

意味の数学モデルによる画像データベース探索方式とその学習機構

正 員 清木 康[†] 非会員 金子 昌史[†] 非会員 北川 高嗣[†]

A Semantic Search Method and its Learning Mechanism for Image Databases Based on a Mathematical Model of Meaning

Yasushi Kiyoki[†], Member, Yoshifumi Kaneko[†], and Takashi Kitagawa[†], Nonmember

あらまし 画像データを対象としたデータベース・システムにおいては、検索者の印象や画像の内容による検索を実現する方法が重要である。我々は、文脈あるいは状況に応じて動的に変化するデータ間の意味的な関係を計算するモデルである意味の数学モデルを提案している。本論文では、意味の数学モデルを用いた意味的画像探索方式を提案し、また、その学習機構を示す。

本方式では、メタデータ空間とよぶ正規直交空間を形成し、その空間上に画像データ群、および、検索に用いるキーワード群を配置する。そして、その空間上での距離計算により、検索者の印象、および、画像の内容の指定に応じた画像探索を実現する。検索対象の各画像は、画像の特徴（印象あるいは内容）を表す言葉（単語群）によって表現されることを前提とする。本方式は、検索者が発行する検索語、および、画像データの特徴を表現する単語間の相関関係の分析により、画像を検索する方式として位置付けられる。

本方式では、検索者が指定する印象あるいは画像の内容を文脈として言葉により与えると、その文脈に対応する画像をメタデータ空間より動的に抽出する。本学習機構では、その文脈から得られた画像が、本来、抽出されるべき画像と異なる場合、抽出されるべき画像を指定することにより、文脈を構成している言葉についての学習を行う。本学習機構を適用することにより、画像の印象表現における個人差に対応することが可能となる。

キーワード マルチメディアデータベース、画像データベース、意味的連想検索、学習機構

1. はじめに

近年、マルチメディア・データベースの分野においては、広域ネットワーク上の様々なサイトに散在する膨大なメディアデータから適切なメディアデータを抽出する方法の開発が重要な課題となっている。その基本的な機能の一つとして、検索者独自の印象および内容の指定による画像データの抽出を実現するデータベースシステムの構築が重要となっている。画像データを対象としたデータベースシステムの実現には、画像データの定義、表現、および、検索のための方式設計が重要である(1),(3),(8),(10)～(12)。

画像探索方式は、部分的なパターンマッチングによる直接的な方法と、画像に付加された抽象情報を利用する

間接的な方法に大別できる。本論文では、後者の方法を対象とした意味的画像探索方式を提案する。この方式は、文脈あるいは状況に応じて動的に変化するデータ間の意味的な関係を計算するモデルとして我々が提案している意味の数学モデル^{(4),(6)}を拡張した新しい画像データベースの検索方式であり、検索者（利用者）の印象、および、画像の内容による画像データの意味的連想検索を実現する。本探索方式では、正規直交空間上に、検索者が獲得したい画像の内容あるいは印象を表現するために使用する検索語に対応するベクトル（検索語ベクトル）、および、各画像の特徴を表す言葉（単語群）に対応するベクトル（画像ベクトル）として写像する。そして、画像の内容あるいは検索者の印象を表す文脈として複数の検索語が与えられると、その文脈に対応する正規直交空間の部分空間を選択する。その部分空間における画像ベクトルの位置により、その文脈と相関の強い画像を動的に抽出する。

[†] 筑波大学電子・情報工学系、つくば市
Institute of Information Science and Electronics, University of Tsukuba, Tsukuba-shi, 305 Japan

本方式では、画像特徴は、画像の印象あるいは内容を表す言葉によって表現されることを前提とする。すなわち、画像特徴の表現では、実際に画像処理によって直接的に獲得される情報を用いるのではなく、データベース作成者が画像の印象や出現オブジェクトを言葉によって表した情報を用いて、その画像に対応するベクトルを生成する。したがって、新たな画像が追加された場合には、その都度、画像の特徴を表す言葉を割り当てる必要がある。また、画像についての解説文から印象、オブジェクトを表す言葉を切り出し、それらによって、画像の特徴を表現するといった研究があり⁽¹³⁾、それらを応用して、画像データの印象、オブジェクトを言葉によって表現することを前提とすることも可能である。また、画像の配色による印象の自動認識の方法により、その印象を表す言葉を選ぶ方式も利用可能である⁽⁷⁾。

さらに、本論文では、印象表現が検索者ごとに異なることを考慮し、各検索者の主観を反映した画像の抽出を実現するための学習機構を示す。本学習機構では、教師としての検索者が、文脈とその文脈に関連が強いと考える画像を指定する。そして、その文脈を与えた場合に、その画像が選択されない状況において、正規直交空間上でのその画像の位置をもとに、文脈を構成している検索語の位置を修正する。この学習機構では、個々の利用者が発行する検索語によって、適切な画像データを抽出できるように、各検索者毎に検索語の定義（検索語ベクトル）を修正する。この修正によって、画像データを表現しているベクトルと検索語ベクトルの関係が修正され、それ以後、適切な画像データを抽出できるようになる。この学習は、画像データの特徴を表すベクトルを修正するものではなく、各検索者が発行する検索語のベクトルを修正することによって、各検索者の意図を正しく反映する機構である。すなわち、ここでの“学習”とは、各検索者の用いる検索語ベクトルの修正により、適切な画像データの抽出のために検索語を適応させることをいう。

本方式は、多変量解析による空間生成を用いた情報検索手法（例えば⁽²⁾）とは、次の点で本質的に異なる。本方式では、直交空間における部分空間の選択を行う演算（意味射影）を定義し、その演算により、言葉の意味を文脈に応じて、曖昧性を排除して解釈する機構を実現している。この機構により、言葉と言葉、あるいは、言葉と画像データの間の意味的な関係を、与えられた文脈に応じて動的に計算することが可能となる。現在の実現システムでは、本モデルで表現できる部分空間、すなわ

ち、文脈の様相の数は、約 2^{800} であり、非常に多くの文脈を表すことが可能である。

画像データベースにおいて、検索者の印象による画像データ抽出に関する研究として、⁽⁹⁾、⁽¹⁰⁾ において提案された方法がある。この方法は、正準相関分析により、各検索者の印象語と画像特徴との相関関係を学習し、印象による画像の検索を可能としている。この方法は、画像の特徴を表す静的に固定されたベクトルと検索者の印象語群を表すベクトル間の相関の強さに関する計算により画像を抽出する方法である。この方式との比較において、本方式の特徴は、各画像について、検索者が与える印象語群に対応する特徴に関する要素群を部分空間の選択という操作により抽出し、その要素群に対応する部分においてのみ強い相関をもつ画像データを抽出する点にある。これにより、本方式では、検索者によって与えられる印象語群に対応する画像の特徴を動的に抽出し、それ以外の特徴を排除することによって、与えられた印象語群に応じた特徴に関する要素群に対応する部分だけに着目した画像の抽出を行うことが可能となる。

2. 意味的画像探索方式の概要

2.1 意味の数学モデルの基本構成

意味の数学モデルは、本来、言葉の意味を扱うためのモデルである。ここでは、印象や画像の内容を表現する文脈に対応した画像を探索するために拡張した意味の数学モデルの基本構成を示す。

(1) メタデータ空間 MDS の設定：

m 個の基本データ（“空間生成用メタデータ”）が m 行 n 列の行列（以下，“データ行列”）の形で与えられる。 m 個の基本データはそれぞれ n 個の特徴 (feature) によって特徴づけられている。このデータ行列から正規直交空間（以下，“メタデータ空間”） MDS を生成する。

(2) 画像データの n 次元ベクトル表現：

(1) で用いた n 個の特徴と同一の特徴を用いて、個々の画像を n 次元ベクトルで表現する。これを画像のメタデータとよぶ。また、検索者の印象や画像の内容を表すキーワードあるいは文脈語を n 次元ベクトルで表現する。これを検索語のメタデータとよぶ。

(3) メタデータ空間 MDS 上への写像：

n 次元ベクトルで表現された 3 種類のメタデータをメタデータ空間上へ写像する。これにより、同じ空間上に言葉と画像が配置されることになり、言葉と画像の動的な関係を空間上での距離として計算することが可能になる。

(4) 意味的連想検索 :

印象や画像の内容を決定づける文脈が検索語の並びとして与えられる。このとき、文脈として与えられた検索語の列により、メタデータ空間から動的に固有部分空間（以下、“意味空間”）を選択する。そして、この部分空間上で、文脈に対応する（最も関連の深い）画像を選び出す。

2.2 画像のメタデータ作成方式

ここでは、画像を表現するメタデータを形成する方法を示す。

(1) STEP1: オブジェクトの抽出

t 個のオブジェクト（あるいは、 t 個の印象語） $\mathbf{o}_1, \mathbf{o}_2, \dots, \mathbf{o}_t$ から成る画像 P を次のように定義する。

$$P = \{\mathbf{o}_1, \mathbf{o}_2, \dots, \mathbf{o}_t\}.$$

ここで、各オブジェクト \mathbf{o}_i は、データ行列の特徴と同一の特徴を用いて表現される特徴付ベクトルである。

$$\mathbf{o}_i = (o_{i1}, o_{i2}, \dots, o_{in})$$

(2) STEP2: 画像 P のベクトル表現

画像 P を構成する t 個のオブジェクト $\mathbf{o}_1, \mathbf{o}_2, \dots, \mathbf{o}_t$ が、それぞれ n 次元のベクトルで定義されている。オブジェクト $\mathbf{o}_1, \mathbf{o}_2, \dots, \mathbf{o}_t$ の和演算子 \oplus を次のように定義し、画像のメタデータを形成する。

$$\bigoplus_{i=1}^t \mathbf{o}_i \equiv (\text{sign}(o_{\ell_1 1}) \max_{1 \leq i \leq t} |o_{i1}|, \text{sign}(o_{\ell_2 2}) \max_{1 \leq i \leq t} |o_{i2}|, \dots, \text{sign}(o_{\ell_n n}) \max_{1 \leq i \leq t} |o_{in}|).$$

ここで $\text{sign}(a)$ は、“ a ” の符号（正、負）を表す。また、 $l_k (k = 1, \dots, t)$ は、特徴が最大となるオブジェクトを示す指標であり、次のように定義する。

$$\max_{1 \leq i \leq t} |o_{ik}| = |o_{l_k k}|.$$

3. メタデータ空間の設定と基本機能

3.1 メタデータ空間 MDS の設定

初めに、 m 個の基本データについて各々 n 個の特徴 (f_1, f_2, \dots, f_n) を列挙した特徴付ベクトル $\mathbf{d}_i (i = 1, \dots, m)$ が与えられているとし、そのベクトルを並べて構成する $m \times n$ 行列を M とおく（図1）。このとき、 M は列ごとに2ノルムで正規化されている。

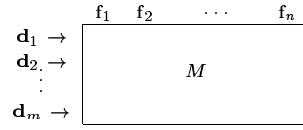


図1 データ行列 M によるメタデータの表現
Fig. 1 Metadata represented in data matrix M

- (1) データ行列 M の相関行列 $M^T M$ を計算する。
- (2) $M^T M$ を固有値分解する。

$$M^T M = Q \begin{pmatrix} \lambda_1 & & & & \\ & \ddots & & & \\ & & \lambda_\nu & & \\ & & & \ddots & \\ & & & & 0 \dots 0 \end{pmatrix} Q^T,$$

$$0 \leq \nu \leq n.$$

ここで行列 Q は、

$$Q = (\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n)$$

である。この $\mathbf{q}_i (i = 1, \dots, n)$ は、相関行列の正規化された固有ベクトル（以下、“意味素”）である。相関行列の対称性から、この固有値は全て実数であり、その固有ベクトルは互いに直交している。

- (3) メタデータ空間 MDS を以下で定義する。

非ゼロ固有値に対応する固有ベクトル（以下、“意味素”と呼ぶ）によって形成される正規直交空間をメタデータ空間 MDS と定義する。この空間の次元 ν は、データ行列のランクに一致する。この空間は、 ν 次元ユークリッド空間となる。

$$MDS := \text{span}(\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_\nu).$$

$\{\mathbf{q}_1, \dots, \mathbf{q}_\nu\}$ は MDS の正規直交基底である。

3.2 意味射影集合 Π_ν の設定

メタデータ空間 MDS から固有部分空間（以下、意味空間）への射影（以下、“意味射影”）の集合 Π_ν を考える。 P_{λ_i} を次の様に定義する。

$$P_{\lambda_i} \stackrel{d}{\longleftrightarrow} \lambda_i \text{ に対応する固有空間への射影} \\ \text{i.e. } P_{\lambda_i} : MDS \rightarrow \text{span}(\mathbf{q}_i).$$

意味射影の集合 Π_ν を次のように定義する。

$$\Pi_\nu := \{ \mathbf{0}, P_{\lambda_1}, P_{\lambda_2}, \dots, P_{\lambda_\nu}, \\ P_{\lambda_1} + P_{\lambda_2}, P_{\lambda_1} + P_{\lambda_3}, \dots, P_{\lambda_{\nu-1}} + P_{\lambda_\nu}, \\ \vdots \\ P_{\lambda_1} + P_{\lambda_2} + \dots + P_{\lambda_\nu} \}.$$

i 次元の意味空間は、 $\frac{\nu(\nu-1)\cdots(\nu-i+1)}{i!}$, ($i = 1, 2, \dots, \nu$) 個存在するので、射影の総数は、 2^ν となる。つまり、このモデルは、 2^ν 通りの意味の様相の表現能力をもつ。

3.3 意味解釈オペレータ S_p の構成

検索者の印象や画像の内容を与える文脈を表す ℓ 個の検索語列

$$s_\ell = (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_\ell)$$

と、しきい値 ε_s ($0 < \varepsilon_s < 1$) が与えられたとき、それに応じた、意味射影 $P_{\varepsilon_s}(s_\ell)$ を構成するオペレータ (以下, “意味解釈オペレータ”) S_p を構成する。 T_ℓ を長さ ℓ の検索語列の集合とすると、 S_p は、次のように定義される。

$$S_p: T_\ell \mapsto \Pi_\nu$$

$$\text{ここで, } T_\ell \ni s_\ell, \Pi_\nu \ni P_{\varepsilon_s}(s_\ell).$$

また、 $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_\ell\}$ の各要素は、特徴付ベクトルであり、データ行列 M の特徴と同一の特徴を用いて表される。

オペレータ S_p は以下の計算を行う。

(1) \mathbf{u}_i ($i = 1, 2, \dots, \ell$) をフーリエ展開する。

検索語列 s_ℓ を構成する ℓ 個の検索語を各々メタデータ空間 MDS へ写像する。

この写像では、 ℓ 個の単語を各々メタデータ空間 MDS 内でフーリエ展開し、フーリエ係数を求める。これは、各検索語と各意味素の相関を求めることに相当する。

\mathbf{u}_i と \mathbf{q}_j の内積 u_{ij} は次のようになる。

$$u_{ij} := (\mathbf{u}_i, \mathbf{q}_j), \quad j = 1, 2, \dots, \nu.$$

ベクトル $\hat{\mathbf{u}}_i \in MDS$ を次のように定める。

$$\hat{\mathbf{u}}_i := (u_{i1}, u_{i2}, \dots, u_{i\nu}).$$

これは、単語 \mathbf{u}_i をメタデータ空間 MDS に写像したものである。

(2) 検索語列 s_ℓ の意味重心 $\mathbf{G}^+(s_\ell)$ を求める。

まず、各意味素ごとに、フーリエ係数の総和を求める。これは、検索語列 s_ℓ と各意味素との相関を求めることに相当する。このベクトルは、 ν 個の意味素があるため、 ν 次元ベクトルとなる。このベクトルを、無限大ノルムによって正規化したベクトルを、以下、検索語列 s_ℓ の意味重心 $\mathbf{G}^+(s_\ell)$ と呼ぶ。

$$\mathbf{G}^+(s_\ell) := \frac{\left(\sum_{i=1}^{\ell} u_{i1}, \dots, \sum_{i=1}^{\ell} u_{i\nu} \right)}{\left\| \left(\sum_{i=1}^{\ell} u_{i1}, \dots, \sum_{i=1}^{\ell} u_{i\nu} \right) \right\|_\infty}$$

ここで、 $\|\cdot\|_\infty$ は無限大ノルムを示す。

(3) 意味射影 $P_{\varepsilon_s}(s_\ell)$ を決定する。

検索語列 s_ℓ の意味重心を構成する各要素において、しきい値 ε_s を越える要素に対応する意味素を、画像のメタデータを射影する意味空間の構成に用いる。意味射影 $P_{\varepsilon_s}(s_\ell)$ を次のように決定する。

$$P_{\varepsilon_s}(s_\ell) := \sum_{i \in \Lambda_{\varepsilon_s}} P_{\lambda_i} \in \Pi_\nu.$$

ただし $\Lambda_{\varepsilon_s} := \{i \mid |(\mathbf{G}^+(s_\ell))_i| > \varepsilon_s\}$ とする。

3.4 意味空間における相関の定量化

文脈により選択された部分空間上で、文脈に対応した画像を選び出す方法を示す。

メタデータ空間に写像された画像メタデータの意味空間におけるノルムを求めることにより、文脈に対応した画像の探索を行う。意味空間における画像メタデータのノルムの大きさをその文脈と画像メタデータとの関連の強さとする。

画像 \mathbf{x} のノルム $\rho(\mathbf{x}; s_\ell)$, $\mathbf{x} \in MDS$ を次のように定める。

$$\rho(\mathbf{x}; s_\ell) = \frac{\sqrt{\sum_{j \in \Lambda_{\varepsilon_s} \cap S} \{c_j(s_\ell)x_j\}^2}}{\|\mathbf{x}\|_2},$$

ただし、 $S = \{i \mid \text{sign}(c_i(s_\ell)) = \text{sign}(x_i)\}$ とする。

ここで、意味空間を構成する軸において、文脈に関係しているのは、正と負のどちらか一方である。そこで、意味空間を構成する軸の符合を考慮するため、意味空間を構成する軸の符合と正負が逆の成分についてはノルムの計算において無視している。

また、画像を特徴づける特徴の数が多いと、どのような意味空間が選ばれても、意味空間における画像のノルムが大きくなる傾向がある。そのため、本来、文脈との相関が強いと考えられる画像メタデータのノルムよりも、特徴の数が多い画像メタデータのノルムが大きくなってしまい、適切な抽出が行なわれないことがある。そのため、メタデータ空間での画像メタデータの2ノルムで正規化を行なっている。

4. 学習機構

異なる検索者間では、文脈による画像の内容や印象の表現においては、同じ表現であっても、抽出が期待さ

れる画像が異なることがある。また、同じイメージを表現する場合に、検索者ごとに異なる表現を行うことがある。

そのため、個々の検索者について、正しい検索を行うことができるように、個人差を扱う機構が必要となる。そこで、次の学習機構を実現する。

これらの個人差は、画像から受ける印象が検索者ごとに異なる、あるいは、画像の内容や印象を表現するとき用いる検索語についての認識が検索者ごとに異なるために生じるものと考えられる。本意味的画像探索方式においては、前者はメタデータ空間での画像の位置に、後者はメタデータ空間での検索語の位置に関して個人差があることに対応する。

本学習機構では、画像のメタデータは、複数の検索者が利用する正しく定義された共有資源であると考え、検索者の印象や画像の内容を表す文脈を構成する際に検索者が利用する検索語群のメタデータを検索者ごとに修正する。つまり、ある文脈から連想される画像が検索者の印象に合う画像になるように、検索語の特徴づけを修正する。

以上の学習機構を、意味重心の移動、および、その移動量の検索語への反映の2段階の処理によって実現する。

4.1 意味重心の移動

検索者の印象や画像の内容を示す文脈を表す文脈語群 s_ℓ とその文脈と最も相関が強くなるべき画像 P_t のメタデータのベクトル \mathbf{g} が与えられたとき、次の2つの前提をもとに、検索語列 s_ℓ の表す文脈と画像 P_t の相関が最も強くなるように意味重心 \mathbf{G}^+ の移動を行う。

[前提]

(1) 意味重心 \mathbf{G}^+ とベクトル \mathbf{g} のなす角が小さいときほど、文脈語群 s_ℓ の表現する文脈と画像 P_t の相関が強くなる。

(2) 検索語を表すベクトルの変化量を最小化するため、意味重心 \mathbf{G}^+ の移動は、最小限にとどめる。

意味重心の移動は、画像 P_t のメタデータのベクトルと意味重心 \mathbf{G}^+ で張られる平面上で以下のアルゴリズムによって行う。

[アルゴリズム]

$\mathbf{G} := \mathbf{G}^+$;

while ($\theta \geq \delta$) **do begin**

新しい意味重心ベクトルの候補 \mathbf{G}_{new}^+ を図2の

図2 新しい意味重心の候補

Fig.2 A candidate for the new semantic center

ように θ のなす角を2分して求める；

\mathbf{G}_{new}^+ を用いて意味空間（部分空間）を再選択する；

if (その部分区間において目的の画像のノルムが最大である) **then**

$\mathbf{G} := \mathbf{G}^+$;

else

$\mathbf{G} := \mathbf{G}^+$;

end

ただし、 δ は、意味重心の移動を行う場合の θ と δ の最小角度とする。

4.2 検索語の位置の修正

検索語列 s_ℓ により定まる意味重心が、新しい意味重心 \mathbf{G}_{new}^+ になるように、検索語列 s_ℓ を構成している検索語のメタデータ空間での位置を修正する2つの方法を示す。これらの方法により、もとの意味重心から新しい意味重心への移動量を検索語列 s_ℓ を構成している検索語に反映させる。

4.2.1 方法1: 均等配分による方法

検索語列 s_ℓ により求められた元の意味重心

$$\mathbf{G}^+ = (g_1, g_2, \dots, g_\nu)$$

から、4.1節で示した方法により求めた新しい意味重心

$$\mathbf{G}_{new}^+ = (g'_1, g'_2, \dots, g'_\nu)$$

への移動量を検索語列の各検索語

$$\mathbf{u}_i = (u_{i1}, u_{i2}, \dots, u_{i\nu}), \quad i = 1, 2, \dots, l$$

に均等に反映させる。この方法では、検索語の平均移動距離が最小となる。学習後の各検索語

$$\mathbf{u}'_i = (u'_{i1}, u'_{i2}, \dots, u'_{i\nu}), \quad i = 1, 2, \dots, l$$

の各軸の成分を次のように定める。

$$u'_{mk} = u_{mk} + \frac{1}{l} \max_{1 \leq j \leq \nu} \sum_{i=1}^l u_{ij}(g'_k - g_k),$$

$$m = 1, 2, \dots, l, \quad k = 1, 2, \dots, \nu.$$

4.2.2 方法2: 比例配分による方法

意味重心の各軸の重みを決定している各検索語の成分の大きさは検索語ごとに異なる。ここでは、軸の成分の大きい検索語、つまり文脈に強く影響を与えている検索語ほど、その軸の成分は、検索者ごとの個人差が大きいと考える。

そこで、学習前の各検索語の各軸の成分の大きさに応じて、意味重心の各軸の移動量を各検索語の成分に反映させる。学習後の各検索語の各軸の成分を次のように定める。

$$u'_{mk} = \begin{cases} u_{mk} + \frac{|u_{mk}|}{\sum_{i=1}^l |u_{ik}|} \max_{1 \leq j \leq \nu} \sum_{i=1}^l u_{ij}(g'_k - g_k), & \sum_{i=1}^l |u_{ik}| \neq 0 \\ u_{mk} + \frac{1}{l} \max_{1 \leq j \leq \nu} \sum_{i=1}^l u_{ij}(g'_k - g_k), & \sum_{i=1}^l |u_{ik}| = 0 \end{cases}$$

$$m = 1, 2, \dots, l, \quad k = 1, 2, \dots, \nu.$$

5. 提案方式の実現

実際に提案方式を Sun SPARC システム上で C 言語により実現した。

5.1 空間生成用メタデータの生成

空間生成用メタデータの生成、すなわち、データ行列 M の生成を行うために、“The General Basic English Dictionary⁽¹⁵⁾”(以下、“BD”と呼ぶ)と、“Longman Dictionary of Contemporary English⁽¹⁴⁾”(以下、“LD”と呼ぶ)の2冊の英英辞典を使用した^{(5),(7)}。このBDは、850語の基本語だけを用いて20,000語の説明をしており、LDは、2000語の基本語だけを用いて56,000語の説明をしている。このBDの基本語に合成語を加え、冠詞、be動詞、代名詞、間投詞、接続詞、前置詞、助動詞を取り除いた単語群(874語)をデータ行列 M の列、すなわち、特徴とした。また、LDの基本語に特徴の単語群を加えた単語群(2329語)を、その説明文と共にBDから取り出した。

次の操作を行うことにより、メタデータ空間の作成に使用するデータ行列 M を自動生成した。空間生成用メタデータの各単語(2329語)について、各単語の説明語としてBDから取り出した基本語群を用いて、2329行874列の行列を作成した。その単語を説明する基本語が肯定の意味に用いられていた場合“1”、否定の場合“-1”、使用されていない場合“0”とし、見出し語自身が特徴である場合その特徴の要素を“1”として自動生成する。その操作後に、列ごとに2ノルムで正規化する。

5.2 検索語メタデータの生成

メタデータ空間へ写像する検索語群のメタデータ生成については、各検索語を説明するBDの基本語群からの行ベクトルの生成において、その単語を説明する基本語が肯定の意味に用いられていた場合、その特徴の要素を“1”、否定の場合“-1”、使用されていない場合“0”とし、検索語自身が特徴語である場合、その特徴の要素を“1”として自動生成する。

5.3 画像メタデータの生成

内容による定義を行なった画像群、および、印象による定義を行なった画像群の2種類の探索対象画像群を用意した。

内容によるメタデータ定義を行う画像群として、“The Oxford Children's Picture Dictionary⁽¹⁶⁾”(以下、“PD”)の21枚の画像を用いた。PDは、絵とそれに対応する単語(空間を構成している事物の名前)によって、日常生活の空間を説明している。空間を構成する事物を示す単語群をその説明文とともにBDから取り出し、これをもとに、2.2節で示した方法によって画像メタデータの作成を行った。各オブジェクトのベクトル化は検索語メタデータの生成と同様の方法で行った。

印象による定義を行う画像群として、30枚の絵画を用意した(Chagall⁽¹⁷⁾, Corot⁽¹⁸⁾, Hokusai⁽¹⁹⁾, Loirand⁽²⁰⁾, Nelson⁽²¹⁾, Renoir⁽²²⁾, Sarthou⁽²³⁾, Hiro⁽²⁴⁾)。複数の印象語で各絵画の特徴づけを行ない、各印象語を1オブジェクトとして画像の内容による定義と同様に画像メタデータの生成を行なった。

6. 実験

6.1 実験方法

表1に示す4検索語群についての学習に関する実験を行った。実験では、それぞれの検索語群の中の検索語のみを用いて検索語列を構成し、本来、その検索語列によって形成される文脈に相関が最も強くあるべき画像を指定し、4.2節に示した2方法によって学習操作を行っ

表 1 学習対象とした検索語

内容		印象	
検索語群 A	検索語群 B	検索語群 c	検索語群 d
vehicle	transport	happy	glad
plant	vegetation	bright	gay
water	water	dark	vague
fire	burn	force	power
sand	sand	slight	feeble
animal	beast	change	change
shop	store	still	rest
food	provision	gentle	gentle
people	person	splendid	grace
cage	cage		

表 2 画像とその画像を構成するオブジェクト

画像 id	画像を構成するオブジェクト
spcib	horse cart donkey sheep goat fence chicken rabbit cow bird hedge pig duck
spcic	flats factory cinema school restaurant post office chemist bank hotel hairdresser butcher dentist shop car-park
spcid	vegetables fruit bread shelf basket freezer meat trolley check-out sweets tea customer coffee bottle
spcih	cooker grill oven can cheese frying-pan pan kettle sink milk egg
spcim	fireman smoke ladder fire-escape fire building water fire-engine ambulance crowd policeman axe bell
spcio	fair canoe shell drink picnic net spade bucket sand
spciu	eagle cage tiger crocodile snake lion fish dolphin camel bear elephant penguin monkey panda

表 3 画像とその画像を構成する印象語

画像 id	画像を定義している印象語群
corotb	beautiful delicate calm
corotd	shine beautiful calm
hirobb	cheer dim quiet
hokusaib	fight motion calm
hokusaid	vivid motion speed
loirandb	delight shine calm
loirandc	delight grand calm
nelsona	grand dynamic motion
renoir	dim tender quiet
renoird	fine strong quiet

た. 検索語群 A, B の学習では, 内容により定義された画像群を対象とした. 学習に用いた各画像を構成するオブジェクトの内容を表 2 に示す. また, 画像の検索語群 C, D の学習では, 印象により定義された画像群を対象とした. 学習に用いた各画像の印象語による定義を表 3 に示す.

ここで検索語として用いた各単語は, 画像の定義語の中に現れていない点が重要である. 本方式では, このような状況においても, 意味空間における検索語の意味解釈により, 文脈と画像との意味的な相関の強さを計算することができる.

6.2 実験結果

与えられた文脈と, その文脈によって, 本来, 抽出されるべき画像との相関について, 学習による変化を表 4, 5, 6, 7 に示す. 各表は, 学習の対象となっている全画像について, それぞれ 1 回の検索語メタデータのベクトルの修正を行なった場合を学習回数 1 回としてカウントし, n 回学習を行った場合における文脈との相関の強さの順位を示す.

画像内容による検索では, 学習前の段階では, 本来, 1 位に選ばれるべき画像が 2 1 枚の比較対象画像群の中で, 1 ~ 6 位に入っている. また, 印象による検索では, 本来, 1 位に選ばれるべき画像が 3 0 枚の比較対象画像群の中で, 1 ~ 8 位に入っている. 本画像探索方式により, 学習を伴わない場合においても, 内容および印象による画像検索において, 本来, 選択されるべき画像が上位に位置付けられている. しかし, 学習が行なわれない段階では, 文脈表現に関する個人差に対応することはできていない.

学習の効果についての実験では, 本学習機構を検索語群に対して繰り返し適用することにより, 本来, 1 位に選ばれるべき画像の順位を最終的に 1 位にすることができることを確認できた.

この実験では, 画像 spciu, および, 画像 renoird を対象とした学習がそれぞれ常に各回の最後に行われるため, 各回の学習の際に, それぞれ与えた文脈と画像 spciu, および, 画像 renoird との相関は常に最も強くなっている. 他の画像については, 各回において, その画像を対象とした学習以降に行われる学習による検索語の特徴付けの変更の影響を受けるため, 学習回数 1 回では相関が最も強くないことがある. しかし, 学習を繰り返すことにより与えた文脈との相関が強くなっていく.

この実験においては, 均等配分による方法との比較に

表4 検索語群Aの学習結果

(a) 均等配分による方法を用いた場合

画像 id	n 回学習後の順位			与えた文脈
	学習前	1 回	2 回	
spcib	2	1	1	vehicle plant animal
spcic	6	1	1	vehicle building shop
spcid	3	3	1	shop food people
spcih	2	1	1	water fire food
spcim	3	1	1	vehicle water fire building
spcio	6	1	1	water food sand
spciu	1	1	1	animal cage

(b) 比例配分による方法を用いた場合

画像 id	n 回学習後の順位						与えた文脈
	学習前	1 回	2 回	3 回	4 回	5 回	
spcib	2	2	2	2	2	1	vehicle plant animal
spcic	6	2	2	2	2	1	vehicle building shop
spcid	3	2	1	1	1	1	shop food people
spcih	2	2	1	1	1	1	water fire food
spcim	3	1	1	1	1	1	vehicle water fire building
spcio	6	1	1	1	1	1	water food sand
spciu	1	1	1	1	1	1	animal cage

において、比例配分による方法では、学習の際に与えた各文脈とそれに対応する画像の相関が最も強くなるまでに要する学習回数が多くなっている。これは、比例配分による方法では均等配分による方法に比べ、学習による各検索語の移動量が大きくなる傾向があるため、文脈語群を構成する際に、同じ検索語を含む他の文脈による画像の検索への影響が強くなるためであると考えられる。したがって、学習においては、均等配分による方法が有効であると考えられる。

7. おわりに

本論文では、意味の数学モデルを適用した意味的画像探索方式を提案し、さらに、その学習機構を示した。また、本学習機構により、検索者の主観を反映させた画像の抽出が可能になることを示した。

本方式により、検索者の印象や画像の内容による画像検索を実現し、また、学習機構により、個々の検索者が同じ印象を異なる表現によって与える場合に、印象表現の違いに対応することを可能とした。

現在の実現では、データベースへ新たな画像データを追加する場合には、新たにその画像データの特徴を表す言葉を割り当てる必要がある。今後、画像データの配色、あるいは、その解説文から画像データの特徴を自動抽出する方式の検討を行っていく予定である。また、学習機構における意味重心の移動量を検索語に反映させる他の方法の開発、本探索方式の動画、音声など他のメディア情報への適用を行っていく予定である。さらに、

マルチメディアデータベース・システム⁽⁸⁾における一機能としての本方式の統合を行っていく予定である。

謝辞 画像検索方式について御助言頂き、また、実験データの作成に御協力頂いた中村恭子氏（お茶の水女子大学理学部）に深謝致します。

文 献

- (1) 有澤 博, 由井 仁, 富井 尚志, "画像データベースシステムの構成の一方式," Proc. Advanced Database System Symposium '93, pp.181-190, 1993.
- (2) S. Deerwester, S. T. Dumais, T. K. Landauer, G. W. Furnas, and R. A. Harshman, "Indexing by Latent Semantic Analysis," Journal of the American Society for Information Science, Vol.41, No.6, pp.391-407, 1990.
- (3) 平田 恭二, 原良 憲, "概略画像を用いた画像検索," 信学技報, DE92-2, 1992.
- (4) T.Kitagawa, and Y.Kiyoki, "The Mathematical Model of Meaning and its Application to Multidatabase Systems," Proc. 3rd IEEE International Workshop on Research Issues on Data Engineering: Interoperability in Multidatabase Systems, pp.130-135, April 1993.
- (5) Y. Kiyoki, and T. Kitagawa, "A Metadatabase System for Supporting Semantic Interoperability in Multidatabases," Information Modelling and Knowledge Bases, Vol. V, IOS Press, 1993.
- (6) Y. Kiyoki, T. Kitagawa and Y. Hitomi, "A Fundamental Framework for Realizing Semantic Interoperability in a Multidatabase Environment," International Journal of Integrated Computer-Aided Engineering (John Wiley & Sons), Jan. 1995.

表 5 検索語群 B の学習結果

(a) 均等配分による方法を用いた場合

画像 id	n 回学習後の順位					与えた文脈
	学習前	1 回	2 回	3 回	4 回	
spcib	1	3	1	2	1	transport vegetation beast
spcic	2	1	2	2	1	transport structure store
spcid	5	4	2	1	1	store provision person
spcih	2	1	1	1	1	water burn provision
spcim	4	1	1	1	1	transport water burn structure
spcio	6	1	1	1	1	water provision sand
spciu	1	1	1	1	1	beast cage

(b) 比例配分による方法を用いた場合

画像 id	n 回学習後の順位									与えた文脈
	学習前	1 回	2 回	3 回	4 回	5 回	6 回	7 回		
spcib	1	5	2	2	2	2	2	2	1	transport vegetation beast
spcic	2	2	2	2	2	2	2	2	1	transport structure store
spcid	5	2	1	1	1	2	1	1	1	store provision person
spcih	2	1	1	1	2	1	1	1	1	water burn provision
spcim	4	2	2	1	1	2	1	1	1	transport water burn structure
spcio	6	1	1	1	1	1	1	1	1	water provision sand
spciu	1	1	1	1	1	1	1	1	1	beast cage

- (7) Y. Kiyoki, T. Kitagawa and T. Hayama, "A Meta-database System for Semantic Image Search by a Mathematical Model of Meaning", ACM SIGMOD Record, Vol. 23, No. 4, pp.34-41, Dec. 1994.
- (8) Y. Kiyoki, and T. Hayama, "The Design and Implementation of a Distributed System Architecture for Multimedia Databases," Proc. 47th Conference of International Federation for Information and Documentation, pp.374-379, Oct. 1994.
- (9) T. Kurita, T. Kato, "Learning A Cognitive Schema for Multimedia Indexing -Experiments on Image Database-", Technical Reports of IEICE, DE93-3, May 1993.
- (10) 栗田多喜夫, 加藤俊一, 福田郁美, 板倉あゆみ, "印象語による絵画データベースの検索," 情処学論, Vol. 33, No. 11, pp.1373-1383, Nov. 1992.
- (11) Y. Masunaga, "A Temporal Expression to the Multimedia Object Model in OMEGA," Proc. of DAS-FAA'95, pp.430-440, April 1995.
- (12) 坂内正夫, "画像検索技術," 信学誌, Vol.71, No.9, pp.911-914, 1988.
- (13) 渡辺靖彦, 中村裕一, 長尾眞, "絵画解説文の対象情報・感性的情報の抽出," 情処研報 93-CH-20, Vol.93, No.105, pp.13-20, Nov. 1993.
- (14) "Longman Dictionary of Contemporary English," Longman, 1987.
- (15) Ogden, C.K., "The General Basic English Dictionary," Evans Brothers Limited, 1940.
- (16) L. A. Hill and Charles Inne, Illustrated by Barry Rowe, "The Oxford Children's Picture Dictionary," Oxford University Press, 1981.
- (17) ミシェル・マカリウス, 小勝 禮子訳, 岩波 世界の巨匠『シャガール』, 岩波書店, 1992.
- (18) マドレーヌ・ウール著, 田中 淳一訳, BSS ギャラリー 世界の巨匠『コロロ』
- (19) 永田 生慈 監修・執筆, 『北斎美術館 第2巻 風景画』, 集英社, 1990.
- (20) アール・ヴィジョン・ア・パリ 24, 『モーリス・ロワラン (Maurice LOIRAND)』,
- (21) ロバート・リン・ネルソン著, 『海からのメッセージ ジ・アート・オブ・ロバート・リン・ネルソン』, マリン企画, 1992.
- (22) ソフィー・モヌレ, 隠岐 由紀子訳, 『ルノワール』, 中央公論社, 1991.
- (23) アール・ヴィジョン・ア・パリ 4, 『モーリス・エリ・サルトウー (Maurice-Elie SARTHOU)』, 京都書院, 1993
- (24) 山形 博導, 室伏 哲郎, ヒロ・ヤマガタ画集『虹の狩人』, 講談社, 1993.

表 6 検索語群 C の学習結果

(a) 均等配分による方法を用いた場合

画像 id	n 回学習後の順位			与えた文脈
	学習前	1 回	2 回	
corotb	5	1	1	splendid slight gentle
corotd	3	2	1	bright splendid gentle
hiroba	1	1	1	happy dark still
hokusaib	5	3	1	force change gentle
hokusaid	7	2	1	bright change
loirandb	1	1	1	happy bright gentle
loirandc	1	1	1	happy splendid gentle
nelsona	1	1	1	splendid force change
renoira	3	2	1	dark slight still
renoird	8	1	1	bright force still

(b) 比例配分による方法を用いた場合

画像 id	n 回学習後の順位					与えた文脈
	学習前	1 回	2 回	3 回	4 回	
corotb	5	4	1	1	1	splendid slight gentle
corotd	3	6	4	1	1	bright splendid gentle
hiroba	1	3	1	1	1	happy dark still
hokusaib	5	1	1	4	1	force change gentle
hokusaid	7	2	1	1	1	bright change
loirandb	1	2	2	1	1	happy bright gentle
loirandc	1	1	1	1	1	happy splendid gentle
nelsona	1	1	1	1	1	splendid force change
renoira	3	1	1	1	1	dark slight still
renoird	8	1	1	1	1	bright force still

表 7 検索語群 D の学習結果

(a) 均等配分による方法を用いた場合

画像 id	n 回学習後の順位				与えた文脈
	学習前	1 回	2 回	3 回	
corotb	1	1	1	1	grace feeble gentle
corotd	1	4	2	1	gay grace gentle
hiroba	1	1	1	1	glad vague rest
hokusaib	6	1	1	1	power change gentle
hokusaid	1	1	1	1	gay change
loirandb	2	1	1	1	glad gay gentle
loirandc	3	1	1	1	glad grace gentle
nelsona	1	1	1	1	grace power change
renoira	6	1	1	1	vague feeble rest
renoird	1	1	1	1	gay power rest

(b) 比例配分による方法を用いた場合

画像 id	n 回学習後の順位								与えた文脈
	学習前	1 回	2 回	3 回	4 回	5 回	6 回	7 回	
corotb	1	9	2	1	1	1	1	1	grace feeble gentle
corotd	1	6	5	4	5	5	1	1	gay grace gentle
hiroba	1	3	2	1	1	1	1	1	glad vague rest
hokusaib	6	3	3	4	4	5	2	1	power change gentle
hokusaid	1	1	1	3	1	1	1	1	gay change
loirandb	2	3	2	1	3	2	1	1	glad gay gentle
loirandc	3	1	1	1	1	1	1	1	glad grace gentle
nelsona	1	1	1	1	1	1	1	1	grace power change
renoira	6	1	1	1	1	1	1	1	vague feeble rest
renoird	1	1	1	1	1	1	1	1	gay power rest

(平成 x 年 xx 月 xx 日受付)

清木康

1978年慶応義塾大学工学部電気工学科卒業。1983年同大学院工学研究科博士課程修了。工学博士。同年、日本電信電話公社武蔵野電気通信研究所入所。1984年より筑波大学電子・情報工学系に勤務。現在同学系助教授。データベースシステム、知識ベースシステム、分散・並列処理システムの研究に従事。ACM, IEEE, 情報処理学会, 日本ソフトウェア科学会各会員。

金子昌史

1994年筑波大学第三学群情報学類卒業。現在同大学院工学研究科修士課程に在籍中。データベースシステムの研究に従事。

北川 高嗣

1978年名古屋大学工学部情報工学科卒業。1983年同大学院工学研究科博士課程修了。工学博士。1990年より筑波大学電子・情報工学系に勤務。現在同学系助教授。関数解析, 逆問題の研究に従事。日本応用数理学会会員。