マージンに基づく問い合わせ学習を用いたオンライン動作認識

森 武 俊* 下 坂 正 倫* 原 田 達 也* 佐 藤 知 正*

Online Action Recognition with Margin-Based Query Learning

Taketoshi Mori*, Masamichi Shimosaka*, Tatsuya Harada* and Tomomasa Sato*

In this paper, we propose an online recognition method for daily actions, such as walking and standing. The proposed method has following characteristics: (1) simultaneous recognition that is able to output multiple action names when human act more than one action, such a situation as *human is waving hand on standing*, (2) modeling action classifiers with kernel methods, (3) effective optimization for the parameters of the recognition system with margin-based query learning. The characteristic (2) unifies the process for modeling and learning the classifiers, and makes us easy to incorporate prior knowledge about action. The characteristic (3) reduces the burden of process for annotating action, which is an inevitable task for supervised learning. The experimental results using real motion capture data show that the proposed margin-based query learning is very effective to achieve high performance of the recognition system with very small sized query and annotation process.

Key Words: Kernel, Support Vector Machines, Active Learning, Spectral Learning, Motion Capture Data

1. は じ め に

機械による人間の運動理解 [1]~[3], 特に人の動作の認識 [4]~[7] は, ロボットを含む知能情報システムが人間と協調して動作するための重要な基盤技術の一つである. 例えば, 人とロボットとの対話, 計算機との直感的なインタフェース, 人の行動の観察によるロボット動作の適応的獲得などに動作認識は重要である. また, マルチメディアデータベースのコンテンツの一つとして期待されている人の生活行動を扱う際にも重要である.

動作認識に関する研究では, 手話・ジェスチャ [5] [6] [8], 茶道動作 [9], 机上作業動作 [10] などが対象として選ばれることが多い. これは, 上記の対象動作では, 動作とその遷移の過程を有限状態オートマトンとして記述しやすいという背景に由来する. 実際, 自然言語処理・音声認識処理研究と同様のスキームで実現されることが多い. 具体的には, 例えば, 動作に関する有限状態オートマトンと観測される運動とのマッピングを確率的な言語モデル [11] として定式化し, その隠れ状態量についての推論問題として実現している. これまで, その定式化と学習の簡便さから, 隠れマルコフモデル [12] などが用いられることが多かった. 近年では, 運動のダイナミクスを考慮し, 「歩いている」や「走っている」などの認識精度を向上させる, 隠れマルコフモデルの拡張といえる手法が Pavlović ら [13] や Bregler

ら [14] から提案されているが, 動作とその遷移過程を有限状態オートマトンとして記述している点では共通である.

一方, 日常人間の生活行動を蓄積し生活支援に役立てるといふ観点にたった動作認識には, 上記の枠組による実現は容易ではない. 上記の枠組では手話・ジェスチャなどの特定の動作に認識対象を絞り込むことで, 有限状態オートマトンの設計を可能としているが, 幅広い日常生活行動を対象とした場合, その設計は困難である. 設計を困難にさせる顕著な現象が, 複数の動作種が共起することである. 例えば「手を振りながら歩いている」などの状況である. 有限状態オートマトンで設計する場合, 「歩いている (が手を振ってない)」と「手を振りながら歩いている」を別の状態として設計する必要がある. 対象動作が増え, 共起する組み合わせが増えた場合設計は極めて難しい. これに対し, 「手を振っている」と「歩いている」が同時に複数行われていると認識した方が扱いが容易である.

このような認識出力を我々は, 「同時認識」と呼び, それを実現するシステム [15] を構築してきた. 本研究でもこの枠組を踏襲する. しかしながら我々の過去のシステムでは, 認識性能の調整は手動で行う必要があるため, 拡張性に乏しかった. よって, システムの構成要素である動作要素認識器と呼ばれる識別器のモデル化と学習の枠組の構築が本稿の主要な貢献である.

認識器のモデル化の観点で考慮すべきことを記す. 本研究で想定する認識対象動作, 例えば, 「ジャンプする」や「立っている」などでは, 識別に有効な運動特性は識別対象動作により異なる. 日常生活の広範な動作を扱えるようにするため, 認識器およびその学習法のフレームワークは統一する必要がある. 本

原稿受付 2005 年 11 月 8 日

* 東京大学大学院情報理工学系研究科

* Graduate School of Information Science and Technology, the University of Tokyo

研究では、カーネル法[16]を基盤とする動作要素認識器のモデル化を行う。カーネルとは、二つの標本間の広義の類似度である。任意のデータ構造に適用可能であり、扱いが容易なため数多くの応用がなされている。カーネル法を採用した理由は、識別対象動作の違いによって生じる、識別に有効な運動特性の差異をカーネルで吸収させるためである。

カーネル法を基盤とする認識器のモデル化を行った場合、その学習法は数多く存在する[17]。中でも、Support Vector Machine (SVM) [18] と呼ばれる識別のための学習法は、性能の高さと安定性から、顔発見処理、テキスト分類など数多くの分野で適用されている。しかしながら SVM を含む教師あり学習法では、高い識別性能を得るために事前にアノテーションされた運動データが大量に必要な。アノテーションとは、入力された運動情報に対し、どのような動作が行われているかの情報を記す作業の総称である。アノテーションは人が行うため、その人的コストは多大である。よって、提案する認識器の学習法の観点で考慮すべきことは、アノテーションに伴う負荷の軽減である。この解決策として、問い合わせ学習法[19]と呼ばれる枠組が存在する。これは、少数のアノテーション済みのデータから学習した識別器が、識別性能を高効率に向上させる未アノテーションなデータを能動的に抽出し、そのラベルを人に問い合わせることでラベル情報を取得し再学習するプロセスをさす。そのプロセスのイメージを動作認識について示したのが Fig. 1 である。問い合わせ学習法と類似した枠組として、準教師つき学習 (Semi-supervised learning) [20] があり、未アノテーションのデータの存在を前提としている点が似ているが、ラベル付けされたデータの量は固定であり、インタラクティブなプロセスは存在しない点で異なる。準教師つき学習は、どのデータをラベル付けするかという根本的な問題も重要となるため、真の意味で少ない労力で頑健な認識器の構築を狙える、問い合わせ学習法の枠組を採用する。ここでは、カーネル法で一般的に用いられる SVM を基盤に問い合わせ学習法を設計する。具体的には、最適化の基準である「マージン」の特性を考慮したものを新たに提案する。

以上のことから、同時認識を実現する動作認識システムの構成要素である動作要素認識器について、(1) カーネルに基づく認識器のモデル化、(2) マージンに基づく問い合わせ学習法の構築、をすることが本研究の目的である。以下、本論文の構成について述べる。2章では、本研究でベースとする同時認識を実現する動作認識システムの入出力と構成法、構成要素である動作要素認識器の機能概要について述べる。3章にて、カーネルに基づいた動作要素認識器のモデル化を述べる。4章では、カー

ネルに基づく認識器の学習法の一つである、SVM について述べ、5章でそれに基づく問い合わせ学習法を提案する。6章では、動作の事前知識を援用する、動作要素認識器におけるカーネルの設計法について述べる。7章では、まず SVM に基づく動作要素認識器の性能評価を他の基準で得られる学習法と比較して行う。次に、本稿で提案する問合せ学習の性質について評価する。8章にて結論を述べる。

2. 同時認識を可能とする動作認識システム

2.1 システムの入出力

本研究でベースとする動作認識システムは、我々の過去の研究[15]に倣い、「同時認識」の機能を有するものとする。同時認識とは、「歩きながら手を振る」を「歩いている」・「手を振っている」の二つの動作が同時に起きているものとして認識する枠組である。多数の認識対象動作の中から、特定の動作一つのみを選択する問題とは異なる。また、本研究では動作の分節化の側面は扱わずオンラインで認識結果は時々刻々と出力するものとする。ここで、分節化の側面を扱わないとは動作ラベルに関する時間的前後方向の依存性を一切考慮しないということである。一般に隠れマルコフモデルなどでは、動作ラベル間の遷移を扱うことで性能を頑健化させるが、本研究の対象動作では、同時に複数の動作ラベルが付与されることを考える必要があるため、隠れマルコフモデルのように動作ラベルにマルコフ性を仮定するのは容易ではない。実際、隠れマルコフモデルの推定に必要なビタビ復号アルゴリズムをオンラインで稼働させるのは、この状況下では容易ではない。本研究で基盤とする動作認識システムの入出力は Fig. 2 のようになる。

2.2 並列処理構造による同時認識の実現

同時認識を実現するため、Fig. 3 のような構成とした。具体的には認識対象動作名の中から割り当てられた動作一つのみに着目して認識する処理部（動作要素認識器）を対象動作名の数だけ用意する。これらの動作要素認識器はシステム内部において、他の動作要素認識器にまったく影響されることなく独立に処理を行う。また、システムの出力は、各動作要素認識器の結果をそのまま統合、換言すれば集積するかたちで生成される。

このようにすることで「横を向く」と「歩く」などの二つ以上

Fig. 2 Example of input / output of the recognition system

Fig. 1 Process of query learning

Fig. 3 Configuration of the recognition system

の動作が同時に認識可能となる。この構成法のメリットは、性質の異なる動作を追加する場合も単純に動作要素認識器を追加するだけでよく、拡張性が高いことである。勿論、動作名による依存関係を考慮し誤識別を防ぐ方法も考えられるが、多様な日常動作を扱う場合、その依存関係の表現は容易ではなく、その認識プロセスも複雑化することが予想される。その点から本研究では完全な並列構造を採用した。

3. 動作要素認識器のモデル化

3.1 動作要素認識器の概要と構成

動作要素認識器は、認識対象動作名のうちの一つの動作のみに着目し入力時系列運動情報に基づき判別処理を行う。例えば「歩いている」が割り当てられた動作要素認識器は「歩いている」か「歩いていない」を時々刻々と判別する。入力は時系列運動情報であり、出力は対象動作が行われているか否かの判定結果の系列である。

ここで人間の動作認識の特性として、人間の認識の曖昧さを考慮する必要がある。認識の曖昧さとは、例えば、起きあがり途中の動作に対して、「寝ている」か「寝ていない」の判断が困難な場合、曖昧だと判定をするのが妥当である状況があることを指す。並列動作認識処理法では動作要素認識器の結果を統合するため、動作要素認識器の出力では認識結果の曖昧さを表現する必要がある。

過去の我々の研究 [15] では、曖昧さを表現するためファジィ推論の結果（実数値）を用いていたが、人間が運動データに認識結果としてラベルを付与する際に困難さを伴うため、動作の生起に関する度合いを離散的（カテゴリ）に扱うこととした。具体的には正否（yes, no）以外にその中間的な認識結果（neutral）の結果を含めた3カテゴリを用意した。その構造はyesとyes以外を判別する2クラス識別器とnoとno以外を判別する2クラス識別器を並列に処理する構造とした。yesとnoが選択される矛盾した結果の場合、noを出力するように設定した。本稿で提案する動作要素認識器のモデルでは、この2クラス識別器の性能によってシステム全体のパフォーマンスが左右される。

3.2 動作要素認識器の性能基準

動作要素認識器には「性能」の定義が必要である。本研究では、人が認識した結果を正しいとして、その結果に類似した結果を出すものを性能が高い認識器とする。

ここで、人がある運動を観測した下で判定した結果を「正解」とするべきであるが、同時認識が起こりえる動作種を対象とするため、単純な判定結果では曖昧性が残る。よって入力運動に対してある特定の動作に着目して正解例を付与したものを正解とする。この正解データは、認識対象動作全種に行う。すなわち対象動作「歩いている」、「座っている」などのそれぞれの判別結果を入力時刻ごとに人が作成することとする。

なお、人による動作の認識は、人個体内での曖昧さだけでなく、人個体間での曖昧さもあるのが当然である。例えば、ある動作が開始するタイミングというのは同じ運動に対しても人によって異なるはずであり、その考慮は重要である。しかしながら、このような人一般の認識の曖昧さを扱う場合、人に共通の認識の難しさの問題と、人によって動作の概念の異なる要素を

それぞれ考慮することで適切にモデル化しなければならないが、そのモデル化は一般に困難である。仮にそれらについて何もモデル化なしに動作の認識を扱った場合、正解ラベル自体に雑音が含まれる可能性が増加し、その結果ある特定の人の認識の概念に合った結果を出力する保証が低下する。この考えに基づき、本稿では、ある特定の一人の価値基準に沿う認識結果を出力することを正解の基準とした。

3.3 動作要素認識器内の2クラス識別器の定式化

動作要素認識器を構成する2クラス識別器では、事前に用意した運動情報とそれに対する人の「正解」ラベル情報を手がかりに、ノンパラメトリックな識別を行う。換言すれば、未知の入力運動と類似する既知の運動データであれば、その既知運動データのラベルと類似した正解ラベルを出力するように設計する。具体的には、識別対象となる時系列運動情報とあらかじめ用意しておいた参照時系列運動とのそれぞれの類似度を計算し、適切な重み付けを行った後、その和を二値化して識別する。アノテーション済みの参照時系列運動情報の集合 $D = \{\mathbf{x}^{(n)}, y^{(n)}\}_{n=1}^N$ が、 N フレーム分あるとする。ここで、 $\mathbf{x}^{(n)} \in \mathcal{X}$ は入力時系列運動情報であり、 $\mathcal{X}$ は、ある時刻から一定期間の過去における人間の時系列運動情報の集合である。 $y^{(n)} \in \{\pm 1\}$ はある瞬間の、識別対象動作についての人による二値の生起判定結果を示すラベルである。例えば、yes か yes 以外の二値である。これは、アノテーションにより付与される。また、適当な重み付けベクトル $\boldsymbol{\alpha} \in \mathbb{R}^N$ ($\alpha_n \geq 0$) により、識別関数 $f(\mathbf{x}) : \mathcal{X} \rightarrow \mathbb{R}$ を

$$f(\mathbf{x}) = \sum_{n=1}^N \{\boldsymbol{\alpha}\}_n y^{(n)} K(\mathbf{x}^{(n)}, \mathbf{x}) + b \quad (1)$$

として、その識別則を $\hat{y} = \text{sgn}(f(\mathbf{x}))$ とする。ただし、

$$\text{sgn}(t) = \begin{cases} +1 & \text{if } t > 0 \\ -1 & \text{if otherwise} \end{cases}$$

であり、運動間の類似度を計算する計量としてカーネル $K(\mathbf{x}, \tilde{\mathbf{x}}) : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ を事前に用意する。ここで重要なのは、 $\mathbf{x}$ がどのような特性を持っているかどうかを動作要素認識器は一切関知せず、その類似度 $K(\mathbf{x}, \tilde{\mathbf{x}})$ にのみ従って判別を行うことである。類似度計量がひとたび与えられれば、動作要素認識器は動作種によらず、その認識・学習のプロセスが統一されることを意味する。このモデルの下での学習とは、データ集合 D を用いて、正解ラベルと類似する出力を行う f を、 $b \in \mathbb{R}$ および $\boldsymbol{\alpha}$ を調整して得るプロセスを指す。換言すれば、学習には N 個のサンプルについて正解ラベルを用意する必要がある。

4. SVM: 最大マージン学習

写像関数の推定結果とラベルが不一致する場合の誤差の定義は多数考えられるが、特徴空間内のクラスの分離性を示すマージンと呼ばれる計量を最大化する基準で性能の最適化を行う。これを具現化するのが、SVM である。

4.1 SVM の定式化

SVM は、二つのクラスのデータの分離性を示すマージンを最大化する学習法である。結果として $\boldsymbol{\alpha}$ に関する制約つき二次

計画法に帰着される [18].

$$\begin{aligned} \text{minimize : } & W(\boldsymbol{\alpha}) = \frac{1}{2} \boldsymbol{\alpha}^T G \boldsymbol{\alpha} - \mathbf{1}_N^T \boldsymbol{\alpha} \\ \text{subject to : } & \begin{cases} \sum_{n=1}^N \{\boldsymbol{\alpha}\}_n y^{(n)} = 0 \\ 0 \leq \{\boldsymbol{\alpha}\}_n \leq C, n = 1, \dots, n \end{cases} \end{aligned} \quad (2)$$

ただし, $\{G\}_{i,j} = y^{(i)} y^{(j)} K(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})$ であり, 正定数 $C > 0$ はマージン領域 ($|f(\cdot)| \leq 1$ を満たす領域) に存在するデータへの罰則量である. また, $\mathbf{1}_N$ は全要素が 1 となる N 次元ベクトルである. b は $\boldsymbol{\alpha}$ の最適化後, 後述する最適性の必要条件から導出される. 二次計画法の特性上, $\boldsymbol{\alpha}$, b が一意に導出できることが SVM の利点の一つである. 最適化の後に得られたベクトル $\boldsymbol{\alpha}$ のうち, その成分が正であるものに相当する学習サンプルをサポートベクタと呼ぶ. サポートベクタの存在する領域をマージンと呼ぶ.

4.2 最大マージン学習の関連手法

カーネル法に基づく 2 クラス識別器は, その定式化が簡単なことから, SVM 以外の学習法が多数存在する. 中でも, パーセプトロンをカーネル識別器に拡張したもの [21] や Bayes 統計に基づくもの [22] が一般的である. 双方とも SVM と競合しえる識別性能を発揮し, 特に前者は逐次学習にも適用可能であり大きなアドバンテージを有しているが, SVM と異なり $\boldsymbol{\alpha}$ がスパースなベクトルにならず, 認識時の処理の負荷が大きい. 後者についてはこの負荷問題を回避するために, 階層 Bayes モデルを導入する試み [23] もあるが, 学習時のメモリコストが大きく, 数千から数万のオーダーの標本数のデータからの学習を行う可能性がある動作要素認識器には適していない.

5. マージンに基づく問い合わせ学習法

本章では問い合わせ学習の問題設定を行い, 問い合わせ学習においてベースになるマージンについての性質について述べる. 次に本研究で提案する問い合わせ学習法について述べる.

5.1 問い合わせ学習の問題設定

まず, 本研究で構築する学習法で用いる変数の説明を行う. まず事前に用意した学習データが N 個あるとして, データ全体の集合を $X = \{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N)}\}$ とおく. そのうち, L 個, ($L \ll N$) についてアノテーションが済んでいるとする. また, $N - L$ 個のデータについてはアノテーションが済んでいないとする. アノテーション済みのインデックスの集合を I とおき, アノテーション済みのデータ集合を D_I , 未アノテーションなデータ集合 $X_{\setminus I}$ とし, インデックスサイズ $L = |I| = N$ の場合, $D_{ALL} = D_I$ とする.

本稿における問い合わせ学習とは, 以下の通りである.

- (1) 初期フェーズ: まず, X の一部を抽出し, 初期インデックス集合 I とデータ集合 D_I を構築後, D_I により識別器を学習させる.
- (2) 問い合わせ主プロセス: $X_{\setminus I}$ から選択的にデータを抽出し I の更新する. 抽出した新データに対するラベルを人に問い合わせる. ラベル付与後 D_I を更新し, $X_{\setminus I}$ からその標本を削除し, D_I で再学習する.

- (3) 性能検証フェーズ: 一定の繰り返し後, 構築された識別器を $X_{\setminus I}$ の一部データを問い合わせ, 性能検証する.

ここで重要なのは, $|I|$ の大きさを抑えながら, $f_{D_I}(\cdot) \approx f_{D_{ALL}}(\cdot)$ を実現することである. ただし $f_D(\cdot)$ とは, データ集合 D を用いて学習した識別関数を指す.

5.2 マージンに基づく学習法の性質

SVM に基づく学習で得られた識別器は, その最適性の必要条件である, KKT 条件 [24] から以下の興味深い性質が成り立つ. まず, 学習データ集合 D_I を用いて最適化した識別関数を $f_{D_I}(\cdot)$ とし, 学習データ集合 D_I からある標本 $\mathbf{x}^\dagger$ を除いた集合を $D_{I \setminus \dagger}$, 新たに運動データ $\mathbf{x}^*$, y^* を入力し, $D_{I \cup *}$ $\leftarrow \{D \cup \{\mathbf{x}^*, y^*\}\}$ とする. ここで $f_{D_I}(\cdot)$ において, データ $\mathbf{x}^\dagger$ がサポートベクタでないとし, $f_{D_I}(\mathbf{x}^*) y^* > 1$ が成立すると

$$f_{D_{I \setminus \dagger}}(\mathbf{x}) = f_{D_I}(\mathbf{x}) = f_{D_{I \cup *}}(\mathbf{x}), \forall \mathbf{x} \in \mathcal{X}$$

が成り立つ.

つまり, $f_{D_I}(\cdot)$ にとって, マージン領域外のサンプルを追加・削除しても, 関数形状は更新されない. 一方で, マージン領域内にデータが存在する限り, 学習の余地があるといえる. 換言すれば, 積極的にマージン領域に存在するデータを取得すれば, D_{ALL} で学習した場合のサポートベクタを早々に抽出できる可能性が高まり, その結果, アノテーションの負荷の軽減が期待できる.

5.3 提案する問い合わせ学習法

本研究で提案する問い合わせ学習法は, 上記の通り, 三つのフェーズからなる. 以下にその概要を示す.

- (1) 初期フェーズ: まず正例のサンプルを一つ用意する ($|I| = 1$). 次に $i^* = \operatorname{argmin}_{i \in I} \{\mathbf{r}^*\}_i$ インデックス i^* を選び正解ラベルを付与した後, I に追加する. $\mathbf{r}^*$ は, 大局的類似度なるものであり, 以下の問い合わせベクトル $\mathbf{q} \in \mathbb{R}^N$

$$\{\mathbf{q}\}_{i \in I} = 1, \{\mathbf{q}\}_{i \notin I} = 0 \quad (3)$$

から算出するものである. その計算過程の詳細は次節で述べる. $\mathbf{r}^*$ の値は, アノテーション済みのデータ集合 I と関連性に依存するものであり, その値が高ければ高いほど集合 I のデータと関連度が高いことを示すものである. この手順を繰り返し, $|I|$ がある規定の少数の標本数 n_b になるまで繰り返す. これを行うことで, 互いに関連性が低いデータが選ばれることになる. これにより, 一度選択された時系列運動データと局所的に類似する運動は選ばれなくなるメリットが生じる.

- (2) 問い合わせ主プロセス: マージンの性質より, $X_{\setminus I}$ のデータのうち, 関数 $f_{D_I}(\cdot)$ の絶対値が 1 未満のデータ (マージン領域内のデータを意味する) を抽出することが問い合わせ学習の基準となる. 本研究では, 学習の効率を上げるため,

$$j^* = \operatorname{argmin}_{j \notin I} |f_{D_I}(\mathbf{x}^{(j)})| \quad (4)$$

なるヒューリスティックを導入する. 直感的には, 「最も識別が難しい」, すなわち識別面に近いデータを問い合わせす

Table 1 Process of the margin-based query learning

Setting: Preparing motion dataset D that includes $\mathbf{x}$ with positive label (the number of this kind of samples is a) and $N - a$ motions without labels, Kernel K , Parameters α , then making index I , calculating transition Matrix S in (6), and setting initial query: $\mathbf{q}$ in (3)
1 Initial phase: Calculating global similarity $\mathbf{r}^*$ in (7), then select $i^* = \arg \min_i \{\mathbf{r}^*\}_i$, then updating ($I \leftarrow I \cup i^*$, $\mathbf{q}$ by (3), and D_I) until $ I < n_b$
2 Main loop: Making action recognizer by SVM from D_I select most significant data $\mathbf{x}_{i^*}$ that satisfies (4) until unlabeled samples exist in margin area of $f_{D_I}(\cdot)$, then, updating I and D_I and returning to 2
3 Validation phase: Calculating global similarity $\mathbf{r}^*$ in (7), then select $i^* = \arg \min_i \{\mathbf{r}^*\}_i$, then updating ($I \leftarrow I \cup i^*$ and D_I), then returning to 2 if the classifier fails to classify, else if the classifier never fails to classify n_v times, terminating the algorithm

ることで識別の安定性を向上させると解釈できる。

この作業を、 $|f_{D_I}(\mathbf{x})| \leq 1$ を満たすデータ $\mathbf{x} \in X_{\setminus I}$ がなくなるまで繰り返す。なくなったときに初めて、識別器の妥当性をチェックする。 $|f_{D_I}(\mathbf{x})| > 1$ であり、 $y f_{D_I}(\mathbf{x}) < 1$ を満たすサンプルが存在する可能性があるためである。

- (3) **性能検証フェーズ:** 性能評価フェーズでも同様に、 $\mathbf{q}$ を式 (3) に基づき用意し、 $X_{\setminus I}$ に含まれるデータのうち、大局的類似度 $\mathbf{r}^*$ の小さいデータを抽出し、性能検証する。選ばれたデータを I に追加・ $\mathbf{q}$ の更新・ $\mathbf{r}^*$ の再計算・ $X_{\setminus I}$ からの抽出、なる一連の手順を適当な回数 n_v 繰り返す。つまり、データ集合 I と関連性の低いデータでその性能を検証するものとなる。 X_I のデータと類似するデータに関しては、 X_I で学習した識別器で良好な性能が得られると期待されるが、一方で、 X_I とは不関与なデータで評価すること、性能検証の意義を高めるものとなる。

以上述べてきた、問い合わせ学習の初期フェーズ・問い合わせ主プロセス・性能評価フェーズをまとめると **Table 1** のようになる。

5.4 大局的類似度の算出法

問い合わせ学習の初期フェーズと性能検証フェーズで用いる $\mathbf{r}^*$ の計算法について述べる。 $\mathbf{r}^*$ は、データ集合 I から見た、 X のそれぞれの点の類似度である。カーネルは二つのサンプル間の局所的なデータ間の類似計量であるため、簡単に $\mathbf{r}^*$ を計算することはできない。

データ集合全体を眺めて類似度を決定するということは、換言すれば、データのクラスタ構造や存在密度を考慮することにはほかならない。データの存在密度を考慮する素直な方法は、混合分布などの密度推定法 [25] などが考えられる。しかしながら混合分布で用いる各コンポーネントの分布の選択の基準や、混合数をどのように設定するかというモデル選択の問題が残る。また、入力時系列運動情報が $\mathbf{x}$ 単純なベクトルデータとして表現できない場合に、どのようにモデル化するかという問題も発生するため、本研究で用いる大局的な類似度の計算においては、任意のデータ構造に適用可能とする必要がある。

よって、本研究では異なるアプローチとして、Google 社に

よる WWW 検索サービスにおいて、その検索結果の表示の順番を決定するために利用されているとされる PageRank [26] や Spectral 学習法 [27] と関連が深い類似度計算法を導入する。これには、局所類似度行列の構築と正規化と、問い合わせベクトルに基づく大局的類似度の算出が必要である。以下にそれぞれの計算の詳細を記す。

5.4.1 局所類似度行列の構築と正規化

まず、二つデータ $\mathbf{x}$, $\tilde{\mathbf{x}}$ 間の局所的な類似性を示す値 $A(\mathbf{x}, \tilde{\mathbf{x}})$ を導出する。そのために、距離を示す計量 $d(\mathbf{x}, \tilde{\mathbf{x}}) \geq 0$ を用いて、

$$A(\mathbf{x}, \tilde{\mathbf{x}}) = \exp(-\lambda^2 d^2(\mathbf{x}, \tilde{\mathbf{x}}))$$

として定義する。 λ は適当な正数である。これにより、すべての標本間の類似度は 0 から 1 に変換され、結果として $A(\cdot, \cdot)$ の値が高ければ高いほど二つのデータ間の類似度が高いということになる。距離計量 $d(\cdot, \cdot)$ は事前に定義したカーネル K を利用し導出する。具体的には $K(\mathbf{x}, \tilde{\mathbf{x}}) = \phi(\mathbf{x})^\top \phi(\tilde{\mathbf{x}})$ とおいて以下のように変形する。

$$\begin{aligned} d^2(\mathbf{x}, \tilde{\mathbf{x}}) &= \left| \frac{\phi(\mathbf{x})}{\|\phi(\mathbf{x})\|} - \frac{\phi(\tilde{\mathbf{x}})}{\|\phi(\tilde{\mathbf{x}})\|} \right|^2 \\ &= 2 - \frac{2K(\mathbf{x}, \tilde{\mathbf{x}})}{\sqrt{K(\mathbf{x}, \mathbf{x})K(\tilde{\mathbf{x}}, \tilde{\mathbf{x}})}} \end{aligned} \quad (5)$$

ここで、類似計量 $A(\cdot, \cdot)$ は、場合によってはカーネルとは異なるが、ガウシアンカーネルのように類似度計算に $\exp(\cdot)$ を用いている場合、 $A(\cdot, \cdot) = K(\cdot, \cdot)$ とする。

次に、 $A(\cdot, \cdot)$ を用いて局所類似度行列を構築する。具体的には、 $X = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(N)}\}$ のデータ間の類似度行列は $W_{ij} = (1 - \delta_{ij})A(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})$ となる。これの正規化処理

$$S = D^{-1/2} W D^{-1/2} \quad (6)$$

を行う。ただし、ただし D は対角行列であり、 i 行 i 列の要素には、 W の i 行目の要素の総和が入る。すなわち $D_{ii} = \sum_{j=1}^N W_{ij}$ である。式 (6) の S の i 行 j 列の要素は i 番目の標本と j 番目の標本の「近さ」を確率的な観点で表現しているといえる。すなわち、各運動データの点を結んだグラフとしてモデルを考え、 i 番目の頂点に状態が存在したとき、次の時刻に頂点 j に遷移する確率が S_{ij} に相当するものとして解釈できる。

5.4.2 問い合わせベクトルに基づく大局的類似度の算出

式 (6) で定義された S と式 (3) で定義される $\mathbf{q}$ を用いて $\mathbf{r}^*$ を導出する手順を述べる。本研究では、 $\mathbf{r}$ に関する次のような力学系を設計する。時刻 t における類似度を $\mathbf{r}_t$ とおくと

$$\mathbf{r}_t = \alpha S \mathbf{r}_{t-1} + (1 - \alpha) \mathbf{q}$$

とするものである。ただし、 $\alpha \in [0, 1)$ として、 $\mathbf{r}_0 = \mathbf{0}$ とする。この力学系は、 X_I と類似する $X_{\setminus I}$ 中のデータに対応する $\mathbf{r}$ の値が大きく、 X_I と異質なデータに対応する $\mathbf{r}$ が小さくなるように作用するものである。 $t \rightarrow \infty$ で $\mathbf{r}$ が収束することを仮定すると、 $\mathbf{r}_\infty = \alpha S \mathbf{r}_\infty + (1 - \alpha) \mathbf{q}$ が成り立つため、解析的に

$$\mathbf{r}^* \propto (I - \alpha S)^{-1} \mathbf{q} \quad (7)$$

として求められる。

5.5 関連手法

5.5.1 マージンに基づく問い合わせ学習

マージンを基準とする問い合わせ学習法としては、Campbellらの手法[28]がある。本稿で提案する問い合わせ学習法の主プロセスにおける式(4)のヒューリスティックは、これを踏襲したものである。しかし、Campbellらの手法は初期データ集合 D_I の構築および性能検証での $X_{\setminus I}$ からの選出には、いずれも実際のデータの存在密度を無視した一様なランダム抽出法を用いているため、偏りあるデータ集合の抽出を招くことがある。これにより初期学習のフェーズで低品質なSVMを構築したり、性能検証の効率が悪くなるという問題がある[29]。本稿で提案した問い合わせ学習法の優位性は、上記手法を基盤にしながらも、問い合わせ学習の初期フェーズおよび性能検証フェーズにて、データの存在密度を考慮したデータの選出法を実現している点にある。

6. 動作の事前知識に基づくカーネルの設計法

これまでの議論は、時系列運動間の類似計量を示すカーネルが与えられた下での認識と学習法の議論であった。この章では、高い性能をもつ動作要素認識器の実現のためのカーネルの設計法について論じる。本研究では、人の経験的な動作に関する事前知識を利用することで解決を図る。基本的な設計指針は、識別対象の動作について、身体部位を用いて定性的に記述し、その記述を手がかりに識別に有効な運動特徴の選出を行う。いわゆるヒューリスティックに基づく特徴選出を行う。次にその特徴量に基づき類似度算出法を設計する。これは、カーネル関数の設定に相当する。以下、身体部位を用いた動作の定性的な記述法、記述に基づく運動特徴の選出とカーネルの設定法について述べる。

6.1 身体部位を用いた動作の定性的記述

本研究では、動作の定性的な記述を、大きく分けて下記3分類に従って行うものとした。

(1) 定常状態の記述

認識対象の動作の特徴を顕著に示す、ある特定の身体部位の定常的な位置・姿勢・その定常的な変動量を用いた記述である。速度についても定常的な性質として扱う。例えば「寝ている」における「頭が低い」・「上体が横になっている」などや、「手を上げている」の「手の位置が高い」や、「歩いている」の「腰が前方向に進む」という記述などである。

(2) 状態変化の記述

上記の定常状態の変化を記述したものである。例えば「転ぶ」についての「頭の高さが低くなる」という記述を指す。

(3) 状態の反復的な変化の記述

上記の身体部位の状態量の変動が繰り返し起こることを示した記述である。例えば「歩く」という動作について「足を交互に着地する」から転じて、「足の地面からの高さが周期的に変動する」という記述を指す。

6.2 経験的記述に基づく運動特徴選出とカーネルの設定法

上記の分類それぞれについて説明する。

(1) 定常状態の記述に基づく方法

記述中に生じる身体部位の位置・姿勢やその速度などを識

別に有効な特徴として選出する。例えば、「頭が高い」や「前に進んでいる」という記述に対し、頭の高さや腰の前進方向に関する速さ情報が識別に有効な特徴として採用する。選出した特徴に基づくカーネルには、本研究では、線形カーネルもしくはガウシアンカーネルを用いる。ガウシアンカーネルとは、入力量 $\mathbf{z}$, $\tilde{\mathbf{z}}$ についての類似度は

$$K(\mathbf{z}, \tilde{\mathbf{z}}) = \exp\left(-\frac{|\mathbf{z} - \tilde{\mathbf{z}}|^2}{\sigma^2}\right)$$

として定義されるものである。 $\sigma > 0$ は適当な調整パラメータである。線形カーネルの場合、 $K(\mathbf{z}, \tilde{\mathbf{z}}) = \mathbf{z}^T \tilde{\mathbf{z}}$ である。ここで重要となるのが、ある一定時間の区間の時系列運動情報を格納した $\mathbf{x}$ から $\mathbf{z}$ への変換法であるが、本研究では、例えば、頭の高さというものをを用いる場合、まず、 $\mathbf{x}$ からその区間に含まれる運動データのうち、時間方向で見て一番最新の地面からの頭の高さ情報を取得する。それをその人間の身長で正規化する。これを $h \in \mathbb{R}$ とする。学習データセットに含まれるこのデータ $\{h^{(1)}, h^{(2)}, \dots, h^{(N)}\}$ すべてを走査し、その最大値と最小値から下記の一様正規化により $\mathbf{z}$ を取得する。この場合 $\dim(\mathbf{z}) = 1$ である。

$$\mathbf{z}^{(n)} = \frac{h^{(n)} - \min_{n'} h^{(n')}}{\max_{n'} h^{(n')} - \min_{n'} h^{(n')}}.$$

このような一様正規化を行うのは、ガウシアンカーネルを用いた場合の σ の設定を比較的直感的に与えるようにするためである。

(2) 状態変化の記述に基づく方法

識別に有用だと思われる注目運動部位の時系列変化を検出することが特定の動作の認識に有効である場合、定常状態の記述に基づくカーネルの設定法と同じく、適当な運動部位の選出と正規化を行い、 $\mathbf{z} \in \mathbb{R}^1$ を得る。 $\mathbf{z}$ の時系列変化に着目した類似度計算を行うものがここでの方法である。そのため、 $\mathbf{z}$ の時系列情報を取得する必要がある。前述の「定常状態の記述に基づく方法」では、 $\mathbf{z}$ の取得に時系列データ $\mathbf{x}$ のうち、最新のデータから計算を行っていたが、これを区間すべてにおいて $\mathbf{z}$ を算出し、次のような類似度計算法を構築する。 $\mathbf{x}$ の区間の長さを T として一番最新のフレームの運動データから計算した $\mathbf{z}$ を $\mathbf{z}_T$ 、一番最初のフレームから計算したものを $\mathbf{z}_1$ として、新たな特徴ベクトル $\boldsymbol{\xi} = [\mathbf{z}_1^T, \dots, \mathbf{z}_T^T]^T$ とおき、 $K(\mathbf{x}, \tilde{\mathbf{x}}) = K_c(\boldsymbol{\xi}, \tilde{\boldsymbol{\xi}})$ とする。ここで K_c には、線形もしくはガウシアンカーネルを用いる。 $\mathbf{z}_t \in \mathbb{R}^1$ であるため、 $\boldsymbol{\xi}$ は実質的に T 次元のベクトルになる。この方法は、例えば、「転ぶ」の記述の「頭の高さが低くなる」から、「頭の高さの時系列情報」を選出し、頭の高さに関する時系列情報の抽出を行った後、上記のカーネルにて類似度を算出する。

(3) 状態の反復変化の記述に基づく方法

「歩いている」についての「足の地面からの高さが周期的に変動する」などの記述に対しては、上記の「状態変化の記述に基づく方法」で行った $\mathbf{z}$ の時系列情報について直接的な類似度算出は行わず、 $\mathbf{z}_{1:T}$ についての周波数特性を利用す

Table 2 Qualitative description for walking action

Description	Selected motion feature
Head to be high	Height of head from ground
Step forward	Forward velocity of hip
Feet are contacting to ground	Time-series of height of feet

Fig. 4 Example of kernel in an action element recognizer

るものとした。この周波数ベクトルに基づき線形もしくはガウシアンカーネルを用いて類似度算出を行うものとする。

6.3 複数記述への対応

上記の記述それ単体だけでは、認識対象の動作を適切に表現することが難しいことがある。複数の記述を組み合わせることにより適切な運動間の類似度計算が得られることが予想される。本研究では複数の経験的記述を用いる場合への対応として、次のような定式化を行う。

$$K(\mathbf{x}, \tilde{\mathbf{x}}) = \text{ops}_{j=1}^B K^{(j)}(\mathbf{x}, \tilde{\mathbf{x}})$$

ここで、ops は、積演算 $\prod$ もしくは和演算 $\sum$ が入り、 B はある動作について記述した記述の数を示し、 j 番目の記述に相当したカーネル計算を $K^{(j)}(\cdot, \cdot)$ としている。「歩いている」に関して、複数記述を行ったものを Table 2 に示し、その記述に基づくカーネルの計算の手順の様子を Fig. 4 に示す。

6.4 関連手法

本研究で採用したカーネルの構築法のうち、特に定常状態の記述に基づくカーネルの構築法は、古典的な特徴抽出・特徴選出問題とも関係が深い。大量の特徴量の候補を列挙し、そこから多変量解析手法 [25] などを適用することで識別に有用な特徴量の選出も可能であり、これらのほうが識別性能が高いということも多い。しかしながら、動作認識の分野では、運動データに偏在性の問題 [30] があることが指摘されており、単純な特徴抽出では、識別に有効な特徴量が抽出されず、新奇な運動データが入ってきた場合に人の直感に沿う識別が行えない可能性がある。よって、本研究では、識別に有効な特徴をあらかじめ動作に関する事前知識に基づき特徴を選出する方法をとった。

7. 性能評価実験

この章では、提案した認識モデル・学習の基本的な性能を評価する。提案する問い合わせ学習では、正解ラベルをアノテーションしたデータ集合 D_I のサイズをできるだけ小さくしながら、 $f_{D_I}(\cdot) \approx f_{D_{ALL}}(\cdot)$ を目標としているが、 $f_{D_{ALL}}$ の性能がどの程度得られるかを評価しておく必要がある。これには、運

Fig. 5 Human as linked joint model and an actor with magnetic sensors

動データすべてに対し正解ラベルを付与したものを用いて評価を行う。ここで、マージンとは異なる他の基準、例えば誤差二乗などを基準にする学習法などと性能を比較し、マージンを利用する妥当性を検証する。

次にマージンに基づく問い合わせ学習の有効性の検証として、アノテーション済み学習サンプルのサイズ、すなわち $|I|$ のサイズに対する識別性能を検証する。どの程度の問い合わせ回数でどれほどの識別性能が得られるかを確認する。また、Campbell らの手法と比較して、大域的類似度の計算を用いた効果を識別性能の観点から検証する。また、可視化の行いやすい、低次元な特徴量の人工データを通じて Campbell らの手法の問題点を明らかにする。

7.1 実験共通のデータ群

認識の評価実験では対物動作ではなく、室内行動の動作から、「寝ている」、「起きあがる」などの 25 種の動作を対象とし、正解ラベル付きモーションキャプチャデータを用いた。実際に用いたデータは ICS Action Database [31] である。これは、1 ファイル当たり 3 秒程度の、20 代男性 1 名による運動が収められた総計 125 個の正解ラベル付きモーションキャプチャデータの集合である。入力される運動は、CG 制作用ソフトウェアで読み書き可能な BVH 形式で記録されており、人間の関節構造を木構造で表現し、各関節に関する姿勢情報をサンプリングレート 30 Hz で、磁気式モーションキャプチャシステム (Ascension 社 MotionStar) により取得したものである。このデータセットに含まれる運動の人間のモデルは腰・胸・頭部・左右上腕・左右前腕・左右大腿・左右脛の 11 の身体部位、計 36 自由度からなるものである。Motion Star では、磁気センサに対応するジョイントに合わせて装着する。その装着の様子と使用したリンクジョイントモデルの関係を Fig. 5 に示す。

これらの運動データに対して日常生活中の動作に関する正解ラベル 25 種 (Fig. 6) を各フレームに付与したものである。各フレームの正解ラベルは、本研究の動作要素認識器の出力と同等な、3 段階で付与されている。また、いずれの動作についても正例および負例の組が存在することを保証している。このデータセットでは 25 種それぞれの動作が 20 代男性により演じられたものである。各動作につき 5 回演じられている。よって、データセットを 5 分割に扱うことが可能である。

7.2 マージンを基準とする学習法を採用する妥当性の検証

動作要素認識器の $f_{D_{ALL}}(\cdot)$ について、学習の最適化基準に

Fig. 6 Thumbnails of action categories

マージンを採用した妥当性を検証する。ここではその方法として、既存の別の基準で学習する手法と比較する。具体的には、C4.5 [32] と呼ばれる、エントロピーを最小化する基準で識別則を生成する決定木と、正解データと予測値のずれを二乗ノルムの観点で最適化する、誤差逆伝播法 (BP 法) と呼ばれる三層ニューラルネット [25] と比較する。

(1) 評価方法

認識精度の定義として、本研究ではその精度を認識処理されるフレーム数に対する人が付与したラベルと動作要素認識器が割り当てたラベルの一致フレーム数との比率を認識精度と定義した。

また、精度の算出には 5 セットに分けたデータ集合について、1 セットを学習データに、4 セットを評価データとする処理を交互に行い算出した。

(2) 実験の条件

前章で述べたカーネルの構築法に従い設定した。カーネルにガウシアンカーネルを用いた場合、そのパラメータである、 $\sigma > 0$ を次のように設定した。本実験では、経験的に σ を、各動作の記述における運動特性から変換した後の特徴量 $\mathbf{z}$ の次元数の総数が d とすると $\sigma = 1.5\sqrt{d}$ として設定した。また、式 (2) におけるパラメータ α の上限である C を $C = 100\sqrt{d}$ として設定した。比較手法である C4.5 はその実装として C4.5 Release8 [34] を利用した。設定は既定値を利用した。また、BP 法ではその隠れ層の神経素子数を $2d$ に設定し、出力素子数は一つとして、3 クラスを表現することとした。最適化に当たってそのパラメータは適当な乱数により初期化するものとし、準ニュートン法により最適化することとした。

(3) 結果

実験の結果を Table 3 に示す。各手法とも、上記実験条件の微調整によりパフォーマンスの変動は起きえるが、多くの認識対象動作について SVM を利用した動作要素認識器の認識精度は高いということが確認された。比較対象とした C4.5 および BP 法でも一定の性能レベルに達成しているといえるが、特に BP 法ではその最適化におけるパラメータの初期値依存性の問題点が再確認された。BP 法のパフォーマンスチューニングにより SVM と同等の性能が達成できる可能性があるが、二次計画法に基づく SVM では最適解は一意に導出できるという扱い

Table 3 Accuracy of action element recognizers

Target Action	SVM	BP	C4.5
Fold Arms	98.3±2.0	94.1±5.3	98.9±1.6
Get Up	99.0±0.6	95.4±1.3	97.2±1.3
Jump	99.4±0.2	99.3±0.0	99.0±0.6
Keep Down	98.4±1.4	96.1±2.3	97.0±1.2
Lie Down	96.5±1.7	96.2±0.8	95.2±2.7
Look Away	92.3±1.4	91.1±2.1	91.3±2.1
Look Down	86.7±2.2	84.8±2.2	84.5±1.6
Look Up	96.6±1.0	94.3±2.0	96.1±2.6
Lying	98.6±0.5	98.4±0.5	99.4±1.3
Lying on Back	99.4±0.8	98.7±1.9	98.5±1.9
Lying on Face	99.8±0.1	99.8±0.1	97.9±1.8
Lying on Side	99.5±0.3	97.9±1.8	97.5±2.8
On Four Limbs	99.6±0.1	98.8±1.0	96.8±2.5
Raise Hand	99.1±0.1	98.6±0.2	98.6±0.4
Running	100.0±0.0	99.8±0.0	99.8±0.2
Showing Hand	99.8±0.4	97.6±0.8	97.3±1.3
Sit Down	99.2±0.1	98.8±0.1	98.8±0.1
Sitting	96.4±0.8	94.9±1.6	97.2±4.5
Sitting on Chair	97.2±1.2	96.0±1.6	92.4±1.8
Sitting on Floor	95.4±1.9	94.9±2.1	92.3±3.3
Standing	98.7±0.5	98.5±0.4	98.3±0.8
Stand Still	94.5±2.4	91.3±2.5	93.3±2.0
Stand Up	98.0±0.1	98.5±0.2	98.2±0.5
Turn	95.5±0.6	95.4±0.7	95.2±1.1
Walking	93.7±1.4	90.9±0.5	90.8±2.0

やすさの観点で優位性がある。C4.5 では、SVM と比較して学習データに対しての認識精度のばらつきが大きいことが確認された。これは学習データへの過剰適合の問題があるのではないかと考えられる。C4.5 の刈り込みと呼ばれる、正則化パラメータの調整により、この変動を抑えられる可能性があるが、どちらの手法が絶対的に性能が良いという議論を行うことが本実験の目的ではないため、この議論は他の研究に譲ることとする。

7.3 問い合わせ学習の効率性の評価

問い合わせ学習を導入した一番の目的はアノテーションの負荷の軽減であるため、実際の動作認識の問題における効果を評価する。具体的には、アノテーション済みのサンプル数に対する識別性能を計算し、どの程度の問い合わせ回数ですべてのデータをアノテーションした場合と同程度の識別性能が得られるかを評価する。ここでは、初期プロセスおよび、性能評価プロセスに関する評価は行わず、ラベル済みのサンプル数に対する性能のみに着目する。

(1) 実験データ

本実験でも前述の実験と同じく ICS Action Database を用いた。動作要素認識器の学習はそこに二つ存在する 2 クラス識別器の学習と等価であるため、本実験では、yes と非 yes に判別する識別器に着目して評価する。具体的には、yes のラベルを +1 に割り当て、neutral, no のラベルを -1 に割り当てることとした。また本実験ではまず上記の運動データ集合のうち、Lying (寝ている) Standing (立っている), Sitting (座っている), Lying (寝ている) の計 3 種の動作を対象とする。

(2) 評価方法

学習の効率のよさを評価として、 $f_{D_I}(\cdot)$ による X_{ALL} での F 値 [33] を学習量 $|I|$ と比較して評価し、どの程度のサンプルで $f_{D_{ALL}}(\cdot)$ の F 値に飽和するかを測った。F 値とは、情報検

Fig. 7 Annotated size $|I|$ vs. F-measure of recognizer for *Sitting*

素などで一般的に用いられる性能指標の一つである。具体的には認識対象動作をもれなく (Recall), そして誤りなく抽出する (Precision) 性能指標であり, 単位はパーセントに相当し, 100 が最大値である。定義的には Recall を R と Precision を P とすれば,

$$F = \frac{1}{\frac{1}{(\beta+1)R} + \frac{\beta}{(\beta+1)P}} \quad (8)$$

である。ただし $\beta > 0$ であり, Precision の重み付けを意味する。本実験では Recall と Precision を等価な重み付けで評価するために $\beta = 1$ として定義した。上の式から分かるとおり, F 値は高ければ高いほど性能が高い。本手法では, 問い合わせ学習の初期フェーズで用いるデータの選出法では, まず最初の一点を正例データからランダム抽出を行い, $n_b - 1$ のデータを, 大局的類似度計算に従い抽出した。

(3) 実験の条件

実験の条件としては, 上記運動データセットを 5 セットに分け, X とした ($N \approx 2000$)。訓練誤差を 5 回評価し, その $|I|$ のサイズに応じた F 値の最小値・最大値・平均値を取得した。カーネルは前節の実験で用いたパラメータを利用した。また, 5.3 節の定数として, $\alpha = 0.99$, $n_b = 10$, $n_v = 10$ とした。

(4) 結果

Lying, Standing, Sitting いずれの動作についても, およそ 150 サンプル以内に識別性能が $f_{D_{ALL}}(.)$ 程度の性能に達していることが確認された。そのうちのひとつとして, Fig. 7 に, $|I|$ のサイズに対する Sitting の動作要素認識器の性能 (F 値) を示す。+ 記号の実線が F 値の平均である。 F 値の平均を囲む点線は実験結果の $|I|$ に対する F 値の最大値と最小値を示している。2,000 サンプルのデータセットに対し, たかだか 50 データ程度のアノテーションで, 識別器の性能が全データを使って学習した場合 (F 値: 99.35) に達していることが分かる。

7.4 大局的類似度を利用する妥当性の検証

本実験では, 問い合わせ学習における初期フェーズにて大局的類似度, すなわちデータの存在密度を考慮したデータの選出法を利用する妥当性を検証する。その妥当性の検証の方法として, 本研究の先行研究である Campbell らの手法における, ランダムサンプリングの方法と比較する。具体的には, 初期データ集合 D_I を構築した際の, 二つの手法における識別性能の差を明らかにし, 有効性を検証する。次に, ランダムサンプリングと提案手法との差異をより明確なものとするために, 二次元の人工データを利用し可視化により検証する。

7.4.1 識別性能での評価

まず, それを識別性能の観点で評価を行う。次にその差異を

Fig. 8 Visualization of global similarity on two dimensional synthetic dataset when $|I| = 1$

分かりやすく表示するために, 実際の動作認識における二つの手法の差異を評価する。

(1) 実験データと評価方法

前節で用いたのと同様の実験データを用いることとした。また, 評価方法も前節の実験と同様に F 値を利用した。実験の条件としては, 初期データを選ぶ際の n_b も前節と同様に 10 とした。このときの F 値を評価する。Campbell らの手法でも初期の SVM を構築するときに用いる標本の数を 10 とした。具体的には, 1 個の正例と $n_b - 1$ 個のランダムに抽出するものとした。識別対象動作についても前節と同様に 3 種の動作に適用した。

(2) 結果

初期データ集合 D_I の構築の段階での SVM の F 値が, 本手法では, 平均で, 88.9% に達していた一方で, Campbell らの手法では, 85.9% にとどまることが分かった。大局的類似度を組み込むことによる SVM の初期学習時の性能の安定性が高まることが確認できた。Campbell らの方法では, $n_b - 1$ 回のうち, 負例が出てこないことがあり, SVM の学習に失敗することが数回あった。なお, 失敗した場合には F 値の計算は行わないこととした。提案する手法でも失敗の可能性は 0 ではないが, 存在密度を考慮したこの方法では $|I| = 2$ の段階で負例が選ばれることが多く, その点でも安定性が高いといえる。

7.4.2 二次元人工データへの適用による定性的評価

次に二次元の人工データを利用し, 提案手法および Campbell らの手法における問い合わせ学習の初期フェーズにおけるデータの抽出の様子を可視化させることで, 本手法の特性および有効性について検証する。

(1) 実験データ

二次元の特徴空間中に 5 個の混合ガウス分布からデータを生成させた。そのうち, 200 点程度の点をデータ集合 X として用いた。

(2) 適用結果

まず, $|I| = 1$ の段階での大局的類似度を Fig. 8 に可視化する。丸で囲われた点がアノテーション済みの運動データであるとする。四角で囲われた点はアノテーションされていない運動データを示し, その大きさが大きいほど, 大局的類似度が高いことを示している。アノテーションされた運動データから比較的近傍にある点はその大局的類似度も大きく離れるにつれてその値

Fig. 9 Sampling result on the initial phase of the query learning for two dimensional synthetic dataset via our method (left) and Campbell et al.'s algorithm (right). In each figure, small size circles represent X , (red) large size and bold circles represent X_I , respectively

が小さくなることが確認される。しかしながらもっと重要な点は、学習サンプルの中心から逸脱した図中右下のクラスタ構造のデータの大局的類似度が極端に小さくなっていることである。これにより前節の実験のように $|I| = 2$ の段階で、正例とは性質が異なるという観点で、負例が選出されやすくなるものと予想される。

次に提案手法および Campbell らのランダム抽出の方法による $n_b = 5$ 程度での初期データの抽出結果を Fig. 9 に示す。丸で囲われたものが問い合わせ学習の初期 SVM の学習で用いられる運動データに相当する。ランダムなデータ抽出では、図中右下のクラスタ構造を無視し、近傍のデータを無駄に抽出してしまう可能性があるといえる。提案手法では、データ構造を考慮しているため、比較的満遍なくそして安定にデータを抽出できるという特性があることが確認され、大局的類似度を導入する妥当性が検証できた。

8. おわりに

本稿では、日常動作を対象とした認識法として、複数動作を同時にアノテーションすること（同時認識）を可能とする認識法を前提に、動作要素認識器と呼ばれる認識対象動作それぞれについての判別器のモデル化およびその学習法を構築し、その評価を行った。この動作要素認識器は、動作認識システムに複数存在するコンポーネントであり、それぞれがオンラインで並列して独立に処理するものである。本研究では、動作要素認識器をカーネル法によりモデル化を行い、その学習法として、新たに提案した、マージンに基づく問い合わせ学習法を適用した。また、動作に関する事前知識を反映したカーネルの設計法について述べた。

本稿で提案した手法は、認識性能の高さだけでなく、他のこれまでの動作認識手法と比較し、以下の点で、大きなアドバンテージがある。動作に関する事前知識としての時系列運動データ間の類似度設計さえ適切に与えることができれば、多様な動作種に対して、少数のアノテーションデータを用意するだけで、頑健な認識器を構築できる点である。

今後の課題を以下に記す。まず本手法では、カーネルは経験

的知識に基づき設計するものとしたが、カーネルに与えるパラメータにより識別性能が変化してしまう。このことから、識別性能をより良くするためのパラメータ調整法を導入する予定である。また、本稿では問い合わせ学習において、識別関数の出力値、すなわちマージンか否かの判定と大局的な類似度情報を計算した方法を、学習のフェーズにより使い分けている。前者では、データの存在密度を一切考慮しない方法であり、後者は運動データのラベル情報を無視したある種の密度推定法に近い枠組である。双方の利点である、識別器の出力情報およびデータの存在密度を明示的、統一的に考慮する問い合わせ学習法を新たに構築し、より効率性と安定性を高めることが将来課題である。

参考文献

- [1] C. Cedras and M. Shah: "Motion-based recognition: a survey," *Image and Vision Computing*, vol.13, no.2, pp.129-155, 1995.
- [2] C. Wren, A. Azarbayejani, T. Darrell and A. Pentland: "Pfinder: real-time tracking of the human body," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.19, no.7, pp.780-785, 1997.
- [3] C. Sminchisescu and A. Jepson: "Generative modeling for continuous non-linearly embedded visual inference," in *Proceedings of the Twenty-First International Conference on Machine Learning*, article no.96, 2004.
- [4] J. Yamato, J. Ohya and K. Ishii: "Recognizing human action in time-sequential images using hidden Markov models," in *Proceedings of the 1992 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp.379-385, 1992.
- [5] T. Starner and A. Pentland: "Real-time American sign language recognition using desk and wearable computer based video," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.20, no.12, pp.1371-1375, 1998.
- [6] A. Wilson and A. Bobick: "Parametric hidden Markov models for gesture recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.21, no.9, pp.884-900, 1999.
- [7] T. Inamura, Y. Nakamura, H. Ezaki and I. Toshima: "Imitation and primitives symbol acquisition of humanoids by integrated mimesis loop," in *Proceedings of the 2001 IEEE International Conference on Robotics and Automation*, pp.4208-4213, 2001.
- [8] A. Bobick and Y. Ivanov: "Action recognition using probabilistic parsing," in *Proceedings of the 1998 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp.196-202, 1998.
- [9] H. Mitomi, F. Fujiwara, M. Yamamoto and T. Sato: "Bayesian classification of human custom based on stochastic context-free grammar," *The IEICE Transactions on Information and Systems*, PT.2 (Japanese Edition), vol.J88-D-II, no.4, pp.716-726, 2005.
- [10] K. Ogawara, S. Iba, T. Tanuki, H. Kimura and K. Ikeuchi: "Recognition of human task by attention point analysis," in *Proceedings of the 2000 IEEE/RJS International Conference on Intelligent Robots and Systems*, vol.3, pp.2121-2126, 2000.
- [11] C. Manning and H. Schutze: *Foundations of Statistical Natural Language Processing*. MIT Press, 1999.
- [12] L. Rabiner: "A tutorial on hidden Markov models and selected applications in speech recognition," in *Proceedings of the IEEE*, vol.77, no.2, pp.257-285, 1989.
- [13] V. Pavlović, J. Reh and J. MacCormick: "Learning switching linear models of human motion," in *Advances in Neural Information Processing Systems 13*, pp.981-987, 2001.
- [14] C. Bregler: "Learning and recognizing human dynamics in video sequences," in *Proceedings of the 1997 IEEE Computer*

- Society Conference on Computer Vision and Pattern Recognition, pp.568–574, 1997.
- [15] T. Mori, K. Tsujioka and T. Sato: “Human-like action recognition system on whole body motion-captured file,” in Proceedings of the 2002 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp.2066–2073, 2001.
- [16] B. Schölkopf and A. Smola: Learning with Kernels. MIT Press, 2002.
- [17] R. Herbrich: Learning Kernel Classifiers. MIT Press, 2002.
- [18] C. Cortes and V. Vapnik: “Support-vector networks,” Machine Learning, vol.20, no.3, pp.273–297, 1995.
- [19] D. Angluin: “Queries and Concept Learning,” Machine Learning, vol.2, no.4, pp.319–342, 1988.
- [20] A. Blum and T. Mitchell: “Combining labeled and unlabeled data with Co-training,” in Proceedings of the Eleventh Annual Conference on Computational Learning Theory, pp.92–100, 1998.
- [21] Y. Freund and R. Schapire: “Large margin classification using the perceptron algorithm,” Machine Learning, vol.37, no.3, pp.277–296, 1999.
- [22] C. Williams and D. Barber: “Bayesian classification with Gaussian processes,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.20, no.12, pp.1342–1351, 1998.
- [23] M. Tipping: “Sparse Bayesian learning and the relevance vector machine,” Journal of Machine Learning Research, vol.1, pp.211–244, 2001.
- [24] J. Nocedal and S. Wright: Numerical Optimization. Springer-Verlag, 1999.
- [25] R. Duda, P. Hart and D. Stork: Pattern classification. John Wiley & Sons, 2000.
- [26] S. Brin and L. Page: “The anatomy of a large scale hypertextual web search engine,” in Proceedings of the Seventh International World Wide Web Conference, 1998.
- [27] D. Zhou, J. Weston, A. Gretton, O. Bousquet and B. Schölkopf: “Ranking on data manifolds,” in Advances in Neural Information Processing Systems 16, pp.321–328, 2004.
- [28] C. Campbell, N. Cristianini and A. Smola: “Query learning with large margin classifiers,” in Proceedings of the Seventeenth International Conference on Machine Learning, pp.111–118, 2000.
- [29] L. Wang, K. Luk and Z. Zhang: “Bootstrapping SVM active learning by incorporating unlabeled images for image retrieval,” in Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol.1, pp.629–634, 2003.
- [30] M. Shimosaka, T. Mori, T. Harada and T. Sato: “Action recognition based on kernel machine encoding qualitative prior knowledge,” in Proceedings of 2004 IEEE International Conference on Systems, Man and Cybernetics, pp.1569–1576, 2004.
- [31] T. Mori, M. Shimosaka and K. Tsujioka: ICS action database, <http://www.ics.t.u-tokyo.ac.jp/action/>, 2003.
- [32] J. Quinlan: “Induction of decision trees,” Machine Learning, vol.1, no.1, pp.81–106, 1986.
- [33] C. Rijsbergen: Information retrieval. Butterworth, 1976.
- [34] C4.5 Release 8, <http://www.cse.unsw.edu.au/quinlan/c4.5r8.tar.gz>
- [35] F. Bach and M. Jordan: “Kernel Independent Component Analysis,” Journal of Machine Learning Research, vol.3, pp.1–48, 2002.
- [36] H. Lütkepohl: Handbook of Matrices. Wiley & Sons, 1996.

付録 A. 大局的類似度計算の効率化

本稿で用いている、式 (7) の大局的類似度計算法は、学習サンプル数の数 N の正平方行列が必要であることから、メモリサイズは $\mathcal{O}(N^2)$ であり、逆行列計算は $\mathcal{O}(N^3)$ 必要となる。本稿での手法は、フレーム数 10^3 から 10^4 程度の時系列運動データに適用することを想定するため、一般的な PC での実行時には特にメモリコストが問題になる。

本稿では、このような問題に対して、近似的な計算法を導入し効率よく大局的類似度計算を実現している。具体的には、二つのデータの局所的な類似度 $A(\cdot, \cdot)$ を Gaussian カーネルと見立てたグラム行列（それぞれの行・列要素にカーネル関数の値が格納されている行列）に対する不完全 Choleskey 分解 [35] を用いる。具体的には不完全 Choleskey 分解により、 $\hat{A} \approx RR^T$, $R \in \mathbb{R}^{N \times M}$, $M \ll N$ に近似し、この情報を用いて解く。ただし、 $\{\hat{A}\}_{ij} = A(\mathbf{x}_i, \mathbf{x}_j)$ である。よって、大局的類似度の計算に必要な $(I - \alpha S)^{-1}$ は

$$\begin{aligned} (I - \alpha S)^{-1} &= (I - \alpha D^{-1/2} W D^{-1/2})^{-1} \\ &= D^{1/2} (D - \alpha W)^{-1} D^{1/2} \\ &= D^{1/2} (D + \alpha L - \alpha \hat{A})^{-1} D^{1/2} \\ &\approx D^{1/2} (C - \alpha R R^T) D^{1/2} \end{aligned}$$

として変形できる。ここで $L = \hat{A} - W$, $C = D + \alpha L$ である。Woodbury の公式 [36] を適用すると、

$$\begin{aligned} (I - \alpha S)^{-1} &\approx D^{1/2} C^{-1} D^{1/2} + \alpha D^{1/2} C^{-1} R \cdot \\ &\quad (I - \alpha R^T C^{-1} R)^{-1} R^T C^{-1} D^{1/2} \end{aligned}$$

を得る。 D および C が対角行列であること、 $R^T C^{-1} R \in \mathbb{R}^{M \times M}$, $R^T C^{-1} D^{1/2} \in \mathbb{R}^{M \times N}$ であることから、上記の計算する手順を工夫すれば、メモリコストは $\mathcal{O}(MN)$ 程度となる。近似による計算誤差は $\hat{A}$ の分解時に調整することができ、非距離関数 $A(\cdot, \cdot)$ のパラメータに依存するが、本研究でのパラメータ設定の場合、 $|\hat{A} - RR^T|_2 \approx 10^{-3}$ を実現するのに必要な M はおよそ $N = 10^4$ に対し $M = 10^2$ 程度で済む。

森 武俊 (Taketoshi Mori)

1967年9月8日生。1990年東京大学工学部機械工学科卒業。1995年同大学大学院工学系研究科博士課程修了(情報工学専攻)。同年東京大学先端科学技術研究センター助手。1998年同講師。2001年米国MIT客員研究員。現在東京大学大学院情報理工学系研究科助教授。動作認識、行動理解、パーソナルコンピューティング、ヒューマンコンテンツの研究をしている。博士(工学)。IEEE, ACM, 電子情報通信学会, 人工知能学会, 計測自動制御学会等の会員。(日本ロボット学会正会員)

原田達也 (Tatsuya Harada)

1972年6月14日生。1996年東京大学工学部機械情報工学科卒業。2001年3月同大学大学院工学系研究科機械工学博士課程修了。2000年1月から2001年12月まで日本学術振興会特別研究員。2001年12月より東京大学大学院情報理工学系研究科助手。2006年4月より同講師。人の動作観察・知覚, 生体計測, ワイヤレスネットワークを活用した支援に興味を持つ。博士(工学)。IEEE, 計測自動制御学会等の会員。(日本ロボット学会正会員)

下坂正倫 (Masamichi Shimosaka)

1978年6月5日生。2001年東京大学工学部機械情報工学科卒業。2006年同大学大学院博士課程修了。同年東京大学大学院情報理工学系研究科助手。2004年より2006年まで独立行政法人日本学術振興会特別研究員。統計的機械学習, コンピュータビジョンの研究を行っている。博士(情報理工学)。IEEE, 電子情報通信学会, 人工知能学会等の会員。(日本ロボット学会正会員)

佐藤知正 (Tomomasa Sato)

1948年9月22日生。1971年東京大学産業機械工学科卒業。1976年同大学博士課程修了後, 電子技術総合研究所入所。1991年東京大学先端科学技術研究センター教授に移籍。1998年東京大学工学系研究科機械情報学専攻に移籍。2000年同大学情報理工学系研究科知能機械情報学専攻に改組。人間共棲ロボットの研究に従事。日本機械学会, IEEEなどの会員。(日本ロボット学会正会員)