価値管理情報において、前記第 1 コンテンツに類似するコンテンツの評価値を更新する。

【発明の効果】

【0013】

本発明の一態様によれば、短い学習時間において、視聴者に対してより適切なコンテンツを選択して提示できる。

【図面の簡単な説明】

【0014】

10

【図 1】実施例 1 に係る情報レコメンドシステムの構成例を示す。

【図 2】実施例 1 に係る情報レコメンドシステムの機能ブロック図を示す。

【図 3】実施例 1 に係る行動データベースの構成例を示す。

【図 4】実施例 1 に係る行動特徴量データベースの構成例を示す。

【図 5】実施例 1 に係る Q 値データベースの構成例を示す。

【図 6】実施例 1 に係る行動選択部のフローチャートを示す。

【図 7】実施例 1 に係る行動類似性計算部のフローチャートを示す。

【図 8】実施例 1 に係る Q 値更新部のフローチャートを示す。

【図 9】実施例 2 に係る、情報レコメンドシステムの構成例を示す。

【図 10】実施例 2 に係る情報レコメンドシステムの機能ブロック図を示す。

20

【図 11】実施例 2 に係る Q 値データベースの構成例を示す。

【図 12】実施例 3 に係る情報レコメンドシステムの構成例を示す。

【図 13】実施例 3 に係る情報レコメンドシステムの機能ブロック図を示す。

【図 14】実施例 3 に係る行動選択部のフローチャートを示す。

【図 15】実施例 3 に係る情報ゲイン計算部のフローチャートを示す。

【発明を実施するための形態】

【0015】

以下、添付図面を参照して本発明の実施形態を説明する。本実施形態は本発明を実現するための一例に過ぎず、本発明の技術的範囲を限定するものではないことに注意すべきである。各図において共通の構成については同一の参照符号が付されている。

30

【0016】

以下に開示される情報レコメンドシステムは、提示されたコンテンツに対する視聴者の反応に基づいて、視聴者により適切なコンテンツを提示する。以下に開示する情報レコメンドシステムは、例えば、空港、ショッピングモール、駅のような、オフライン環境において動作する。提示されるコンテンツは、バナー、サイネージ、音楽、映像等である。情報レコメンドシステムは、提示するコンテンツそれぞれの評価値（各コンテンツを提示する場合の評価値）を管理し、評価値に基づいて提示するコンテンツを選択する。

【0017】

情報レコメンドシステムは、提示されたコンテンツに対する視聴者の反応に基づいて、提示したコンテンツの評価値を更新すると共に、提示したコンテンツに類似するコンテンツの評価値を更新する（学習）。これにより、提示したコンテンツに加え、実際に提示をしていないコンテンツを提示した場合の効果を、効果的に学習することができ、短い学習時間で、より適切なコンテンツを提示することができる。

40

【0018】

一例において、情報レコメンドシステムは、コンテンツの複数の特徴量を管理し、特徴量に基づいてコンテンツ間の類似性を判定する。たとえば、情報レコメンドシステムは、値が一致する特徴量が規定数以上のコンテンツを、互いに類似するコンテンツと判定する。

【0019】

例えば、情報レコメンドシステムは、提示したコンテンツにおいて、類似するコンテン

50

ツと同一値の特徴量を特定し、当該特徴量に対する評価値に基づいて、類似するコンテンツの評価値を更新する。これにより、提示したコンテンツと類似するコンテンツとの間の類似度に応じて、類似するコンテンツの評価値を更新できる。

【実施例１】

【００２０】

図１～８を参照して、実施例１を説明する。情報レコメンドシステムは、サーバ１００、表示端末２００、そして観測端末３００を含む。表示端末２００は、推奨コンテンツ（お勧めコンテンツ）を視聴者に表示する。観測端末３００は、表示されたコンテンツに対する視聴者の反応を認識し、サーバ１００にその情報を送信する。

【００２１】

サーバ１００は、その情報を使用して報酬を計算し、Ｑ値データベース１３３を更新する。Ｑ値データベース１３３は、行動（表示コンテンツ）と状態のペアのＱ値管理する。Ｑ値は、評価値である。行動選択プログラム１２１は、Ｑ学習アルゴリズムに従って次に表示するコンテンツを決定し、そのコンテンツを表示端末２００に送信する。

【００２２】

図１は、実施例１に係る情報レコメンドシステムの構成例を示す。表示端末２００を、まず説明する。表示端末２００は、プロセッサ２１０、メモリデバイス２２０、表示デバイス２３０、通信インタフェース２４０を含む。各構成要素はバスによって互いに接続されている。デジタルサイネージシステムは、表示端末２００の一例である。しかし、表示端末２００は、デジタルサイネージシステムに限定されず、サーバ１００からの要求に応じて、特定のコンテンツを視聴者に対して提示できればよい。

【００２３】

プロセッサ２１０は、メモリデバイス２２０に格納されたプログラムに従って、様々な演算処理を実行する。メモリデバイス２２０は、コンテンツ表示プログラム２２１を格納する。コンテンツ表示プログラム２２１は、特定のコンテンツを表示デバイス２３０において表示させる。表示デバイス２３０は、表示パネルに加え、スピーカ及び／又はタッチパネルを含んでもよく、表示デバイス２３０に代えてスピーカが実装されてもよい。

【００２４】

通信インタフェース２４０は表示端末２００をネットワーク１５０に接続する。コンテンツ表示プログラム２２１は、メモリデバイス２２０に格納され、ネットワーク１５０及び通信インタフェース２４０を介して、サーバ１００から、選択されたコンテンツを受信する。

【００２５】

次に、観測端末３００を説明する。観測端末３００は、プロセッサ３１０、メモリデバイス３２０、カメラ３３０、通信インタフェース３４０を含む。各構成要素はバスによって互いに接続されている。観測端末３００の一例は、高機能カメラシステムであり、画像処理技術を記録された視聴者の映像に適用して、観測結果から情報を抽出する。

【００２６】

観測端末３００は、高機能カメラシステムに限定されず、表示されたコンテンツに対する視聴者の反応を捉えることができればよい。例えば、観測端末３００は、表示端末２００に対する視聴者の入力（例えばタッチパネルでの選択）、視聴者が保持するスマートフォンから得られた情報、等の情報を使用できる。

【００２７】

プロセッサ３１０は、メモリデバイス３２０に格納されたプログラムに従って、様々な演算処理を実行する。メモリデバイス３２０は記憶装置であって、反応認識プログラム３２１を格納する。反応認識プログラム３２１は、カメラ３３０により補足された画像データを介して、公知の画像処理技術により、視聴者の反応を抽出する。反応の情報は、サーバ１００に、通信インタフェース３４０を介して送信される。通信インタフェース３４０は、観測端末３００をネットワーク１５０に接続する。

【００２８】

10

20

30

40

50

次に、サーバ１００を説明する。サーバ１００は、プロセッサ１１０、メモリデバイス１２０、補助記憶デバイス１３０、通信インタフェース１４０を含む。構成要素はバスで接続されている。通信インタフェース１４０は、サーバ１００をネットワーク１５０に接続する。メモリデバイス１２０、補助記憶デバイス１３０又はこれらの組み合わせは、記憶装置である。

【００２９】

プロセッサ１１０は、メモリデバイス１２０が格納するプログラムに従って、様々な演算処理を実行する。メモリデバイス１２０は、行動選択プログラム１２１、Ｑ値更新プログラム１２２、行動類似性計算プログラム１２３、報酬計算プログラム１２４、を格納する。一意のコンテンツの選択は、一意の行動と考えられる。従って、コンテンツの選択は、行動の選択と言える。

10

【００３０】

行動選択プログラム１２１は、パラメータを有する。行動選択プログラム１２１は、表示端末２００に送信して視聴者に表示するための行動（コンテンツ）を選択する二つのストラテジを有する。行動選択プログラム１２１は、パラメータの値（０～１）に基づいて、ストラテジを選択する。一方のストラテジは、Ｑ値と独立に、行動データベースから行動を選択する。本実施例はランダムに行動を選択する。もう一方のストラテジは、Ｑ値データベース１３３において最も高いＱ値の行動を選択する。

【００３１】

Ｑ値更新プログラム１２２は、Ｑ値データベース１３３の更新において、Ｑ値計算式と、行動類似性計算プログラム１２３の出力と、を使用する。行動類似性計算プログラム１２３は、選択された行動（コンテンツ）と他の残りの行動（コンテンツ）との間の類似性を、行動特徴量データベース１３２に基づいて比較する。報酬計算プログラム１２４は、観測端末３００から受信した、表示されたコンテンツに対する視聴者の反応に関する情報に基づいて、報酬値を計算する。

20

【００３２】

補助記憶デバイス１３０は、例えば不揮発性記憶媒体を含み、行動データベース１３１、行動特徴量データベース１３２、Ｑ値データベース１３３を格納する。行動データベース１３１は、視聴者に推奨するために使用される（表示される）コンテンツを集めたものである。そのデータサイズは、通常大きなものであり、多くのコンテンツが含まれる。行動データベース１３１の構成例は、後に、図３を参照して詳述される。

30

【００３３】

行動特徴量データベース１３２は、各コンテンツの特徴量情報を格納しており、特徴量の組み合わせは、一意のコンテンツを示す。行動特徴量データベース１３２の構成例は、後に、図４を参照して詳述される。

【００３４】

Ｑ値データベース１３３は、状態及び行動のペアと関連付けられる、総計Ｑ値と部分Ｑ値とを格納する。定義される状態の例は、「コンテンツ表示の開始」（０）と「コンテンツ表示の終了」（１）、である。「コンテンツ表示の開始」（０）と「コンテンツ表示の終了」（１）は、「コンテンツ表示の開始」及び「コンテンツ表示の終了」の状態が、０及び１で表わされることを示す。

40

【００３５】

一つの行動は、行動データベース１３１からの一意のコンテンツの選択又は表示であり、一つの行動は一つのコンテンツに対応する。総計Ｑ値は、状態と行動のペアの評価値であり、つまり、ある状態で表示するコンテンツの評価値である。総計Ｑ値が高い程、そのコンテンツを視聴者に対して表示することがより適切であることを意味する。

【００３６】

部分Ｑ値は、コンテンツの特徴量に対して与えられる。部分Ｑ値と総計Ｑ値との関係は予め定義されている。なお、部分Ｑ値は使用しなくてもよく、部分Ｑ値の定義によっては、Ｑ値データベース１３３に格納されていなくてもよい。部分Ｑ値、部分Ｑ値と総計Ｑ値

50

との関係の詳細は後述される。Q 値データベース 133 の構成例は、後に、図 5 を参照して詳述される。

【0037】

図 2 は、実施例 1 に係る情報レコメンタシステムの機能ブロック図を示す。例えば、観測端末 300 の不図示のプログラムが、画像データから、表示端末 200 に近づいている潜在的視聴者を検知すると、それを行動選択部 1210 に通知する。行動選択部 1210 は、状態が、「コンテンツ表示の終了」(1) から「コンテンツ表示の開始」(0) に変化したと判定する。

【0038】

行動選択部 1210 は、ランダムに、又は、Q 値データベース 133 における最高 Q 値に基づいて、行動データベース 131 からコンテンツを選択する。行動選択部 1210 は、行動選択プログラム 121 を実行するプロセッサ 110 により実現される。コンテンツデータは、コンテンツ表示部 2210 に送信される。行動選択部 1210 の処理フローは、後に、図 6 を参照して詳述される。

【0039】

コンテンツ表示部 2210 は、行動選択部 1210 からコンテンツデータを受信し、画面において視聴者にコンテンツを可視化する。コンテンツ表示部 2210 は、コンテンツ表示プログラム 221 を実行するプロセッサ 210 により実現される。

【0040】

コンテンツが視聴者に表示された後、視聴者が表示されたコンテンツに対して反応する。視聴者は、推奨されているコンテンツを気に入り、推奨されているコンテンツに従って行動を起こすことを決定するかもしれない。または、視聴者は、推奨されているコンテンツを気に入らず、推奨されているコンテンツに対して何も行動を起こさないかもしれない。

【0041】

例えば、コメディのコンテンツが視聴者に表示された場合、視聴者が気に入れば、視聴者は微笑むであろうし、そうでなければ、微笑まないであろう。画面に表示されたポストによって視聴者に行先が推奨される場合、視聴者が気に入れば、推奨された場所に行くであろうし、そうでないならば、推奨された場所に行かないであろう。

【0042】

反応認識部 3110 は、そのような視聴者からの反応を認識することができる。例えば、反応認識部 3110 は、視聴者の画像データに対して公知の画像処理を行うことで、視聴者の動作を特定し、当該動作から視聴者のコンテンツへの興味の有無を判定する。または、反応認識部 3110 は、視聴者のスマートフォンから取得した情報から、視聴者が訪れた場所の情報を取得する。

【0043】

反応認識部 3110 は、反応認識プログラム 321 を実行するプロセッサ 310 により実現される。反応認識部 3110 が取得する情報は、例えば、コンテンツを視聴している視聴者の数、そのコンテンツを気に入った視聴者の数、そのコンテンツが気に入らない視聴者の数、等である。なお、これらに限定はされない。反応認識部 3110 が取得する情報は、報酬計算部 1240 に送信される。

【0044】

報酬計算部 1240 は、反応認識部 3210 から、表示されたコンテンツを見た視聴者の反応についての情報データを受信する。報酬計算部 1240 は、受信したデータを変換して、表示されたコンテンツがどれほど有効であるかを決定するための、報酬と呼ばれる一つのメトリックを生成する。

【0045】

変換の一例は、視聴者の総数に対する、ポジティブな反応を示した者の比である。報酬 r は、0 から 1 の範囲の値を有する。なお、これは一例であって、報酬 r の決定方法はこれに限定されるものではない。また、報酬計算部 1240 が、反応認識部 3210 からデ

10

20

30

40

50

ータを受信することは、システムが視聴者から反応を得たことを示し、システムは、状態を「コンテンツ表示の開始」(0)から「コンテンツ表示の終了」(1)に変更する。

【0046】

行動類似性計算部1230は、表示されたコンテンツと、行動特徴量データベース132における残りのコンテンツとを、比較する。行動類似性計算部1230は、プロセッサ110が行動類似性計算プログラム123を実行することで実現される。各コンテンツにおける特徴量の値を比較することによって、行動類似性計算部1230は、表示されたコンテンツに類似するコンテンツを見出すことができる。

【0047】

例えば、行動類似性計算部1230は、規定数(例えば一つ)以上の特徴量の値が同一である場合に、二つのコンテンツが類似していると判定する。これにより、簡便かつ適切に類似コンテンツを見る蹴ることができる。行動類似性計算部1230は、行動特徴量データベース132を参照して、表示されたコンテンツに対して、各類似コンテンツのどの特徴量が同一であるか、特定することができる。行動類似性計算部1230の処理フローは、後に、図7を参照して詳述される。

10

【0048】

Q値更新部1220は、Q値データベース133における総計Q値及び部分Q値を更新する。Q値更新部1220は、Q値更新プログラム122を実行するプロセッサ110により実現される。視聴者に対して表示されたコンテンツの総計Q値は、次の式に従って更新できる。

20

【0049】

【数1】

$$Q_{total}(s,a) \leftarrow Q_{total}(s,a) + \alpha[r + \gamma \max_{a'} Q_{total}(s',a') - Q_{total}(s,a)] \quad (1)$$

【0050】

ここで、 $Q_{total}(s, a)$ は、表示されたコンテンツの総計Q値である。 s は、現在の状態であり、視聴者にコンテンツが表示されている場合、その状態は、「コンテンツ表示の開始」(0)である。 a は現在の行動(コンテンツ)である。 α は、学習レートである。0から1の値を取り、システムが、どの程度、最新の情報から学習すべきかを示す。この値は、システム設計者によって、システムに応じて予め設定される。一般に、最初の行動において1に設定され、その後、0.5が選択される。

30

【0051】

r は表示されたコンテンツに対する視聴者の反応から計算される報酬である。 γ は、ディスクカウントファクタである。0から1の値を取り、より先の未来における報酬よりも、より近い未来における報酬に対してより多くの重みを与える。この値は、システム設計者によって、システムに応じて予め設定される。一般に、1に設定される。

【0052】

s' は次の状態であり、 a' は、次の状態における行動(次の行動)である。 $\max_{a'} Q_{total}(s', a')$ は、次の状態 s' において最も高い総計Q値のコンテンツ(行動) a' を選択した場合の総計Q値である。「コンテンツ表示の開始」(0)と「コンテンツ表示の終了」(1)が定義されている例において、「コンテンツ表示の開始」(0)の次の状態は「コンテンツ表示の終了」(1)である。

40

【0053】

以下に記載する例において、状態「コンテンツ表示の終了」(1)における全ての総計Q値は0である。つまり、 $\max_{a'} Q_{total}(s', a')$ は0である。例えば、状態「コンテンツ表示の終了」(1)に対して、数式(1)は適用されない、又は、 r が0である。状態「コンテンツ表示の終了」(1)において視聴者が存在しないため、報酬 r は0である。

【0054】

上述のように、視聴者は、コンテンツ全体に対して反応を示し、その反応に対する報酬

50

rに基づいて、総計Q値が更新される。表示するコンテンツは、総計Q値に基づいて選択される。コンテンツが初めて表示されるとき、表示されたコンテンツに類似するコンテンツとして、更新前の総計Q値が初期値から更新されていることがある。この場合、更新前の総計Q値が維持される、又は、初期値にリセットされてもよい。類似コンテンツとしてのQ値を維持することで、これまでの学習を利用することができる。リセットすることで、当該コンテンツに対する実際の視聴者の反応に応じたQ値を得ることができる。

【0055】

次に、一つの特徴に対して与える部分Q値は、次の式で決定される。

【0056】

【数2】

$$Q_{feature}(s, a) \leftarrow \frac{Q_{total}}{k} \quad (2)$$

【0057】

ここで、kは、各行動（コンテンツ）における特徴量の数である。視聴者は、コンテンツ全体に対して反応を示すため、総計Q値が特徴量それぞれに等分に割り当てられる。異なる特徴量に対して異なる部分Q値が割り当てられてもよい。表示されたコンテンツの総計Q値から部分Q値を決定することで、適切に部分Q値を決定できる。

【0058】

次に、表示されたコンテンツに類似するコンテンツの総計Q値と部分Q値の更新で使用する数式を説明する。類似コンテンツの部分Q値は、次の式に従って更新される。更新される部分Q値は、表示されたコンテンツと類似するコンテンツとの間において類似する特徴量の部分Q値である。本例においては、表示コンテンツと類似コンテンツとにおいて、同一の値を有する特徴量の部分Q値が更新される。

【0059】

【数3】

$$Q_{feature}(s, a_{similar}) \leftarrow Q(s, a_{similar}) + \alpha [Q_{feature}(s, a) + \gamma \max_{a'} Q_{feature}(s', a') - Q_{feature}(s, a_{similar})] \quad (3)$$

【0060】

$Q'_{feature}(s, a_{similar})$ は、類似コンテンツの一つの特徴量に対する部分Q値である。 $a_{similar}$ は、表示されたコンテンツに類似するコンテンツ（行動）である。表示コンテンツの部分Q値（ $Q_{feature}$ ）の値は、視聴者の反応に応じた更新後の値である。

【0061】

$\max_{a'} Q'_{feature}(s', a')$ は、次の状態 s' において、当該特徴量の値が同一であるコンテンツ（類似するコンテンツ）（行動）の内、最も高い部分Q値（最も高い部分Q値のコンテンツ（行動）を選択した場合の部分Q値）である。上述のように、以下に記載する例において、この値は0である。

【0062】

類似コンテンツの総計Q値は、次の数式に従って更新される。

【0063】

【数4】

$$Q_{total}(s, a_{similar}) \leftarrow \sum_{i=1}^k Q_{feature_i}(s, a_{similar}) \quad (4)$$

【0064】

$Q_{feature_i}$ は、特徴量iの部分Q値である。数式（4）において、各部分Q値に重み係数が与えられてもよい。数式（3）が示すように、類似するコンテンツの特徴量の部分Q値は、表示コンテンツの同一特徴量の部分Q値に基づき更新される。数式（4）が示すように、類似するコンテンツの総計Q値は、更新された部分Q値に基づき更新される。特徴量単位で類似コンテンツのQ値を更新することで、表示コンテンツと類似コン

10

20

30

40

50

テンツとらの間の類似度に応じて類似コンテンツの総計Q値を更新することができる。

【0065】

Q値更新部1220は、他の方法によって、類似するコンテンツの総計Q値を更新してもよい。例えば、Q値更新部1220は、類似するコンテンツにおける、表示されたコンテンツと同一の特徴量の数（割合）と、表示されたコンテンツの更新された総計Q値に基づき、類似するコンテンツの総計Q値を更新してもよい。

【0066】

Q値を更新した後、状態が「コンテンツ表示の開始」（1）に切り替わると、行動選択部1210が呼び出される。Q値更新部1220の処理フローは、後に、図8を参照して詳述する。

10

【0067】

図3は、実施例1に係る行動データベース131の構成例を示す。行動データベース131は、行動IDカラム301、コンテンツ名カラム302、表示コンテンツカラム303を有する。

【0068】

行動IDカラム301は、表示されるコンテンツを一意に同定する整数を示す。コンテンツ名カラム302は、表示されるコンテンツの名前を示す。例えば、東京タワーや、明神宮と言った名称である。表示コンテンツカラム303は、例えば、視聴者に表示される実際の画像を格納する、又は、画像ファイルの格納位置（パス）を示す。これにより、コンテンツ表示プログラム221は、関連するコンテンツを認識し、表示できる。

20

【0069】

図4は、実施例1に係る行動特徴量データベース132の構成例を示す。行動特徴量データベース132は、行動IDカラム401、特徴量1カラム402、特徴量2カラム403、特徴量3カラム404を有する。

【0070】

行動IDカラム401は、行動データベース131の行動IDカラム301の値と関連付けられている、整数を示す。特徴量1カラム402は、例えば、表示されるコンテンツの画像についての特徴量であり、例えば、画像における対象の種類を示す。特徴量2カラム403は、例えば、表示されるコンテンツにおけるテキストを示す。特徴量3カラム404は、例えば、表示されるコンテンツにおける主題の時間を示す。

30

【0071】

特徴量の値は、例えば、数値コード又は文字列で示すことができる。コンテンツが同一の値を有するか比較できれば、どのようなタイプの値でもよい。コンテンツを正確に記述できれば、特徴量の数や種類は限定されない。

【0072】

図5は、実施例1に係るQ値データベース133の構成例を示す。Q値データベース133は、状態IDカラム501、行動IDカラム502、部分Q1カラム503、部分Q2カラム504、部分Q3カラム505、総計Qカラム506、を有する。

【0073】

状態IDカラム501は、システム状態を示す。「コンテンツ表示の開始」（0）と、「コンテンツ表示の終了」（1）が定義されている。状態の数及び種類は、設計や表示するコンテンツに依存する。例えば、レストランの日替わりメニューが表示コンテンツである場合、状態は曜日を示すことができる。行動IDカラム502は、行動データベース131における行動ID301と関連付けられている整数を示す。

40

【0074】

部分Q1カラム503は、行動特徴量データベース132における特徴量1と関連付けられている部分Q値を示す。部分Q2カラム504は、行動特徴量データベース132における特徴量2と関連付けられている部分Q値を示す。部分Q3カラム505は、行動特徴量データベース132における特徴量3と関連付けられている部分Q値を示す。総計Qカラム506は、関連付けられている状態及び行動の総計Q値を示す。システム動作開始

50

時（学習開始時）、いずれの行動も選択されておらず、全ての総計Q値と全ての部分Q値は、0に初期化されている。

【0075】

なお、Q値データベース133は、各レコード（状態IDと行動IDのペア）が選択済み（表示済み）が示す情報を保持してもよい。例えば、Q値データベース133は、各レコードが選択済みフラグを格納するカラムをさらに有してもよい。

【0076】

図6は、実施例1に係る行動選択部1210のフローチャートを示す。状態の変化に応じて、行動選択処理が開始される。行動選択処理は、Q学習モデルに従い、サーバ100のプロセッサ110により実行される。まず、プロセッサ110は、0から1の範囲においてランダム数を生成する（S601）。

10

【0077】

プロセッサ110は、ランダム数とと比較する（S602）。は、システム設計者により、システムに応じて予め設定されたパラメータである。は、システムが、最善行動を利用する（Exploit）よりも、いくつかの行動を探索する（Explore）割合を示す。例えば、は0.5に設定され、50%の割合（確率）で探索がなされ、50%の割合（確率）で最善行動が利用される。

【0078】

ステップS602の処理により、ランダム数がより大きいと判定されると（S602：YES）、プロセッサ110は、Q値データベース133において、システムの状態と同一の状態において最も高い総計Q値を、検索する。最も高い総計Q値に関連付けられている行動IDが、選択される（S604）。複数の行動IDが最も高い総計Q値に関連付けられている場合は、例えば、その中からランダムに一つが選択される。

20

【0079】

一方、ステップS602の処理により、ランダム数がより大きくないと判定されると（S602：NO）、プロセッサ110は、ランダムに、行動IDを行動データベース131から選択する（S603）。

【0080】

次に、ステップS603又はS604において行動IDを選択した後、プロセッサ110は、行動データベース131から、表示するコンテンツを、選択した行動IDに基づいて読み出す（S605）。最後に、プロセッサ110は、表示端末200に、表示するコンテンツのデータを送信して、その表示を要求する（S606）。以上により、行動選択処理が終了する。上記説明から理解されるように、「コンテンツ表示の終了」（1）の状態においては、ランダムにコンテンツが選択される。

30

【0081】

図7は、実施例1に係る行動類似性計算部1230のフローチャートを示す。行動類似性計算処理は、サーバ100のプロセッサ110により実行される。まず、行動選択部1210により選択された行動ID（そのコンテンツは表示端末200により表示されている）は、基準行動として設定される（S701）。この後のステップにおける全ての比較は、この基準行動に対して行われる。

40

【0082】

次に、プロセッサ110は、行動特徴量データベース132における各行動をチェックするループを形成する（S702）。ただし、基準行動は除かれる。行動を検査するため、プロセッサ110は、検査されている行動に属する特徴量それぞれをチェックするもう一つのループを形成する（S703）。

【0083】

プロセッサ110は、検査されている行動の検査されている特徴量の値が、基準行動の特徴量の値と同一であるか判定する（S704）。例えば、基準行動が行動ID「1」を有し、検査されている行動が行動ID「2」を有するとする。行動ID「1」の特徴量1の値が、行動ID「2」の特徴量1の値と比較され、それらが同一であるか判定される。

50

例えば、双方の特徴量 1 が同一の「タワー」（図 4 の例を参照）を有する場合、それらは同一の値を有していると判定される。

【 0 0 8 4 】

検査対象の行動と基準行動とが、特徴量の同一の値を有していると、ステップ S 7 0 4 で判定されると（S 7 0 4 : Y E S）、プロセッサ 1 1 0 は、検査対象の行動が、基準行動に類似していると判定し（S 7 0 5）、同一値の特徴量を類似行動のリストに記録する（メモリデバイス 1 2 0 に格納する）（S 7 0 6）。

【 0 0 8 5 】

検査対象行動と基準行動の特徴量が同一値を有していないと、ステップ S 7 0 4 で判定されると（S 7 0 4 : N O）、プロセッサ 1 1 0 は、何も記録せず、検査対象行動の、全ての特徴量を検査したか判定する（S 7 0 7）。検査対象行動の全ての特徴量の検査が終了していないと、ステップ S 7 0 7 で判定されると（S 7 0 7 : N O）、プロセッサ 1 1 0 は、次の特徴量のために、ステップ S 7 0 3 に戻る。

【 0 0 8 6 】

検査対象行動の全ての特徴量の検査が終了していると、ステップ S 7 0 7 で判定されると（S 7 0 7 : Y E S）、プロセッサ 1 1 0 は、全ての行動の検査を終了したか判定する（S 7 0 8）ステップ S 7 0 8 において、全ての行動の検査は終了していないと判定されると（S 7 0 8 : N O）、プロセッサ 1 1 0 は、次の行動のために、ステップ S 7 0 2 に戻る。ステップ S 7 0 8 において、全ての行動の検査が終了していると判定されると（S 7 0 8 : Y E S）、プロセッサ 1 1 0 は、行動類似性計算処理を終了する。

【 0 0 8 7 】

上記処理においては、表示されているコンテンツと少なくとも一つの特徴量の値が同一であるコンテンツが、類似コンテンツと判定される。2 以上の特徴量が同一値であることが類似と判定される条件でもよく、特徴量に重みがつけられており、重みの総計に基づいて類似が判定されてもよい。

【 0 0 8 8 】

図 8 は、実施例 1 に係る Q 値更新部 1 2 2 0 のフローチャートを示す。Q 値更新処理は、サーバ 1 0 0 のプロセッサ 1 1 0 により実行される。プロセッサ 1 1 0 は、まず、選択されている行動（コンテンツ）の総計 Q 値を、数式（1）に従って計算し、Q 値データベース 1 3 3 を更新する（S 8 0 1）。選択されている行動のコンテンツが表示端末 2 0 0 において表示され、観測端末 3 0 0 から応答が得られている。次に、プロセッサ 1 1 0 は、選択されている行動の部分 Q 値を、数式（2）に従って計算し、Q 値データベース 1 3 3 を更新する（S 8 0 2）。

【 0 0 8 9 】

次に、プロセッサ 1 1 0 は、行動類似性計算部 1 2 3 0 により類似する行動が検出されているか判定する（S 8 0 3）。ステップ S 8 0 3 において、行動類似性計算部 1 2 3 0 により類似する行動が検出されていないと判定されると（S 8 0 3 : N O）、プロセッサ 1 1 0 は、Q 値更新処理を終了する。

【 0 0 9 0 】

ステップ S 8 0 3 において、行動類似性計算部 1 2 3 0 により少なくとも一つの類似する行動が検出されたと判定されると（S 8 0 3 : Y E S）、プロセッサ 1 1 0 は、類似行動のリストをチェックする。類似行動の内の一つの行動が、過去に選択されたことがある場合、つまり、行動選択部 1 2 1 0 に選択され、関連するコンテンツが表示されたことがある場合、プロセッサ 1 1 0 は、その行動を類似行動リストから削除される（S 8 0 4）。

【 0 0 9 1 】

プロセッサ 1 1 0 は、全ての残っている類似行動の部分 Q 値を、数式（3）に従って計算し、Q 値データベース 1 3 3 を更新する（S 8 0 5）。最後に、プロセッサ 1 1 0 は、全ての残っている類似行動の総計 Q 値を、数式（4）に従って計算し、Q 値データベース 1 3 3 を更新する（S 8 0 6）。以上により、Q 値更新処理が終了する。

【 0 0 9 2 】

以上のように、本実施例は、表示されたコンテンツのQ値の更新に応じて、表示されたコンテンツの類似するコンテンツのQ値を更新する。これにより、実際に表示されていないコンテンツのQ値を更新することができ、適切なコンテンツを表示するための学習時間を短縮することができる。

【 実施例 2 】

【 0 0 9 3 】

図9から11を参照して、実施例2を説明する。以下において、実施例1との相違点を主に説明する。実施例1において、全ての視聴者は、同一の嗜好をもっているとして見なされている。しかし、異なる視聴者のグループは、異なる背景及び異なる嗜好を有している。情報レコメンドシステムが、視聴者の各グループに対して、より適切なコンテンツを提供するために、本実施例は、状態空間を視聴者の特徴量（年齢範囲、性別、民族等）により表わし、視聴者認識プログラムを観測端末300に実装する。

10

【 0 0 9 4 】

図9は、実施例2に係る、情報レコメンドシステムの構成例を示す。観測端末300における視聴者認識プログラム332を除き、他の全て構成要素は、実施例1に係るシステムと同様である。観測端末300は、例えば、年齢範囲、性別、民族等の観点から、視聴者の特徴量を認識する。視聴者の一意のグループは、これら特徴量の一意の組み合わせに関連付けられる。

【 0 0 9 5 】

20

Q値データベース133は、状態IDカラム501において、「コンテンツ表示の開始」と「コンテンツ表示の終了」それぞれに、他の情報を追加して格納する。視聴者認識プログラム332により認識される特徴量は、ベクトルで表わすことができる。例えば、年齢範囲が10から20、男性、ラテン系の視聴者のグループは、〔2、1、5〕と表わされる。一方、年齢範囲が20から30、女性、ベトナム人の視聴者のグループは、〔3、2、10〕と表わされる。

【 0 0 9 6 】

図10は、実施例2に係る情報レコメンドシステムの機能ブロック図を示す。視聴者認識部3220以外の構成要素は、実施例1と同様である。潜在的視聴者のグループが、表示端末200に近づいているとき、視聴者認識部3220が起動され、画像処理技術により視聴者の映像を解析する。

30

【 0 0 9 7 】

視聴者認識部3220は、視聴者認識プログラム322を実行するプロセッサ310により実現される。視聴者認識部3220は、視聴者の例えば、年齢範囲、性別、民族を認識し、ベクトルで表わす。システムの状態は、視聴者を認識済みであって、「コンテンツ表示の開始」の状態〔2、1、5、0〕に変化する。最後の数字が、「コンテンツ表示の開始」を示す。

【 0 0 9 8 】

図11は、実施例2に係るQ値データベース133の構成例を示す。状態IDカラム1101が実施例1の図5の状態IDカラム501と異なる点を除き、カラム1102からカラム1106は、図5のカラム502からカラム506と同様である。

40

【 0 0 9 9 】

実施例1の状態カラムID501は、「コンテンツ表示の開始」又は「コンテンツ表示の終了」の状態を、0又は1で示す。本例の状態IDカラム1101は、視聴者の特徴量について情報を、さらに追加して示す。例えば、状態〔2、1、5、0〕は、「年齢範囲が10から20、男性、ラテン系の視聴者に、コンテンツ表示開始」を示し、状態〔2、1、5、1〕は、「年齢範囲が10から20、男性、ラテン系の視聴者に、コンテンツ表示終了」を示す。

【 0 1 0 0 】

行動選択部1210が、最も高い総計Q値の行動を利用する場合、行動選択部1210

50

は、認識されている視聴者の特徴量と同じベクトルを含む状態IDのみを検索する。上述のように、Q値更新部1220は、Q値データベース133において、状態IDの値が、選択された状態IDの値と同一であって、表示コンテンツと同一又は類似するコンテンツのレコードを更新する。

【0101】

本例において、Q値更新部1220は、認識された視聴者の特徴量（状態）と関連付けられており、実際に表示されたコンテンツの総計Q値と部分Q値を更新する。Q値更新部1220は、さらに、認識された視聴者の特徴量（状態）と関連付けられており、表示されたコンテンツと類似するコンテンツ（行動）の総計Q値と部分Q値を更新する。

【0102】

より具体的には、Q値更新部1220は、認識されている視聴者の特徴量を含み「コンテンツ表示の開始」である状態IDと、表示されたコンテンツ（実行された行動）と、のペアの、総計Q値及び部分Q値を更新する。Q値更新部1220は、さらに、当該状態IDと、表示されたコンテンツ（実行された行動）に類似するコンテンツ（行動）と、のペアの、総計Q値及び部分Q値を更新する。

【0103】

以上のように、本実施例によれば、視聴者の特徴に応じてQ値を管理することで、視聴者に嗜好により合致したコンテンツを表示することができる。

【実施例3】

【0104】

図12から図15を参照して、実施例3を説明する。実施例1において、行動選択部1210が行動を探索する場合、行動データベース131においてランダムに行動が選択される。しかし、ランダムに選択される行動が、最も高いQ値の行動である（最善行動の利用と同じである）可能性がある。または、既に選択済みの行動（表示済みのコンテンツ）である可能性もある。

【0105】

したがって、情報ゲインに基づいて行動を探索することが学習のために効率的である。なぜなら、これまで未選択の特徴量の値を有する、これまで未選択の行動を選択することで、類似特徴量を有する他の行動についても、より多くの知見を得ることができるからである。

【0106】

図12は、実施例3に係る情報レコメンドシステムの構成例を示す。情報ゲイン計算プログラム125が追加されている点を除き、実施例1の情報レコメンドシステムの構成と同様である。情報ゲイン計算プログラム125は、どの行動が最も未知の情報を持しているか、計算する。

【0107】

コンテンツが選択され、視聴者に表示される場合、そのコンテンツの全ての部分Q値が更新され、さらに、類似コンテンツ（類似行動）の多くの部分Q値も更新される。したがって、更新されていない最も多くの部分Q値を有するコンテンツ（行動）が、最も探索すべきコンテンツ（行動）である。そのようなコンテンツ（行動）を探索することで、Q値データベース133は、ランダムに行動を選択する場合より早く、集束する。

【0108】

図13は、実施例3に係る情報レコメンドシステムの機能ブロック図を示す。情報ゲイン計算部1250が追加されている点を除き、実施例1の機能ブロック図と同様である。行動選択部1210は、情報ゲイン計算部1250の出力、又は、Q値データベース133における最も高いQ値に基づいて、行動データベース131からコンテンツ（行動）を選択する。

【0109】

情報ゲイン計算部1250は、情報ゲイン計算プログラム125を実行するプロセッサにより実現される。コンテンツデータは、コンテンツ表示部2210に送信される。実施

10

20

30

40

50

例 3 に係る行動選択部 1 2 1 0 の処理フローは、後に、図 1 4 を参照して詳述される。

【 0 1 1 0 】

情報ゲイン計算部 1 2 5 0 は、全てのこれまでに未選択の行動について情報ゲインを計算し、最も高い情報ゲインのコンテンツ（行動）を選択して、返す。情報ゲイン計算部 1 2 5 0 の処理フローは、後に、図 1 5 を参照して詳述される。

【 0 1 1 1 】

図 1 4 は、実施例 3 に係る行動選択部 1 2 1 0 のフローチャートを示す。行動選択処理は、サーバ 1 0 0 のプロセッサ 1 1 0 により実行される。まず、プロセッサ 1 1 0 は、0 から 1 の範囲においてランダム数を生成する（S 1 4 0 1）。ランダム数は、と比較される（S 1 4 0 2）。は、システム設計者により予め設定されたパラメータである。は、システムが、最善行動を利用するよりも、いくつかの行動を探索する割合を示す。例えば、は 0 . 5 に設定され、5 0 % の割合（確率）で探索がなされ、5 0 % の割合（確率）で最善行動が利用される。

10

【 0 1 1 2 】

ステップ S 1 4 0 2 の処理により、ランダム数がより大きいと判定されると（S 1 4 0 2 : Y E S）、プロセッサ 1 1 0 は、Q 値データベース 1 3 3 において、システムの状態と同一の状態において最も高い総計 Q 値を、検索する。最も高い総計 Q 値に関連付けられている行動 I D が、選択される（S 1 4 0 4）。

【 0 1 1 3 】

一方、ステップ S 1 4 0 2 の処理により、ランダム数がより大きくないと判定されると（S 1 4 0 2 : N O）、プロセッサ 1 1 0 は、情報ゲイン計算プログラム 1 2 5 を呼び出し、最も情報ゲインが高い行動の行動 I D を受信する（S 1 4 0 3）。

20

【 0 1 1 4 】

次に、ステップ S 1 4 0 3 又は S 1 4 0 4 において行動 I D を選択した後、プロセッサ 1 1 0 は、行動データベース 1 3 1 から表示するコンテンツを、選択した行動 I D に基づいて読み出す（S 1 4 0 5）。最後に、プロセッサ 1 1 0 は、表示端末 2 0 0 に、表示するコンテンツのデータを表示端末 2 0 0 に送信して、その表示を要求する（S 1 4 0 6）。以上により、行動選択処理が終了する。

【 0 1 1 5 】

図 1 5 は、実施例 3 に係る情報ゲイン計算部 1 2 5 0 のフローチャートを示す。情報ゲイン計算処理は、サーバ 1 0 0 のプロセッサ 1 1 0 により実行される。まず、プロセッサ 1 1 0 は、これまでに未選択のコンテンツ（行動）のそれぞれにおいて、いくつかの部分 Q 値が 0 と等しいかカウントする（S 1 5 0 1）。0 の部分 Q 値は、その部分 Q 値が更新されたことがないことを示す。

30

【 0 1 1 6 】

Q 値データベース 1 3 3 は、例えば、各レコード（状態 I D と行動 I D のペア）が選択済み（表示済み）か示す、ば選択済みフラグをさらに保持する。情報ゲイン計算部 1 2 5 0 は、選択済みフラグを参照して、未選択のコンテンツ（行動）を特定できる。フラグは、行動選択部 1 2 1 0 により更新される。

【 0 1 1 7 】

カウントが終了すると、プロセッサ 1 1 0 は、最もカウント数が多い行動を決定する（S 1 5 0 2）。最もカウント数が多い行動は、複数の場合もある。次に、プロセッサは、最も高いカウント数の行動が、複数であるか判定する（S 1 5 0 3）。ステップ S 1 5 0 3 において、最も高いカウント数の行動が、複数であると判定されると（S 1 5 0 3 : Y E S）、プロセッサ 1 1 0 は、最も高いカウント数の複数の行動から、ランダムに選択した行動 I D を、行動選択部 1 2 1 0 に返す（S 1 5 0 4）。

40

【 0 1 1 8 】

ステップ S 1 5 0 3 において、最も高いカウント数の行動が、一つのみであると判定されると（S 1 5 0 3 : N O）、プロセッサ 1 1 0 は、最も高いカウント数の行動の行動 I D を、行動選択部 1 2 1 0 に返す（S 1 5 0 5）。以上により、情報ゲイン計算処理が終

50

了する。以上のように、本実施例によれば、更新度合が最も小さい未選択のコンテンツを選択して表示することで、学習効率を向上することができる。

【0119】

なお、本発明は上記した実施例に限定されるものではなく、様々な変形例が含まれる。例えば、上記した実施例は本発明を分かりやすく説明するために詳細に説明したものであり、必ずしも説明したすべての構成を備えるものに限定されるものではない。また、ある実施例の構成の一部を他の実施例の構成に置き換えることが可能であり、また、ある実施例の構成に他の実施例の構成を加えることも可能である。また、各実施例の構成の一部について、他の構成の追加・削除・置換をすることが可能である。

【0120】

また、上記の各構成・機能・処理部等は、それらの一部又は全部を、例えば集積回路で設計する等によりハードウェアで実現してもよい。また、上記の各構成、機能等は、プロセッサがそれぞれの機能を実現するプログラムを解釈し、実行することによりソフトウェアで実現してもよい。各機能を実現するプログラム、テーブル、ファイル等の情報は、メモリや、ハードディスク、SSD(Solid State Drive)等の記録装置、または、ICカード、SDカード等の記録媒体に置くことができる。

【0121】

また、制御線や情報線は説明上必要と考えられるものを示しており、製品上必ずしもすべての制御線や情報線を示しているとは限らない。実際には殆どすべての構成が相互に接続されていると考えてもよい。

【符号の説明】

【0122】

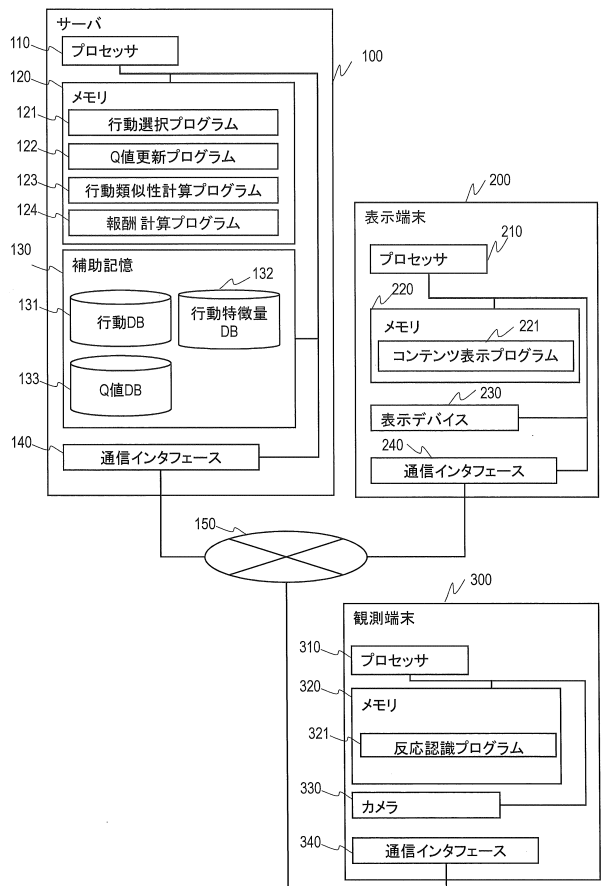
100 サーバ、110 プロセッサ、120 メモリデバイス、121 行動選択プログラム、122 Q値更新プログラム、123 行動類似性計算プログラム、124 報酬計算プログラム、130 補助記憶デバイス、131 行動データベース、132 行動特徴量データベース、133 Q値データベース、140 通信インタフェース、200 表示端末、210 プロセッサ、220 メモリデバイス、221 コンテンツ表示プログラム、230 表示デバイス、240 通信インタフェース、300 観測端末、310 プロセッサ、320 メモリデバイス、330 カメラ、340 通信インタフェース、1210 行動選択部、1220 Q値更新部、1230 行動類似性計算部、1240 報酬計算部、1250 情報ゲイン計算部、2210 コンテンツ表示部、3210 反応認識部

10

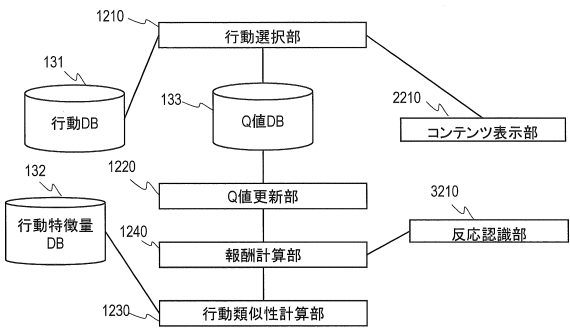
20

30







【図 1】



【図 2】



【図 3】

301 行動ID	302 コンテンツ名	303 表示コンテンツ
1	Tokyo Tower	 Welcome To Tokyo Tower
2	Tokyo Tower	 Why not visit Tokyo Tower?
3	Tokyo Tower	 Welcome To Tokyo Tower
4	Meiji Jingu	 Meiji Jingu
5	Meiji Jingu	 The biggest shrine in Tokyo
6	Meiji Jingu	 Meiji Jingu
...

131 行動DB

【図 4】

401 行動ID	402 特徴量1 (画像)	403 特徴量2 (テキスト)	404 特徴量3 (時刻)
001	Tower	Welcome to Tokyo Tower!	Morning
002	Tower	Why not visit Tokyo Tower?	Night
003	Tower	Welcome to Tokyo Tower!	Night
004	Beautiful gate	Meiji Jingu	Noon
005	Simple gate	The biggest shrine in Tokyo!	Noon
006	Simple gate	Meiji Jingu	Noon
...

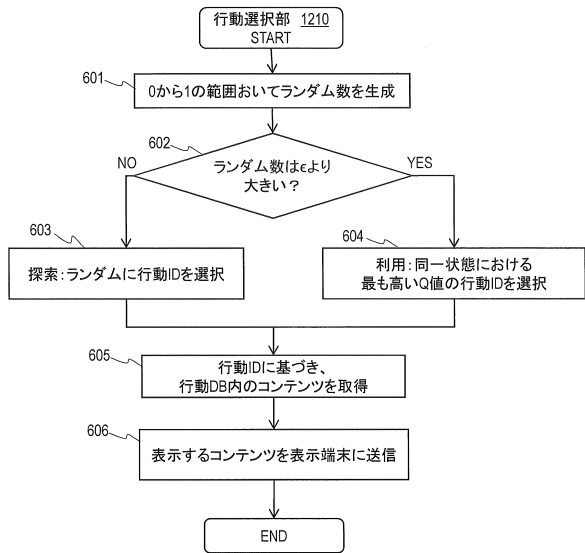
132 行動特徴量DB

【図 5】

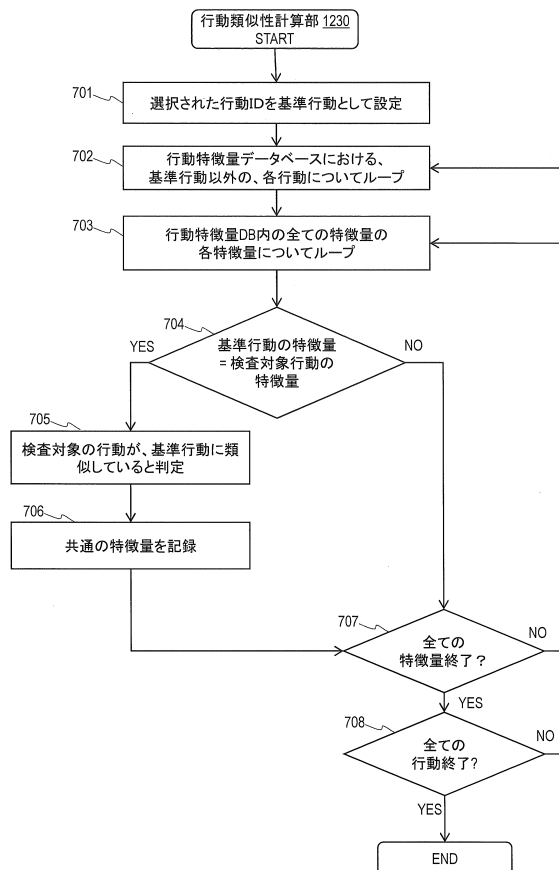
501 状態ID	502 行動ID	503 Q _{partial 1}	504 Q _{partial 2}	505 Q _{partial 3}	506 Q _{total}
0	1	0.3	0.3	0.3	0.9
0	2	0.2	0.2	0.2	0.6
0	3	0.25	0.25	0.2	0.7
0	4	0.3	0.3	0.3	0.9
0	5	0	0	0.3	0.3
0	6	0	0.3	0	0.3
1	1	0	0	0	0
1	2	0	0	0	0
1	3	0	0	0	0
1	4	0	0	0	0
1	5	0	0	0	0
1	6	0	0	0	0
...

Q値DB 133

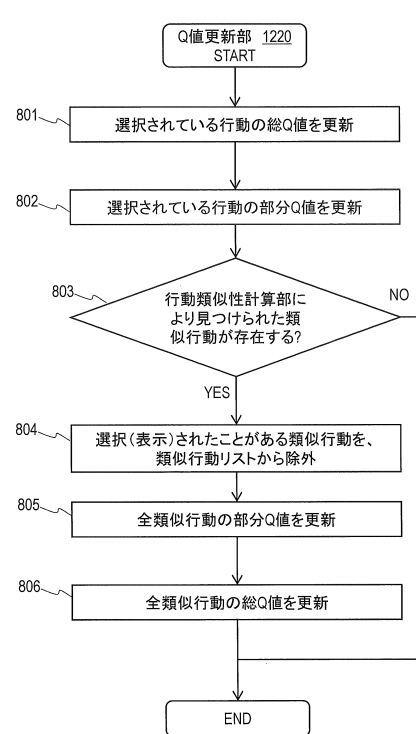
【図 6】



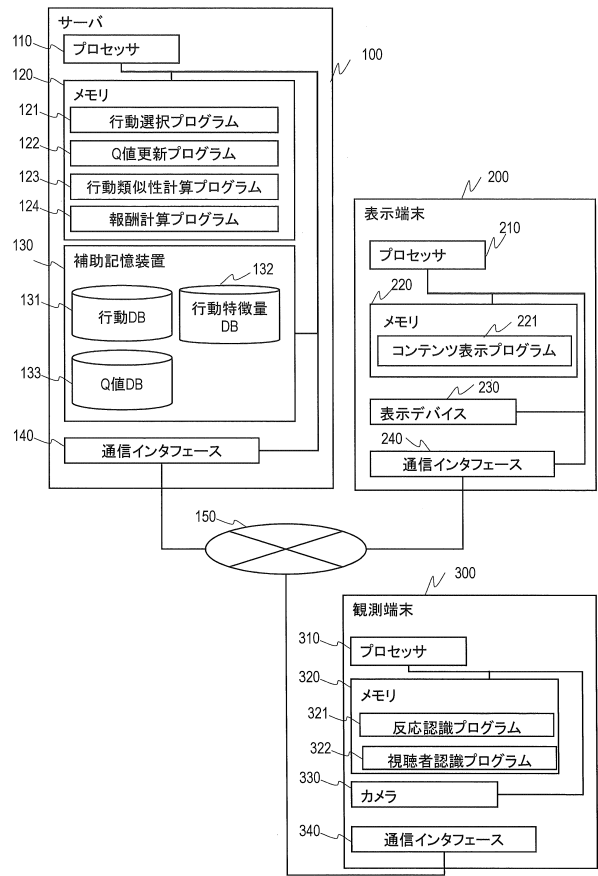
【図 7】



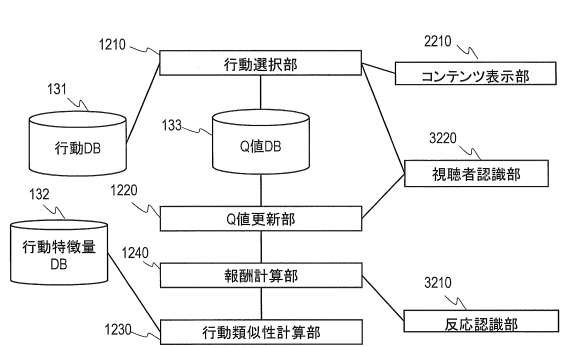
【図 8】



【図 9】



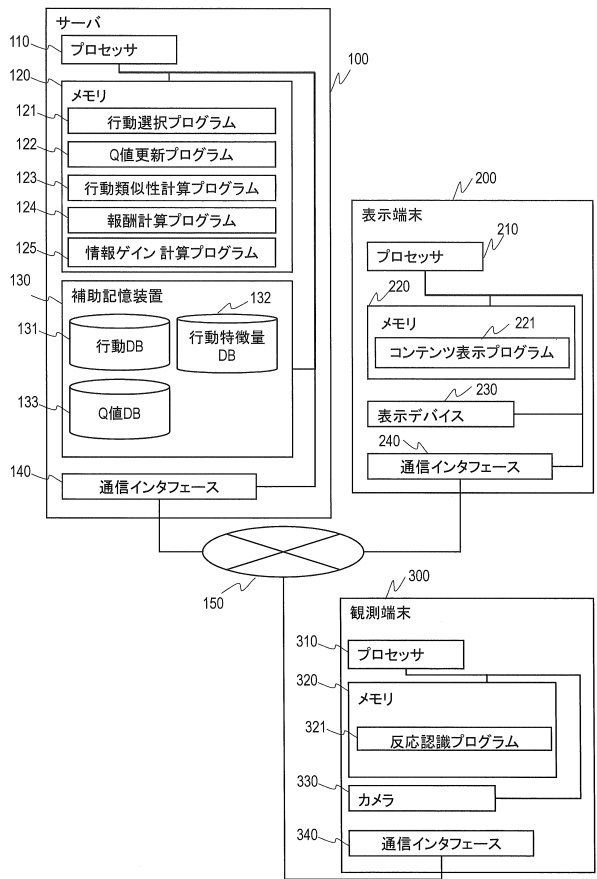
【図 10】



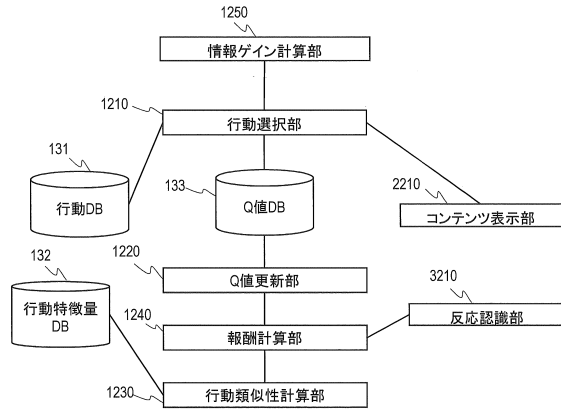
【図 11】

1101 状態ID	1102 行動ID	1103 $Q_{\text{partial } 1}$	1104 $Q_{\text{partial } 2}$	1105 $Q_{\text{partial } 3}$	1106 Q_{total}
[2,1,5,0]	1	0.3	0.3	0.3	0.9
[2,1,5,0]	2	0.2	0.2	0.2	0.6
[2,1,5,0]	3	0.25	0.25	0.2	0.7
[2,1,5,0]	4	0.3	0.3	0.3	0.9
[2,1,5,0]	5	0.1	0.1	0.1	0.3
[2,1,5,0]	6	0.2	0.2	0.1	0.5
[2,1,5,1]	1	0	0	0	0
[2,1,5,1]	2	0	0	0	0
[2,1,5,1]	3	0	0	0	0
[2,1,5,1]	4	0	0	0	0
[2,1,5,1]	5	0	0	0	0
[2,1,5,1]	6	0	0	0	0
...

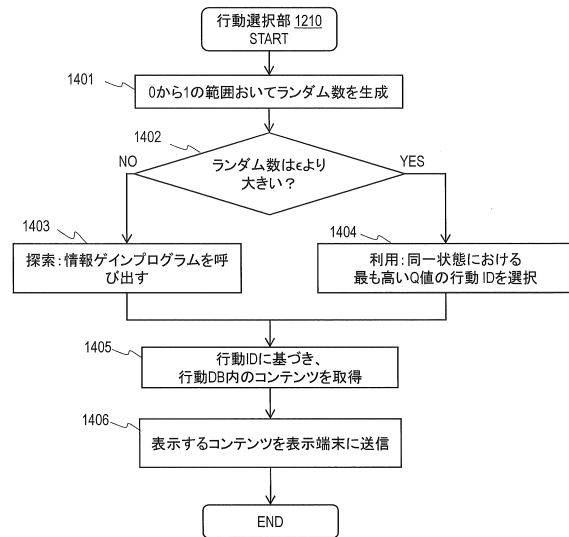
【図 12】



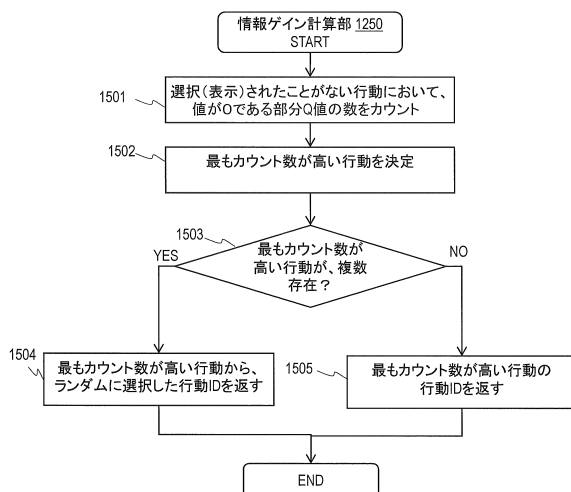
【図 13】



【図 14】



【図 15】



フロントページの続き

(56)参考文献 特開2004-054768(JP,A)
特開2012-231320(JP,A)

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