1G3-5

二段階モチーフ発見アルゴリズムに基づく連続インタラクションデータからのジェスチャパターンの抽出とロボットナビゲーションへの応用

Two Layered Constrained Gesture Motif Discovery from Continuous Interaction Data and It's Application for Robot Navigation

岡田　将吾*1　伊豆蔵　拓也*2　西田　豊明*2
Shogo Okada　Takuya Izukura　Toyoaki Nishida

東京工業大学　大学院総合理工学研究科　知能システム科学専攻
Dept.of Computational Intelligence and Systems Science, Tokyo Institute of Technology

京都大学　大学院情報学研究科　知能情報学専攻
Dept.of Intelligence Science and Technology, Graduate School of Informatics, Kyoto University

Human-Robot Interaction using free hand gestures and speaking word is more importance for humans which are operating robots in home or office environments. In this paper, we propose an unsupervised learning system which detects gesture command and spoken language command, action corresponding to these commands by just observing interaction behavior with the robot operated by a human operator. The main contribution of this paper is the introduction of a novel discovery algorithm that has two layered structure for detection the two kind of gestures. The proposed algorithm finds gesture patterns and action patterns accurately. The Experimental result shows that gesture patterns and action patterns are able to discovered with 75.0% ,90.0% (NMI×100) respectively by using proposed pattern discovery algorithm.

1. はじめに

ジェスチャを認識するインターフェイスはロボットを制御する上で重要な機能である．多くの研究で提案されているジェスチャ認識システムでは認識対象となるジェスチャの種類や数は予め定義されている [Mitra 07]．上記のシステムをロボットに搭載した場合，インタラクション時に，ユーザは定義されたジェスチャのみを使うよう強いられる．ユーザの任意のジェスチャを用いたインターフェイスを実現するべく，[Mohammad 09] ではユーザとロボットのインタラクションを観測し，得られた時系列データからユーザのジェスチャとそれに対応するロボットの行動パターンを教師なし学習により獲得する手法を提案した．坂本ら [Sakamoto 10] は，[Mohammad 09] の提案した時系列データの教師なし学習アルゴリズムを用いてユーザ２人とロボット１体とのポスター撮影タスクに適用した．さらに複数のモダリティのパターンの組み合わせからマルチモーダルイベントを規定し，インタラクション状況の変化を推定した．[Mohammad 09]，[Sakamoto 10] において根幹を担うアルゴリズムは Distance Graph Constrained Motif Discovery (DGCMD) と呼ばれる，時系列データから頻出パターンの検出を行う手法である．この手法はユーザのジェスチャパターンを発見するために重要であり，この手法の精度が後のジェスチャ認識精度に影響を及ぼす．

DGCMD は一般的な時系列パターンの発見アルゴリズムの一種であり，ナビゲーションに用いられるジェスチャパターンの性質を加味していない．このため以下の二点の問題が生じる．

[問題点 1] ロボットのナビゲーションに用いられるジェスチャには軌跡に意味を持つジェスチャや，手招きのようなビートジェスチャが挙げられる．一般的なモチーフ発見アルゴリズムでは出来る限り長い類似モチーフを抽出するように設計されており，ビートジェスチャのように再帰構造を持つジェスチャパターンを抽出する機構を備えていない．

[問題点 2] DGCMD ではパターン (モチーフ) のクラスタリング部において部分グラフの全探索に基づくクラスタリングを用いていた．この手法はモチーフ候補が増加した場合に計算量が増加する上，ノイズによりエッジが生成された場合，サブグラフ同士を誤ってチャンキングしてしまう．

連絡先: 岡田　将吾，東京工業大学　大学院総合理工学研究科　知能システム科学専攻，横浜市緑区長津田町 4259, J2-53, 電話/FAX 045-924-5218

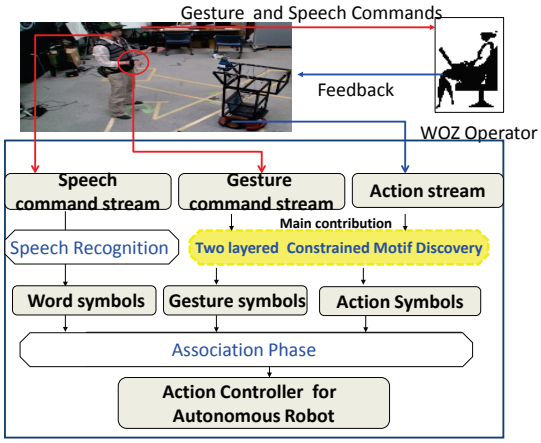

図 1: WoZ 法により取得した HRI データからのジェスチャ・駆動パターンの教師なし学習システム

上記の問題に対し，本研究では，DGCMD を以下の二点で拡張し，モーションセンサより得られた連続時系列データから高精度にジェスチャパターン (モチーフ) を発見可能な Two Layered Constrained Motif Discovery(TLCMD) アルゴリズムを提案する．

再帰的パターンをグラフの形状から特定して，パターンを抽出する処理に加え，二段階でパターン抽出を行い，再

帰構造を持つビートジェスチャと軌跡に意味のあるジェスチャの両方を抽出する．([**問題点 1**] に対応)

スペクトルクラスタリングを適用することにより，計算量の軽減とノイズによるチャンキングを防ぐことで，クラスタリング精度を向上させる．([**問題点 2**] に対応)

本研究では，上記の TLCMD を提案し，非言語情報の指示パターンとそれに対応するロボットの駆動パターンを抽出可能なシステム (図 1) を提案する．実験では TLCMD が高精度にジェスチャパターンを発見・獲得出来る事を示す．

2. 関連研究

TLCMD に関連するモチーフ発見アルゴリズムはデータマイニングの分野で提案されている．近年までに多くの研究で用いられる手法として PROJECTIONS アルゴリズム [Chiu 03] が挙げられる．PROJECTIONS アルゴリズムは SAX と呼ばれるアルゴリズム [Chiu 03] を用いて時系列データを有限個のシンボルの列に変換し，シンボルの組み合わせとしてモチーフを発見する手法である．PROJECTIONS ではシンボルの数やモチーフの系列長を指定する必要がある．

[Minnen 07] ではモチーフ候補を特定するために密度推定を行い，高い密度を与えるモチーフを選択し，これを HMM で学習して再度モチーフ候補を探索する．

[Vahdatpour 09] は多次元の時系列データからモチーフを発見する手法を提案している．各次元のモチーフ発見には PROJECTIONS が用いられており，各次元で得られたモチーフのペア同士の共起関係に基づき，統合していく手法を取っている．また次元間のモチーフの統合の後，この結果を各次元の処理にフィードバックし，この共起関係を制約とし再度，モチーフ発見を各次元で行う．これを繰り返すことで，高精度に多次元モチーフの発見が可能である事を示した．

一章で述べたように [Mohammad 09, Chiu 03, Minnen 07, Vahdatpour 09], のアルゴリズムは一般的な時系列パターンの発見に焦点を当てており，ナビゲーションに用いられるジェスチャパターンの性質を加味していない．

3. 実験環境

本論文ではロボットのナビゲーションタスクを想定し，以下のように問題を設定した．

3.1 データの取得方法

ユーザのジェスチャを観測するため，リアルタイム光学式モーションキャプチャである， MAC3D Motion capture system を用いた．このシステムではマーカーが反射する光を複数のカメラで測定することにより，マーカーの三次元座標を 50fps で取得するする．このマーカーを指示者であるユーザの右腕に 1 つと動作を行うロボットに 1 つ取り付ける．また，指示者に従って動作するロボットとして，Mobilerobots 社製の Pioneer3 を基盤とした移動型ロボットを使用する．ロボットが駆動可能な自由度は，向きの変更（左回転と右回転）と前進，後退の三つである．

3.2 ロボットナビゲーションタスクの概要

本実験環境ではロボットに指示をするユーザとロボット，またユーザに見えない場所でロボットを操作する操作者の三者によりデータ取得が行われる．ユーザは自由にジェスチャと言語を用いて指示を行い，ロボットを誘導する．ユーザの指示に対して，もう 1 人の操作者は遠隔からロボットを操作する．ロボットのナビゲーションタスクを通じて，ユーザとロボットに取り付けられたマーカの位置座標の多次元時系列データ (インタラクションの履歴データ) が収集される．

4. 提案システム

WoZ を介したロボットのナビゲーションタスクの後，ユーザの発話した音声時系列データと，ジェスチャパターンを含む動作時系列データ，ロボットの駆動時系列データを得る．音声データに対しては予め，ナビゲーションに使われる単語セットを辞書として用意した音声認識システムにより音声区間検出・単語認識を行い単語シンボルに変換する　(図 1 の左列)．ジェスチャ指示を含む時系列データとロボットの駆動時系列データからモチーフ発見アルゴリズムを用いてパターンの発見を行い，クラスタリングされたジェスチャ・アクション群を HMM で学習した後，シンボル化する．最後に単語・ジェスチャ・アクションをベイジアンネットワークでモデリング (図 1 の Association Phase) し，このネットワークがロボットのモーションコントローラとなる．本研究では，頑健にジェスチャパターンを抽出するためのモチーフ発見アルゴリズム:TLCMD の提案・評価に焦点を当てる．

Algorithm1: Two layered Constrated Motif Discovery Algorithm

入力: 一次元の連続時系列データ (X), 最大モチーフ長 (l_{max}), 分節化に係る閾値 (Th), 尤度の閾値 (Th_H)

出力: 発見したモチーフセット

1. 時系列データ X に対して Sliding Window and Bottom Up (SWAB) アルゴリズム [Keogh 01] を実行し，変化点を検出する．検出した変化点集合を $\mathcal{C}$ とする 各分節点の系列番号を $c_m \in \mathcal{C}$ と定義する．

2. ここで c_m を始点として c_{m+1} を終点とするモチーフ候補を $y_m \in \mathcal{Y}$ とする．

3. モチーフ集合 $\mathcal{Y}$ の各要素 y 同士の距離行列を作成する．距離には,Dynamic Time Warping (DTW) を用いる．距離行列を類似度行列 S に以下の方法で変換する．

$$\mathbf{S}_{i,j} = \begin{cases} \exp(-\frac{dtw(\mathrm{y}_i,\mathrm{y}_j)^2}{\sigma_{i,k_i}\sigma_{j,k_j}}), \text{if } i \neq j \\ 0, \qquad\qquad\qquad \text{if } i = j, \end{cases} \tag{1}$$

ここで $\mathrm{y}_i, \mathrm{y}_j$ は $\mathcal{Y}$ の要素であり，σ_{i,k_i} は y_i の k_i 番目の近傍要素である．k_i　はパラメータであり 1 と定義した．

4. 類似度行列から再帰的なモチーフを探索する．ビートジェスチャなどの再帰的なジェスチャはプリミティブパターンの繰り返し構造を持つ．例えば，ある変化点 $c(m)$ を始点とするモチーフ $y(m)$ と，変化点 $c(m+2n)$（n は自然数）を始点とするモチーフ $y(m+2n)$ の類似度は高くなったとする．S において，ビートジェスチャを含む系列から得られる部分類似度行列は図 2 の右上図のように市松模様を形成する．一方で，軌跡に再帰構造を持たないジェスチャを含む系列から得られる部分類似度行列は図 2 の右下図のように市松模様を形成せずランダムな模様を形成する．この特徴を利用し，ビートジェスチャを抽出する．図 2 の右上図・右下図において黒い領域の類似度は 0 に近く，白い領域の類似度は 1 に近い事を示

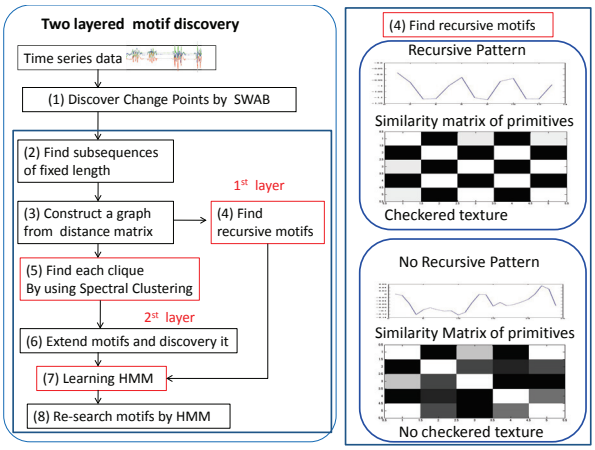

図 2: TLCMD アルゴリズムの処理手順

す．この市松模様を抽出するための条件式を式 (2) に示す．式 (2) が成り立つ時，始点 c_i，終点 c_{i+2l-1} のモチーフをビートジェスチャとして抽出する．

$$min(S^*_{i+j,i+j}, S^*_{i+j,i+j+l}) > \sigma^*_{i+j,k_{i+j}}, \forall j = \{0 \dots l-1\} \qquad (2)$$

5. ビートジェスチャとして抽出された部分行列を S から除外する．次に類似度行列 S を用いてスペクトラルクラスタリング [Ng 01] を行い，モチーフ集合 $\mathcal{Y}$ の各要素をクラスタリングする．類似度行列に対してラプラシアン L を定義し，固有値分解を行った後，固有ベクトルの直行性を利用して，k-means の初期値を選択する．ここでクラスタの数 (K) の値の選択には [Yin 05] で用いられた手法を用いた．

6. 各クラスタに分類されたモチーフ群に対し，各モチーフ同士の距離の最大値をフレーム数で割って正規化した値（正規化距離）を d_m として算出する．次にモチーフ群から 1 つのモチーフを選び，各々のモチーフを終点から系列方向に，始点から時系列を遡る方向に 1 フレームずつ伸ばしながら正規化距離を算出する．この操作を正規化距離が d_m を超えるまで繰り返す．最大の系列長を持つモチーフペアを最終的なモチーフとして抽出する．

7. 次にクラスタ内で抽出されたモチーフ群から 1 つの HMM を学習し，全てのクラスタについて同様に HMM の学習を行う．この結果 K 個の HMM が構築され，これらがジェスチャの認識器となる．ここでビートジェスチャのモチーフ群は全結合型 HMM で学習し，軌跡に基づくジェスチャのモチーフ群は left-to-right 型の HMM で学習を行う．

8. 学習された HMM 群を用いて，時系列データ X を最初のフレームから最後のフレームまで入力し，尤度を算出する．Th_H を超える部分系列をモチーフと認定し，このモチーフを訓練データとして用いて再度各 HMM を学習する．この操作を Th_H を超える部分系列がなくなるまで繰り返す．

上記の手法を全ての次元の時系列データに対して実行し，各次元の時系列データからモチーフセットを抽出する．これらのモチーフセットを [Arita 02] の手法により統合し最終的なモチーフを得る．

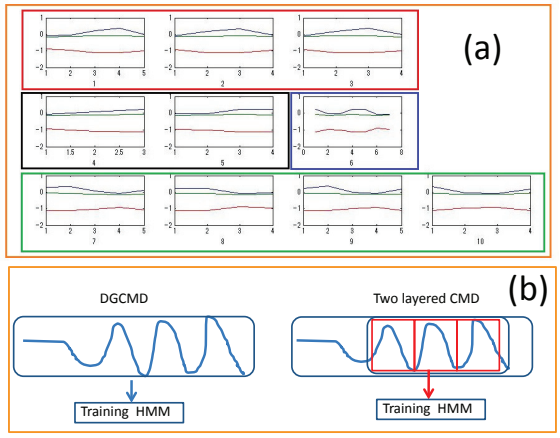

図 3: ジェスチャ指示に関するモチーフ発見の精度の比較 (図 (a) はスペクトラルクラスタリングによりクラスタリング精度が向上した例を示している．異なる軌跡の系列データは別のクラスタに分類されている事がわかる．図 (b) はビートジェスチャの抽出をアルゴリズムに組み込むことにより，分節化精度が向上した例を示す．)

本研究ではクラスタリング部にスペクトラルクラスタリングを採用した．DGCMD では，全てのサブグラフ（部分行列）で分割した場合の組み合わせを試して，最後に共通のモチーフを持つサブグラフを統合する．これらのクラスタリング方法の違いによる精度の違いを予備実験によって確認する．図 3 (a) は後の実験で得られる 4 種類の手の軌跡を観測した 10 個の系列データをクラスタリングした結果を示している．DGCMD におけるグラフクラスタリングでは，全てのパターンをマージしたのに対し，スペクトラルクラスタリングでは 4 クラスに正しく分割出来ている事がわかる．

DGCMD は TLCMD のように再帰構造を持つジェスチャを抽出する機能を有しないため，図 3(b) の左図のように，全てのモチーフについて出来るだけ長いモチーフを抽出するため，ビートジェスチャを誤って系列方向に伸ばしたモチーフとして抽出する．このモチーフを HMM の学習に用いた場合に，ビートジェスチャに対する認識精度が低下する．TLCMD ではビートジェスチャを図 3 (b) の右図のように，プリミティブの組み合わせとして分節区間を正確に抽出出来るため，HMM の認識精度の低下を防ぐ事が可能である．

5. 評価実験

提案した TLCMD の評価実験を行った．

5.1 実験設定

1 人の実験協力者をユーザとし，2 回のロボットナビゲーション実験を行った．2 回の実験で計 33500 フレームのジェスチャ・ロボットの駆動に関する時系列データを取得した．この時系列データに対し TLCMD を適用した．パラメータは $l_{max} = 600$, $Th = 0.1$, $Th_H = 0.75$ と設定した．ここでロボットの駆動データとして，ロボットに装着した 2 つのマーカの三次元座標の差分を系列データとして抽出した．

5.2 実験結果

実験でナビゲーションに用いられたジェスチャ指示の種類は “こちらに来い”, “向こうへ 1 ”, “向こうへ 2”, “左へ”, “右へ”, “止まれ”, “左に回れ”, “右に回れ”の計 8 種類，駆動パターンは “前進”, “後退”, “左へ”, “右へ”, “右回り”, “左回り”, “止まる” の

表 1: モチーフ発見の精度比較

		再現率	適合率	クラスタ数	NMI
DGCMD	Gesture	0.80	0.78	9	0.65
	Action	0.81	0.85	7	0.88
TLCMD	Gesture	0.81	0.91	12	0.75
	Action	0.85	0.90	7	0.91

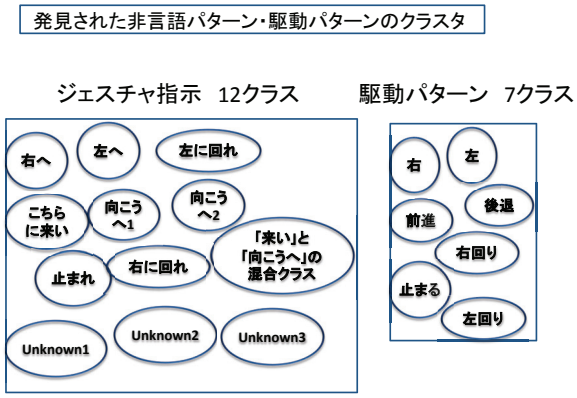

図 4: 獲得されたジェスチャパターンと駆動パターン

計 7 種類であった. モチーフ発見を行った後 12 種類のジェスチャと 7 種類の動作モチーフが抽出された. ここで抽出されたジェスチャ・アクションパターンの再現率 (R)・適合率 (P) は, ビデオ分析による分節点とのフレームの誤差として計算され, これがビデオ分析によるジェスチャ区間と 10% 以下の誤差であれば, 正確に抽出されたと見なす. 計算方法を以下に述べる.

$$R = \frac{\text{Correct Discovered Motifs}}{\text{Total of Correct Discovered Motifs}} \quad (0 \leq R \leq 1)$$
$$P = \frac{\text{Correct Discovered Motifs}}{\text{Total of Discovered Motifs}} \quad (0 \leq P \leq 1) \qquad (3)$$

精度は Normalized Mutual Information(NMI)[Manning 07] を用いて評価した. NMI は以下の式で算出される.

$$NMI = \frac{I(\Omega, \mathcal{C})}{H(\Omega) + H(\mathcal{C})} \qquad (4)$$

式 (4) において, I は相互情報量であり以下のように定義される.

$$I(\Omega, \mathcal{C}) = \sum_k \sum_j \frac{|\omega_k \cap c_j|}{N} log \frac{N|\omega_k \cap c_j|}{|\omega_k||c_j|} \qquad (5)$$

$$H(\Omega) = \sum_k \frac{|\omega_k|}{N} log \frac{|\omega_k|}{N} \qquad (6)$$

$H(\mathcal{C})$ も $H(\Omega)$ と同様に定義出来る.

図 1 に TLCMD と DGCMD の再現率・適合率・クラスタリング精度を示す. 実験結果より TLCMD は DGCMD より高い精度でジェスチャを分節化・抽出した上でモチーフ発見が行えている事を確認した. 図 4 に獲得されたジェスチャパターンのクラスタと駆動パターンのクラスタを記載する.

6. 結論

本研究では, 二段階の制約付きモチーフ発見手法である TLCMD アルゴリズムを提案した. ロボットナビゲーションタスクの実験の結果, 従来手法である DGCMD アルゴリズムより正確にジェスチャパターンを発見することが可能であることが示された.

参考文献

[Arita 02] Arita, D., Yoshimatsu, H., and Taniguchi, R.: Frequent motion pattern extraction for motion recognition in real-time human proxy, in *Proc. of JSAI Workshop on Conversational Informatics*, pp. 25–30 (2002)

[Chiu 03] Chiu, B., Keogh, E., and Lonardi, S.: Probabilistic discovery of time series motifs, in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 493–498ACM New York, NY, USA (2003)

[Keogh 01] Keogh, E., Chu, S., Hart, D., and Pazzani, M.: An Online Algorithm for Segmenting Time Series, *Proc. of IEEE ICDM*, Vol. 0, p. 289 (2001)

[lab] MAC 3D

[Manning 07] Manning, C. D., Raghavan, P., and Schütze, H.: *Introduction to Information Retrieval*, Cambridge University Press (2007)

[Minnen 07] Minnen, D., Isbell, C. L., Essa, I., and Starner, T.: Discovering multivariate motifs using subsequence density estimation, in *AAAI* (2007)

[Mitra 07] Mitra, S. and Acharya, T.: Gesture Recognition: A Survey, *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, Vol. 37, No. 3, pp. 311–324 (2007)

[Mohammad 09] Mohammad, Y. F. O., Nishida, T., and Okada, S.: Unsupervised simultaneous learning of gestures, actions and their associations for Human-Robot Interaction, in *Proc. of IEEE/RSJ IROS*, pp. 2537–2544 (2009)

[Ng 01] Ng, A. Y., Jordan, M. I., and Weiss, Y.: On Spectral Clustering: Analysis and an algorithm, in *Proc. of NIPS*, pp. 849–856, MIT Press (2001)

[Sakamoto 10] Sakamoto, K., Okada, S., and Nishida, T.: MultimodalProcessing for Acquisition of Interaction Behavior inSocial Robot, in *Social Intelligence Design (SID)*, pp. 1261–1266 (2010)

[Vahdatpour 09] Vahdatpour, A., Amini, N., and Sarrafzadeh, M.: Toward Unsupervised Activity Discovery Using Multi-Dimensional Motif Detection in Time Series, in *IJCAI*, pp. 1261–1266 (2009)

[Yin 05] Yin, J. and Yang, Q.: Integrating Hidden Markov Models and Spectral Analysis for Sensory Time Series Clustering, in *Proc. of IEEE ICDM*, pp. 506–513 (2005)