

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第6511455号
(P6511455)

(45) 発行日 令和1年5月15日(2019.5.15)

(24) 登録日 平成31年4月12日(2019.4.12)

(51) Int.Cl. F I
G06F 16/00 (2019.01) G O 6 F 17/30 3 4 O A
 G O 6 F 17/30 4 1 9 B

請求項の数 20 (全 35 頁)

(21) 出願番号 特願2016-545218 (P2016-545218)
 (86) (22) 出願日 平成26年9月24日 (2014.9.24)
 (65) 公表番号 特表2016-536725 (P2016-536725A)
 (43) 公表日 平成28年11月24日 (2016.11.24)
 (86) 国際出願番号 PCT/US2014/057279
 (87) 国際公開番号 W02015/048171
 (87) 国際公開日 平成27年4月2日 (2015.4.2)
 審査請求日 平成29年8月30日 (2017.8.30)
 (31) 優先権主張番号 201310447654.4
 (32) 優先日 平成25年9月25日 (2013.9.25)
 (33) 優先権主張国 中国 (CN)

(73) 特許権者 510330264
 アリババ・グループ・ホールディング・リ
 ミテッド
 ALIBABA GROUP HOLDI
 NG LIMITED
 英国領、ケイマン諸島、グランド・ケイマ
 ン、ジョージ・タウン、ワン・キャピタル
 ・プレイス、フォース・フロア、ビー・オ
 ー、ボックス 847
 (74) 代理人 110001243
 特許業務法人 谷・阿部特許事務所

最終頁に続く

(54) 【発明の名称】 ユーザ行動の特徴を抽出して推奨を個人化する方法及びシステム

(57) 【特許請求の範囲】

【請求項1】

コンピュータにより実施される方法であって、
 カレントユーザのウェブページクリックストリームデータを取得することと、
 前記クリックストリームデータを用いて前記カレントユーザと他のユーザとの間のクリ
 ックパス相関を計算することと、
 前記カレントユーザとのクリックパス相関が最も高く順位付けられたX人の他のユーザ
 を選択し、Xは正整数であることと、
 前記選択されたX人の他のユーザが訪問したウェブページの予め設定されたタグのそれ
 ぞれに関連する総合加重を構成することと、
 前記予め設定されたタグ及び前記総合加重に基づいて、前記カレントユーザと前記X人
 の他のユーザとの間のユーザ相関を計算することと
 を含む、ユーザの行動に基づいてユーザの特徴を抽出する方法。

【請求項2】

前記クリックストリームデータが複数のクリックストリームデータポイントを有し、そ
 れぞれの前記クリックストリームデータポイントが、ユーザ識別子、カレントウェブペ
 ージ、ソースウェブページ、クリック時間、及び滞在の長さを有し、
 前記カレントユーザと他のユーザとの間の前記クリックパス相関を計算することが、
 前記複数のクリックストリームデータポイントを1または複数のセッションに分割し、
 前記セッションの前記それぞれのクリック時間によって順に分類された任意の2つの近接

したセッションの前記クリック時間の差が、第1の予め設定された時間閾値より少ないかあるいは等しいことと、

それぞれのセッションで、前記それぞれのカレントウェブページ及びそのソースウェブページを用いてクリックストリームパスツリーを生成し、前記クリックストリームパスツリーがノード及びパスを有し、前記ノードが前記カレントウェブページであり、前記パスが前記カレントウェブページとそのソースウェブページとの間の関連を示すことと、

前記クリックストリームパスツリーをマージし、それぞれのマージされた前記ノードに階層加重を加え、それぞれのマージされたパスに比率加重を加えることにより加重有向グラフを生成することと、

前記カレントユーザと最高数の共通パスを共有するY人の他のユーザを選択し、Yは正整数であることと、

前記加重有向グラフを用いて前記カレントユーザと前記Y人の他のユーザとの間の前記クリックパス相関を計算することと

を含む、請求項1に記載の方法。

【請求項3】

それぞれのマージされたノードへの前記階層加重が前記ノードの滞在の長さの平均によって決定され、それぞれのマージされたパスへの前記比率加重が、それぞれのページビュー数 p によって決定され、前記滞在の長さの平均が前記ノードの滞在の長さの合計を前記ノードのページビューの合計で除算することにより計算され、前記合計が、前記複数のクリックストリームデータポイントにわたって加算される、請求項2に記載の方法。

【請求項4】

それぞれのマージされたノードに階層加重を加えることが、

前記ノードの滞在の長さの平均によって分割された複数の加重レベルのうちの1つに前記ノードを配置することと、

前記ノードが配置された前記加重レベルにより前記ノードに前記階層加重を加えることと

を含む、請求項2に記載の方法。

【請求項5】

それぞれのノードがルートノードまたは子ノードであってもよく、

前記カレントウェブページ及び前記ソースウェブページを用いて前記クリックストリームパスツリーを生成することが、

予め設定されたマッチング基準により、カレントクリックストリームデータポイントの前記ソースウェブページを、前記カレントクリックストリームデータポイントのクリック時間前にクリック時間を有する他のクリックストリームデータポイントの前記カレントウェブページとマッチングすることと、

前記マッチングに成功した場合、前記カレントクリックストリームデータポイントの前記カレントウェブページを、前記他のマッチングされたクリックストリームデータポイントの前記カレントウェブページの子ノードにすることと、

前記マッチングが不成功である場合、新規のクリックストリームパスツリーの前記ルートノードとして前記カレントクリックストリームデータポイントの前記カレントウェブページを用いて、前記新規のクリックストリームパスツリーを生成することと

を含む、請求項2に記載の方法。

【請求項6】

前記予め設定されたマッチング基準が、予め設定された最大マッチング、及び/または前記カレントクリックストリームデータポイントと前記マッチングするクリックストリームデータポイントのクリック時間の予め設定された最大の差を含む、請求項5に記載の方法。

【請求項7】

前記加重有向グラフが、

10

20

30

40

【数 1】

$$U = \{\dots, \langle u_i, \lambda_i \rangle, \dots, \langle u_j, \lambda_j \rangle, \dots\}$$

式中、 $\langle u_i, \lambda_i \rangle$ がウェブページ u_i 及びそのユーザクリックステータス、 u_i が前記ユーザが訪問した前記カレントウェブページ、 λ_i がウェブページ u_i の前記階層加重をそれぞれ表すこと

及び

【数 2】

$$T = \{\dots, \langle u_i, u_j, \delta_{ij} \rangle, \dots, \langle u_i, u_k, \delta_{ik} \rangle, \dots\}$$

式中、 $\langle u_i, u_j, \delta_{ij} \rangle$ がウェブページ u_j からウェブページ u_i へのユーザクリックパス方向、 u_i が前記カレントウェブページ、 u_j が前記ソースウェブページ、 δ_{ij} が前記それぞれのユーザクリックパス方向の前記比率加重をそれぞれ表すこと

により規定される、請求項 2 に記載の方法。

【請求項 8】

前記カレントユーザと前記 Y 人の他のユーザとの間の前記クリックパス相関が、式

【数 3】

$$\text{相関} = \frac{\sum \text{共通パスの最小加重}}{\sum \text{非共通パスの加重} + \sum \text{共通パスの最大加重}}$$

を用いて計算され、

式中、

共通パスの最小加重が、前記共通パスの前記比率加重の最小値 × 前記共通パスに示される前記カレントウェブページの前記階層加重の最小値 × 前記共通パスに示される前記ソースウェブページの前記階層加重の最小値の乗算の積であり、

共通パスの最大加重が、前記共通パスの前記比率加重の最大値 × 前記共通パスに示される前記カレントウェブページの前記階層加重の最大値 × 前記共通パスに示される前記ソースウェブページの前記階層加重の最大値の乗算の積であり、

非共通パスの加重が、前記非共通パスの前記比率加重の値 × 前記非共通パスに示される前記カレントウェブページの前記階層加重の値 × 前記非共通パスに示される前記ソースウェブページの前記階層加重の値の乗算の積である、

請求項 2 に記載の方法。

【請求項 9】

それぞれのウェブページがページ加重を有し、それぞれのウェブページ上の前記それぞれの予め設定されたタグが初期加重を有し、前記予め設定されたタグの前記総合加重が、式

【数 4】

$$\bar{w} = \frac{\sum_{i=0}^{\sigma k} (w_i \times p v_i \times w p_i)}{\sum_{i=0}^{\sigma k} p v_i}$$

を用いて計算され、

式中、

【数 5】

$$\bar{w}$$

が予め設定されたタグの前記総合加重であり、 k が前記予め設定されたタグが出現するウェブページの総数であり、 w_i が前記ウェブページ i の前記ページ加重であり、 $p v_i$ が前記ウェブページ i のページビューの総数であり、 $w p_i$ が前記ウェブページ i 上の前記予め設定されたタグの初期加重である、

請求項 1 に記載の方法。

10

20

30

40

50

【請求項 10】

前記カレントユーザと X 人の他のユーザとの間の前記ユーザ相関が、
式

【数 6】

$$Sim(i, j) = \frac{K(\bar{Y}_i, \bar{Y}_j)}{\sqrt{K(\bar{Y}_i, \bar{Y}_i)}\sqrt{K(\bar{Y}_j, \bar{Y}_j)}} = \frac{\sum_{n=1}^k (\bar{w}_{in} \times \bar{w}_{jn})}{\sqrt{\sum_{n=1}^k \bar{w}_{in}^2} \sqrt{\sum_{n=1}^k \bar{w}_{jn}^2}}$$

10

を用いて計算され、

式中、 $Sim(i, j)$ がユーザ i とユーザ j との間の前記ユーザ相関であり、それぞれのユーザがそれぞれのベクトル

【数 7】

$$\bar{Y} = \{(\sigma_1, \bar{w}_1), (\sigma_2, \bar{w}_2), \dots, (\sigma_k, \bar{w}_k)\}$$

で表され、 σ が予め設定されたタグを示し、

【数 8】

$$\bar{W}$$

20

が前記予め設定されたタグの前記総合加重を示す、
請求項 9 に記載の方法。

【請求項 11】

ユーザ識別子を含むユーザ情報を取得することと、

前記ユーザとのユーザ類似性が最も高く順位付けられた Z 人の他のユーザを選択し、Z は正整数であることと、

前記選択された Z 人の他のユーザの情報に基づいて前記ユーザに推奨を作成することとを含み、

前記ユーザ類似性が、

前記ユーザのウェブページクリックストリームデータを取得することと、

前記クリックストリームデータを用いて前記ユーザと他のユーザとの間のクリックパス相関を計算することと、

前記ユーザとクリックパス相関が最も高く順位付けられた X 人の他のユーザを選択し、X は正整数であることと、

前記選択された X 人の他のユーザが訪問したウェブページの予め設定されたタグのそれぞれに関連する総合加重を構成することと、

前記予め設定されたタグ及び前記総合加重に基づいて前記ユーザと前記 X 人の他のユーザとの間のユーザ相関を計算することと

を含むプロセスにより生成される、

ユーザの行動に基づいて個人化された推奨を作成する方法。

40

【請求項 12】

プロセッサ、コンピュータ可読メモリ及び記憶媒体、ならびに I/O 装置を有するコンピュータであって、

カレントユーザのウェブページクリックストリームデータを取得するクリックストリームデータ獲得モジュールと、

前記クリックストリームデータを用いて前記カレントユーザと他のユーザとの間のクリックパス相関を計算するクリックパス相関計算モジュールと、

前記カレントユーザとクリックパス相関が最も高く順位付けられた X 人の他のユーザを選択し、X は正整数である第 1 の選択モジュールと、

前記選択された X 人の他のユーザが訪問したウェブページの予め設定されたタグのそれ

50

それに関連する総合加重を構成する総合加重構成モジュールと、

前記予め設定されたタグ及び前記総合加重に基づいて前記カレントユーザと前記X人の他のユーザとの間のユーザ相関を計算するユーザ相関計算モジュールとを含む機能モジュールを有するようプログラムされた、

コンピュータを備える、

ユーザの行動に基づいてユーザ特徴を抽出するコンピュータベースの装置。

【請求項13】

前記クリックストリームデータが複数のクリックストリームデータポイントを有し、それぞれのクリックストリームデータポイントが、ユーザ識別子、カレントウェブページ、ソースウェブページ、クリック時間、及び滞在の長さを有し、

10

前記カレントユーザと他のユーザとの間の前記クリックパス相関を計算することが、前記複数のクリックストリームデータポイントを1または複数のセッションに分割し、前記セッションのそれぞれのクリック時間によって順に分類された任意の2つの近接したセッションの前記クリック時間の差が、第1の予め設定された時間閾値より少ないかあるいは等しいことと、

それぞれのセッションで、前記それぞれのカレントウェブページ及びそのソースウェブページを用いてクリックストリームパスツリーを生成し、前記クリックストリームパスツリーがノード及びパスを有し、前記ノードが前記カレントウェブページであり、前記パスが前記カレントウェブページとそのソースウェブページとの間の関連を示すことと、

前記クリックストリームパスツリーをマージし、それぞれのマージされたノードに階層加重を加え、それぞれのマージされたパスに比率加重を加えることにより加重有向グラフを生成することと、

20

前記カレントユーザと最高数の共通パスを共有するY人の他のユーザを選択し、Yは正整数であることと、

前記加重有向グラフを用いて前記カレントユーザと前記Y人の他のユーザとの間の前記クリックパス相関を計算することと

を含む、請求項12に記載のコンピュータベースの装置。

【請求項14】

それぞれのマージされたノードへの前記階層加重が前記ノードの滞在の長さの平均によって決定され、それぞれのマージされたパスへの前記比率加重が、それぞれのページビュー数 p によって決定され、前記滞在の長さの平均が前記ノードの滞在の長さの合計を前記ノードのページビューの合計で除算することにより計算され、前記合計が前記複数のクリックストリームデータポイントにわたって加算される、

30

請求項13に記載のコンピュータベースの装置。

【請求項15】

それぞれのマージされたノードに階層加重を加えることが、

前記ノードの滞在の長さの平均によって分割された複数の加重レベルのうちの1つに前記ノードを配置することと、

前記ノードが配置された前記加重レベルにより前記ノードに前記階層加重を加えることと

40

を含む、請求項13に記載のコンピュータベースの装置。

【請求項16】

それぞれのノードがルートノードまたは子ノードであってもよく、

前記カレントウェブページ及び前記ソースウェブページを用いて前記クリックストリームパスツリーを生成することが、

予め設定されたマッチング基準により、前記複数のクリックストリームデータポイントのカレントクリックストリームデータポイントの前記ソースウェブページを、前記カレントクリックストリームデータポイントのクリック時間の前にクリック時間を有する前記複数のクリックストリームデータポイントの他のクリックストリームデータポイントの前記カレントウェブページとマッチングすることと、

50

前記マッチングに成功した場合、前記カレントクリックストリームデータポイントの前記カレントウェブページを、前記他のマッチングされたクリックストリームデータポイントの前記カレントウェブページの子ノードにすることと、

前記マッチングが不成功である場合、新規のクリックストリームパスツリーの前記ルートノードとして前記カレントクリックストリームデータポイントの前記カレントウェブページを用いて、前記新規のクリックストリームパスツリーを生成することとを含む、請求項 1 3 に記載のコンピュータベースの装置。

【請求項 1 7】

前記加重有向グラフが、

【数 9】

$$U = \{\dots, \langle u_i, \lambda_i \rangle, \dots, \langle u_j, \lambda_j \rangle, \dots\}$$

式中、 $\langle u_i, \lambda_i \rangle$ がウェブページ u_i 及びそのユーザクリックステータス、 u_i が前記ユーザが訪問した前記カレントウェブページ、 λ_i がウェブページ u_i の前記階層加重をそれぞれ表すこと

及び

【数 1 0】

$$T = \{\dots, \langle u_i, u_j, \delta_{ij} \rangle, \dots, \langle u_i, u_k, \delta_{ik} \rangle, \dots\}$$

式中、 $\langle u_i, u_j, \delta_{ij} \rangle$ がウェブページ u_j からウェブページ u_i へのユーザクリックパス方向、 u_i が前記カレントウェブページ、 u_j が前記ソースウェブページ、 δ_{ij} が前記それぞれのユーザクリックパス方向の前記比率加重をそれぞれ表すこと

により規定される、請求項 1 3 に記載のコンピュータベースの装置。

【請求項 1 8】

前記カレントユーザと前記 Y 人の他のユーザとの間の前記クリックパス相関が、式

【数 1 1】

$$\text{相関} = \frac{\sum \text{共通パスの最小加重}}{\sum \text{非共通パスの加重} + \sum \text{共通パスの最大加重}}$$

を用いて計算され、

式中、

共通パスの最小加重が、前記共通パスの前記比率加重の最小値 × 前記共通パスに示される前記カレントウェブページの前記階層加重の最小値 × 前記共通パスに示される前記ソースウェブページの前記階層加重の最小値の乗算の積であり、

共通パスの最大加重が、前記共通パスの前記比率加重の最大値 × 前記共通パスに示される前記カレントウェブページの前記階層加重の最大値 × 前記共通パスに示される前記ソースウェブページの前記階層加重の最大値の乗算の積であり、

非共通パスの加重が、前記非共通パスの前記比率加重の値 × 前記非共通パスに示される前記カレントウェブページの前記階層加重の値 × 前記非共通パスに示される前記ソースウェブページの前記階層加重の値の乗算の積である、

請求項 1 3 に記載のコンピュータベースの装置。

【請求項 1 9】

それぞれのウェブページがページ加重を有し、それぞれのウェブページ上のそれぞれの予め設定されたタグが初期加重を有し、前記予め設定されたタグの前記総合加重が、式

【数 1 2】

$$\bar{w} = \sum_{i=0}^{\sigma k} (w_i \times p u_i \times w p_i) / \sum_{i=0}^{\sigma k} p u_i$$

を用いて計算され、

10

20

30

40

50

式中、

【数 1 3】

\vec{w}

が前記予め設定されたタグの前記総合加重であり、 k が前記予め設定されたタグが出現するウェブページの総数であり、 w_i がウェブページ i の前記ページ加重であり、 p_i が前記ウェブページ i のページビューの総数であり、 $w p_i$ が前記ウェブページ i 上の前記予め設定されたタグの初期加重である、

請求項 1 2 に記載のコンピュータベースの装置。

【請求項 2 0】

前記カレントユーザと前記 X 人の他のユーザとの間の前記ユーザ相関が、

式

$$Sim(i, j) = \frac{K(\vec{Y}_i, \vec{Y}_j)}{\sqrt{K(\vec{Y}_i, \vec{Y}_i)}\sqrt{K(\vec{Y}_j, \vec{Y}_j)}} = \frac{\sum_{n=1}^k (\vec{w}_{in} \times \vec{w}_{jn})}{\sqrt{\sum_{n=1}^k \vec{w}_{in}^2} \sqrt{\sum_{n=1}^k \vec{w}_{jn}^2}}$$

を用いて計算され、

式中、 $Sim(i, j)$ がユーザ i とユーザ j との間の前記ユーザ相関であり、それぞれのユーザはそれぞれのベクトル

【数 1 5】

$$\vec{Y} = \{(\sigma_1, \vec{w}_1), (\sigma_2, \vec{w}_2), \dots, (\sigma_k, \vec{w}_k)\}$$

で表され、 σ_k が予め設定されたタグを示し、

【数 1 6】

\vec{w}

は前記予め設定されたタグの前記総合加重を示す、

請求項 1 9 に記載のコンピュータベースの装置。

【発明の詳細な説明】

【技術分野】

【0001】

(関連出願)

本願は、2013年9月25日出願の中国特許出願第201310447654.4号「METHOD AND SYSTEM FOR EXTRACTING USER BEHAVIOR FEATURES TO PERSONALIZE RECOMMENDATIONS」の外国優先権を主張し、中国特許出願の全ての内容が参照により本明細書に組み込まれる。

【0002】

本願は、データ処理に関し、より詳細にはユーザ行動の特徴を抽出し、ユーザの行動に基づいて個人化された推奨を作成することに関する。

【背景技術】

【0003】

インターネットの急速な発展は、人々を情報社会及びネットワーク経済の時代へ導き、企業の発展及び個人生活の両方に大きく影響している。しかし同時に、情報過多により、人々が必要な情報を効率的に得るのが困難となり、情報の使用効率を下げている。

【0004】

例えば電子商取引は、オープンネットワーク環境における新規のビジネスモデルであり

10

20

30

40

50

、ブラウザ/サーバベースのアプリケーションを提供して、オンラインショッピング、業者間のオンライン取引、及びオンライン電子決済を可能にする。インターネットの爆発的な発展で、電子商取引は一層広がっている。しかし、サプライチェーン及びロジスティクスの発展により、インターネット上の商品の種類及び量が非常に多いため、それが消費者のショッピングの時間費用を大きく増やし、電子商取引プラットフォームの売上転換率を減少させる。明らかに消費者は、果てしなく存在する商品についてインターネットを検索する時間を過度に費やしたくない。さらにオンラインショッピングは、実際の生活のように商品の質を顧客がチェックすることを可能にしない。顧客は、彼ら自身の興味及び満足感により提案を作成可能な、自動推奨システムを望む。個々のユーザプロファイルにより、またはユーザを異なるユーザグループにクラスタ化することによるパーソナル推奨の対象化は、現在の傾向の適用のうちの1つである。

10

【0005】

先行技術においては、ユーザのクラスタ化は、ユーザのアクセスパスまたはユーザの検索キーワードからの一連のウェブページのヒットに基づく。ユーザのアクセスは一般的に繰り返され中断されるので、それぞれの訪問者のクリックパスは毎回正確に同一ではありえない。結果として、既存の技術は、複数のユーザの訪問に起因する違いを平衡させることが不可能で、ユーザのクラスタ化効果の乏しさと非効率的サービスをもたらす。

【0006】

従って現在、ユーザ間の相関を正確に測定可能で、ユーザグループを効果的及び正確に形成し、効率が改善された対象化サービスを提供可能な個別サービスの推奨を提供するという技術的問題を解決することが差し迫って必要とされている。

20

【発明の概要】**【0007】**

本概要は、簡略化された形態の概念の選択を紹介するために提供し、それらを詳細な記載でさらに後述する。この要約は、主張する主題の全ての主な機能または重要な機能を確認することを意図するものではなく、主張する主題の範囲を決定する補助として単独で使用されることも意図しない。

【0008】

本開示は、ユーザの行動に基づいてユーザの特徴を抽出する方法及び装置を提供する。当該方法は、コンピュータシステムを使用して、カレントユーザのウェブページクリックストリームを取得し、クリックストリームデータを用いてカレントユーザと他のユーザとの間のクリックパス相関を計算し、カレントユーザとのクリックパス相関が最も高く順位付けられるX人の他のユーザを選択し、Xは正整数である。当該方法は、選択されたX人の他のユーザが訪問したウェブページの予め設定されたタグのそれぞれに関連する総合加重を構成し、予め設定されたタグ及び総合加重に基づいて、カレントユーザとX人の他のユーザとの間のユーザ相関を計算する。

30

【0009】

いくつかの実施形態では、当該方法は、クリックストリームデータが複数のクリックストリームデータポイントを有し、それぞれのクリックストリームデータポイントが、ユーザ識別子、カレントウェブページ、ソースウェブページ、クリック時間、及び滞在の長さを有する点でさらに特徴付けられる。カレントユーザと他のユーザとの間のクリックパス相関を計算するため、当該方法は複数のクリックストリームデータポイントを1または複数のセッションに分割し、セッションのそれぞれのクリック時間によって順に分類された任意の2つの近接したセッションのクリック時間の差は、第1の予め設定された時間閾値より少ないかあるいは等しく、それぞれのセッションで、それぞれのカレントウェブページ及びそのソースウェブページを用いてクリックストリームパスツリーを生成する。クリックストリームパスツリーは、ノード及びパスを有し、ノードはカレントウェブページであり、パスはカレントウェブページとそのソースウェブページ間の関連を示す。当該方法は次いで、クリックストリームパスツリーをマージし、それぞれのマージされたノードに階層加重を加え、それぞれのマージされたパスに比率加重を加えることにより加重有向

40

50

グラフを生成し、カレントユーザと最高数の共通パスを共有するY人の他のユーザを選択し、Yは正整数であり、加重有向グラフを用いてカレントユーザとY人の他のユーザとの間のクリックパス相関を計算する。

【0010】

それぞれのマージされたノードへの階層加重を、ノードの滞在の長さの平均により決定してもよく、一方でそれぞれのマージされたパスへの比率加重を、それぞれのページビュー数pにより決定してもよい。滞在の長さの平均はノードの滞在の長さの合計をノードのページビューの合計で除算することにより計算され、合計は複数のクリックストリームデータポイントにわたって加算される。

【0011】

それぞれのマージされたノードに階層加重を加えることを、ノードの滞在の長さの平均によって分割された複数の加重レベルのうちの一つにノードを配置することにより、及びノードがおかれた加重レベルによりノードに階層加重を加えることにより行ってもよい。

【0012】

それぞれのノードはルートノードまたは子ノードであってもよく、カレントウェブページ及びソースウェブページを用いてクリックストリームパスツリーを生成することを、予め設定されたマッチング基準により、複数のクリックストリームデータポイントのカレントクリックストリームデータポイントのソースウェブページを、カレントクリックストリームデータポイントのクリック時間前にクリック時間を有する複数のクリックストリームデータポイントの他のクリックストリームデータポイントのカレントウェブページとマッチングすることにより、行ってもよい。マッチングに成功した場合、当該方法は、カレントクリックストリームデータポイントのカレントウェブページを、他のマッチングされたクリックストリームデータポイントのカレントウェブページの子ノードにする。マッチングが不成功である場合、当該方法は、新規のクリックストリームパスツリーのルートノードとしてカレントクリックストリームデータポイントのカレントウェブページを用いて、新規のクリックストリームパスツリーを生成する。

【0013】

予め設定されたマッチング基準は、予め設定された最大マッチング、及び/またはカレントクリックストリームデータポイントとマッチングするクリックストリームデータポイントのクリック時間の間の予め設定された最大の差を含んでもよい。

【0014】

一実施形態においては、加重有向グラフは、

【0015】

【数1】

$$U = \{\dots, \langle u_i, \lambda_i \rangle, \dots, \langle u_j, \lambda_j \rangle, \dots\}$$

【0016】

式中、 $\langle u_i, \lambda_i \rangle$ はウェブページ u_i 及びそのユーザクリックステータス、 u_i はユーザが訪問したカレントウェブページ、 λ_i はウェブページ u_i の階層加重をそれぞれ表す、

及び、

【0017】

【数2】

$$T = \{\dots, \langle u_i, u_j, \delta_{ij} \rangle, \dots, \langle u_i, u_k, \delta_{ik} \rangle, \dots\}$$

【0018】

式中、 $\langle u_i, u_j, \delta_{ij} \rangle$ はウェブページ u_j からウェブページ u_i へのユーザクリックパス方向、 u_i はカレントウェブページ、 u_j はソースウェブページ、 δ_{ij} はそれぞれのユーザクリックパス方向の比率加重をそれぞれ表す、により規定される。

【0019】

10

20

30

40

50

カレントユーザとY人の他のユーザとの間のクリックパス相関が、式

【0020】

【数3】

$$\text{相関} = \frac{\sum \text{共通パスの最少加重}}{\sum \text{比共通パスの加重} + \sum \text{共通パスの最大加重}}$$

【0021】

を用いて計算される。

【0022】

式中、共通パスの最小加重は、共通パスの比率加重の最小値×共通パスに示されるカレントウェブページの階層加重の最小値×共通パスに示されるソースウェブページの階層加重の最小値の乗算の積であり、共通パスの最大加重は、共通パスの比率加重の最大値×共通パスに示されるカレントウェブページの階層加重の最大値×共通パスに示されるソースウェブページの階層加重の最大値の乗算の積であり、非共通パスの加重は、非共通パスの比率加重の値×非共通パスに示されるカレントウェブページの階層加重の値×非共通パスに示されるソースウェブページの階層加重の値の乗算の積である。

10

【0023】

それぞれのウェブページがページ加重を有し、それぞれのウェブページ上のそれぞれの予め設定されたタグが初期加重を有し、そしてこれらの初期加重に基づいて、予め設定されたタグの総合加重を、式

20

【0024】

【数4】

$$\bar{w} = \frac{\sum_{i=0}^{\sigma k} (w_i \times pv_i \times wp_i)}{\sum_{i=0}^{\sigma k} pv_i}$$

【0025】

を用いて計算してもよい。

【0026】

式中、 \bar{w} は予め設定されたタグの総合加重であり、 σk は予め設定されたタグが出現するウェブページの総数であり、 w_i はウェブページ i のページ加重であり、 pv_i はウェブページ i のページビューの総数であり、 wp_i はウェブページ i 上の予め設定されたタグの初期加重である。

30

【0027】

さらに、カレントユーザとX人の他のユーザとの間のユーザ相関が、式

【0028】

【数5】

$$\text{Sim}(i, j) = \frac{K(\bar{Y}_i, \bar{Y}_j)}{\sqrt{K(\bar{Y}_i, \bar{Y}_i)} \sqrt{K(\bar{Y}_j, \bar{Y}_j)}} = \frac{\sum_{n=1}^k (\bar{w}_{in} \times \bar{w}_{jn})}{\sqrt{\sum_{n=1}^k \bar{w}_{in}^2} \sqrt{\sum_{n=1}^k \bar{w}_{jn}^2}}$$

40

【0029】

を用いて計算される。

【0030】

式中、 $\text{Sim}(i, j)$ はユーザ i とユーザ j 間のユーザ相関であり、それぞれのユーザはそれぞれのベクトル $\vec{Y} = \{(\sigma_1, \vec{w}_1), (\sigma_2, \vec{w}_2), \dots, (\sigma_k, \vec{w}_k)\}$ で表され、 σ は予め設定されたタグを示し、 \vec{w} は特定の予め設定されたタグの総合加重を示す。

【 0 0 3 1 】

開示の他の態様は、ユーザの行動に基づいて個人化された推奨を作成する方法である。当該方法は、ユーザ識別子を含むユーザ情報を取得し、ユーザとのユーザ類似性が最も高く順位付けられた Z 人の他のユーザを選択し、 Z は正整数である。当該方法は次いで、選択された Z 人の他のユーザの情報に基づいてユーザに推奨を作成する。ユーザ類似性は、前述の方法に記載されたプロセスにより計算され、当該方法は、ユーザのウェブページクリックストリームデータを取得することと、クリックストリームデータを用いてユーザと他のユーザの間のクリックパス相関を計算することと、カレントユーザとクリックパス相関が最も高く順位付けられた X 人の他のユーザを選択し、 X は正整数であることと、選択された X 人の他のユーザが訪問したウェブページの予め設定されたタグのそれぞれに関連する総合加重を構成することと、予め設定されたタグ及び総合加重に基づいてカレントユーザと X 人の他のユーザとの間のユーザ相関を計算することとを含む。

10

【 0 0 3 2 】

さらに開示の他の態様は、ユーザの行動に基づいてユーザ特徴を抽出するコンピュータベースの装置である。装置は、プロセッサ、コンピュータ可読メモリ及び（コンピュータ可読媒体のような）記憶媒体、及び I/O 装置を有するコンピュータを有し、コンピュータは以下の機能モジュールを有するようプログラムされる。

20

【 0 0 3 3 】

カレントユーザのウェブページクリックストリームデータを取得するクリックストリームデータ獲得モジュール

【 0 0 3 4 】

クリックストリームデータを用いてカレントユーザと他のユーザとの間のクリックパス相関を計算するクリックパス相関計算モジュール

【 0 0 3 5 】

カレントユーザとクリックパス相関が最も高く順位付けられた X 人の他のユーザを選択し、 X は正整数である第 1 の選択モジュール

30

【 0 0 3 6 】

選択された X 人の他のユーザが訪問したウェブページの予め設定されたタグのそれぞれに関連する総合加重を構成する総合加重構成モジュール

【 0 0 3 7 】

予め設定されたタグ及び総合加重に基づいてカレントユーザと X 人の他のユーザとの間のユーザ相関を計算するユーザ相関計算モジュール

【 0 0 3 8 】

いくつかの実施形態では、コンピュータベースの装置は、クリックストリームデータが複数のクリックストリームデータポイントを有し、それぞれのクリックストリームデータポイントは、ユーザ識別子、カレントウェブページ、ソースウェブページ、クリック時間、及び滞在の長さを有し、カレントユーザと他のユーザとの間のクリックパス相関が、本明細書に記載された方法を用いて計算されることをさらに特徴とする。例えばコンピュータベースの装置は、複数のクリックストリームデータポイントを 1 または複数のセッションに分割することができ、セッションのそれぞれのクリック時間によって順に分類された任意の 2 つの近接したセッションのクリック時間の差は、第 1 の予め設定された時間閾値より少ないかあるいは等しく、それぞれのセッションで、それぞれのカレントウェブページ及びそのソースウェブページを用いてクリックストリームパスツリーを生成し、クリックストリームパスツリーはノード及びパスを有し、ノードはカレントウェブページであり

40

50

、パスはカレントウェブページとそのソースウェブページ間の関連を示し、クリックストリームパスツリーをマージすることにより、及びそれぞれのマージされたノードに階層加重を加え、それぞれのマージされたパスに比率加重を加えることにより、加重有向グラフを生成し、カレントユーザと最高数の共通パスを共有するY人の他のユーザを選択し、Yは正整数であり、加重有向グラフを用いてカレントユーザとY人の他のユーザとの間のクリックパス相関を計算してもよい。

【0039】

同様に、それぞれのマージされたノードへの階層加重をノードの滞在の長さの平均によって決定してもよく、それぞれのマージされたパスへの比率加重をそれぞれのページビュー数 p によって決定してもよい。滞在の長さの平均はノードの滞在の長さの合計をノードのページビュー数の合計で除算することにより計算され、合計は複数のクリックストリームデータポイントにわたって加算される。

10

【0040】

モジュールを、ノードの滞在の長さの平均によって分割された複数の加重レベルのうちの1つにノードを配置することにより、及びノードが配置された加重レベルによりノードに階層加重を加えることにより、それぞれのマージされたノードに階層加重を加えるようプログラムしてもよい。

【0041】

いくつかの実施形態では、それぞれのノードがルートノードまたは子ノードであってもよく、モジュールを、カレントクリックストリームデータポイントのソースウェブページを、カレントクリックストリームデータポイントのクリック時間の前にクリック時間を有する他のクリックストリームデータポイントのカレントウェブページとマッチングすることにより、カレントウェブページ及びソースウェブページを用いてクリックストリームパスツリーを生成してプログラムしてもよい。マッチングは、予め設定されたマッチング基準により実行される。マッチングに成功した場合、モジュールは、カレントクリックストリームデータポイントのカレントページを、他のマッチングされたクリックストリームデータポイントのカレントウェブページの子ノードにする。マッチングが不成功である場合、モジュールは、そのルートノードとしてカレントクリックストリームデータポイントのカレントウェブページを用いて、新規のクリックストリームパスツリーを生成する。

20

【0042】

いくつかの実施形態では、加重有向グラフが、

30

【0043】

【数6】

$$U = \{\dots, \langle u_i, \lambda_i \rangle, \dots, \langle u_j, \lambda_j \rangle, \dots\}$$

【0044】

式中、 $\langle u_i, \lambda_i \rangle$ はウェブページ u_i 及びそのユーザクリックステータス、 u_i はユーザが訪問した前記カレントウェブページ、 λ_i はウェブページ u_i の階層加重をそれぞれ表す、

及び、

40

【0045】

【数7】

$$T = \{\dots, \langle u_i, u_j, \delta_{ij} \rangle, \dots, \langle u_i, u_k, \delta_{ik} \rangle, \dots\}$$

【0046】

式中、 $\langle u_i, u_j, \delta_{ij} \rangle$ はウェブページ u_j からウェブページ u_i へのユーザクリックパス方向、 u_i はカレントウェブページ、 u_j はソースウェブページ、 δ_{ij} はそれぞれのユーザクリックパス方向の比率加重を表す、により規定される。

【0047】

一実施形態においては、カレントユーザとY人の他のユーザとの間のクリックパス相関

50

が、式

【 0 0 4 8 】

【 数 8 】

$$\text{相関} = \frac{\sum \text{共通パスの最少加重}}{\sum \text{比共通パスの加重} + \sum \text{共通パスの最大加重}}$$

【 0 0 4 9 】

を用いて計算される。

【 0 0 5 0 】

式中、共通パスの最小加重は、共通パスの比率加重の最小値 × 共通パスに示されるカレントウェブページの階層加重の最小値 × 共通パスに示されるソースウェブページの階層加重の最小値の乗算の積であり、共通パスの最大加重は、共通パスの比率加重の最大値 × 共通パスに示されるカレントウェブページの階層加重の最大値 × 共通パスに示されるソースウェブページの階層加重の最大値の乗算の積であり、非共通パスの加重は、非共通パスの比率加重の値 × 非共通パスに示されるカレントウェブページの階層加重の値 × 非共通パスに示されるソースウェブページの階層加重の値の乗算の積である。

10

【 0 0 5 1 】

いくつかの実施形態では、それぞれのウェブページがページ加重を有し、それぞれのウェブページ上のそれぞれの予め設定されたタグが初期加重を有し、コンピュータベースの装置は、予め設定されたタグの総合加重を、式

20

【 0 0 5 2 】

【 数 9 】

$$\bar{w} = \frac{\sum_{i=0}^{\sigma k} (w_i \times pv_i \times wp_i)}{\sum_{i=0}^{\sigma k} pv_i}$$

【 0 0 5 3 】

を用いてこれらの初期加重に基づいて計算するようにプログラムされている。

【 0 0 5 4 】

式中、 \bar{w} は予め設定されたタグの総合加重であり、 σk は予め設定されたタグが出現するウェブページの総数であり、 w_i はウェブページ i のページ加重であり、 pv_i はウェブページ i のページビューの総数であり、 wp_i はウェブページ i 上の予め設定されたタグの初期加重である。

30

【 0 0 5 5 】

コンピュータベースの装置を、式

【 0 0 5 6 】

【 数 1 0 】

$$\text{Sim}(i, j) = \frac{K(\bar{Y}_i, \bar{Y}_j)}{\sqrt{K(\bar{Y}_i, \bar{Y}_i)} \sqrt{K(\bar{Y}_j, \bar{Y}_j)}} = \frac{\sum_{n=1}^k (\bar{w}_{in} \times \bar{w}_{jn})}{\sqrt{\sum_{n=1}^k \bar{w}_{in}^2} \sqrt{\sum_{n=1}^k \bar{w}_{jn}^2}}$$

40

【 0 0 5 7 】

を用いて、カレントユーザと X 人の他のユーザとの間のユーザ相関を計算するよう、さらにプログラムしてもよい。

【 0 0 5 8 】

式中、 $\text{Sim}(i, j)$ はユーザ i とユーザ j 間のユーザ相関であり、それぞれのユーザはそれぞれのベクトル $\vec{Y} = \{(\sigma_1, \vec{w}_1), (\sigma_2, \vec{w}_2), \dots, (\sigma_k, \vec{w}_k)\}$ で表され、 σ は予め設定されたタグを示し、 \vec{w} は特定の予め設定されたタグの総合加重を示す。

【0059】

本開示の他の機能及び利点を後述の記載で説明し、一部は記載から明らかになり、または本願の実践により理解される。本願の目的及び他の利点を、明細書、請求項及び特に実現及び達成されて挙げられた構造の図によって得ることができる。

10

【図面の簡単な説明】

【0060】

【図1】ユーザの行動に基づいてユーザの特徴を抽出する開示された方法のブロックフローチャートである。

【図2】本開示の第1の例のセッション(S1)のクリックストリームパスツリーの一例の図である。

【図3】本開示の第2の例のセッション(S2)によるクリックストリームパスツリーのマージの一例の図である。

【図4】本開示の第1の例のセッション(S1)による加重有向グラフの一例の図である

20

。【図5】本開示の第3の例のセッション(S3)による加重有向グラフの一例の図である。

【図6】ユーザの行動に基づいて個人化された推奨を作成するプロセスのブロックフローチャートである。

【図7】ユーザの行動に基づいてユーザの特徴を抽出するシステムのブロック図である。

【図8】ユーザの行動に基づいて個人化された推奨を作成するシステムのブロック図である。

【発明を実施するための形態】

【0061】

30

本開示の前述の目的、特徴及び利点の理解を容易にするため、添付の図及び実施形態例と併せて本開示をさらに詳述する。この記載中、「技術(複数の技術)」の語は、例えば前述及び本開示全体を通じた文脈によって可能な方法、装置、システム、及び/またはコンピュータ可読な命令を言い得る。

【0062】

この記載で、プロセスが記載される順番は限定として解釈されることを意図せず、方法、または別の方法を実行するために任意の数の記載した処理ブロックを任意の順番で組み合わせてもよい。説明の利便性のみの目的で、実施形態を逐次ステップで記載する。矛盾を起こす場合を除いて、本開示に記載される例及び実施形態、及びその特徴及び機能を自由に組み合わせてもよい。さらに、本開示の技術を実践するため、実施形態に記載された全てのステップが必要とはされない。

40

【0063】

図1は、ユーザの行動に基づいてユーザの特徴を抽出する方法の実施形態のブロックフローチャートである。例示的な実施形態は、ブロック内に記載された後述の動作を含む。

【0064】

ブロック101で、カレントユーザのウェブページクリックストリームデータを取得するために、コンピュータベースのシステムが使用される。

【0065】

クリックストリームデータはユーザのウェブサイト閲覧及びクリック行動を記録するために使用されるウェブサイトログから来るかもしれないが、一方でクリックストリームデ

50

ータはユーザ閲覧ウェブページのトレースされたパスを示し得ることに留意すべきである。ユーザが訪問したウェブページは1つのウェブサイトからかもしれないが、また複数のウェブサイトにもわたりクロスプラットフォームサービスが提供されるとき等は、異なるウェブサイトからかもしれない。そのような相違は、本開示内で限定しない。

【0066】

特定の適用で、クリックストリームデータは、時間間隔または複数の時間間隔の間に元のクリックストリームデータを捕捉するApacheサーバのウェブサイトログであってもよい。クリックストリームデータは、ユーザが訪問したIPアドレス、エンリページ（訪問中のカレントページ）、参照元ページ（カレントページのソースページ）、ユーザが訪問したときのカレントページのクリック時間、HTTPコード、HTTPトラフィック、HTTP応答時間、ブラウザ（またはユーザエージェント）特性、訪問中のユーザ固有のcookie ID等を含んでもよい。

10

【0067】

クリックストリームデータの例は以下の通りである。

【0068】

61.18.186.132 - - - [23/Feb/2013:00:00:03+0800] "GET/search.china.alibaba.com/business/k-300_y.html HTTP/1.1" 200 20406 275420 "/search.china.alibaba.com/business/k-300_sortPromotion-false_n-y.html" "Mozilla/4.0 (compatible; MSIE 8.0; Windows NT 5.1; Trident/4.0; .NET CLR 1.1.4322; .NET CLR 2.0.50727)" 61.15.72.11.1305814706754.4 "a=b=mid=world;c=c_key=300|c_sefilter=0" -

20

【0069】

サーバは、ユーザ訪問ログ記録を収集し、時折（例えば周期的に）ログ記録からクリックストリームデータを取得してもよい。クリックストリームデータ取得の周期的な時間間隔は、1時間ごと、または1日ごと、またはリアルタイムでさえよく、制限はない。

【0070】

ウェブサイト訪問中のユーザは頻繁に、後戻り、ページの切り替え、ホームページに戻る、またはページリンクを直接クリックするかもしれない。結果として、ユーザ訪問パスは無作為及び無秩序であるかもしれない。

30

【0071】

ウェブページを訪問するユーザの元のクリックストリームデータを取得すると、システムはデータを処理することができ、構造化されたクリックストリームデータ

$\Psi = \{V_1, V_2, \dots, V_i, \dots, V_{\Psi_n}\}$ をそれぞれのユーザについて出力してもよく、式中、 Ψ_n はカレントユーザの総数を表し、 V_i は i 人目のユーザのクリックストリームデータである。

40

【0072】

上記処理は、誤伝送データ、及びウェブサイトの企業内の内部訪問、または他の無効の一般的な記録のフィルタリング等、元のクリックストリームデータのデータクリーニングを含んでもよい。

【0073】

構造化されたクリックストリームデータは、ユーザ識別子、カレントウェブページ、ソースウェブページ、クリック時間及び滞在の長さ等を含んでもよい。

【0074】

これらの構造化されたクリックストリームデータの例は、以下のようなものである。

【0075】

50

I P = 6 1 . 1 8 . 1 8 6 . 1 3 2
 D A T E T I M E = 2 3 / F e b / 2 0 1 3 : 0 0 : 0 0 : 0 3
 E N T R Y = / s e a r c h . c h i n a . a l i b a b a . c o m / b u s i n e s s / k - 3 0 0 \ x d 4 \ x a a \ x c a \ x c e \ x c 6 \ x b 7 \ x b b \ x e c \ x c 5 \ x f a _ p - 3 _ o f f s e t - 7 _ s o r t P r o m o t i o n - f a l s e _ n - y . h t m l
 R E F E R E R = / s e a r c h . c h i n a . a l i b a b a . c o m / b u s i n e s s / k - 3 0 0 % D 4 % A A % C A % C E % C 6 % B 7 % B B % E C % C 5 % F A _ p - 2 _ o f f s e t - 7 _ s o r t P r o m o t i o n - f a l s e _ n - y . h t m l

10

C O O K I E I D = 6 1 . 1 5 . 7 2 . 1 1 . 1 3 0 5 8 1 4 7 0 6 7 5 4 . 4
 【 0 0 7 6 】

ユーザ識別子がユーザのラベル付けに効果を有することができる限り、ユーザ識別子は、IPアドレス、cookie ID、及び/またはエージェントであってもよいことを述べておく。この点で、本開示はいかなる制限も有さない。カレントウェブページは、訪問の間ユーザが訪問中のウェブページである。ソースウェブページは、そこからカレントウェブページに訪問があった起点のページである。例えば、ユーザがウェブページA上のリンクをクリックしてウェブページBに行く場合、そのときAがソースウェブページであり、一方でBがカレントウェブページである。ユーザが訪問するカレントウェブページのリンクまたはURLを直接入力することによりユーザがウェブページを訪問するような特定の場合、このカレントウェブページはソースウェブページを有さず、このケースではソースウェブページはヌルである。

20

【 0 0 7 7 】

クリック時間は、ユーザがカレントページを訪問したときの時間である。滞在の長さは、ユーザがカレントウェブページを閲覧した持続時間である。

【 0 0 7 8 】

ブロック102で、システムは、クリックストリームデータを用いてカレントユーザと他のユーザとの間のクリックパス相関を計算する。

【 0 0 7 9 】

訪問されたウェブページのクリックパスは、ユーザが閲覧したウェブページの軌跡であり、ユーザの行動傾向または性向にある程度反映することを述べておく。相関、または類似は、2人のユーザが関連する程度を示す。高度の相関を有する2人のユーザは、同一または類似のことに興味を示す機会が多い。

30

【 0 0 8 0 】

実際には、サーバで利用可能なクリックストリームデータは、複数のクリックストリームデータポイントを含んでもよく、これらはそれぞれ、ユーザ識別子、カレントウェブページ、ソースウェブページ、クリック時間、及び滞在の長さ等を含んでもよい。

【 0 0 8 1 】

ブロック102は、以下のステップを有してもよい。

【 0 0 8 2 】

ステップS11は、クリックストリームデータポイントを1または複数のセッションに分割する。セッションのそれぞれのクリック時間順により分類された任意の2つの近接したセッションのクリック時間の差は、第1の予め設定された時間閾値より少ないかあるいは等しい。特に、クリックストリームデータポイントを、それぞれのクリック時間の順番により順次並べてもよい。2つの連続したクリックストリームデータポイントのクリック時間の差が第1の予め設定された時間閾値より少ないかあるいは等しい場合、2つのクリックストリームデータポイントを1つのセッションにマージしてもよい。

40

【 0 0 8 3 】

セッションに分割された後のクリックストリームデータを、

【 0 0 8 4 】

50

【数 1 1】

$$\zeta = \{l_1, l_2, \dots, l_i, \dots, l_{\zeta n}\}$$

【0085】

と表すことができる。

【0086】

ここでユーザ ζ のクリックストリームデータは ζ n個のセッションに分割され、 l_i は
i番目のセッションである。 ζ n及びiの両方は正整数である。

10

【0087】

好適な一実施形態においては、第1の予め設定された時間閾値は30分に設定される。

【0088】

ステップS12で、それぞれのセッションについて、システムはカレントウェブページ及びソースウェブページを用いてクリックストリームパスツリーを確立する。クリックストリームデータポイントパスツリーは、ノード及びパスを有する。ノードはカレントウェブページであり、一方でパスはカレントウェブページ及びそのソースウェブページの間に関連を示す。

【0089】

クリックストリームパスツリーは、ユーザが訪問するウェブページの軌跡の表現である。

20

【0090】

いくつかの実施形態では、ノードはルートノードまたは子ノードのいずれかであってもよい。従って、ステップS12はさらに、以下のサブステップを有してもよい。

【0091】

サブステップS121は、カレントクリックストリームデータポイントのソースウェブページを、カレントクリックストリームデータポイントのクリック時間前にクリック時間を有する他のクリックストリームデータポイントのカレントウェブページとマッチングする。マッチングは、予め設定されたマッチング基準が満たされるまで続けてもよい。マッチングが成功すれば、システムはサブステップS122を実行する。不成功であれば、システムはサブステップS123を実行する。

30

【0092】

マッチングは、2つの原則を監視してもよく、その1つはカレントクリックストリームデータポイントの参照元ページ(ソースページ)が前のクリックストリームデータポイントのエントリページ(カレントページ)と同一であることを必要とする。

【0093】

実際には、複数のクリックストリームデータポイントは、クリック時間による順に並べられる。(最新のクリック時間を有する)最後のクリックストリームデータポイントから開始し、マッチングがセッションの第1のクリックストリームデータポイントに到達するまで、もっと早いクリック時間を有する他のクリックストリームデータポイントにマッチングする。

40

【0094】

カレントウェブページ及びソースウェブページは、URL(ユニバーサルリソースロケータ)の形態であってもよく、URLはさまざまな長さを有し得る。マッチングの間、文字ストリームを元のURLまたは他の識別情報に基づいて生成し、マッチングを向上し、必要とされる永続的記憶装置のサイズを減少してもよい。

【0095】

例えば、URLの長さが16文字以下の場合、URLを直接マッチングに使用してもよい。しかしURLの長さが16文字を超える場合、MD5(メッセージダイジェストアルゴリズム5)アルゴリズムを使用して16文字列を生成し、それを次いでマッチングに使

50

用してもよい。

【0096】

例えば、以下のURLに基づく。

【0097】

```
http://sale.suning.com/images/advertise/
zyn/130722jiuyang/index.html?utm_source=
union&utm_medium=C&utm_campaign=1028&utm_
content=1027,
```

【0098】

MD5アルゴリズムを使用して、B11A4C0B627B8FE4の16文字列を生成する。

10

【0099】

短い文字ストリームを生成する前述の方法は、例示の目的のみである。実際には、同様の目的を達成するために他のどのような適切な方法を使用してもよい。この点で、本開示は限定を有さない。

【0100】

他のマッチング原則は時間の近さであり、これは2つのマッチングするクリックストリームデータポイントの間で時間が最も近い成功したマッチングを探す。

【0101】

例えば、セッションS1の例で、ユーザはウェブページAを訪問し、そこに5秒滞在し、ウェブページAからウェブページBをクリックしてそこに2秒滞在し、次いでウェブページBからウェブページCをクリックしてそこに3秒滞在し、次いでウェブページCからウェブページEをクリックしてそこに7秒滞在し、その後ウェブページBからウェブページDをクリックしてそこに3秒滞在し、ウェブページDからウェブページCをクリックしてそこに1秒滞在し、次いでウェブページCからウェブページEをクリックしてそこに5秒滞在する。

20

【0102】

前述のクリックストリームデータは同一のセッションに属し、それはユーザの実際のクリック軌跡によりA B C(1) E(1) D C(2) E(2)と表現でき、C(1)及びC(2)はウェブサイトCへの第1の訪問及び第2の訪問をそれぞれ示し、E(1)及びE(2)は、ウェブサイトEへの第1の訪問及び第2の訪問をそれぞれ示す。

30

【0103】

カレントページとしてE(2)を用いて上方にマッチングすると、C(1)及びC(2)を見つける。しかし、C(2)のクリック時間がE(2)のクリック時間に最も近いので、C(2)が成功したマッチングと考えられる。

【0104】

ステップS122は、カレントクリックストリームデータポイントのカレントページが、正常にマッチングした新規のカレントクリックストリームデータポイントのカレントページの子ノードであるとみなす。相対的に言って、正常にマッチングした新規のカレントクリックストリームデータポイントのカレントページを、子ノードの親ノードと呼ぶことができる。

40

【0105】

親ノード及び子ノードが正常にマッチングされた場合、ソースページ(親ノード)からカレントページ(子ノード)への訪問軌跡を示すパスも出現する。

【0106】

ステップS123は、ルートノードとしてカレントクリックストリームデータポイントのカレントウェブページを用いて、新規のクリックストリームパスツリーを生成する。

【0107】

別の例においては、セッションS1の例で、ウェブページAのクリックストリームデータポイントを用いたマッチングが失敗する場合、システムは親ノードとしてウェブページ

50

Aを用いて新規のクリックストリームパスツリーを生成する。

【0108】

前述で、ステップS121のマッチングを停止する基準を予め設定してもよく、基準は以下の条件を有してもよい。

【0109】

条件1：マッチングの数が予め設定された最大閾値数を超えた場合、マッチングはない。予め設定された閾値数を、実際の適用条件及び必要条件によって技術者が決定してもよく、この点で、本開示はいかなる制限をも有さない。1つの例では閾値数が5000に設定され、マッチングの数がこの数を超える場合、クリックストリームデータポイントは機械訪問の結果とみなされ、これはクリックトラフィックを増加するため特定のウェブサイトを繰り返し訪問するよう設計された特殊なソフトウェアによって通常行われ、正常でないといみなされる。

10

【0110】

条件2：カレントクリックストリームデータポイントと新規のマッチングクリックストリームデータポイントのクリック時間の差が第2の時間閾値より大きい場合、マッチングはない。この第2の時間閾値を、ステップS121のマッチングを終了すべきかの判断に使用してもよい。

【0111】

同様に、前述の第2の時間閾値を、実際の適用条件及び必要条件によって技術者が決定してもよく、本開示では制限されない。好適な一実施形態においては、第2の時間閾値は30分に設定される。

20

【0112】

別の実施形態では、ブロックS12は以下のステップを有してもよい。

【0113】

ステップS124はあらゆる孤立ノードを除去し、孤立ノードは単一のノードのみを有するクリックストリームパスツリーである。孤立ノード内で、単一のノードは親ノード及び子ノードの両方である。

【0114】

セッション内のクリックストリームパスツリーは、マッチングがなされた後で決定される。

30

【0115】

セッションS1について、クリックストリームパスツリーの例示的な表現が図2で示される。クリックストリームパスツリーを、他の方法を用いて示してもよいことを述べておく。

【0116】

例えばクリックストリームパスツリーを、前順走査出力を用いて

【0117】

【数12】

$$\ell = \{ \langle v_1, p_1, t_1 \rangle, \dots, \langle v_i, p_i, t_i \rangle, \dots, \langle v_{ln}, p_{ln}, t_{ln} \rangle \}$$

式(1)

40

【0118】

と表すことができる。

【0119】

式中、 ℓ はクリックストリームパスツリーを示し、クリックストリームパスツリーは

ℓn 個のノードを有する。 $\langle v_i, p_i, t_i \rangle$ は i 番目のクリックストリームデータポイントで $1 \leq i \leq \ell n$ であり、 ℓn 及び i は両方とも正整数である。 v_i はクライアントウェブページ、 p_i は v_i のソースページ、 t_i は v_i 訪問滞在の長さである。

特に、 $i = 1$ のとき、ノードはクリックストリームパスツリー ℓ

10

のルートノードである。このルートノードは実際のソースページを持たない（すなわち親ノードがない）ので、その記号的なソースページは “-” で表され、つまり $p_1 = \text{“-”}$ である。

【 0 1 2 0 】

例えば、セッション S_1 で前述の式 (1) を、

【 0 1 2 1 】

【 数 1 3 】

$\ell = \{ \langle A, -, 5 \rangle, \langle B, A, 2 \rangle, \langle C, B, 3 \rangle, \langle E, C, 7 \rangle, \langle D, B, 3 \rangle, \langle C, D, 1 \rangle, \langle E, C, 5 \rangle \}$

20

【 0 1 2 2 】

と表すことができる。

【 0 1 2 3 】

前述のタイプの表現で、クリックストリームパスツリー内のパスは直観的に示されない。

【 0 1 2 4 】

ステップ S_{13} は、クリックストリームパスツリーをマージする。特にシステムは、利用可能なクリックストリームパスツリーをマージし、加重を構成し、ユーザウェブサイトクリックパスについて加重有向グラフモデルを構成する。

30

【 0 1 2 5 】

クリックストリームパスツリーのマージは、全てのノードに対して行われることに注目すべきである。同一のノードを有するクリックストリームパスツリーを、これらの同一のノードによってマージしてもよい。共通のノードを共有しないクリックストリームパスツリーについては、検討する必要があるツリーとツリーの関係がないので、マージはさらに単純である。しかしこれは、クリックストリームパスツリーのマージのただ 1 つの特殊なケースである。

【 0 1 2 6 】

例えば、セッション S_2 が 3 つのクリックストリームパスツリー、具体的には、 $A \rightarrow B \rightarrow C \rightarrow D$ 、 $A \rightarrow C \rightarrow D$ 及び $H \rightarrow J$ を有すると仮定する。ここで $A \sim D$ 、 H 及び J はそれぞれ異なるウェブページを示し、一方で矢印はそれぞれパスを示す。パスツリー $A \rightarrow B \rightarrow C \rightarrow D$ 及び $A \rightarrow C \rightarrow D$ をマージしてもよいが、一方で $H \rightarrow J$ はマージ後も同じままである。上記処置結果が、図 3 に示される。

40

【 0 1 2 7 】

異なる観点から、クリックストリームパスツリーに関して統計分析を行うことが可能である。例えば 1 つは、ユーザウェブサイトクリックステータスに関して行ってもよく、他はウェブページクリック方向に関してよい。前者はグラフ内で点で表され、一方で後者はグラフ内で角で表される。

【 0 1 2 8 】

ユーザウェブサイトクリックステータスは、ユーザまたは複数のユーザによる、あるウ

50

ウェブページのページビューの数 p 、及びウェブページの滞在の長さの平均を通常計算する。直観的に、ページビューの数 p 、及び滞在の長さは両方とも、ウェブページ内のユーザの興味のレベルを反映する。 p が大きいほど、及び滞在の長さが長いほど、ウェブページ内のユーザの興味が大きいことを示し得る。

【0129】

ウェブページクリック方向を、クリック傾向及びその割合を記載するために使用してもよい。カレントページの訪問後、ユーザは退出するかもしれないし、またはこのページ及び他のページ上でクリックし続けるかもしれない。そのような指向情報は、通常有益である。

【0130】

ステップS14は、それぞれのマージされたノードに階層加重を加える。それぞれの結び付けられたノードへの階層加重を、ノードの滞在の長さの平均により決定してもよく、これは、ノードの滞在の長さの合計をノードのページビューの総数で除算することによって計算され、当該合計は複数のクリックストリームにわたって加算される。

【0131】

ステップS15は、それぞれのマージされたパスに比率加重を加える。それぞれのマージされたパスへの比率加重は、そのページビュー数 p 、具体的にはマージされたパス内のカレントページのソースページのページビューの総数を、ソースページからクリックされたカレントページのページビューの総数で除算することにより決定される。

【0132】

例えば、マージされたパスがソースページN及びカレントページMを有すると仮定する。ソースページNのページビューの総数 p が10であり、ソースページNからクリックされたカレントページMのビューの総数が5である場合、そのときソースページN及びカレントページMを示すクリックパスは、 $5 / 10 = 0.5$ の比率加重を有する。

【0133】

カレントページは、複数のソースページを有するかもしれない。しかし、1つのパスのみが、それぞれのソースページからカレントページまで存在してもよい。すなわち、それぞれのパスは、特定の対のカレントページ及びソースページに対応する。

【0134】

ユーザのクリック状況を、

【0135】

【数14】

$$U = \{\dots, \langle u_i, pv_i, \bar{t}_i \rangle, \dots, \langle u_j, pv_j, \bar{t}_j \rangle, \dots\} \quad \text{式 (2)}$$

【0136】

と記述することができる。

【0137】

式中、 u_i は i 番目のカレントページであり、 u_i 及び i は両方とも正整数であり、

pv_i は u_i ページビューの合計で、 \bar{t}_i は u_i の滞在の長さの平均である。

そして

【0138】

【数15】

$$T = \{\dots, \langle u_i, u_j, pv_{ij} \rangle, \dots, \langle u_i, u_k, pv_{ik} \rangle, \dots, \langle u_j, u_k, pv_{jk} \rangle, \dots\} \quad \text{式 (3)}$$

【0139】

式中、 $\langle u_i, u_j, pv_{ij} \rangle$ は、ソースページ u_j からカレントページ u_i へクリックされたページビューの合計が pv_{ij} であることを示す。

10

20

30

40

50

【 0 1 4 0 】

式 (2) を用いて前述のセッション S 1 を、

【 0 1 4 1 】

【 数 1 6 】

$$U = \{ \langle u_A, 1, 5 \rangle, \langle u_B, 1, 2 \rangle, \langle u_C, 2, 2 \rangle, \langle u_D, 1, 3 \rangle, \langle u_E, 2, 6 \rangle \}$$

【 0 1 4 2 】

によって表すことができる。

【 0 1 4 3 】

式 (3) を用いてセッション S 1 を、

【 0 1 4 4 】

【 数 1 7 】

$$T = \{ \langle u_A, u_B, 1 \rangle, \langle u_B, u_C, 1 \rangle, \langle u_C, u_E, 2 \rangle, \langle u_D, u_C, 1 \rangle \}$$

【 0 1 4 5 】

によって表すことができる。

【 0 1 4 6 】

個々のユーザが個々の読む速度を有することを考えると、ある人は速くある人は遅く、滞在の長さの平均はユーザによって階層的に加重及び順位付けされる。例えば、K 個の異なる階層順位 (K は正整数) を使用してもよい。K の例は、7 及び 1 3 である。

【 0 1 4 7 】

従って、式 (2) は、下記

【 0 1 4 8 】

【 数 1 8 】

$$U = \{ \dots, \langle u_i, pv_i, \lambda_i \rangle, \dots, \langle u_j, pv_j, \lambda_j \rangle, \dots \} \quad \text{式 (4)}$$

【 0 1 4 9 】

におそらく変換される。式中、 u_i は i 番目のカレントページ、 p_i は u_i の合計ページビュー、 λ_i は u_i の階層加重である。

【 0 1 5 0 】

— 実施形態において、前述のステップ S 1 4 は、以下のサブステップを有してもよい。

【 0 1 5 1 】

サブステップ S 1 4 1 は、ノードの滞在の長さの平均により分けられた複数の加重レベルを用いてノードを順位付けする。

【 0 1 5 2 】

サブステップ S 1 4 1 は、それぞれの加重レベルに階層加重を加える。

【 0 1 5 3 】

例えば、前述のセッション S 1 のクリックストリームパスツリーを、 $K = 5$ の加重レベルに分けてもよく、それぞれの与えられた階層加重は以下の通りである。

【 0 1 5 4 】

【 数 1 9 】

$$\lambda_1 = 1.02, \lambda_2 = 1.01, \lambda_3 = 1, \lambda_4 = 0.99, \lambda_5 = 0.98$$

【 0 1 5 5 】

実際には、クリックストリームデータポイントは、その滞在の長さの平均に基づいて最長から最短へ順に並べられ、5 つの領域に均等に分けられ、それぞれの領域が加重レベルに対応する。第 1 の領域内のウェブサイトは階層加重 λ_1 を与えられ、第 2 の領域内のウェブサイトは階層加重 λ_2 を与えられる等である。

【 0 1 5 6 】

このケースでは、セッション S 1 を、式 (4) を用いて

10

20

30

40

50

【 0 1 5 7 】

【数 2 0】

$$U = \{ \langle u_A, 1, 1.01 \rangle, \langle u_B, 1, 0.99 \rangle, \langle u_C, 2, 0.99 \rangle, \langle u_D, 1, 1 \rangle, \langle u_E, 2, 1.02 \rangle \}$$

【 0 1 5 8 】

と表すことができる。

【 0 1 5 9 】

比率加重を構成するとき、カレントウェブページ u_i 、おそらくクリックストリームパスツリー内の子ノード、及び直接訪問可能であることを考慮に入れてもよい。このケースでは、 $\sum pv_{ij} \leq pv_j$ である。従って、 pv_{ij} / pv_j は u_j から u_i へ転換するクリック比率であり、比率は δ_{ij} 、すなわち比率加重で表され、 pv_j は u_j の合計ページビューである。このケースでは、 $\delta_{ij} \leq 1$ である。 $i = j$ である特殊な状況は、 u_i から u_i 自身へ向けられたクリックを示し、比率は δ_{ij} で表される。

10

【 0 1 6 0 】

ステップ S 16 は、加重有向グラフを生成する。

【 0 1 6 1 】

一実施形態においては、ウェブページのクリックステータスを、式 (3) 及び式 (4) を組み合わせることにより、

20

【 0 1 6 2 】

【数 2 1】

$$U = \{ \dots, \langle u_i, \lambda_i \rangle, \dots, \langle u_j, \lambda_j \rangle, \dots \} \quad \text{式 (5)}$$

【 0 1 6 3 】

式中、 $\langle u_i, \lambda_i \rangle$ はウェブページ u_i (訪問中のカレントウェブページ) のユーザクリック状況、 λ_i は階層加重、 u_i は i 番目のカレントページ、 λ_i は u_i の階層加重である、

及び、

30

【 0 1 6 4 】

【数 2 2】

$$T = \{ \dots, \langle u_i, u_j, \delta_{ij} \rangle, \dots, \langle u_i, u_k, \delta_{ik} \rangle, \dots \} \quad \text{式 (6)}$$

【 0 1 6 5 】

式中、 $\langle u_i, u_j, \delta_{ij} \rangle$ はウェブページ u_j からウェブページ u_i へクリックするユーザのクリックパスであり、 u_i はカレントウェブページであり、 u_j はソースウェブページ、 δ_{ij} は比率加重である。 δ_{ij} は、ソースウェブページ u_j からのクリックの比率 δ_{ij} はカレントウェブページ u_i に向けられたことを示す、と表すことができる。

40

【 0 1 6 6 】

ユーザのウェブページ加重有向グラフを、前述の式 (5) 及び式 (6) を用いて生成してもよい。例えば、前述のセッション S 1 を、式 (5) を用いて

【 0 1 6 7 】

【数 2 3】

$$U = \{ \langle u_A, 1.01 \rangle, \langle u_B, 0.99 \rangle, \langle u_C, 0.99 \rangle, \langle u_D, 1 \rangle, \langle u_E, 1.02 \rangle \}$$

【 0 1 6 8 】

と表すことができる。

【 0 1 6 9 】

50

セッション S 1 を、式 (6) を用いて

【 0 1 7 0 】

【 数 2 4 】

$$T = \{ \langle u_A, -, 1 \rangle, \langle u_B, u_A, 1 \rangle, \langle u_C, u_B, 0.5 \rangle, \langle u_D, u_B, 0.5 \rangle, \langle u_C, u_D, 1 \rangle, \langle u_E, u_C, 1 \rangle \}$$

【 0 1 7 1 】

と表すことができる。

【 0 1 7 2 】

図 4 は、上記の記述の結果としてのセッション S 1 のウェブページ加重有向グラフを示す。ウェブページ加重有向グラフは、構成された階層加重及び比率加重の両方を有する。 10

【 0 1 7 3 】

前述のように、ユーザのクリックストリームデータを、前述の式 (5) 及び式 (6) の記述に基づいて、加重有向グラフモデルに変換可能である。個々のユーザは、ユーザの自身のクリックストリームデータから変換された個々の加重有向グラフを有し得る。結果として、2人のユーザの相関の計算の数量が、同一のユーザの加重有向グラフの相関の計算の数量に変換される。

【 0 1 7 4 】

ステップ S 1 7 で、システムはカレントユーザと最高数の共通パスを共有する Y 人の他のユーザを選択し、Y は正整数である。2人のユーザがそれぞれ同一のカレントウェブページ及び同一のソースウェブページを有するパスを持つとき、その2人のユーザは共通パスを共有するとみなされる。 20

【 0 1 7 5 】

一実施形態においては、 $Y = 3N$ であり、N は正整数であり、この意味は後述の記載で明らかになる。

【 0 1 7 6 】

図 5 は、前述の他のセッション、すなわちセッション S 3 のウェブページ加重有向グラフを示す。セッション S 3 の加重有向グラフとセッション S 1 の加重有向グラフとの間の比較は、2つのグラフが3つの共通パス、A B、B D 及び D C を共有することを示す。

【 0 1 7 7 】

ステップ S 1 8 は、ユーザの加重有向グラフを使用して、カレントユーザと前述の選択された Y 人の他のユーザの間のウェブページ訪問パス相関を計算する。 30

【 0 1 7 8 】

加重有向グラフの相関の計算の第 1 のタスクは、グラフが共有する共通パスの数を見つけることである。2人の異なるユーザについて、彼らが共有する共通パスが多いほど、彼らの加重有向グラフ間の重複が多く存在し、その2人のユーザ間の高い相関を示唆する。同時に、共通パス及びそれらのノードの加重が高いほど、ユーザ相関は高いであろう。

【 0 1 7 9 】

好適な一実施形態においては、カレントユーザと前述の選択された Y 人の他のユーザの間のクリックパス相関は、以下の式を用いて計算される。 40

【 0 1 8 0 】

【 数 2 5 】

$$correlation = \frac{\sum \text{共通パスの最少加重}}{\sum \text{比共通パスの加重} + \sum \text{共通パスの最大加重}}$$

【 0 1 8 1 】

式中、共通パスの最小加重は、共通パスの比率加重の最小値 × 共通パスに示されるカレントウェブページの階層加重の最小値 × 共通パスに示されるソースウェブページの階層加重の最小値の乗算の積である。共通パスの最大加重は、共通パスの比率加重の最大値 × 共通パスに示されるカレントウェブページの階層加重の最大値 × 共通パスに示されるソース 50

ウェブページの階層加重の最大値の乗算の積である。非共通パスの加重は、非共通パスの比率加重の値×非共通パスに示されるカレントウェブページの階層加重の値×非共通パスに示されるソースウェブページの階層加重の値の乗算の積である。

【0182】

2人のユーザの共通パスは異なる比率加重を有してもよく、2人の異なるユーザの共通パスのカレントウェブページ及びソースウェブページもまた異なる階層加重を有してもよいことに留意すべきである。相関を計算するとき、比率加重及び階層加重はフェイスバリュとみなされ、ユーザの加重有向グラフが属する特定の値は特に留意する必要はない。

【0183】

図4及び図5を参照すると、2人のユーザ間のウェブページ訪問相関は以下のように計算される。

【0184】

共通パス：A B、B D及びD Cである。

【0185】

共通パスの最小加重は、 $0.5 \times 1 \times 0.99 + 0.5 \times 0.99 \times 1 + 0.5 \times 1 \times 0.99 = 1.485$ と計算される。

【0186】

共通パスの最大加重は、 $1 \times 1.01 \times 0.99 + 1 \times 0.99 \times 1.02 + 1 \times 1.02 \times 1.01 = 3.05$ と計算される。

【0187】

非共通パス：図4はB C、C Eを有し、一方で図5はD A、A Cを有する。

【0188】

図4の非共通パスの加重は、 $0.5 \times 0.99 \times 0.99 + 1 \times 0.99 \times 1.02 = 1.500$ と計算される。

【0189】

図5の非共通パスの加重は、 $0.5 \times 1.02 \times 1 + 0.5 \times 1 \times 1.01 = 1.015$ と計算される。

【0190】

従って、図4及び図5の加重有向グラフ間の相関は、以下ようになる。

$1.485 / (1.500 + 1.015 + 3.05) = 0.267$

【0191】

ウェブページ訪問パス相関を計算する前述のプロセスは、例示の目的のみである。実際の適用の必要によって、他の適切なパス相関の計算方法もまた使用してもよい。関連後のそのような計算方法の例は、コサイン相関計算、ピアソン相関計算等を含む。この点で、本開示は限定を置かない。

【0192】

ブロック103で、システムはカレントユーザとのクリックパス相関が最も高く順位付けられたX人の他のユーザを選択し、Xは正整数である。

【0193】

一実施形態においては、 $X = 2N$ であり、Nは正整数である。X人の他のユーザは予め選択されたY人の他のユーザから選択されるので、通常、数Xは、数Y（前述の例で $Y = 3N$ ）より小さい。

【0194】

ブロック104でシステムは、選択されたX人の他のユーザが訪問したウェブページの予め設定されたタグに関連する総合加重を構成する。

【0195】

実際には、ウェブページタグライブラリを作成して、種々のウェブページのタグを収集してもよい。それぞれのウェブページは1または複数のタグを有してもよく、それぞれの予め設定されたタグは自身の初期加重を有してもよい。

【0196】

10

20

30

40

50

ウェブページを、タグリストによって以下のように説明し得る。

【0197】

【数26】

$$\mathcal{G} = \{(\sigma_1, wp_1) \cdots (\sigma_i, wp_i) \cdots (\sigma_{g_n}, wp_{g_n})\} \text{ 式 (7)}$$

【0198】

式中、タグリスト \mathcal{G} は g_n 個のタグの合計を有し、 σ_i は i 番目のタグで $1 \leq i \leq$

g_n であり、 g_n 及び i は両方とも正整数であり、 wp_i はタグ σ_i の初期加重である

10

。

【0199】

タグの項目は、2つの構成要素、タグ及びタグ加重を含む。タグは、ウェブページのタイプ、キーワード、ウェブページ分類、ID等のインジケータであってもよい。ここでウェブページのタイプはホームページ、ナビゲーションページ、検索ページ、商品ホームページ、企業ホームページ、商品記載ページ、注文ページ、ログオンページ等であってもよく、IDは商品ID、企業ID、(化粧品のよう)な提供物カテゴリ等であってもよい。

【0200】

前述に挙げたタグは、例示のみである。実際には、タグがウェブサイトの独自性を識別する限り、他のどのようなタグを使用してもよい。この点で、本開示は限定を設けない。

20

【0201】

販売ルータの電子商取引ウェブページを取り上げると、以下のタグを紹介し得る。

【0202】

商品モデル = DIR - 616

商品ブランド = D - Link

商品タイプ = ルータ

商品ID = 22540068533

販売者名 = dlink specialty store

30

【0203】

上記において、等号の左側はタグの名前であり、右側はタグの内容である。

【0204】

結果として、前述のウェブサイトは、

【0205】

【数27】

$$\mathcal{G} = \{("DIR-616", 1) ("D-Link", 0.5) ("router", 0.3) ("22540068533", 1) ("dlink specialty store", 0.7)\}$$

【0206】

と表すことができる。

40

【0207】

ここで、商品IDの初期加重及び商品モデルの初期加重は両方とも1であり、商品ブランドの初期加重は0.5であり、商品タイプルータの初期加重は0.3であり、一方で販売者名の初期加重は0.7である。

【0208】

適用状況により、初期加重を技術者が決定してもよいことが理解される。この点で、本開示は限定を設けない。

【0209】

また、ウェブページは総合ページ加重を有してもよい。

【0210】

50

ウェブページを表すのに u_i を用いて、このウェブページをウェブページ加重及びタグリストにより表してもよい。

【 0 2 1 1 】

【 数 2 8 】

$$u_i = \{w_i, g_i\} \text{ 式 (8)}$$

【 0 2 1 2 】

式中、 w_i はウェブページ u_i のページ加重であり、 g_i はタグリストである。

10

【 0 2 1 3 】

タグの総合加重は、以下の式を用いて計算される。

【 0 2 1 4 】

【 数 2 9 】

$$\bar{w} = \sum_{i=0}^{\sigma k} (w_i \times pv_i \times wp_i) / \sum_{i=0}^{\sigma k} pv_i$$

【 0 2 1 5 】

式中、 \bar{w} は予め設定されたタグの総合加重であり、 σk は予め設定されたタグが出現するウェブページの総数であり、 w_i は i 番目のウェブページのページ加重であり、 pv_i は i 番目のウェブページのページビューの総数であり、 wp_i は i 番目のウェブ P A G 内の予め設定されたタグの初期加重である。

20

【 0 2 1 6 】

ブロック 1 0 5 でシステムは、予め設定されたタグ及び総合加重に基づいて、カレントユーザと X 人の他のユーザとの間のユーザ相関を計算する。

【 0 2 1 7 】

式 (8) を式 (7) に代入し、式 (2) と組み合わせると k 次元のタグ加重ベクトル空間が得られる。

【 0 2 1 8 】

【 数 3 0 】

$$\bar{Y} = \{(\sigma_1, \bar{w}_1), (\sigma_2, \bar{w}_2), \dots, (\sigma_k, \bar{w}_k)\}$$

【 0 2 1 9 】

式中、 k は正整数であり、 σ はタグであり、 \bar{w} はそれぞれのタグの総合加重である

40

【 0 2 2 0 】

それぞれのタグは、 k 次元のタグ加重ベクトル空間内の次元である。 k 次元のタグ加重ベクトル空間内のユーザは、ベクトルがそれぞれのタグ次元内で連動するとき、総合加重を有するベクトルで表される。ユーザ間の相関の計算は、タグ加重ベクトル空間内で表現するベクトル間の距離の計算に変換される。

【 0 2 2 1 】

一実施形態においては、カレントユーザと X 人の他のユーザとの間のユーザ相関は、以下の設問を用いて予め設定されたラベル及び総合加重に基づいて計算される。

【 0 2 2 2 】

50

【数 3 1】

$$Sim(i, j) = \frac{K(\bar{Y}_i, \bar{Y}_j)}{\sqrt{K(\bar{Y}_i, \bar{Y}_i)}\sqrt{K(\bar{Y}_j, \bar{Y}_j)}} = \frac{\sum_{n=1}^k (\bar{w}_{in} \times \bar{w}_{jn})}{\sqrt{\sum_{n=1}^k \bar{w}_{in}^2} \sqrt{\sum_{n=1}^k \bar{w}_{jn}^2}}$$

【0 2 2 3】

式中、Sim (i , j) はユーザ i とユーザ j との間の相関であり、

$\bar{Y} = \{(\sigma_1, \bar{w}_1), (\sigma_2, \bar{w}_2), \dots, (\sigma_k, \bar{w}_k)\}$ σ はタグであり、 \bar{w} は対応するタグの総合加重である。

10

【0 2 2 4】

前述のユーザ相関の計算例は、例示の目的のみである。実際の適用の必要により、他の適切なユーザ相関計算方法もまた使用してもよい。関係後のそのような計算方法例は、コサイン相関計算、ピアソン相関計算等を含む。この点で、本開示は限定を設けない。

【0 2 2 5】

ユーザ相関を拡張すると、異なる使用の関心の間の類似性を示す。ユーザ相関が高いと、ユーザの関心の間の類似性のレベルが高いことを示す。

20

【0 2 2 6】

本開示は、クリックストリームデータを使用してユーザウェブサイトクリックパスの加重有向グラフモデルを構成し、ユーザ相関の計算を異なるユーザの加重有向グラフの類似性の計算に変換する。本開示はさらに、ウェブページタグの内容相関の計算をユーザ相関に組み込むのにウェブページタグライブラリを導入する。開示された方法は、ユーザのクラスタ化の精度及び効率を向上するため、ユーザのクリック習性及び個別の行動性向を効果的に発見し、次にユーザの推奨の成功率及び効率を向上する。

【0 2 2 7】

クリックストリームパスツリー内のノードに階層加重を構成すると、個々のユーザの読み出し習性及び能力の影響を減少し、個人化の精度を上げる。

30

【0 2 2 8】

図 6 は、ユーザの行動に基づいて個人化された推奨を作成するプロセスのブロックフローチャートである。上記プロセスは、以下のブロックに記載される。

【0 2 2 9】

ブロック 6 0 1 でシステムは、ユーザ識別子を含むユーザ情報を取得する。

【0 2 3 0】

ブロック 6 0 2 でシステムは、そのユーザとのユーザ類似性が最も高く順位付けられた Z 人の他のユーザを選択し、Z は正整数である。

【0 2 3 1】

ブロック 6 0 3 でシステムは、選択された Z 人の他のユーザの情報に基づいてそのユーザに推奨を作成し、ユーザ類似性は以下に記載されるプロセスにより生成される。

40

【0 2 3 2】

ユーザのウェブページクリックストリームデータを取得する。

【0 2 3 3】

クリックストリームデータを用いてユーザと他のユーザの間のクリックパス相関を計算する。

【0 2 3 4】

カレントユーザとのクリックパス相関が最高水準順位の X 人の他のユーザを選択し、X は正整数である。

【0 2 3 5】

50

選択された X 人の他のユーザが訪問したウェブページの予め設定されたタグのそれぞれに関連する総合加重を構成する。

【 0 2 3 6 】

予め設定されたタグ及び総合加重に基づいてカレントユーザと X 人の他のユーザとの間のユーザ相関を計算する。

【 0 2 3 7 】

Z は通常 X より小さい。一実施形態においては、 $Z = N$ で、一方で $X = 2N$ であり、N は正整数である。

【 0 2 3 8 】

ユーザのクラスタ化を、ユーザ情報内へ予め構成してもよい。ユーザがウェブサイトを訪
10
訪問するとき、システムは取得されたユーザ ID に基づいて特定のユーザの個人化された
推奨情報を直接取得してもよく、よって計算リソースの使用効率及び個人化された推奨の
効率を向上する。

【 0 2 3 9 】

記載のように、本開示による方法はまず、カレントユーザと最高数の共通パスを共有す
る Y 人の他のユーザを選択し、カレントユーザと選択された他のユーザとの間のクリック
パス相関を計算し、次いでカレントユーザとの最も高いクリックパス相関を有する X 人の
ユーザを選択し、次いでカレントユーザとの最も高いユーザ相関を有する Z 人のユーザを
20
選択する。その選択はレイヤ構造であり、それぞれのレイヤが計算に最も関連する情報を
抽出する。このアプローチは精度を上げ、データ処理要求及び計算資源の使用量を減らす
。

【 0 2 4 0 】

実際の適用で、カレントユーザとの最も高い相関を有する選択された X 人の他のユーザ
に利用可能な推奨情報に基づいて、友達推奨、ブログ推奨、製品推奨等を作成してもよい
。

【 0 2 4 1 】

種々の予め設定された閾値のような規則及びパラメータを、適用の実際の必要に基づい
て技術者が設定してもよい。例えば、ユーザ相関を時折、または周期的に計算してもよく
、この計算は利用可能なクリックストリームデータの全てまたは一部のいずれかに基づい
てもよい。数人のユーザのクリックストリームデータを、特定の時間間隔または複数の時
30
時間間隔の間に収集されたデータに限定してもよい。この点で、本開示は限定を設けない。

【 0 2 4 2 】

いくつかの実施形態では、現在信頼できないユーザ及び信頼できるユーザを記録するた
めにシステムはブラックリスト及びホワイトリストを有してもよい。ブラックリスト及び
ホワイトリストは、推奨できるまたは推奨できないユーザ、ブログまたは製品を含んでも
よい。

【 0 2 4 3 】

システムは、ユーザに向けてシステムが作成した推奨へのユーザのフィードバックを受
40
信し、フィードバックを使用して初期加重、ウェブサイト加重、及び/または階層加重を
調整してもよい。

【 0 2 4 4 】

ユーザクラスタ化空間内の特徴次元の抽出に使用される方法はユーザ行動の特徴の抽出
に使用される方法と同様であり、そのような方法は本開示では重複して記載せず、本明細
書の関連する記載を参照することができる。

【 0 2 4 5 】

前述の技術を、コンピュータ実行可能な命令を有する 1 または複数のコンピュータ可読
媒体の補助のもとで実行してもよい。コンピュータ実行可能な命令は、コンピュータプロ
セッサが本明細書に記載された技術により動作を実行するのを可能にする。コンピュータ
可読媒体は、コンピュータデータを格納するために適切等のようなメモリ装置であっても
50
よいことが理解される。そのようなメモリ装置は、ハードディスク、フラッシュメモリ装

置、光学データ記憶装置、及びフロッピー（登録商標）ディスクを含むが、それらに限定されない。さらに、コンピュータ実行可能な命令を有するコンピュータ可読媒体はローカルシステム内の構成要素からなってもよく、または複数のリモートシステムのネットワーク上に分散された構成要素からなってもよい。コンピュータ実行可能な命令のデータを、有形の物理メモリ装置内に配信、または電子的に送信のいずれかでもよい。本明細書で定義するように、コンピュータ可読媒体は、変調されたデータ信号及び搬送波等のような一時的な媒体を含まない。

【0246】

本明細書中に開示される方法に関連して、本開示はまた、本明細書に記載された方法を実行するコンピュータベースの装置を提供する。

10

【0247】

本開示で、「モジュール」は一般に特定のタスクまたは機能を実行するよう設計された機能をいう。モジュールは、特定のタスクまたは機能に関連する目的を達成するハードウェア、ソフトウェア、計画またはスキームの一部、またはそれらの組み合わせであることが可能である。また、分離したモジュールの説明は、物理的に分離した装置が使用されることを必ずしも示唆しない。代わりに上記説明は機能のみであってもよく、いくつかのモジュールの機能を単一の組み合わせた装置または構成要素により実行してもよい。コンピュータベースのシステムで使用されるとき、プロセッサ、記憶装置及びメモリのような標準のコンピュータ構成要素を、種々のそれぞれの機能を実行する1または複数のモジュールとして機能するようプログラムしてもよい。

20

【0248】

図7は、双方向通信での情報交換方法を実行するサーバシステムの機能ブロックの概略図である。

【0249】

サーバコンピュータシステム700は、1または複数のプロセッサ790、I/O装置792、アプリケーションプログラム（複数可）780を格納するコンピュータ可読メモリ及び記憶媒体794を有する一般的なスマートフォンハードウェアに基づいてよい。サーバコンピュータシステム700は、以下の機能モジュールを有するようプログラムされる。

【0250】

30

クリックストリームデータ獲得モジュール701は、カレントユーザのウェブページクリックストリームを取得するようプログラムされる。

【0251】

クリックパス関連計算モジュール702は、クリックストリームデータを用いてカレントユーザと他のユーザとの間のクリックパス関連を計算するようプログラムされる。

【0252】

ユーザ選択モジュール703は、何故ならカレントユーザとのクリックパス関連が最も高く順位付けられたX人の他のユーザを選択することをプログラムされ、Xは正整数である。

【0253】

40

総合加重構成モジュール704は、選択されたX人の他のユーザが訪問したウェブページの予め設定されたタグのそれぞれに関連する総合加重を構成するようプログラムされる。

【0254】

ユーザ関連計算モジュール705は、予め設定されたタグ及び総合加重に基づいてカレントユーザとX人の他のユーザとの間のユーザ関連を計算するようプログラムされる。

【0255】

実際には、サーバ上で利用可能なクリックストリームデータは複数のクリックストリームデータポイントを有してもよく、それらはそれぞれ、ユーザ識別子、カレントウェブページ、ソースウェブページ、クリック時間、及び滞在の長さ等を有してもよい。

50

【 0 2 5 6 】

モジュールは、本開示に記載された方法の機能を実行するようプログラムされる。さらに前述のモジュールは、開示された方法の文脈で本明細書に記載した種々の機能を実行するようプログラムされたサブモジュールを有してもよい。これらのモジュール及びサブモジュールの詳細は、繰り返さない。

【 0 2 5 7 】

例えば、クリックパス関連計算モジュール 7 0 2 は、以下のサブモジュールを有してもよい。

【 0 2 5 8 】

セッション分割サブモジュールは、クリックストリームデータポイントを 1 または複数のセッションに分割する。セッションのそれぞれのクリック時間によって順に分類された任意の 2 つの近接したセッションのクリック時間の差は、第 1 の予め設定された時間閾値より少ないかあるいは等しい。特に、クリックストリームデータポイントを、それぞれのクリック時間の順によって順に並べてもよい。2 つの連続したクリックストリームデータポイントのクリック時間の差が第 1 の予め設定された時間閾値より少ないかあるいは等しい場合、その 2 つのクリックストリームデータポイントを 1 つのセッションにマージしてもよい。

10

【 0 2 5 9 】

クリックストリームのこれまでの 3 つの確立されたサブモジュールは、それぞれのセッションについて、カレントウェブページ及びソースウェブページを用いてクリックストリームパスツリーを確立する。クリックストリームデータポイントパスフリーは、ノード及びパスを有する。ノードはカレントウェブページであり、一方でパスはカレントウェブページとそのソースウェブページの間に関連を示す。

20

【 0 2 6 0 】

クリックストリームパスツリーマージサブモジュールは、利用可能なクリックストリームパスツリーをマージし、加重を構成し、及びユーザウェブサイトクリックパスの加重有向グラフモデルを構成する。

【 0 2 6 1 】

階層加重割り当てサブモジュールモジュールを、ノードの滞在の長さの平均によって分割された複数の加重レベルのうちの一つにノードを配置することにより、及びノードが配置された加重レベルによりノードに階層加重を加えることにより、それぞれのマージされたノードに階層加重を加えるようプログラムしてもよい。

30

【 0 2 6 2 】

加重有向グラフ生成サブモジュールは、加重有向グラフを生成する。

【 0 2 6 3 】

第 2 の選択サブモジュールは、カレントユーザと最高数の共通パスを共有する Y 人の他のユーザを選択し、Y は正整数である。

【 0 2 6 4 】

計算サブモジュールは、カレントユーザと選択された Y 人の他のユーザの間のクリックパス関連を計算する。

40

【 0 2 6 5 】

特に図 8 は、ユーザの行動に基づいて個人化された推奨を作成するシステムのブロック図を示す。推奨システム 8 0 0 は、1 または複数のプロセッサ 8 0 1、ネットワークインタフェース 8 0 2、メモリ 8 0 3 及び入力/出力 (I/O) インタフェース 8 0 4 を有してもよいが、それらに限定されない。メモリ 8 0 3 は、前述のようなコンピュータ可読媒体を有してもよい。メモリ 8 0 3 は、プログラムモジュール 8 0 5 及びプログラムデータ 8 0 6 を有してもよい。実施形態で、プログラムモジュール 8 0 5 は、ユーザ識別子を有し得るユーザ情報を獲得するようプログラムされたユーザ情報獲得モジュール 8 0 7、カレントユーザとの最も高い類似性を有する Z 人の他のユーザを選択し、Z は正整数であるユーザ選択モジュール 8 0 8、及び選択された Z 人の他のユーザの情報に基づいてカレン

50

トユーザに個人化された推奨を作成するようプログラムされた推奨モジュール 809 を有してもよい。

【0266】

ユーザの類似性は、本明細書に記載された方法を用いて、プログラムされたモジュール及びサブモジュールにより決定される。

【0267】

前述の装置の実施形態は、本明細書に記載された方法の実施形態に緊密に関連し、従って当該方法の実施形態の詳細な記載はまた装置の実施形態に適用可能であり、繰り返さない。

【0268】

本開示に記載された技術を、パーソナルコンピュータ、サーバコンピュータ、ハンドヘルド装置または持ち運び可能な装置、タブレット装置、マルチプロセッサシステム、マイクロプロセッサベースシステム、セットトップボックス、プログラム可能な消費者装置、ネットワークPC、マイクロコンピュータ及び大規模メインフレームコンピュータ、または前述の例の1または複数を含む任意の分散環境を含むが、それらに限定されない、汎用計算装置または環境、または専用計算装置または環境内で実行してもよい。

【0269】

特にモジュールを、機械で実行可能なコマンド及びコードに基づいてコンピュータプログラムモジュールを用いて実行してもよい。一般に、コンピュータプログラムモジュールは、特定のタスクを実行、または特定の抽象データ型のルーチン、プログラム、オブジェクト、構成要素、データ構造等を実施し得る。本開示に記載された技術を、分散コンピューティング環境で実践することもまた可能であり、そのような分散コンピューティング環境は、通信ネットワークを通して接続されたりリモート処理装置によりタスクを実行する。分散コンピューティング環境で、プログラムモジュールを、メモリ装置を含むローカルまたはリモートコンピュータ記憶媒体のいずれかに設けてもよい。

【0270】

本明細書の種々の実施形態を、例及び環境と共に徐々により詳細に記載した。それぞれの実施形態は本開示の特定の態様に焦点を当ててもよく、従って異なる実施形態は互いに異なり得るが、同様の部分を共有してもよい。

【0271】

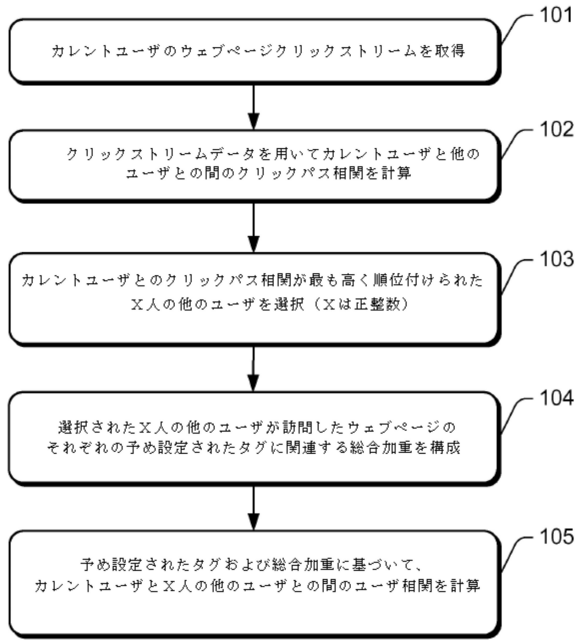
情報検証の方法及び装置を、本開示内の前述の開示で詳細に記載した。本開示において、本発明の概念及び実施を説明するために、例示的な実施形態を採用している。例示的な実施形態は、本開示の方法及び主要な概念のより明瞭な理解のためにのみ使用される。本開示の概念に基づいて、当業者は例示的な実施形態及び適用分野を変更することができる。

10

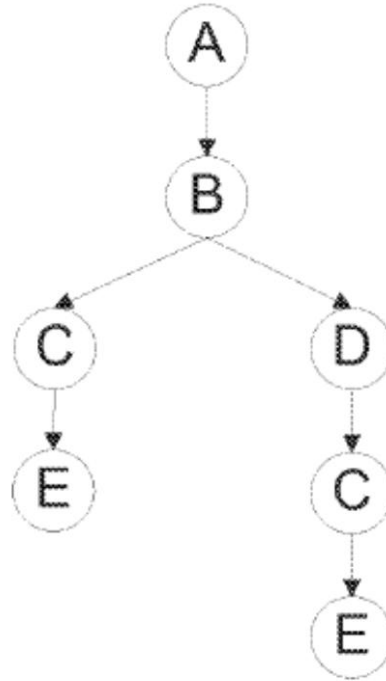
20

30

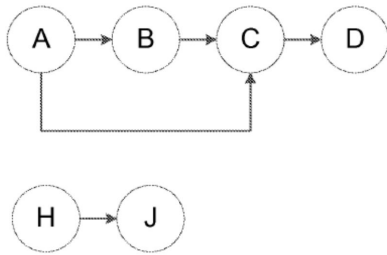
【図1】



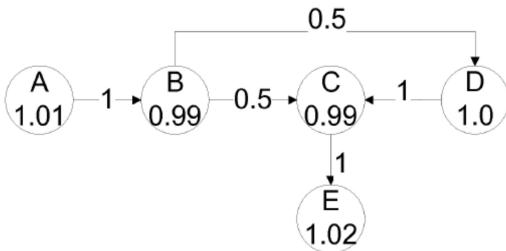
【図2】



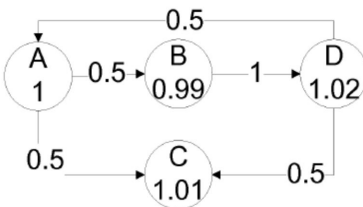
【図3】



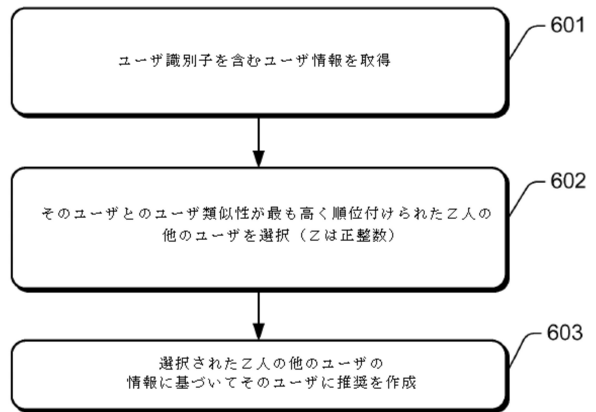
【図4】



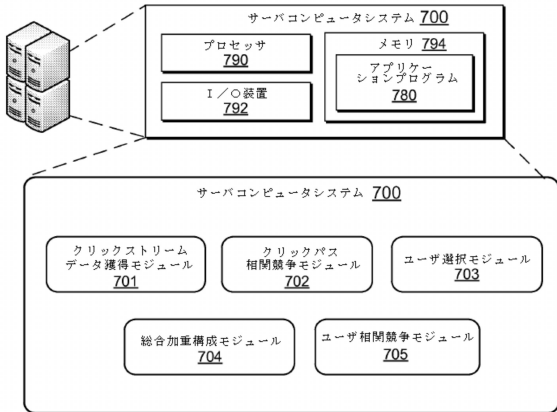
【図5】



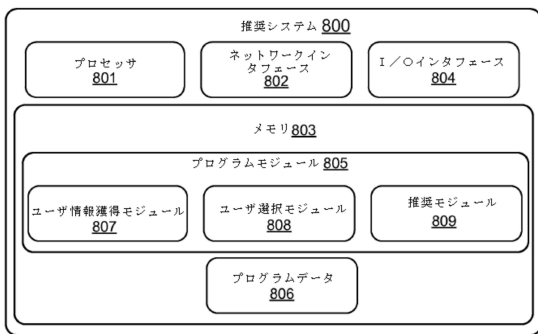
【図6】



【図7】



【図8】



フロントページの続き

(72)発明者 ピン チアオ

中華人民共和国 31112 ハンチョウ ユーハン ディストリクト ウェスト ウェン イ
ー ロード ナンバー 969 ビルディング 3 5 / エフ アリババ グループ リーガル
デパートメント内

(72)発明者 ミンガン ウー

中華人民共和国 31112 ハンチョウ ユーハン ディストリクト ウェスト ウェン イ
ー ロード ナンバー 969 ビルディング 3 5 / エフ アリババ グループ リーガル
デパートメント内

審査官 吉田 誠

(56)参考文献 特開2009-193465(JP,A)

特開2006-331014(JP,A)

特表2004-527019(JP,A)

特表2006-524009(JP,A)

(58)調査した分野(Int.Cl., DB名)

G06F 16/00