

# 自然言語による身振り動画の動的意味付け\*

太田 陽輔, 唐澤 博†

山梨大学 工学部‡

E-mail: {yohsuke, karasawa}@jewel.yamanashi.ac.jp

## 1 はじめに

人間の日常生活での対話におけるジェスチャーには、対話における発話主の発言の意味との関係が付加されているものと考えられる。

先行研究 [1] では、曲率に基づく記号化手法 [2] の考え方をベースとして、ビデオカメラから取得される被験者の身振り動作データの記号列への変換を行った。本研究は、先行研究 [1] の考え方にに基づき、身振り動作の動的意味付けを行う。

本研究のシステムは、被験者の動作に対する記号列のみが入力された場合は、その意味を回答することができ、記号列と自然言語文が同時に入力された場合には、動的にそれらの関係を学習 (意味付け) する事が出来る。

## 2 システムの概要

### 2.1 システムの中核部

本研究のシステムは、要素として動作記号列と意味を有し、図1の様に完全グラフにより連結されている。この完全グラフにおける各リンクには、重みがつけ

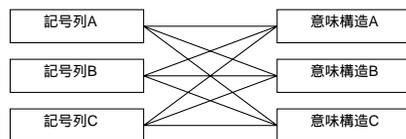


図1: 意味付けによる完全グラフ

られる。システムはこのリンクをたどることにより、重みの一番高いリンク先を動作記号列に対応する対の意味とする。

\*Dynamic association of meaning of gesture

†Yosuke Ota, Hiroshi Karasawa

‡Yamanashi University, 4-3-11 Takeda, Kofu, Yamanashi 400-8511, Japan

### 2.2 学習モードと連想想起モード

本研究のシステムは、2つのモードを有する。

システムに動作記号列と自然言語文の両方を同時に入力すると、システムは学習モードになる。システムに動作記号列一方のみを入力すると、システムは連想想起モードになり、対の他方を出力する。

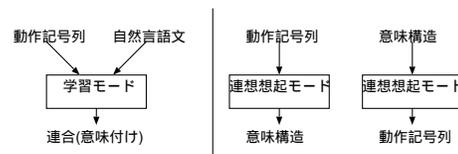


図2: 学習モードと連想想起モード

## 3 学習モード

### 3.1 要素へのデータの格納方法

#### 3.1.1 動作記号列における DP マッチング

時系列データである動作記号列を、どの様に意味付けの要素として格納するかについて、本研究では、動作記号列モデルを用意し、そのモデルと、得られた動作記号列間において DP マッチングを行い、距離の一番近いモデルを要素とする。

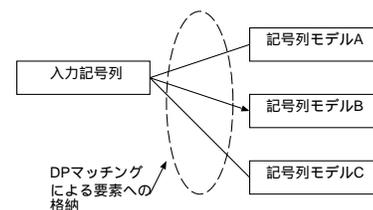


図3: 動作記号列による DP マッチング

### 3.1.2 自然言語文の意味解析

自然言語文をシステムに入力すると、システムはそれを意味解析し、意味構造を出力する。本研究では、この意味構造を身振り動作の意味付けの要素とする。

### 3.2 重み付けによる学習

学習モードにおいて、動作記号列と意味構造の意味付けに用いる数値は、1セッションの重み付けにおいて、1だけ単調増加するものとする。

$$W_i^t = W_i^{t-1} + 1 \quad (1 \leq t) \quad \text{ただし, } W_i^0 = 0 \quad (1)$$

$W_i^t$  は、時刻  $t$  におけるリンク  $i$  の重みを表す。

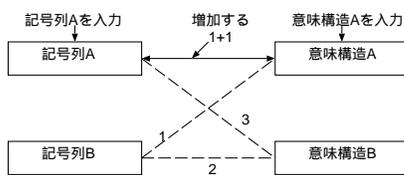


図 4: 重みの更新

## 4 連想想起モード

### 4.1 連想想起

連想想起モードにおける連想想起処理は以下の式を用いる。

$$A_i = \frac{W_j}{\sum_i W_i} \quad (2)$$

$A_i$  はノード  $i$  の活性値であり、 $W_i$  は入力ノードと対をなすリンクの重み、 $\sum_i W_i$  は入力された記号列の要素から出ているリンクの重みの総和である。この式を全てのリンクの重みに対して計算し、一番値の大きいリンクをたどり、その先の要素を入力記号列に対応する対とする。

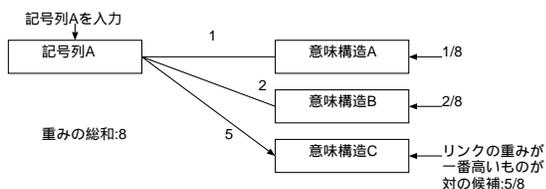


図 5: 連想想起

## 4.2 重みの等価競合の問題

意味付けによる学習の進行により、2つ、またはそれ以上のリンクの重みが等しくなった場合、候補の競合が発生する。一つの入力記号列に対し、一つの候補を選択すべきと考えられるが、本研究では競合が起こった場合、連想想起される複数の候補全てを選択し、対応する意味とする。

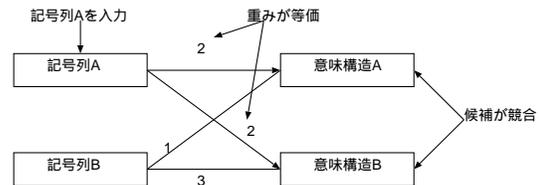


図 6: 候補の競合問題

## 5 考察

本研究のシステムでは、単純に身振りの意味を教え込む手法と比較して、常に動作記号列と意味構造との関係を動的に更新出来ることが強みである。常に学習を監視していく事により、一つの身振りに対する意味が、教え込む回数に応じて変化する。

自然言語文を入力した場合、対応する動作記号列を回答する逆方向の処理については、将来的に入力された自然言語文に対応した身振りを、システムがインタフェースを介してコンピュータグラフィックス表示する事に応用できると考えている。それにより、システムの対話コミュニケーション能力の向上が期待出来ると考えられる。

今後の課題として、

- 動作記号列の取り扱いにおける、DP マッチング以外の手法の考案
  - 重み付け手法の改善
  - 重みの競合問題の対処法の考案
- 等が挙げられる。

## 参考文献

- [1] 青木由香: 身振り動画像認識とその自然言語表現への変換, 山梨大学工学部コンピュータメディア工学科卒業論文, p5-20, 山梨大学卒業論文 (2002)
- [2] 西川 敦, 大西 映生, 宮崎文夫: 連続動画像からのオプティカルフローを用いた身振り動作の曲率に基づく分割と認識, 画像の認識・理解シンポジウム (MIRU'98), p37-38, 1998