(19)日本国特許庁(JP)

再公表特許(A1)

(11)国際公開番号
WO2012/077818

発行日　平成26年5月22日(2014.5.22)

(43)国際公開日　平成24年6月14日(2012.6.14)

(51)Int.Cl.		F I		テーマコード（参考）
G06F 17/30	(2006.01)	G06F 17/30	350C	
		G06F 17/30	412	

審査請求 未請求　予備審査請求 未請求　（全 22 頁）

出願番号	特願2012-547940 (P2012-547940)
(21)国際出願番号	PCT/JP2011/078702
(22)国際出願日	平成23年12月12日(2011.12.12)
(31)優先権主張番号	特願2010-276013 (P2010-276013)
(32)優先日	平成22年12月10日(2010.12.10)
(33)優先権主張国	日本国(JP)

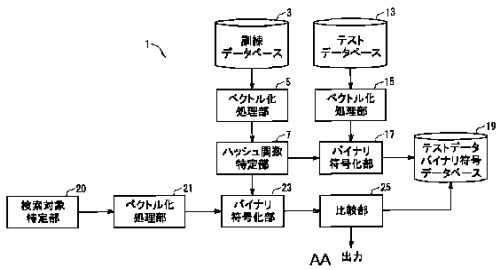
(71)出願人　304027349
国立大学法人豊橋技術科学大学
愛知県豊橋市天伯町雲雀ヶ丘１－１
(74)代理人　100095577
弁理士　小西　富雅
(72)発明者　青野　雅樹
愛知県豊橋市天伯町雲雀ヶ丘１－１　国立大学法人豊橋技術科学大学内
(72)発明者　立間　淳司
愛知県豊橋市天伯町雲雀ヶ丘１－１　国立大学法人豊橋技術科学大学内

最終頁に続く

(54)【発明の名称】ハッシュ関数の変換行列を定める方法、該ハッシュ関数を利用するハッシュ型近似最近傍探索方法、その装置及びそのコンピュータプログラム

(57)【要約】
インターネット上に存在するデータの量は拡大の一途をたどっているので、検索サイトに用いられる検索アルゴリズムには更なる精度の向上が期待されている。
データによる多様体上の局所的な近傍構造で表される非線形な関係を保持し、短いバイナリ符号に変換する新しいハッシュ型近似最近傍探索手法を提案する。

[図1]



3　Training database
5, 15, 21　Vectorization processing unit
7　Hash function designation unit
13　Test database
17, 23　Binary encoding unit
19　Test data binary code database
20　Search target designation unit
25　Comparison unit
AA　Output

【特許請求の範囲】
【請求項１】
データベースに含まれる第１のベクトルデータｘ（ｎ次元）をバイナリ符号であるｙ＝[y_1,y_2,…y_d]、ただし、ｎ＞＞ｄに変換するハッシュ型近似最近傍探索方法において、
下記式（Ａ）～式（Ｃ）より前記バイナリ符号ｙを得る下記ハッシュ関数ｈ（ｘ）に適用する変換行列ｒを定める方法であって、
【数６】

$$h_i(\mathbf{x}) = \mathrm{sign}(\mathbf{r}_i^T \mathbf{x}) \qquad \text{式（A）}$$

ここに、前記バイナリ符号ｙは次のように表わされる、
【数７】

$$\mathbf{y} = [y_1, y_2, ..., y_d]^T \qquad \text{式（B）}$$

【数８】

$$y_i = (1 + h_i(\mathbf{x}))/2 \qquad \text{式（C）}$$

前記第１のベクトルデータｘを前記バイナリ符号ｙのビット数である前記ｄの次元に射影したときの第２のベクトルデータｈを下記式（Ｄ）で規定したとき、
【数９】

$$\mathbf{h} = F^T(\mathbf{x} - \bar{\mathbf{x}}) \qquad \text{式（D）}$$

ただし、
【数１０】

$$\bar{\mathbf{x}} = \frac{1}{n}\sum_i \mathbf{x}_i$$

下記式（Ｅ）が最小になる変換行列Ｆを求め、この変換行列Ｆを式（Ａ）の変換行列ｒとする、
【数１１】

$$\Phi(W) = \sum_i ||\mathbf{h}_i - \sum_{j \in \mathcal{N}_i} w_{ij}\mathbf{h}_j||^2 \qquad \text{式（E）}$$

ただし、w_{ij}は前記データベースの第１のベクトルデータｘにつき、
【数１２】

$$\mathcal{E}(W) = \sum_i ||\mathbf{x}_i - \sum_{j \in \mathcal{N}_i} w_{ij}\mathbf{x}_j||^2 \qquad \text{式（F）}$$

式（Ｆ）を最小とする重みである、ハッシュ関数の変換行列を定める方法。
【請求項２】
前記変換行列Ｆを求める際に前記重みw_{ij}を正規化する、請求項１に規定の方法。
【請求項３】
前記正規化は、前記データベースにおける第１のベクトルデータｘでの近傍の分布を下記式（Ｇ）のスコアで定義し、
【数３５】

$$d_i = \sum_{j \in \mathcal{N}_i} \exp(-||\mathbf{x}_i - \mathbf{x}_j||/\sigma^2)$$

この分布スコアの対角行列Ｄ＝diag[d_1,……,d_m] をもとめて、グラフラプラシアン理論

に基づき行なう、請求項２に記載の方法。
【請求項４】
原データベースの原データから訓練データを抽出する訓練データ生成部と、
前記訓練データに基づき第１のハッシュ関数を特定するハッシュ関数特定部と、
前記第１のハッシュ関数を用いて前記原データベースの原データをバイナリ符号化するバイナリ符号化部と、
バイナリ符号化された前記原データを保存するバイナリ符号化データ保存部と、を備える検索サーバと、
入力された検索対象データを前記第１のハッシュ関数を用いてバイナリ符号化する第２のバイナリ符号化部を備えるクライアント端末と、
前記クライアント端末の第２のバイナリ符号化部でバイナリ符号化された検索対象データと前記バイナリ符号化された原データとを比較する比較部と、
を備える、ハッシュ型近似最近傍探索装置において、
前記第１のハッシュ関数は請求項１～請求項３にいずれかの方法で定められる、ハッシュ型近似最近傍探索装置。
【請求項５】
ハッシュ型近似最近傍探索装置に用いられる検索サーバであって、
原データベースの原データから訓練データを抽出する訓練データ生成部と、
前記訓練データに基づき第１のハッシュ関数を特定するハッシュ関数特定部と、
前記第１のハッシュ関数を用いて前記原データベースの原データをバイナリ符号化するバイナリ符号化部と、
バイナリ符号化された前記原データを保存するバイナリ符号化データ保存部と、を備え、
前記第１のハッシュ関数は請求項１～請求項３にいずれかの方法で定められる、検索サーバ。
【請求項６】
原データベースの原データから訓練データを抽出する訓練データ生成部と、
前記訓練データを第１のベクトル化方法に基づきベクトル化処理する第１のベクトル化処理部と、
ベクトル化処理された前記訓練データに基づき第１のハッシュ関数を特定するハッシュ関数特定部と、
前記原データを前記第１のベクトル化方法に基づきベクトル化処理する第２のベクトル化処理部と、
前記第１のハッシュ関数を用いて前記ベクトル化処理された原データをバイナリ符号化するバイナリ符号化部と、
バイナリ符号化された前記原データを保存するバイナリ符号化データ保存部と、を備える検索サーバと、
入力された検索対象データを前記第１のベクトル化方法に基づきベクトル化処理する第３のベクトル化処理部と、
前記ベクトル化処理された検索対象データを前記第１のハッシュ関数を用いてバイナリ符号化する第２のバイナリ符号化部を備えるクライアント端末と、
前記クライアント端末の第２のバイナリ符号化部でバイナリ符号化された検索対象データと前記バイナリ符号化された原データとを比較する比較部と、
を備える、ハッシュ型近似最近傍探索装置において、
前記第１のハッシュ関数は請求項１～請求項３にいずれかの方法で定められるハッシュ型近似最近傍探索装置。
【請求項７】
ハッシュ型近似最近傍探索装置に用いられる検索サーバであって、
原データベースの原データから訓練データを抽出する訓練データ生成部と、
前記訓練データを第１のベクトル化方法に基づきベクトル化処理する第１のベクトル化

処理部と、
ベクトル化処理された前記訓練データに基づき第１のハッシュ関数を特定するハッシュ関数特定部と、
前記原データを前記第１のベクトル化方法に基づきベクトル化処理する第２のベクトル化処理部と、
前記第１のハッシュ関数を用いて前記ベクトル化処理された原データをバイナリ符号化するバイナリ符号化部と、
バイナリ符号化された前記原データを保存するバイナリ符号化データ保存部と、を備え、
前記第１のハッシュ関数は請求項１～請求項３にいずれかの方法で定められる、検索サーバ。
【請求項８】
請求項１～請求項３のいずれかの方法により定められた変換行列を備えるハッシュ関数を利用するハッシュ型近似最近傍探索方法。
【請求項９】
前記データベースとして所定の訓練データベースを用いて、請求項１～請求項３のいずれかの方法により定められた変換行列を備えるハッシュ関数を特定するステップと、
特定されたハッシュ関数をテストデータベースに適用し、該テストデータベースのベクトルデータをテストデータバイナリ符号に変換するステップと、
検索対象のベクトルデータへ前記ハッシュ関数を適用して検索対象バイナリ符号を作成するステップと、
該検索対象バイナリ符号を前記テストデータバイナリ符号と比較するステップと、
備えるハッシュ型近似最近傍探索方法。
【請求項１０】
前記データベースとして所定の訓練データベースを用いて、請求項１～請求項３のいずれかの方法により定められた変換行列を備えるハッシュ関数を特定するハッシュ関数特定部と、
特定されたハッシュ関数をテストデータベースに適用し、該テストデータベースのベクトルデータをテストデータバイナリ符号に変換する変換部と、
検索対象のベクトルデータへ前記ハッシュ関数を適用して検索対象バイナリ符号を作成するバイナリ符号作成部と、
該検索対象バイナリ符号を前記テストデータバイナリ符号と比較する比較部と、
備えるハッシュ型近似最近傍探索装置。
【請求項１１】
コンピュータを、
前記データベースとして所定の訓練データベースを用いて、請求項１～請求項３のいずれかの方法により定められた変換行列を備えるハッシュ関数を特定するハッシュ関数特定部と、
特定されたハッシュ関数をテストデータベースに適用し、該テストデータベースのベクトルデータをテストデータバイナリ符号に変換する変換部と、
検索対象のベクトルデータへ前記ハッシュ関数を適用して検索対象バイナリ符号を作成するバイナリ符号作成部と、
該検索対象バイナリ符号を前記テストデータバイナリ符号と比較する比較部と、
備えるハッシュ型近似最近傍探索装置として機能させるコンピュータプログラム。

【発明の詳細な説明】
【技術分野】
【０００１】
本発明はハッシュ関数の変換行列を定める方法、該ハッシュ関数を利用するハッシュ型近似最近傍探索方法、その装置及びそのコンピュータプログラムに関する。

【背景技術】

【０００２】

現在、インターネット上には、文書、画像、音楽、動画など様々なデータが大量に存在している。これら大量のデータを有効利用するため、大規模データベースを対象とし、高速に検索質問と類似するものを見つけ出す技術が、コンピュータビジョンやテキストマイニングの分野で注目されている。例えば、画像であれば、カメラ付き携帯電話で商品を撮影し、それと見た目が類似した商品を、大量の商品データから瞬時に検索することができる。また、特定の風景画像を検索したい場合に、大規模画像データベースから類似する画像を高速に検索できれば、どこで撮影された画像かを即座に判定することができる。

【０００３】

一般的に、文書ベクトルや画像の特徴ベクトルは、数百から数千の高次元なものとなる。高次元特徴ベクトルで、大規模データベースを検索対象とした場合、線形探索では実用的な検索速度を得ることは難しい。この問題に対して、近似最近傍探索という、大規模データベースで高速な検索を実現する技術が注目されている。

近似最近傍探索は、木構造型とハッシュ型（非特許文献１）の二つに大きく分類される。木構造型近似最近傍探索は、特徴空間上に張られる軸の分割を繰り返して木構造を生成し、探索の際に探索範囲を狭めることで高速に探索を行う。探索範囲は、検索クエリとの暫定的な距離と許容誤差で定義される半径による超球で定義されるが、高次元ベクトルデータを対象とした場合に、次元の呪いの影響をうける。ハッシュ型近似最近傍探索は、高次元ベクトルデータを、短いバイナリ符号に変換し、これをハッシュテーブルのキーとすることで、高速に検索を行う。特徴空間上での距離関係をとらえ、類似するベクトルデータ同士では、バイナリ符号間のハミング距離が小さくなるように変換することで、次元の呪いの影響を小さくできる。また、短いバイナリ符号に変換することで、検索インデックスに必要な容量を抑えることもできる。

【０００４】

ハッシュ型近似最近傍探索で焦点となるのは、高次元ベクトルデータをバイナリ符号に変換するアルゴリズムである。その目的は、 n次元の m個のベクトルデータ集合 $X=[x_1, x_2, \ldots, x_m] \in R^{n \times m}$が与えられた場合に、ハッシュ関数 hを用いて、ベクトルデータ間の類似関係を保持したまま、dビットのバイナリ符号集合 $Y=[y_1, y_2, \ldots, y_m] \in B^{d \times m}$ へと変換することである。

【０００５】

Locality Sensitive Hashing (LSH)は、最も知られているハッシュ型近似近傍探索アルゴリズムである。 LSHのハッシュ関数は以下の特性を満たすことを条件としている。

$Pr[h(x_i)= h(x_j)] = sim(x_i, x_j)$

ここで、$sim(x_i, x_j) \in [0, 1]$は、類似度を表す関数である。これは、類似するベクトルデータ同士は、同じハッシュ値になることを示している。 Charikarは、内積による類似度 $sim(x_i, x_j)= x_i^T x_j$を考え、データxと同じ次元の標準正規分布 N(0,I)によるランダムな超平面（変換ベクトル）rとの積によるハッシュ関数を提案した。

$h_i(x) = sign(r_i^T x)$ここで、signは、与えられた数値の符号を返す関数である。バイナリ符号yは、以下のようにして得る。

$y=[y_1, y_2, \ldots, y_d]^T$

$y_i = (1+ h_i(x))/2$

このハッシュ関数が、 LSHの特性を満たすことは、最大カット問題の近似解法で示される。

【数１】

$$\Pr[\mathrm{sign}(\mathbf{r}^T\mathbf{x}_i) = \mathrm{sign}(\mathbf{r}^T\mathbf{x}_j)] = 1 - \frac{1}{\pi}\cos^{-1}\left(\frac{\mathbf{x}_i^T\mathbf{x}_j}{||\mathbf{x}_i||||\mathbf{x}_j||}\right)$$

【０００６】
また、Kulisらは、非線形な写像 Φ(x)の内積による類似度 $sim(x_i, x_j)= k(x_i, x_j)= \Phi(x_i)^T\Phi(x_j)$を用いた Kernelized Locality Sensitive Hashing (KLSH)を提案した。
Salakhutdinovらは、ユニット数を徐々に減少させた複数の Restricted Boltzmann Machines（RBM）によるネットワーク構造を用いてバイナリ符号を得る Semantic Hashingを提案した。Semantic Hashingのアルゴリズムでは、教師なしの事前訓練フェーズと、教師ありの微調整フェーズの二段階の学習からなる。事前訓練フェーズでは、ある層での出力は次の層の入力となるように、段階を追って各層ごとに訓練が実施される。微調整フェーズでは、ラベル付きデータを用いて、誤差逆伝搬法により、事前訓練フェーズで得られた重みを調整する。Torralbaらは、この Semantic Hashingを類似画像検索に応用し、 LSHよりも高い検索精度を得た。
【０００７】
Weissらが提案した Spectral Hashing (SH)は、グラフの分割問題を応用してバイナリ符号を求める（非特許文献２）。Weissらは、ハッシュ型近似最近傍探索における有効なバイナリ符号を求めるため、(1)新規データのバイナリ符号計算が容易であること (2)わずかなビット数で全データセットを表現すること (3)類似するデータは類似するバイナリ符号となること、の三つの条件を設定した。これら条件を満たすバイナリ符号を求めるため、 Weissらは、以下の最小化問題を考えた。
目的関数：
【数２】

$$\sum_{i,j} W_{i,j}||\mathbf{h}_i - \mathbf{h}_j||^2 \qquad \text{式（1）}$$

制約条件
【数３】

$$\mathbf{h}_i = [h_1(\mathbf{x}_i), h_2(\mathbf{x}_i), ..., h_d(\mathbf{x}_i)]^T,\ h_k(\mathbf{x}_i) \in \{-1, 1\},\ 1 \le k \le d \qquad \text{式（2）}$$

【数４】

$$\sum_i \mathbf{h}_i = 0 \qquad \text{式（3）}$$

【数５】

$$\frac{1}{n}\sum_i \mathbf{h}_i\mathbf{h}_i^T = I \qquad \text{式（4）}$$

ここで、 $W_{i,j} = \exp(-||x_i - x_j||^2/\sigma^2)$であり、式 (1)は、特徴空間の局所的な類似関係をバイナリ符号に反映することを示す。制約条件において、式 (2)は、バイナリ符号が -1と 1からなることを、式 (3)は、各ビットは偏りなく -1もしくは 1を取り得ることを、式 (4)は、異なるビット間は互いに独立であることを表す。 Weissらは、式 (2)の制約条件を緩和することで、グラフラプラシアンの固有ベクトルを求める問題とした。 SHは、 LSH、KLSH、Semantic Hashingなどよりも検索精度が高いことが知られている。
【０００８】
この他に、逐次学習を応用して頑健なバイナリ符号を得る手法、ラベルが付与されたデータを用いてデータの類似・相違を教師する手法、平行移動不変カーネルで表されるデータ間の関係を保持したバイナリ符号を得る手法、ベクトルデータの分布とバイナリ符号の分布によるカルバック・ライブラー情報量が最小となるようにバイナリ符号を求める手法などがある。
【先行技術文献】
【特許文献】

【０００９】
【特許文献１】特開２０１０－３９７７８号公報
【特許文献２】特開２００９－７５６０３号公報
【特許文献３】特開２００６－３０９７１８号公報
【非特許文献】
【００１０】
【非特許文献１】Charikar,M., Similarity estimation techniques from rounding algorithms, In Proceedings of the thirty-fourth annual ACM symposium on Theory of computing, pp.280-288, 2002.
【非特許文献２】Weiss,Y.,Torralba,A.,Fergus,R., Spectral hashing, In The Naural Information Processing Systems, Vol.21, pp.1753-1760,2008.
【非特許文献３】Wang,B.,Li,Z.,Li,M.,Ma,W-Y., Large-Scale Duplicate Detection for Web Image Search, In Proceedings of IEEE International Conference on Multimedia and Expo, pp.353-356, 2006.
【発明の概要】
【発明が解決しようとする課題】
【００１１】
インターネット上に存在するデータの量は拡大の一途をたどっているので、検索サイトに用いられる検索アルゴリズムには更なる精度の向上が期待されている。
そこで本発明者らは、検索サイト用の検索アルゴリズムとして現在利用されている上記ＬＳＨに着目し、その改良、即ち検索精度の向上を検討した。
その結果、ＬＳＨでは、そこで用いられる下記ハッシュ関数
$h_i(x) = \mathrm{sign}(r_i^T x)$
において変換ベクトルｒがランダムに決められているため、ベクトルデータをバイナリ符号に変換したとき、もととなるベクトルデータ間の類似関係が充分に維持されないおそれがあると考えた。
非特許文献３には、ベクトルデータが含まれるデータベースの特性を当該変換ベクトルに反映させ、もってバイナリ符号に変換する際にベクトルデータ間の類似関係維持を図ろうとする試みが示されている。
しかしながら非特許文献３に記載の方法は、主成分分析法を利用しているため、非線形構造をなす多次元空間のデータベースの特性を充分に変換ベクトルに反映させることには無理がある。
【課題を解決するための手段】
【００１２】
そこで本発明者らは、変換ベクトルにデータベースがなす非線形構造を反映させるべく鋭意検討を重ねた結果、この発明に想到した。
即ち、この発明の第１の局面は次のように規定される。
データベースに含まれる第１のベクトルデータｘ（ｎ次元）をバイナリ符号である$y=[y_1,y_2,\cdots y_d]$、ただし、ｎ＞＞ｄに変換するハッシュ型近似最近傍探索方法において、
下記式（Ａ）～式（Ｃ）より前記バイナリ符号を得る下記ハッシュ関数ｈ（ｘ）に適用する変換行列ｒを定める方法であって、
【数６】

$$h_i(\mathbf{x}) = \mathrm{sign}(\mathbf{r}_i^T \mathbf{x}) \qquad \text{式（A）}$$

ここに、前記バイナリ符号ｙは次のように表わされる、
【数７】

$$\mathbf{y} = [y_1, y_2, ..., y_d]^T \qquad \text{式（B）}$$

【数８】

$$y_i = (1 + h_i(\mathbf{x}))/2 \qquad \text{式（C）}$$

前記第１のベクトルデータｘを前記バイナリ符号ｙのビット数である前記ｄの次元に射影したときの第２のベクトルデータｈを下記式（Ｄ）で規定したとき、

【数９】

$$\mathbf{h} = F^T(\mathbf{x} - \bar{\mathbf{x}}) \qquad \text{式（D）}$$

ただし、

【数１０】

$$\bar{\mathbf{x}} = \frac{1}{n}\sum_i \mathbf{x}_i$$

下記式（Ｅ）が最小になる変換行列Ｆを求め、この変換行列Ｆを式（Ａ）の変換行列ｒとする、

【数１１】

$$\Phi(W) = \sum_i ||\mathbf{h}_i - \sum_{j \in \mathcal{N}_i} w_{ij}\mathbf{h}_j||^2 \qquad \text{式（E）}$$

ただし、w_{ij}は前記データベースの第１のベクトルデータｘにつき、

【数１２】

$$\mathcal{E}(W) = \sum_i ||\mathbf{x}_i - \sum_{j \in \mathcal{N}_i} w_{ij}\mathbf{x}_j||^2 \qquad \text{式（F）}$$

式（Ｆ）を最小とする重みである、ハッシュ関数の変換行列を定める方法。

【００１３】

第１の局面に規定の発明の基本原理を以下に説明する。

高次元ベクトルデータをバイナリ符号に変換するアルゴリズムの多くが、次元削減のアルゴリズムに基づいている。次元削減の目的は、高次元空間上でベクトルデータがなす、低次元の部分空間を推定することである。近年、全体では非線形構造を成していても、局所的には通常のユークリッド空間と同じ構造をなす、多様体の性質を利用した非線形次元削減手法が幾つか提案されている。その内、Locally Linear Embedding (LLE)は、局所的な範囲で低次元の線形モデルをあてはめ、それらが滑らかに繋がるように全体の多様体を推定する。このLLEによる多様体構造の推定方法を用いて、ベクトルデータがなす非線形な関係をとらえたバイナリ符号を得る。以下、Weissらと同様に、問題の簡単化のため、式(2)の制約条件を緩和する。

【００１４】

局所的な範囲で低次元の線形モデルをあてはめるため、それぞれのベクトルデータx_iを、近傍のベクトルデータ $x_j \in N_i$を用いて再構成することを考える。これは、以下の再構成誤差を最小化することで表される。

【数１３】

$$\mathcal{E}(W) = \sum_i ||\mathbf{x}_i - \sum_{j \in \mathcal{N}_i} w_{ij}\mathbf{x}_j||^2$$

ここで、w_{ij}は、再構成する際の重みである。再構成には近傍のベクトルデータを用いるため、

【数１４】

$$j \notin \mathcal{N}_i \Rightarrow w_{ij} = 0,$$

重みの大きさの任意性を解決するため$\Sigma_j w_{ij} = 1$とする。各ベクトルデータでの再構成誤差は、

【数１５】

$$\begin{aligned}\mathcal{E}_i(W) =& ||\mathbf{x}_i - \sum_{j\in\mathcal{N}_i} w_{ij}\mathbf{x}_j||^2 \\ =& ||\mathbf{x}_i \sum_{j\in\mathcal{N}_i} w_{ij} - \sum_{j\in\mathcal{N}_i} w_{ij}\mathbf{x}_j||^2 \\ =& \sum_{j,k\in\mathcal{N}_i} w_{ij}w_{ik}(\mathbf{x}_i - \mathbf{x}_j)^T(\mathbf{x}_i - \mathbf{x}_k) \\ =& \sum_{j,k\in\mathcal{N}_i} w_{ij}w_{ik}C_{jk}\end{aligned}$$

と表わせる。ここで、$C_{jk} = (x_i - x_j)^T (x_i - x_k)$とした。この再構成誤差は、ラグランジュ乗数$\eta_i$を用いて

【数１６】

$$\mathcal{E}_i(W) = \sum_{j,k\in\mathcal{N}_i} w_{ij}w_{ik}C_{jk} + \eta_i\left(\sum_{j\in\mathcal{N}_i} w_{ij} - 1\right)$$

となる。極致を求めるため、w_{ij}について偏微分し０とおくことで、以下の線形方程式を解く問題となる。

【数１７】

$$\sum_{k\in\mathcal{N}_i} C_{jk}w_{ik} = 1$$

以上の重みを求める計算については、 de Ridderらや Panらの研究で詳述されている。

【００１５】

近傍による再構成の重みw_{ij}で表される、それぞれのベクトルデータにおける局所的な関係を、バイナリ符号に反映させるため、Weissらの最小化問題において、式(1)の目的関数を以下のものに置き換える（式（Ｅ））。

【数１８】

$$\Phi(W) = \sum_i ||\mathbf{h}_i - \sum_{j\in\mathcal{N}_i} w_{ij}\mathbf{h}_j||^2$$

さらに、新規データのバイナリ符容易に求められるよう、変換行列 $F \in R^{n\times d}$による線形変換を考える（式Ｄ）。

【数１９】

$$\mathbf{h} = F^T(\mathbf{x} - \bar{\mathbf{x}})$$

ここで、式 (3)の制約条件を満たすため、平均ベクトル $\bar{x}$で引いた。

【数２０】

$$\bar{\mathbf{x}} = \frac{1}{n}\sum_i \mathbf{x}_i$$

【００１６】

バイナリ符号の集合を $H = [h_1, \ldots, h_m]$と表すと、式（Ｅ)の目的関数は、以下のように表される。

【数２１】

$$\begin{aligned}\Phi(W) &= \mathrm{tr}[(H-HW)^T(H-HW)] \\ &= \mathrm{tr}[(H-HW)(H-HW)^T] \\ &= \mathrm{tr}[H(I-W)(I-W)^T H^T] \\ &= \mathrm{tr}[F^T X(I-W)(I-W)^T X^T F] \\ &= \mathrm{tr}[F^T XMX^T F]\end{aligned}$$

ここで、Ｍ＝(Ｉ－Ｗ)(Ｉ－Ｗ)^{T}とした。さらに、式(4)の制約条件は、

【数２２】

$$\frac{1}{n}HH^T = \frac{1}{n}F^T XX^T F = I$$

となり、式（６）の目的関数の最小化は、

【数２３】

$$\underset{F^T XX^T F=nI}{\underset{F}{\mathrm{argmin}}} \ \mathrm{tr}[F^T XMX^T F]$$

と表わせる。これは、ラグランジュの未定乗数法から、

【数２４】

$$\mathcal{L}(F) = F^T(XMX^T)F + \lambda(nI - F^T XX^T F)$$

となるので、Ｆで偏微分して０としておくことにより、

【数２５】

$$\frac{\partial \mathcal{L}}{\partial F} = 2(XMX^T)F - 2\lambda XX^T F = 0$$

以下の一般固有値問題へと帰着する。

【数２６】

$$XMX^T F = \lambda XX^T F \qquad \text{式（H）}$$

ここで、変換行列 $\mathbf{F}$ は、式(H)の一般化固有値問題を解くことで得られる、固有値$\lambda_1 \leq \lambda_2 \leq \cdots\cdots \leq \lambda_d$に対応する固有関数からなる。

【数２７】

$$F = [\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_d]$$

【００１７】

上記において、ベクトルデータｈは、ベクトルデータｘをバイナリ符号ｙのビット数であるｄの次元に射影したものである。変換行列Ｆを式（Ａ）の変換行列ｒへ当てはめる。

上記において、局所的な近傍構造を表す重みW_{ij}を、それぞれのベクトルデータの近傍から求める。しかし、注目しているベクトルデータに対して、近傍が密集して位置するものもあれば、遠く離れて位置するものもある。そこで、この分布の偏りによる影響を軽減するため、各ベクトルデータでの近傍の分布を表すスコアを定義する。

【数２８】

$$d_i = \sum_{j \in \mathcal{N}_i} \exp(-||\mathbf{x}_i - \mathbf{x}_j||/\sigma^2)$$

近傍の分布を考慮した変換行列 Fは、分布スコアからなる対角行列 D = diag[d_1, ..., d_m]により、以下の最小化問題を解くことで得られる。

【数２９】

$$\underset{F^T X X^T F = nI}{\operatorname{argmin}} \ \operatorname{tr}[F^T X M X^T F]$$

【００１８】

これは、ラグランジュの未定乗数法から、

【数３０】

$$XDMX^T F = \lambda X X^T F$$

と表わすことができる。さらに、Xがフルランク行列であれば、 X^T Xは正則となることから、

【数３１】

$$X^T X D M X^T F = \lambda X^T X X^T F$$
$$D M X^T F = \lambda X^T F$$

分布スコアは必ず正数であるので、 $\mathbf{S} = \mathbf{D}^{-1} = \mathbf{diag}[1/d_1, ..., 1/d_m]$とおくと、上記の最小化問題は、以下の一般化固有値問題に帰着する。

【数３２】

$$XMX^T F = \lambda X S X^T F$$

Xがフルランク行列ではない場合は、特異値分解を用いて、行列 Xのランク数 lと等しい次元の部分空間に射影する。

【数３３】

$$X = URV^T$$
$$\tilde{X} = U^T X = RV^T$$

ここで、 Rは要素が特異値 $\eta_1 \geq \ ... \geq \ \eta_l$の大きさ l × l の対角行列であり、 Uは大きさ n × lの、 Vは大きさ m × lの直行行列である。 $\tilde{\mathbf{X}}$ はフルランク行列となるので、以下の一般化固有値問題を解くことで、変換行列 **F** を得る。

【数３４】

$$\tilde{X} M \tilde{X}^T F = \lambda \tilde{X} S \tilde{X}^T F$$

【００１９】

上記の処理はいわゆるグラフラプラシアンの理論を用いて、重みW_{ij}を正規化したことを意味し、ｎ次空間のベクトルデータｘをｄ次空間への射影するとき、ｎ次空間におけるベクトルデータｘの分布の偏りの影響を軽減できる。よって、ベクトルデータ得られた変

換行列Ｆはデータベースの特性をより正確に反映したものとなる。

【図面の簡単な説明】

【００２０】

【図１】図１は、この発明のハッシュ型近似最近傍検索システムの構成を示すブロック図である。

【図２】図２は、20-newsgroupsでビット数を8から64まで変化させた場合の、検索結果上位400件における適合率(Precision)を表したグラフである。

【図３】図３は、20-newsgroupsで検索結果の上位件数を変化させた場合の、ビット数64における再現率 (Recall)を表したグラフである。

【図４】図４は、20-newsgroupsにおける再現率と適合率を表したグラフである。

【図５】図５は、CIFAR-10でビット数を8から64まで変化させた場合の、検索結果上位1,000件における適合率(Precision)を表したグラフである。

【図６】図６は、 CIFAR-10で検索結果の上位件数変化させた場合の、ビット数 64における再現率 (Recall)を表したグラフである。

【図７】図７は、 CIFAR-10における再現率と適合率を表したグラフである。

【図８】図８は、CIFAR-10で、SHとNSHによる 32ビットのバイナリ符号を用いた、自動車の画像での検索結果上位20件を並べたものである。

【図９】図９は、実施例のハッシュ型近時最近傍探索装置を示すブロック図である。

【発明を実施するための形態】

【００２１】

このようにして定められた変換行列を有するハッシュ関数を用いてハッシュ型近似最近傍検索を行なうシステム１を図１に示す。

図１において、訓練データベース３、テストデータベース１３、テストデータバイナリ符号データベース１９はサーバのメモリ装置の所定の領域が対応される。

訓練データベース３のデータはベクトル化処理部５において所定の方法でベクトル化される。

ハッシュ関数特定部７は、ベクトル化された訓練データベース３のデータ（第１のベクトルデータｘ）の一部又は全部を用いて、既述の処理を行ない、変換ベクトルＦを特定し、もって、第１のベクトルデータｘをバイナリ符号化するハッシュ関数を特定する。

【００２２】

テストデータベース１３のデータは、ベクトル化処理部１５において、ベクトル化処理部５と同一の方法でベクトル化される。

バイナリ符号化部１７は、ハッシュ関数特定部７で特定されたハッシュ関数を用いてベクトル化されたテストデータベースのデータの一部又は全部をバイナリ符号に変換し、テストデータバイナリ符号データベース１９に保存する。

【００２３】

クライアントＰＣが検索対象特定部２０に対応し、この検索対象特定部２０においてユーザが検索対象を特定する。特定された検索対象はベクトル化処理部２１においてベクトル化処理される。このベクトル化処理の方法は、既述のベクトル化処理部５において、訓練データベース３のデータを第１のベクトルデータに変換したベクトル化処理法と同一である。

このようにしてベクトル化された検索対象はバイナリ符号化部２３において、ハッシュ関数特定部７で特定されたハッシュ関数を用いてバイナリ符号化される。比較部２５は、検索対象のバイナリ符号をテストデータバイナリ符号データベース１９に保存されているテストデータバイナリ符号と比較し、例えば、検索対象のバイナリ符号に対する距離が所定の閾値以内のものを、近い順に出力する。

【００２４】

この発明の探索方法を評価するために、ベンチマークに20-newsgroupsと CIFAR-10を用いて、従来手法との比較実験を行った。従来手法には、Locality-Sensitive Hashing (LSH)、Kernelized Locality-Sensitive Hashing (KLSH)、Spectral Hashing (SH)を選択し

た。この内、アルゴリズム中に乱数生成を含むLSHとKLSHについては、5回実行して平均をとった。

20-newsgroupsは、Usenet newsgroupから取得した18,845件のニュースグループの文書からなる。各文書は、異なる20個のニュースグループのいずれかに分類され、11,314件が訓練データセットとして、7,531件がテストデータとして与えられる。各アルゴリズムの訓練には、訓練データセットからランダムに選択した5,000件を用いた。実験では、前処理として単語のステミングとストップワードの除去を行ったのち、文書頻度の大きいほうから 2,000語を選択し、tf-idfにより重み付けを行った文書ベクトルを作成した。本発明のパラメータは、事前実験により求めた最適値、近傍数 k = 205、分布スコアのガウスカーネル幅 λ =4.0を用いた。

【００２５】

図１は、20-newsgroupsでビット数を8から64まで変化させた場合の、検索結果上位400件における適合率(Precision)を表したグラフである。全てのビット数で、本件発明（NSH）が最も大きな適合率となっている。図３は、20-newsgroupsで検索結果の上位件数を変化させた場合の、ビット数64における再現率 (Recall)を表したグラフである。全ての上位件数で、本件発明（NSH）が最も大きな再現率となっているのがわかる。図４は、20-newsgroupsにおける再現率と適合率を表したグラフである。曲線が右上に伸びるほど検索精度が高いと言える。本件発明（NSH）が、従来手法と比較して、正確性・網羅性ともに高いことがわかる。

【００２６】

20-newsgroupsは、20個のカテゴリに分類されているが、それらは、コンピュータの話題やスポーツの話題など、大きく分類することもできる。これらは高次元の文書ベクトル空間で、類似した話題のものが集まりながらも、粗密をなして複雑な構造を成していると予想する。ベクトルデータの非線形構造を推定し、分布の偏りを重みにより考慮する本件発明（NSH）は、従来手法と比較して、この複雑な文書ベクトル空間を正しくとらえられたと考える。

【００２７】

CIFAR-10は、飛行機や自動車、イヌなどの 10種類のラベルが付与されクラス分けされた、60,000個の 32 × 32のカラー画像が含まれており、50,000個が訓練データセットとして、10,000個がテストデータセットとして与えられる。各アルゴリズムの訓練には、訓練データセットからランダムに選択した 5,000個を用いた。実験では、RGBごとに 4×4領域 6方向 4スケールで 3×4×4×6×4=1，152次元の GIST特徴ベクトルを抽出し、それをバイナリ符号に変換した。本件発明（NSH）のパラメータは、事前実験により求めた最適値、近傍数 k = 90、分布スコアのガウスカーネル幅 λ =0.5を用いた。

【００２８】

図５は、CIFAR-10でビット数を8から64まで変化させた場合の、検索結果上位1,000件における適合率(Precision)を表したグラフである。全てのビット数で、本件発明（NSH）が最も大きな適合率となっている。図６は、 CIFAR-10で検索結果の上位件数変化させた場合の、ビット数 64における再現率 (Recall)を表したグラフである。全ての上位件数で、本件発明（NSH）が最も大きな再現率となっているのがわかる。図７は、 CIFAR-10における再現率と適合率を表したグラフである。本件発明（NSH）が、従来手法と比較して、正確性・網羅性ともに高いことがわかる。

【００２９】

CIFAR-10には、飛行機や自動車などからなる乗り物画像と、イヌやシカなどからなる生物画像に、大きく分けられる。しかし、画像全体では特定の色が多く使用されている、背景が類似しているなどの要因で、特徴空間上では、乗り物画像と生物画像がはっきりと区別されることなく、粗密をなして分布していると予想する。分布の偏りを重みにより抑えつつ、ベクトルデータがなす非線形な構造を推定する NSHは、この特徴空間上の複雑なデータ関係を、従来手法と比較して正しくとらえられたと考える。

【００３０】

図８は、CIFAR-10で、SHとNSHによる 32ビットのバイナリ符号を用いた、自動車の画像での検索結果上位20件を並べたものである。左上端の画像が検索質問画像であり、正解の画像は緑色の枠で、不正解の画像は赤色の枠で囲んだ。SHと比較して、本件発明（NSH）では適合画像が多く、検索精度が高いことがわかる。

【００３１】

本件発明（NSH）は、特徴空間上の局所的な近傍構造に基づいて、ベクトルデータがなす非線形構造をとらえた短いバイナリ符号を得る。文書データベンチマーク20-newsgroupsと画像データベンチマーク CIFAR-10を用いた比較実験から、 Spectral Hashingなどの従来手法よりも高い検索精度を得られることが確認できた。

【００３２】

実施例のハッシュ型近似近最傍探索装置１００を図９に示す。

この装置１００は、検索サーバ１０１とクライアント端末２００がネットワークＮ１を介して接続されている。任意の数のクライアント端末２００をネットワークＮ１へ接続可能で、ネットワークＮ１はインターネット３００へ開いていてもよい。

【００３３】

検索サーバ１０１はデータ保存部１１０とデータ処理部１２０とを備える。データ保存部１１０は原データ取得・保存部１１１を備える。この原データ取得・保存部１１１は被検索対象となるデータを、インターネット３００を介して外部のデータベースから取得し、保存する。外部のデータベースは特定のデータが整理保存された商用、または検索用のデータベースはもとより、法人若しくは個人が運営するホームページ、ブログ、ツイッター等も含まれるものとする。

訓練データ抽出・更新部１１３は、原データ取得・保存部１１１に保存されたデータから無作為にデータを抽出し、訓練データとする。抽出するデータ数は特に限定されるものではないが、既述の試験例に準じて５０００個程度とすることが好ましい。なお、この訓練データは周期的に、若しくは任意のタイミングで更新することが好ましい。

【００３４】

ベクトル化処理部１２１及びハッシュ関数特定部１２３の動作は、図１のベクトル化処理部５及びハッシュ関数特定部７と同じである。即ち、訓練データ抽出・更新部１２１で抽出されたデータを特徴量次元削減法等の汎用的な手法（第１のベクトル化方法）でベクトル化処理し、ハッシュ関数特定部１２３において、本発明の手法に従いハッシュ関数（第１のハッシュ関数）を特定する。

ベクトル化処理部１２１は原データ取得・保存部１１１のデータの全部をまたは所定のルールで選択されたその一部を訓練データと同様にベクトル化処理する。ベクトル化処理され原データは、バイナリ符号化部１２５においてバイナリ符号化データに次元削減される（前処理１３０）。このとき、ハッシュ関数特定部１２３において特定されたハッシュ関数が利用される。バイナリ符号化されたデータはバイナリ符号化データ保存部１１５に保存される。

【００３５】

クライアント端末２００において、検索対象となるデータが入力部２１０で指定される。指定されたデータはベクトル化処理部２２１において、検索サーバ１０１のベクトル化処理部１２１と同一の方法によりベクトル化される。バイナリ符号化部２２３には、検索サーバ１０１のハッシュ関数特定部１２３で特定されたハッシュ関数が提供され、ベクトル化された検索対象データをバイナリ符号化する。このようにしてバイナリ符号化された検索対象データは検索サーバ１０１の比較部１２７へ送られる。比較部１２７はバイナリ符号化された検索対象データとバイナリ符号化データ保存部１１５に保存されているバイナリ符号化された原データとを比較し、所定のルールに従い近似するデータを抽出する。

【００３６】

比較部１２７で抽出されたデータはクライアント端末２００の出力部２３０へ送られ、ここで出力される。なお、出力部２３０はバイナリ符号化されたデータをデコード（逆ハッシュ関数処理、逆ベクトル処理）し、原データの状態で表示できる。

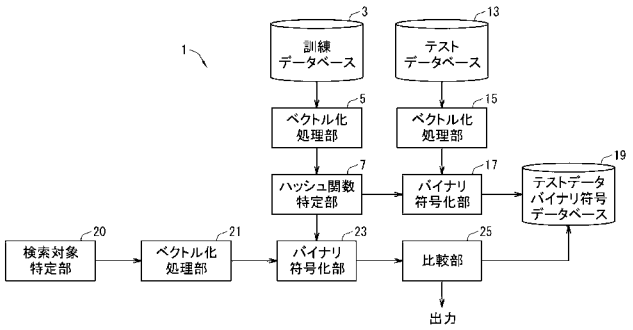
この比較部を端末側に配置することができる。また、比較部を、検索サーバ及び端末と独立して設置することもできる。

【符号の説明】

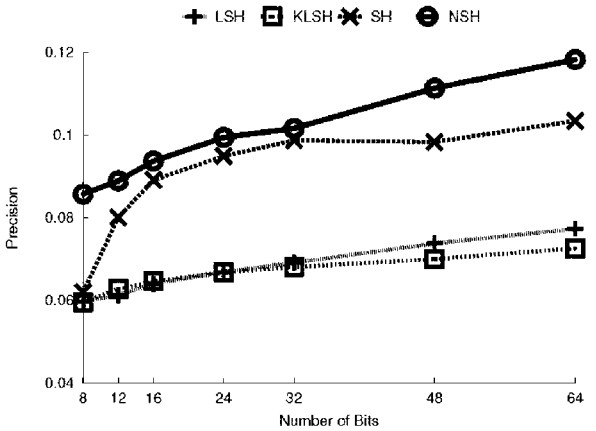
【００３７】

１　ハッシュ型近似最近傍検索システム
３、１３、１９データベース
５、１５、２１　ベクトル化処理部
７　ハッシュ関数特定部
１７、２３　バイナリ符号化部
２５　比較部

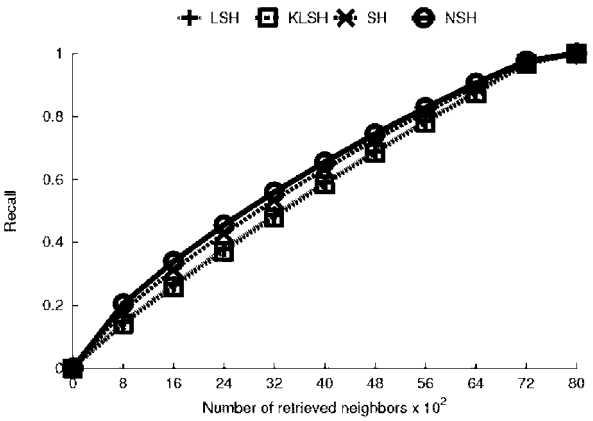
【図１】



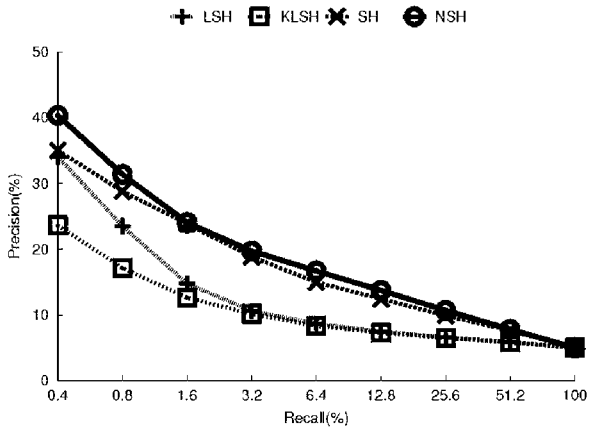
【図２】



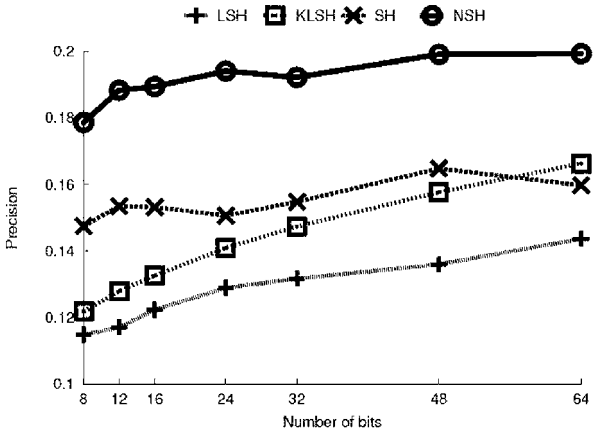
【図３】



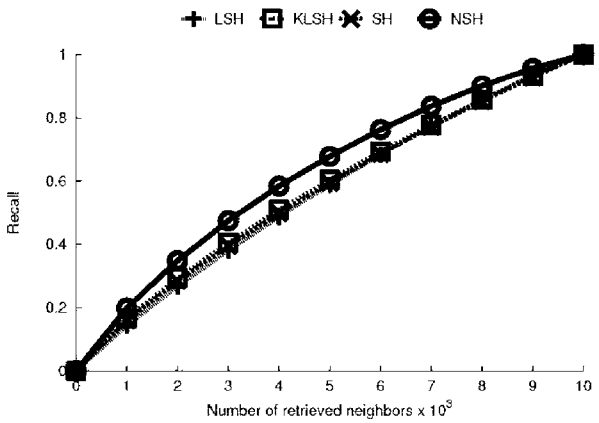
【図４】


LSH
KLSH
SH
NSH
Precision(%)
Recall(%)
50
40
30
20
10
0
0.4
0.8
1.6
3.2
6.4
12.8
25.6
51.2
100

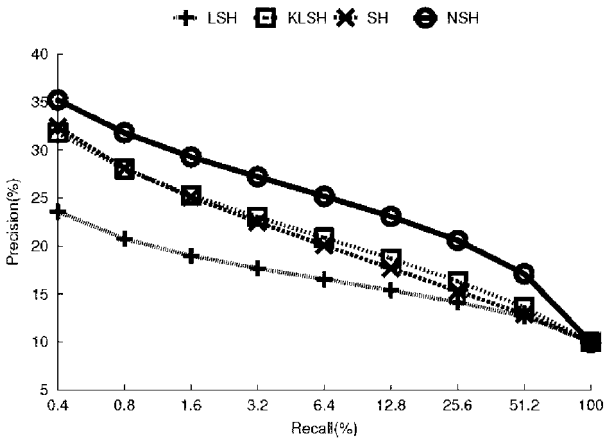
【図５】


LSH
KLSH
SH
NSH
Precision
Number of bits
0.2
0.18
0.16
0.14
0.12
0.1
8
12
16
24
32
48
64

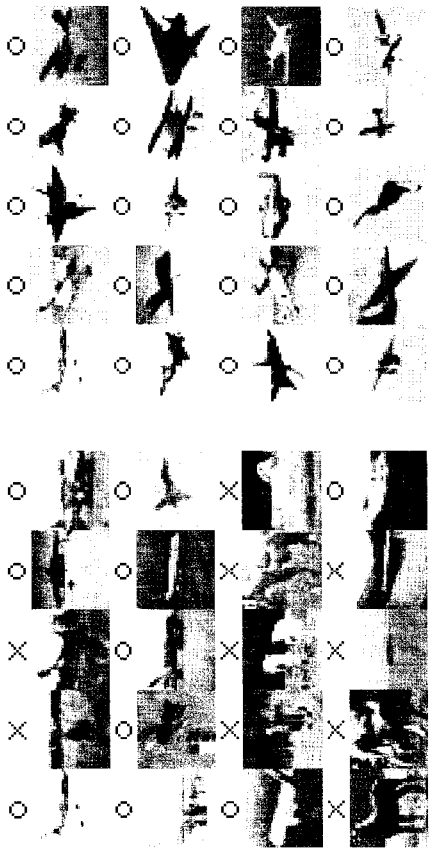
【図６】


LSH
KLSH
SH
NSH
Recall
Number of retrieved neighbors x 10^3
1
0.8
0.6
0.4
0.2
0
0
1
2
3
4
5
6
7
8
9
10

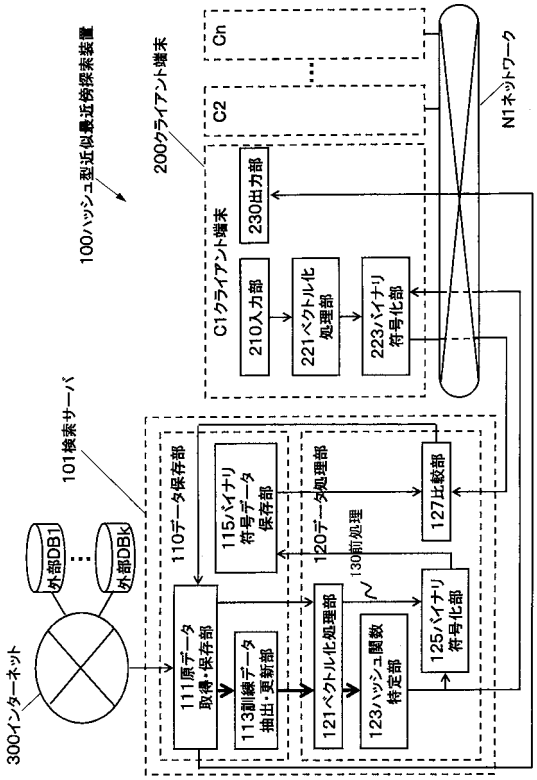
【図７】


LSH
KLSH
SH
NSH
Precision(%)
Recall(%)
40
35
30
25
20
15
10
5
0.4
0.8
1.6
3.2
6.4
12.8
25.6
51.2
100

【図８】



【図９】


100ハッシュ型近似最近傍探索装置
200クライアント端末
C1クライアント端末
C2
Cn
N1ネットワーク
210入力部
221ベクトル化処理部
223バイナリ符号化部
230出力部
101検索サーバ
外部DB1
外部DBk
300インターネット
110データ保存部
111原データ取得・保存部
113訓練データ抽出・更新部
115バイナリ符号データ保存部
120データ処理部
130前処理
121ベクトル化処理部
123ハッシュ関数特定部
125バイナリ符号化部
127比較部

【国際調査報告】

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2011/078702

A. CLASSIFICATION OF SUBJECT MATTER
G06F17/30(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G06F17/30

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
Jitsuyo Shinan Koho 1922-1996 Jitsuyo Shinan Toroku Koho 1996-2012
Kokai Jitsuyo Shinan Koho 1971-2012 Toroku Jitsuyo Shinan Koho 1994-2012

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
JSTPlus (JDreamII)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2003-141160 A (International Business Machines Corp.), 16 May 2003 (16.05.2003), entire text; all drawings & US 2003/0159106 A1	1-11
A	JP 2010-256951 A (Dehenken Ltd.), 11 November 2010 (11.11.2010), entire text; all drawings (Family: none)	1-11
A	JP 2010-277522 A (Nippon Telegraph and Telephone Corp.), 09 December 2010 (09.12.2010), entire text; all drawings (Family: none)	1-11

☒ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
10 February, 2012 (10.02.12)	21 February, 2012 (21.02.12)
Name and mailing address of the ISA/ Japanese Patent Office Facsimile No.	Authorized officer Telephone No.

Form PCT/ISA/210 (second sheet) (July 2009)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2011/078702

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	Shuhei FURUKAWA et al., "Document Similarity Search and Reranking using Adjacency Information", The First Forum on Data Engineering and Information Management DEIM Forum 2009 Ronbunshu, 09 May 2009 (09.05.2009), no.A9-2, [online] URL:http://db-event.jpn.org/deim2009/proceedings/files/A9-2.pdf	1-11
A	Masaru NAKANO et al., "Linearized Diffusion Maps and its Application to Documents", Dai 72 Kai (Heisei 22 Nen) Zenkoku Taikai Koen Ronbunshu, 08 March 2010 (08.03.2010), vol.2, pages 2-465 to 2-466	1-11

国際調査報告

国際出願番号 ＰＣＴ／ＪＰ２０１１／０７８７０２

Ａ． 発明の属する分野の分類（国際特許分類（ＩＰＣ））

Int.Cl. G06F17/30(2006.01)i

Ｂ． 調査を行った分野

調査を行った最小限資料（国際特許分類（ＩＰＣ））

Int.Cl. G06F17/30

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	１９２２－１９９６年
日本国公開実用新案公報	１９７１－２０１２年
日本国実用新案登録公報	１９９６－２０１２年
日本国登録実用新案公報	１９９４－２０１２年

国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）

JSTPlus(JDreamII)

Ｃ． 関連すると認められる文献

引用文献のカテゴリー＊	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する請求項の番号
A	JP 2003-141160 A (インターナショナル・ビジネス・マシーンズ・コーポレーション) 2003.05.16, 全文、全図 & US 2003/0159106 A1	１－１１
A	JP 2010-256951 A (株式会社データ変換研究所) 2010.11.11, 全文、全図 (ファミリーなし)	１－１１
A	JP 2010-277522 A (日本電信電話株式会社) 2010.12.09, 全文、全図 (ファミリーなし)	１－１１

☑ Ｃ欄の続きにも文献が列挙されている。 ☐ パテントファミリーに関する別紙を参照。

＊ 引用文献のカテゴリー
「Ａ」特に関連のある文献ではなく、一般的技術水準を示すもの
「Ｅ」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
「Ｌ」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）
「Ｏ」口頭による開示、使用、展示等に言及する文献
「Ｐ」国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献
「Ｔ」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの
「Ｘ」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
「Ｙ」特に関連のある文献であって、当該文献と他の１以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
「＆」同一パテントファミリー文献

国際調査を完了した日	国際調査報告の発送日
１０．０２．２０１２	２１．０２．２０１２

国際調査機関の名称及びあて先
日本国特許庁（ＩＳＡ／ＪＰ）
郵便番号１００－８９１５
東京都千代田区霞が関三丁目４番３号

特許庁審査官（権限のある職員） ５Ｍ ９２８９
野崎 大進
電話番号 ０３－３５８１－１１０１ 内線 ３５９９

様式ＰＣＴ／ＩＳＡ／２１０（第２ページ）（２００９年７月）

国際調査報告

国際出願番号 ＰＣＴ／ＪＰ２０１１／０７８７０２

Ｃ（続き）． 関連すると認められる文献

引用文献のカテゴリー＊	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する請求項の番号
A	古川 修平 他，隣接情報を用いた類似文書検索とリランキング，第1回データ工学と情報マネジメントに関するフォーラムDEIMフォーラム2009論文集，2009.05.09，No.A9-2，[online]ＵＲＬ：http://db-event.jpn.org/deim2009/proceedings/files/A9-2.pdf	１－１１
A	仲野 将 他，線形化拡散写像手法の提案とその文書データへの適用，第７２回（平成２２年）全国大会講演論文集，2010.03.08，Vol.2，pp.2-465～2-466.	１－１１

フロントページの続き

(81)指定国 AP(BW,GH,GM,KE,LR,LS,MW,MZ,NA,RW,SD,SL,SZ,TZ,UG,ZM,ZW),EA(AM,AZ,BY,KG,KZ,MD,RU,TJ,TM),EP(AL,AT,BE,BG,CH,CY,CZ,DE,DK,EE,ES,FI,FR,GB,GR,HR,HU,IE,IS,IT,LT,LU,LV,MC,MK,MT,NL,NO,PL,PT,RO,RS,SE,SI,SK,SM,TR),OA(BF,BJ,CF,CG,CI,CM,GA,GN,GQ,GW,ML,MR,NE,SN,TD,TG),AE,AG,AL,AM,AO,AT,AU,AZ,BA,BB,BG,BH,BR,BW,BY,BZ,CA,CH,CL,CN,CO,CR,CU,CZ,DE,DK,DM,DO,DZ,EC,EE,EG,ES,FI,GB,GD,GE,GH,GM,GT,HN,HR,HU,ID,IL,IN,IS,JP,KE,KG,KM,KN,KP,KR,KZ,LA,LC,LK,LR,LS,LT,LU,LY,MA,MD,ME,MG,MK,MN,MW,MX,MY,MZ,NA,NG,NI,NO,NZ,OM,PE,PG,PH,PL,PT,QA,RO,RS,RU,RW,SC,SD,SE,SG,SK,SL,SM,ST,SV,SY,TH,TJ,TM,TN,TR,TT,TZ,UA,UG,US,UZ,VC,VN

（注）この公表は、国際事務局（ＷＩＰＯ）により国際公開された公報を基に作成したものである。なおこの公表に係る日本語特許出願（日本語実用新案登録出願）の国際公開の効果は、特許法第１８４条の１０第１項(実用新案法第４８条の１３第２項）により生ずるものであり、本掲載とは関係ありません。