MC-GP-HSMM を用いたマルチモーダル情報の分節化による インタラクションのルール学習

木村大河* 長野匡隼* 中村友昭*

Learning Rules of Interaction by Segmenting Multimodal Data Using MC-GP-HSMM

Taiga Kimura*, Masatoshi Nagano* and Tomoaki Nakamura*

Humans can learn rules of interactions by observing others' interactions. To learn various interactions flexibly, we believe such an ability is also important for robots. To this end, we proposed Coupled Gaussian Process-Hidden Semi-Markov Models (C-GP-HSMM) that enabled robots to learn rules of interactions by observing motions of two persons. However, not only motions but also multimodal information is used in the actual human interactions. Therefore, in this paper, we extend C-GP-HSMM into a method to learn rules from multimodal data. In experiments, we show the proposed model can estimate the rules of a game, which contains multimodal interactions.

Key Words: Human-robot Interaction, Multimodal Interaction, Unsupervised Learning, Rule Learning, Motion Segmentation, Gaussian Process, Hidden Semi-Markov Model

1. はじめに

近年、家庭環境で人と協調しながら作業を行う家庭用ロボットが期待されている。このようなロボットには、人間と自然にインタラクションを行い、日常生活をサポートすることが求められる。しかし、人間が行うインタラクションのルールは多種多様であり、それらすべてを事前にロボットに教示することは困難である。

人間は、他者と自分または他者どうしが行うインタラクションを観察することで、その中に存在するルールを学習することが可能である。ロボットも同様に、観察により多種多様なルールを学習する能力を持つことが、人間との共生を実現するためには非常に重要であるといえる。そのための第一段階として本論文では、インタラクションのルールを単純化し、複数の人間が実施する行動間の共起関係をインタラクションのルールとして、その学習手法を提案する。

このようなルールを学習する手法として階層型 RNN を利用した方法が提案されている [1]。しかし、この手法では学習に多量のデータを必要とするため、人のデモンストレーションからの学習には適していない。また、複数人が行うインタラクション内に存在するルールを教師なしで推定可能なモデルとして Coupled Gaussian Process-Hidden Semi-Markov Model (C-GP-HSMM) [2] が提案されている。このモデルは、時系列データを意味のある単位ごとに分節・分類するためのモデルである GP-HSMM [3] を拡張したものであり、分節化と同時に複数人

の単位動作間の相互関係および時間差を推定することで、分節化とルール推定の精度を向上させている。しかし、このモデルでは動作情報のみを観測としていることから、人間が行うマルチモーダルなインタラクションのルール推定には対応していない。

本論文では、これらの問題を解決するため、C-GP-HSMM を拡張し、マルチモーダル情報を用いたインタラクションのルール推定を可能とした Multimodal Coupled Gaussian Process-Hidden Semi-Markov Model (MC-GP-HSMM) を提案する。さらに、実際に人間が行うマルチモーダルなインタラクションが含まれるゲームのルール推定を行い、推定されたインタラクションのルールに対する定量的な評価を行うことで、提案手法の有効性を示す。

なお、文献 [4] において本論文で示す提案手法の概要は述べており、人工データを用いた予備実験について定性的な結果のみを報告している。本論文では、提案手法をより詳細に説明するとともに、実際に人どうしのインタラクションデータを用いた実験を通じて、より詳細な定量的結果を報告する。

2. 提案手法

2.1 MC-GP-HSMM

Fig. 1 が MC-GP-HSMM のグラフィカルモデルである。このモデルでは 2 人のエージェント A, B の動作と発話を用いたインタラクションを想定しており、 S_α , S_β はそれぞれの連続動作を、 L_α , L_β はそれぞれの連続発話を表している。これらの観測系列に対して分節化を行い、単位行動ごとに教師なしで分類することで、クラス c_α , c_β が得られる。 C をクラス数、 M を観測系列数として、提案モデルでは C-GP-HSMM と同様に、以下のプロセスを M 回繰り返すことで M 個の観測系列が生成されることを仮定している。まず、エージェント A の j'

原稿受付 2020 年 8 月 18 日

*電気通信大学

*The University of Electro-Communications

■ 本論文は有用性(要素分野)で評価されました。

Fig. 1 Graphical model of MC-GP-HSMM

番めのクラス $c_{\alpha j'}$ は自身の直前のクラス $c_{\alpha, j'-1}$ により生成され、エージェント B の j 番めのクラス $c_{\beta j}$ は自身の直前のクラス $c_{\beta, j-1}$ とエージェント A の j' 番めのクラス $c_{\alpha j'}$ により生成される。

$$c_{\alpha j'} \sim P(c | c_{\alpha, j'-1}) \quad (1)$$

$$c_{\beta j} \sim P(c | c_{\beta, j-1}, c_{\alpha j'}) \quad (2)$$

$\mathbf{x}_{\alpha j'}$, $\mathbf{x}_{\beta j}$ はそれぞれエージェント A, B の単位動作であり, $\mathbf{X}_c$ をパラメータとしたガウス過程から生成される。

$$\mathbf{x}_{\alpha j'} \sim \mathcal{GP}(\mathbf{x} | \mathbf{X}_{c_{\alpha j'}}) \quad (3)$$

$$\mathbf{x}_{\beta j} \sim \mathcal{GP}(\mathbf{x} | \mathbf{X}_{c_{\beta j}}) \quad (4)$$

$\mathbf{w}_{\alpha j'}$, $\mathbf{w}_{\beta j}$ はそれぞれエージェント A, B の単位発話であり, π_c をパラメータとした多項分布から生成される。

$$\mathbf{w}_{\alpha j'} \sim P(\mathbf{w} | \pi_{c_{\alpha j'}}) \quad (5)$$

$$\mathbf{w}_{\beta j} \sim P(\mathbf{w} | \pi_{c_{\beta j}}) \quad (6)$$

生成された単位動作と単位発話をそれぞれ結合することで、観測系列である連続動作 $\mathbf{S}_\alpha$, $\mathbf{S}_\beta$ と連続発話 $\mathbf{L}_\alpha$, $\mathbf{L}_\beta$ が生成される。式 (1), (2) に示すように、提案モデルでは相手のクラスが自身のクラスの推定に影響を及ぼす。これにより、それぞれの連続行動を分節化すると同時に互いの行動間に存在する相互関係を推定することができる。

2.2 単位動作の生成モデル

本論文では、エージェント A, B それぞれの連続動作に含まれる j' , j 番めの単位動作 $\mathbf{x}_{\alpha j'}$, $\mathbf{x}_{\beta j}$ をガウス過程回帰で学習する。以降、本節では必要な場合を除き各添字を省略し、単位動作を $\mathbf{x} = \{x_1, x_2, \dots\}$ と、単位動作内の i 番めの観測値を x_i と表記する。 i を入力として、 x_i を出力する分布 $P(x_i | i)$ をガウス過程により学習することで、単位動作を連続的な分布として表現する。

ガウス過程では、クラス c に分類された単位動作の観測値の集合 $\mathbf{X}_c$ と、それに対応するタイムステップ $\mathbf{I}_c$ がそれぞれ与えられたとき、タイムステップ i' における出力 $x_{i'}$ の予測分布は式 (7) のガウス分布で表される。

$$P(x_{i'} | i', \mathbf{X}_c, \mathbf{I}_c) \propto \mathcal{N}(\mathbf{k}^T \mathbf{C}^{-1} \mathbf{X}_{c, c} - \mathbf{k}^T \mathbf{C}^{-1} \mathbf{k}) \quad (7)$$

ただし、 $\mathbf{I}_c$ の p , q 番めを i_p , i_q とすると、 $\mathbf{C}$ は p 行 q 列に式 (8) を要素に持つ行列である。

$$\mathbf{C}(i_p, i_q) = k(i_p, i_q) + \epsilon^{-1} \delta_{pq} \quad (8)$$

ここで、 $k(\cdot, \cdot)$ はカーネル関数を表し、式 (7) において、 $\mathbf{k}$ は $k(i_p, i')$ を要素に持つベクトルである。また、 c は $k(i', i')$ であり、式 (8) における ϵ は観測に含まれるノイズを表すハイパーパラメータである。ガウス過程では、カーネル関数を用いることで複雑に変化する時系列データを学習することができる。カーネル関数には様々な関数が存在するが、本論文ではガウス過程回帰において一般的に用いられている式 (9) を用いる。

$$k(i_p, i_q) = \theta_0 \exp\left(-\frac{1}{2} \theta_1 \|i_p - i_q\|^2\right) + \theta_2 + \theta_3 i_p i_q \quad (9)$$

ただし θ_* はカーネルのパラメータであり、実験ではすべて固定値を用いた。

連続動作が多次元ベクトルの時系列で表現されるとき、タイムステップ i における D 次元の出力値 $\mathbf{x}_i = (x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(D)})$ がクラス c に対応するガウス過程から生成される確率 $P(\mathbf{x}_i | i, \mathbf{X}_c)$ を式 (10) のように計算する。

$$P(\mathbf{x}_i | i, \mathbf{X}_c) = \prod_{d=1}^D P(x_i^{(d)} | i, \mathbf{X}_c^{(d)}) \quad (10)$$

最終的に多次元の時系列である長さ J の分節 $\mathbf{x} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_J\}$ がクラス c のガウス過程から生成される確率を次式のように計算する。

$$\mathcal{GP}(\mathbf{x} | \mathbf{X}_c) = \prod_{i=1}^J P(\mathbf{x}_i | i, \mathbf{X}_c) \quad (11)$$

2.3 単位発話の生成モデル

本論文では、連続発話 $\mathbf{L}$ は単位発話 $\mathbf{w}$ が連結されて生成され、単位発話 $\mathbf{w}$ は単語 $w_1, w_2, \dots$ が連結されて生成されることを仮定する。各クラス c に対応した単位発話内の単語 w_n は π_c をパラメータとする多項分布から生成されるとし、単位発話 $\mathbf{w}$ が生成される確率 $P(\mathbf{w} | \pi_c)$ を次式のように計算する。

$$P(\mathbf{w} | \pi_c) = \prod_n P(w_n | \pi_c) \quad (12)$$

$$P(w_n | \pi_c) = \frac{N_{cw_n} + \phi}{N_c + W \phi} \quad (13)$$

ここで、 N_c はクラス c に分類された総単語数、 N_{cw_n} はクラス c に分類された単語 w_n の個数、 W は単語の種類数をそれぞれ表している。また、 ϕ はディリクレ事前分布のパラメータである。これにより、単位発話を各クラスに分類することが可能となる。

2.4 学習

MC-GP-HSMM では、Blocked Gibbs Sampler を用いて隠れ変数をサンプリングすることで、モデルのパラメータを推定する。また、二者の行動は同時に行われるわけではなく時間差が生じるため、文献 [2] と同様に、二者の行動間に存在する時間差も隠れ変数として推定する。これにより行動間の誤対応を低減し、ルールの推定誤りを抑制することができる。すなわち学習では、隠れ変数である観測系列内の単位行動（単位動作と単位発話）とそのクラス、二者の行動の時間差を同時にサンプリングしパラメータを推定する。

まず、 M 個の観測系列 $\mathbf{S}_\alpha^{(m)}$, $\mathbf{S}_\beta^{(m)}$, $\mathbf{L}_\alpha^{(m)}$, $\mathbf{L}_\beta^{(m)}$ ($m =$

Algorithm 1 Backward Sampling

```

1:  $t = T, j' = J', c_{\alpha'} = \text{EOS}$ 
2: while  $t > 1$  do
3:    $k, c, d \sim A_{\alpha}[t][k][c_{\alpha}][d]P(c'_{\alpha}|c_{\alpha})$ 
4:    $\mathbf{x}_{\alpha j'} = \mathbf{S}_{\alpha, t-k:t}^{(m)}, \mathbf{w}_{\alpha j'} = \mathbf{L}_{\alpha, t-k:t}^{(m)}, c_{\alpha j'} = c, d_{\alpha j'} = d$ 
5:    $j' = j' - 1, t = t - k, c'_{\alpha} = c$ 
6: end while

```

1, 2, ..., M) があるとし、これらをランダムに分節・分類することで初期化する。次に、ある m 番目の観測系列の単位行動を、それらが分類されているクラスからそれぞれ取り除き、各クラスのガウス過程のパラメータ $\mathbf{X}_c$ 、多項分布のパラメータ π_c 、遷移確率 $P(c'|c)$ 、互いの行動間の相互関係を表現する確率 $P(c_{\alpha}|c_{\beta})$ を更新する。続けて、エージェント A について、文献 [5] と同様に Forward filtering-Backward sampling を用いて単位動作 $\mathbf{x}_{\alpha j'}$ 、単位発話 $\mathbf{w}_{\alpha j'}$ 、クラス $c_{\alpha j'}$ 、二者の行動の時間差 $d_{\alpha j'}$ をサンプリングする。Forward filtering では、観測系列のうちタイムステップ t を終点とし長さが k の部分系列 $\mathbf{S}_{\alpha, t-k:t}^{(m)}, \mathbf{L}_{\alpha, t-k:t}^{(m)}$ がクラス c_{α} の単位行動となり、そのときの相手との行動の時間差が d となる確率 $A_{\alpha}[t][k][c_{\alpha}][d]$ を次のように計算する。

$$\begin{aligned}
& A_{\alpha}[t][k][c_{\alpha}][d] \\
&= \mathcal{GP}(\mathbf{S}_{\alpha, t-k:t}^{(m)} | \mathbf{X}_{c_{\alpha}}) P(c_{\alpha} | \bar{c}_{\beta, t+d}) P(\mathbf{L}_{\alpha, t-k:t}^{(m)} | \pi_{c_{\alpha}}) P(d) \\
&\quad \times \sum_{k'=1}^K \sum_{c'=1}^C \sum_{d'=-D}^D P(c'_{\alpha} | c_{\alpha}) A_{\alpha}[t-k][k'][c'][d']
\end{aligned} \tag{14}$$

ただし、 C はクラス数、 K は単位行動の最大長、 D は時間差の最大値である。また、 $P(d)$ は 0 を平均としたガウス分布であり、 $\bar{c}_{\beta, t+d}$ は時刻 $t+d$ におけるエージェント B のクラスを表す。さらに、遷移確率 $P(c'_{\alpha} | c_{\alpha})$ と相互関係確率 $P(c_{\alpha}|c_{\beta})$ は、次式に従い計算する。

$$P(c'_{\alpha} | c_{\alpha}) = \frac{N_{c_{\alpha}c'_{\alpha}} + \lambda}{N_{c_{\alpha}} + C\lambda} \tag{15}$$

$$P(c_{\alpha} | c_{\beta}) = \frac{N_{c_{\beta}c_{\alpha}} + \gamma}{N_{c_{\beta}} + C\gamma} \tag{16}$$

ただし、 $N_{c_{\alpha}}$ はクラス c_{α} に分類された単位行動の数を、 $N_{c_{\alpha}c'_{\alpha}}$ はクラスが c_{α} から c'_{α} へ遷移した回数をそれぞれ表している。また、 $N_{c_{\beta}}$ はクラス c_{β} に分類された単位行動の数を表しており、 $N_{c_{\beta}c_{\alpha}}$ はエージェント B がクラス c_{β} の行動をしたときにエージェント A がクラス c_{α} の行動をした回数を表している。次に、Forward filtering により求めた確率に従い Backward sampling (Algorithm 1) を行い、単位動作 $\mathbf{x}_{\alpha j'}$ 、単位発話 $\mathbf{w}_{\alpha j'}$ 、クラス $c_{\alpha j'}$ 、二者の行動の時間差 $d_{\alpha j'}$ をサンプリングする。このサンプルされた隠れ変数から、各クラスのガウス過程のパラメータ $\mathbf{X}_c$ 、多項分布のパラメータ π_c を更新する。

この処理を M 個すべての観測系列に対して行い、さらに α と β を入れ替え、エージェント B についても同様の処理を行う。以上を繰り返すことにより、すべての観測系列を分節・分類し、各クラスのガウス過程のパラメータ $\mathbf{X}_c$ 、多項分布のパラメータ π_c を推定することができる。

3. 実験

本実験では、提案手法である MC-GP-HSMM の有効性を検

Fig. 2 Training data that represents behaviors of two persons

証するために、二者間で行う「旗揚げゲーム」を対象としてインタラクションのルール推定を行った。旗揚げゲームでは、指示者が「赤上げて」「赤下げて」「白上げて」「白下げて」のいずれかの指示を出し、動作者がその指示に従い旗を上下に動かす。ただし、指示者の姿勢は一定であり、動作者は発話を行わないものとする。

3.1 実験に使用したデータ

エージェント A を指示者、エージェント B を動作者として設定し、両者から取得した動作情報および発話情報を学習データとして用いた。なお、エージェント A には一定の姿勢を表す動作情報を、エージェント B には発話がない状態を表す発話情報をそれぞれ入力として与えた。

動作情報：動作者の動作をビデオカメラで撮影後、学習時の計算量を削減するため毎秒 10 フレームにダウンサンプリングした。その後、動画の各フレームに対して OpenPose [6] を用いて骨格推定を行い、得られた動作者の左右手首の横 (x)、縦 (y) 座標で構成される四次元ベクトルを動作の観測情報とした。**Fig. 2** (下) が、実際に実験に用いたエージェント B の動作情報の一例である。

発話情報：指示者の発話をマイクで録音後、Julius [7] を用いて音声認識を行った。音声認識により得られた単語列を One-Hot ベクトルに変換することで、発話のタイミングを保持した時系列データとした。**Fig. 2** (上) が、エージェント A の発話情報の一例である。グラフ上にプロットされた点は、それぞれの単語が発話された時刻と単語の種類を表している。

実験では、このように取得した長さが 193~204 サンプル (平均 198 サンプル) の 4 個 ($M = 4$) の時系列データを用いた。

3.2 実験結果**3.2.1 分節化結果**

Fig. 3 に、エージェント A とエージェント B の行動の分節化結果および想定した正解を示す。図において、横軸はタイムステップを、バーの色は単位行動のクラスをそれぞれ示している。ただし、単位動作については、動作者の直前の姿勢によって動作後の姿勢が異なるため、各動作に対して二つのクラスに分類された状態を正解とした。例えば、赤旗を上げるという動作に対して、赤旗と白旗の両方を上げているクラスと、赤旗だけを上げているクラスの 2 クラスに分類されることを想定している。

エージェント A については、末尾の三つの単位行動を除いて正解に類似した分節化結果を得ることができた。またエージェント B については、クラス 5 の「白旗を下げる動作」がクラス 12 の「赤旗を上げる動作」に誤分類されている。これは、この

Fig. 3 Segmentation results

Table 1 Learned probabilities $P(C_{\beta j} | C_{\alpha j'})$

		Class of B's action $C_{\beta j}$							
Class of A's action $C_{\alpha j'}$		5	6	7	8	9	10	11	12
	1	0.00	0.09	0.14	0.03	0.26	0.10	0.00	0.37
	2	0.00	0.29	0.00	0.01	0.00	0.44	0.11	0.14
	3	0.00	0.03	0.32	0.06	0.12	0.00	0.37	0.10
	4	0.00	0.11	0.04	0.39	0.08	0.02	0.00	0.35

Table 2 Learned probabilities $P(C_{\beta, j+1} | C_{\beta j})$

		Class of B's action $C_{\beta j+1}$										
Class of B's action $C_{\beta j}$		5	6	7	8	9	10	11	12			
	5	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08			
	6	0.02	0.02	0.40	0.02	0.02	0.02	0.02	0.40			
	7	0.02	0.02	0.02	0.40	0.40	0.02	0.02	0.02			
	8	0.02	0.02	0.40	0.02	0.02	0.02	0.02	0.40			
	9	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.74			
	10	0.02	0.02	0.02	0.40	0.40	0.02	0.02	0.02			
	11	0.02	0.02	0.02	0.02	0.02	0.79	0.02	0.02			
	12	0.01	0.45	0.01	0.01	0.01	0.01	0.45	0.01			

二つの動作の最終的な姿勢が類似していることや、データの偏りによりクラス5の前後のクラス遷移がクラス12と類似していたことが原因として考えられる。

さらに分節化の定量的評価として、正規化ハミング距離を計算した。正規化ハミング距離は分類の誤り率に相当し、値が低いほど正解に近いことを意味する。その結果、エージェントAでは0.14、エージェントBでは0.15となり、一部分類誤りはあるものの、全体としてはおおむね正しい分節が得られていることが確認できた。

3.2.2 ルール推定結果

分節化された単位行動から相互関係を表す確率を計算した結果がTable 1, 2である。Fig. 3に示すように、エージェントAの行動はクラス1~4に、エージェントBの行動はクラス5~12にそれぞれ分類されている。

Table 1より、例えばエージェントAがクラス2の行動（「白上げて」と発話）をするとき、エージェントBがクラス10の行動（白旗を上げる動作）をする確率が44%となっている。このように、指示者の発話に応じて動作者が旗を動かすルールを学習できていることが確認できる。一方で、エージェントAがクラス1の行動（「白下げて」と発話）をするとき、エージェントBがクラス12の行動（赤旗を上げる動作）をする確率が37%となっており、誤った相互関係がルールとして学習されている。これは前述の分類誤りによるものであり、クラス5に分類されると想定していた単位行動がクラス12に分類されたことにより、ルール推定にも誤りが生じている。

またTable 2より、例えばエージェントBがクラス7の行動（白旗を上げた状態で、赤旗を下げる動作）をした後に、クラス8の行動（赤旗を上げる動作）をする確率および、クラス9の

行動（白旗を下げる動作）をする確率がともに40%となっている。このように、エージェントBが行う動作の遷移がルールとして学習されていることが分かる。

さらにルール推定の定量的評価として、正しいルールを表現した確率分布と学習された確率分布の間のコサイン類似度を計算した。その結果、相互関係確率 $P(C_{\beta j} | C_{\alpha j'})$ では0.81、遷移確率 $P(C_{\beta, j+1} | C_{\beta j})$ では0.86となり、誤分類に起因する誤りはあるものの、全体としてはおおむね正しいルールが学習されていることが確認できた。

4. ま と め

本論文では、これまでに提案されてきた、複数人が行うインタラクションのルールを学習可能なモデルであるC-GP-HSMMを拡張し、時系列マルチモーダル情報を用いた学習を可能とするMC-GP-HSMMを提案した。提案モデルでは、各クラスに分類された単位発話内の単語の発生確率を、単位発話の生成モデルにより導出することで、人間の発話を含めた行動を各クラスに分類することを可能とした。実験では、二者のマルチモーダル情報を用いたインタラクションが含まれる旗揚げゲームを対象としたルール学習を行い、学習データとして実際に人間の行動から取得した動作情報および発話情報を用いることで、提案モデルの実データに対する有効性を検証した。実験結果より、提案手法を用いることで、人間の動作情報および発話情報を分節化すると同時に、行動間の相互関係を表す確率からインタラクションのルールを学習可能であることを示した。

しかし、本論文で想定したインタラクションのルールは限定的であることから、より複雑なインタラクションに対応することが今後の課題である。そこで次のステップとして、ルールの表現に条件付き確率場を導入し、多様な行動間の相互関係を推定することで、より複雑なインタラクション内に存在するルールも学習可能なモデルへ拡張することを考えている。

参 考 文 献

- [1] T. Yamada, S. Murata, H. Arie and T. Ogata: "Dynamical Linking of Positive and Negative Sentences to Goal-Oriented Robot Behavior by Hierarchical RNN," International Conference on Artificial Neural Networks, pp.339-346, 2016.
- [2] S. Oshikawa, T. Nakamura, T. Nagai, K. Funakoshi, N. Iwahashi, M. Nakano and M. Kaneko: "Interaction Modeling Based on Segmenting Two Persons Motions Using Coupled GP-HSMM," IEEE International Symposium on Robot and Human Interactive Communication, pp.288-293, 2018.
- [3] T. Nakamura, T. Nagai, D. Mochihashi, I. Kobayashi, H. Asoh and M. Kaneko: "Segmenting Continuous Motions with Hidden Semi-markov Models and Gaussian Processes," Frontiers in Neurobotics, vol.11, article 67, 2017.
- [4] 押川慧, 中村友昭, 長井隆行, 岩橋直人, 船越孝太郎, 竹内誉羽, 中野幹生, 金子正秀: "Coupled GP-HSMMを用いた時系列マルチモーダル情報の分節化に基づくインタラクションのモデル化," 言語処理学会第25回年次大会, A4-1, 2019.
- [5] K. Uchiumi, H. Tsukahara and D. Mochihashi: "Inducing Word and Part-of-Speech with Pitman-Yor Hidden Semi-Markov Models," The 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, pp.1774-1782, 2015.
- [6] Z. Cao, G. Hidalgo, T. Simon, S.E. Wei and Y. Sheikh: "OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.43, issue 1, pp.172-186, 2021.
- [7] Julius, <https://julius.osdn.jp/>