

意味の数学モデルによる異種メディア間検索の実現方式

An Implementation Method of Heterogonous Media to Media Search utilizing Mathematical Model of Meaning

中西 崇文^{*} 北川 高嗣^{*} 清木 康

Takafumi NAKANISHI Takashi KITAGAWA
Yasushi KIYOKI

本稿では、意味の数学モデルを用いた異種メディア間検索の実現方式を示す。本方式は、メディアデータを対象としたメタデータ自動抽出方式を言葉の相関を計量できる意味の数学モデルのメタデータ空間により連結することで異種メディア間の検索を実現する。本方式により、これまで独立に実装された異種メディアデータを統一的に扱うことが可能になる。これにより、新しい情報生成が可能となり、既存のデータの利用機会を増大させ、データベース群の利用価値を飛躍的に増大させることが可能となると考えられる。本稿では、画像メディアデータから顔の表情の抽出することを対象とした異種メディア間検索の実現方式を示す。

This paper presents an implementation method of heterogonous media to media search utilizing mathematical model of meaning. This method is realized by measuring correlation of the impression words extracted from automatic metadata extraction method for each mediadata. This method realizes connection of the heterogeneous mediadata mounted independently and can work these different-species mediadata systematically. Thereby, this method increases the use opportunity of the existing data, and it is possible to increase the utility value of a database group by leaps and bounds. In this paper, we show an implementation method of heterogeneous media to media search accordingly the media transformation like picture to facial expression.

1. はじめに

近年、広域高速ネットワーク上の多数のサイトに散在している膨大な量のメディアデータ群を対象として、適切な検索方式、および異種メディア間を統一的に扱い、編集・統合する方式の実現が重要となってきた。

我々は、これまでメディアデータに対応するメタデータを言葉によって表現し、検索者の与える文脈に応じた意味的解釈を伴う間接的な検索方式として、メディアデータを対象とした意味の数学モデルによる、意味的連想検索方式[1][2][3]

を提案している。この方式は、言葉の意味を文脈に応じて解釈する機構より、言葉と言葉、あるいは、言葉とメディアデータ間の意味的な関係を与えられた文脈や状況に応じて動的に計算することを可能としている。また、我々は、文献[2][5][6]でメディアデータのメタデータを自動抽出するための実現方式について示している。特に文献[5]では、メディアデータの印象を表す語をメタデータとして自動抽出する枠組みとして、Media-lexco Transformation Operatorを示している。さらに、文献[6]ではメタデータを抽出する際に人間の感性や感覚に基づいた関数を適用する方式を示している。

本稿では、画像メディアデータと人間の顔の表情を対象とした異種メディア間検索の実現方式を示す。本方式はメディアデータから印象語を抽出するメタデータ自動抽出する方式[2][5][6]を意味の数学モデルのメタデータ空間により連結することにより、印象と合致した異種のメディアデータ間の検索が可能となる。これにより、異種メディア間を統一的に扱うことが可能となる。また、異種メディアの連結による新しい情報生成が可能となり、既存の情報利用機会を増大させ、データベース群の利用価値を飛躍的に増大させることが可能になると考えられる。

本稿では、画像メディアデータから顔の表情を抽出することを対象とした異種メディア間検索の実現方式を示す。

2. 意味の数学モデルの基本構成

本節では、人間が様々な印象を表す際に用いられる様々な単語(以下、印象語)によって表現した問い合わせに対応したメディアデータを検索することを目的とした意味の数学モデルによるメディアデータを対象とした意味的連想検索方式の概要を示す。詳細は、文献[1][2][3]に述べられている。

(1) メタデータ空間 MDS の設定

検索対象となるメディアデータをベクトルで表現したデータをマッピングするための正規直交空間(以下、メタデータ空間 MDS)を設定する。

(2) メタデータをメタデータ空間 MDS へ写像

設定されたメタデータ空間 MDS へメディアデータのメタデータをベクトル化し写像する。これにより、同じ空間に検索対象データのメタデータがメタデータ空間上に配置されることになり、検索対象データ間の意味的な関係を空間上でのノルムとして計算することが可能となる。

(3) メタデータ空間 MDS の部分空間(意味空間)の選択

検索者は与える文脈を複数の単語を用いて表現する。検索者が与える単語の集合をコンテキストと呼ぶ。このコンテキストを用いてメタデータ空間 MDS に各コンテキストに対応するベクトルを写像する。これらのベクトルは、メタデータ空間 MDS において合成され、意味重心を表すベクトルが生成される。意味重心から各軸への射影値を相関とし、閾値を超えた相関値(以下、重み)を持つ軸からなる部分空間(以下、意味空間)が選択される。

(4) メタデータ空間 MDS の意味空間における相関の定量化

選択されたメタデータ空間 MDS の部分空間(意味空間)において、メディアデータベクトルのノルムを検索語列との相関として計量する。これにより、与えられたコンテキストと各メディアデータとの相関の強さを定量化している。この意味空間における検索結果は、各メディアデータを相関の強さについてソートしたリストとして与えられる。

^{*} 筑波大学大学院システム情報工学研究科

takafumi@nalab.is.tsukuba.ac.jp

^{*} 筑波大学電子・情報工学系 takashi@is.tsukuba.ac.jp

慶應義塾大学環境情報学部 kiyoki@sfc.keio.ac.jp

3. Media-lexco Transformation Operatorの実現方式

文献[5]において、メディアデータと言葉との相関を求めることによって、メタデータを抽出するMedia-lexco Transformation Operatorについて示している。本節では、Media-lexco Transformation Operatorの概要を示し、さらに、画像メディアデータ、顔の表情を対象とした実現方式について示す。

3.1 Media-lexco Transformation Operatorの概要

Media-lexco Transformation Operator ML [5]は、メディアデータからその分野の専門家による研究や評論、統計などによる人間がそのメディアデータから受ける印象を表す単語をメディアデータのメタデータとして抽出を実現する。

3.2 静止画像を対象とした印象語抽出方式

ここでは、静止画像を対象とした色彩情報によるMedia-lexco Transformation Operator ML の実現方式を示す。

(1) 色印象行列 C の作成

色彩とその色彩から想起される印象の関係を示した統計データとしてカラーイメージスケール[7]がある。カラーイメージスケールを用いて、有彩色 120 色、及び無彩色 10 色の基本色 130 色とその印象を表す 180 語により 180 行 130 列の色印象行列 C とする。色印象行列 C の各要素はカラーイメージスケールによる色と印象語の関連の強さを示す数値データである。

(2) 対象静止画像から色彩情報抽出

静止画像から色彩情報が抽出し、その色彩情報は静止画像全体における基本色130色の占める割合で構成される画像色彩ベクトル m によって表現される。画像色彩ベクトル m を次に示す。

$$m = (m_1, m_2, \dots, m_{130})^T$$

(3) 静止画像のメタデータを抽出

色印象行列 C 、及び画像色彩ベクトル m を用いて、静止画像メタデータ I の抽出を行う。画像メタデータ I は、色印象行列180個の印象語と同一の印象語で特徴付けられるベクトルである。

$$I = Cg(m)$$

上記の式の g は特徴量をどれくらい反映させるかの関数を示す。文献[6]では、刺激と人間の感覚の関係を示したFechnerの法則[8]に基づいた感性作用素を適用する方式を示している。本方式においても感性作用素を適用する。具体的に関数 g は、以下のように表される。

$$x = (x_1, x_2, \dots, x_n)$$

$$g(x) := (\gamma_1, \gamma_2, \dots, \gamma_n)$$

$$\gamma_j = \begin{cases} k \log_{\alpha} |x_j| + 1 & (x_j > 0) \\ 0 & (x_j = 0) \\ -(k \log_{\alpha} |x_j| + 1) & (x_j \leq 0) \end{cases}$$

上式の k を感覚ボリューム係数とし、メディアデータの違いや検索者個々の感性の違いによって設定するパラメータとする。また、対数の底の値 α についても同様の役割を果たすパラメータとする。これら二つのパラメータ (k, α) を感性パラメータとし感性のボリュームとして設定できるパラメータとする。これらのパラメータはメディアデータの種類ごとに設定するパラメータとして、メディアデータ毎に設定する。ここでは、感性パラメータを

$\alpha = 8, k = 1$ とする。

3.3 印象語による顔の表情の自動合成方式

本節では、任意の印象語からAction Unitを抽出することによって表情を自動合成する方式について示す。本方式は、与えられた任意の印象語を対象として印象に合致する顔の表情の合成を顔の表情関連する研究において最も広く利用されている表情記述法であるFacial Action Coding System (FACS)[9]を用いる。FACS[9]はP.Ekmanらの研究による顔の行動の解剖分析に基づいて顔の動きを説明する方法であり、筋肉ではなくAction Unit(AU)という外見から識別可能な表情筋の動きの最小単位の組み合わせとして記述する方式である。表1は、AU番号とその動作の説明の一部を示している。さらに、文献[10][11]で、基本的な感情である基本情動が、どのようなAUの組み合わせで構成されているかが示されている。図1は、それを模式的に示したものである。図1の数字はAUの番号を表す。また、背後に示す顔は特に感情を表さない無表情の顔を表す。図1の表情の作成には、イメージ情報科学研究所のソフトウェアである「Face Tool」[12]を使用した。

表1 Action Unitの一部。
Fig.1 A part of Action Unit.

| AU | Description |
|----|----------------------|
| 1 | Inner Brow Raiser |
| 2 | Outer Brow Raiser |
| 4 | Brow Lower |
| 5 | Upper Lid Raiser |
| 6 | Cheek Raiser |
| 7 | Lid Tighter |
| 9 | Nose Wrinkler |
| 10 | Upper Lip Raiser |
| 12 | Lip Corner Puller |
| 15 | Lip Corner Depressor |
| 17 | Chin Raiser |
| 20 | Lip stretcher |
| 23 | Lip Tighter |
| 24 | Lip Pressor |
| 25 | Lips part |
| 26 | Jaw Drop |
| 27 | Mouth Stretch |



図1 FACSによる書く基本情動とAUの関係。

Fig.1 Relation between basic emotions and AU by FACS.

以下に具体的な実現方式を示す。

(1) 変換行列 T の作成

文献[10][11]により、6つの基本情動 c_k (happiness,

surprise, fear, anger, disgust, sadness)とそれに関連するAUの図1のような関連が示されている。これより、6つの基本情動とn種類のAUとの関連を表すn行6列の変換行列Tとする。

(2) 基本情動量抽出

入力として任意の印象語が与えられたとき、その印象語群を対象として6つの各基本情動がどのくらいの割合で混合しているかを示す基本情動量を抽出する。これは、表情において、6つの基本情動の様々な組み合わせにより複雑な表情を構成することは文献[11]により示されており、各基本情動がどのくらいの割合で混合することで与えられた印象語に近い印象を表すことができるかを計量することを意味する。基本情動量の抽出手順を以下に示す。

印象語と基本情動との相関を表すノルム値を導出
与えられた任意の印象語と顔の表情を決定するための6つの基本情動語(happiness, surprise, fear, anger, disgust, sadness)との意味的な相関を求める。意味的な相関は、2節で示す意味の数学モデルを用いて与えられた印象語をコンテキストとして6つの基本情動語のノルム cn_k ($k=1, \dots, 6$)を導出することにより求める。6つの基本情動語のノルム値でなる基本情動相関ベクトル fev を以下のように表す。

$$fev = (cn_1, cn_2, \Lambda, cn_6)^T$$

基本情動相関ベクトルの正規化

意味の数学モデルにより抽出された6つの基本情動のノルム cn_k ($k=1, \dots, 6$)を cn'_k に正規化を行う。

$$fev' = (cn'_1, cn'_2, \Lambda, cn'_6)^T,$$

$$cn'_k = \frac{cn_k}{\sum_{i=1}^6 cn_i}.$$

これは、表情において、6つの基本情動の様々な組み合わせにより複雑な表情を構成することは文献[11]により示されており、表情を合成するための基本情動量として、各基本情動をそれぞれどのくらいの割合で混合するのかを表す値にするための正規化である。

基本情動量の小さい基本情動の排除

手順 で求めた基本情動量は表情の特徴を表す特徴量とみなすことができる。一般的に特徴量が小さい特徴は、表示結果を悪化させるノイズの可能性がある。これらのノイズとなりうる値を排除することによって、表情の特徴をよりの確に表す。

一般的に、排除後の値を c_k ($k=1, \dots, 6$)として基本情動量の小さい基本情動の排除方法を以下のように表される。

$$c_k = \begin{cases} cn'_k & (cn'_k \geq \epsilon) \\ 0 & (cn'_k < \epsilon) \end{cases}$$

本方式では を各基本情動量の平均値とした。最適な の設定方法については今後の課題である。以上により、6つの基本情動量 c_k ($k=1, \dots, 6$)で構成される、基本情動量ベクトル em が抽出される。

$$em = (c_1, c_2, \Lambda, c_6)^T.$$

(3) AUメタデータ生成

変換行列 T と基本情動ベクトル em を用いて、与えられた任意の印象語とAUの相関を表すAUメタデータauを抽出する。

$$au = Tg(em)$$

また、g は3.2節の静止画像の場合と同様に感性作用素を適用する。感性パラメータについては、予備実験により仮に $\alpha = 8, k = 10$ に設定する。AUメタデータは、各AUにおける動作の大きさを表している。AUメタデータを「Face Tool」[12]に受け渡すことにより、メディアデータの印象に合致する顔の表情の表示が可能となる。

4. 静止画像メディアデータと顔の表情を対象とした異種メディア間検索の実現

4.1 異種メディア間検索の概要

異種メディア間検索の概要図を図2に示す。様々な異種のメディアデータから特徴量としてメディアデータの印象を表す印象語を抽出し、言葉の相関を計量できる意味の数学モデルによるメタデータ空間でメディアデータのメタデータ間の計量を実現することにより、異種のメディアデータ間でも統一的に検索が可能となる。

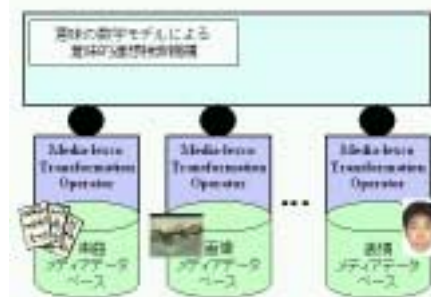


図2 異種メディア間検索方式概要.

Fig.2 Fundamental framework for heterogeneous media to media search.

4.2 静止画像メディアデータと顔の表情を対象とした異種メディア間検索の実現方式

本節は、静止画像メディアデータの色彩情報の印象に合致した顔の表情を自動合成する静止画像メディアデータと顔の表情を対象とした異種メディア間検索の実現方式を示す。手順は以下の通りになる。

- (1) 静止画像メディアデータからメタデータを抽出
静止画像メディアデータから3.2節で示した方式により、その色彩の印象語からなるメタデータを抽出する。
- (2) 表情の基本情動語とメタデータの相関を計量
意味の数学モデルによるメタデータ MDSにおいて(1)で抽出された印象語と表情の基本情動語の相関を求める。これにより、静止画像と顔の表情とを統一的に扱うことが可能となる。
- (3) AUメタデータ抽出
3.3節で示した方式により AUメタデータを抽出する。

5. 実験

本方式に基づく実験システムの構築を行った。なお、自動抽出したAUメタデータについて、実際の顔の表情の合成部については、「Face Tool」[12]を利用して実装を行った。実験結果例を図3,4に示す。但し、背後に示す顔は特に感情を表さない無表情の顔を表す。



図3 実験結果 1.
Fig.3 Experiment result 1.



図4 実験結果 2.
Fig.4 Experiment result 2.

実験結果1の入力静止画像の色彩は全体的にオレンジが占めるため、カラーイメージスケール[7]では「まぶしい、楽しい」などの明るいイメージである。本実験システムの出力では、眉の内側及び頬が上がっている等から、喜びの表情が多分に入っている。しかし、唇の端や眉の外側が幾分下がり気味なことから、少し物悲しいような様子も見て取れる。また、実験結果2の入力画像の色彩が全体的に黒、灰色、濃紺であることから、「枯れた、高尚な」など荘厳なイメージである。本実験システムの出力では、全体的には悲しみの表情である。しかし、純粋な悲しみの表情ではなく、様々な表情が混ざった表情となっている。

これらにより、比較的印象に合致した表情が合成されると考えられる。さらに本実験により、様々な表情が混ざった複雑な表情まで再現できることを示している。

6. まとめと今後の課題

本稿では、メディアデータの印象を表すメタデータと意味の数学モデルによる異種メディアデータ間の検索方式を示し、画像メディアデータからその印象に合致する顔の表情を自動合成する方式を導入し、有効性の検証を行った。

本方式により、異種メディアデータ間において、統一的に扱うことが可能となる。これにより、異種メディアデータを統合による新しい情報生成が可能となり、情報資源の有効利用になると考えられる。

今後の課題は、本方式の定量的な評価方式の確立、本方式の各種メディアデータへの適用、メタデータ自動抽出方式への個人差の計量方式の導入、大規模データベースへの適用、有効性の検証が挙げられる。

[謝辞]

本研究の一部は、文部科学省、日本学術振興会科学研究費補助金（学術創成研究費）における研究課題“人文社会科学と自然科学を連携するメタレベル知識ベースシステムの開発”（課題番号：13GS0020）によっております。ここに記して謝意を表します。

[文献]

[1] Kitagawa, T. and Kiyoki, Y.: "The mathematical model

of meaning and its application to multidatabase systems", Proceedings of 3rd IEEE International Workshop on Research Issues on Data Engineering: Interoperability in Multidatabase Systems, pp.130-135(1993).

- [2] Kiyoki, Y., Kitagawa, T. and Hayama, T.: "A Metadatabase System for Semantic Image Search by a Mathematical Model of Meaning", Multimedia Data Management-- using metadata to integrate and apply digital media --, McGrawHill, A. Sheth and W. Klas(editors), Chapter 7 (1998).
- [3] 清木康, 金子昌史, 北川高嗣: "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", 電子情報通信学会論文誌, D-II, Vol.J79-D-II, No. 4, pp. 509-519 (1996).
- [4] Michael, W. B., Susan, T. D., Gavin, W. O.: "Using linear algebra for intelligent information retrieval", SIAM Review Vol. 37, No.4, pp.573-595 (1995).
- [5] Kitagawa, T. and Kiyoki, Y.: "Fundamental framework for media data retrieval system using media lexco transformation operator", Information Modeling and Knowledge Bases, IOS Press (2000).
- [6] 北川高嗣, 中西崇文, 清木康: "楽曲メディアデータを対象としたメタデータ自動抽出方式の実現とその意味的楽曲検索への適用", 電子情報通信学会論文誌, Vol.J85-D-I, No.6, pp.512-526 (2002).
- [7] 小林重順: "カラーイメージスケール", 講談社, (1984).
- [8] 新編 感覚・知覚心理学ハンドブック, 誠信書房 (1994).
- [9] P.Ekman, W.V.Friesen: "Facial Action Coding System", Consulting Psychologist Press(1978).
- [10] P.Ekman, W.V.Friesen, 工藤 力訳編: "表情分析入門—表情に隠された意味をさぐる", 誠信書房(1987).
- [11] 千葉 浩彦, 佐伯 胖 佐々木 正人編: "アクティブ・マインド", 東京大学出版会(1990).
- [12] <http://www.tokyo.image-lab.or.jp/ipa/>

中西 崇文 Takafumi NAKANISHI

筑波大学大学院システム情報工学研究科在学中。2001年筑波大学第三学群情報学類を卒業。マルチメディアシステムに関する研究に興味を持つ。情報処理学会学生会員。電子情報通信学会学生会員。日本データベース学会学生会員。

北川 高嗣 Takashi KITAGAWA

筑波大学電子・情報工学系教授。1978年名古屋大学工学部卒業。1983年同大学院工学研究科博士過程修了。工学博士。スタンフォード大学計算機科学科客員研究員、愛媛大学理学部数学科講師、筑波大学電子・情報工学系助教授を経て現在に至る。数値解析、逆問題、マルチメディア情報システムの研究に従事。日本応用数学会会員。

清木 康 Yasushi KIYOKI

慶應義塾大学環境情報学部教授。1978年慶應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～1995年筑波大学電子・情報工学系講師、助教授、1996年、慶應義塾大学環境情報学部助教授を経て、現在に至る。データベースシステム、知識ベースシステム、マルチメディア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。