

# 言語モデルを用いた人の動作を説明するテキスト生成 およびその評価

樺山 絵里<sup>†</sup> 麻生 英樹<sup>‡</sup> 小林 一郎<sup>†</sup> 持橋 大地<sup>¶</sup> Muhammad Attamimi<sup>§</sup>

中村友昭<sup>§</sup> 長井隆行<sup>§</sup>

<sup>†</sup>お茶の水女子大学大学院 <sup>‡</sup>産業総合技術研究所 <sup>¶</sup>統計数理研究所 <sup>§</sup>電気通信大学大学院

<sup>†</sup>{kabayama.eri,koba}@is.ocha.ac.jp, <sup>‡</sup>h.asoh@aist.ne.jp, <sup>¶</sup>daichi@ism.ac.jp,

<sup>§</sup>m.att@apple.ee.uec.ac.jp, <sup>§</sup>tnakamura@uec.ac.jp, <sup>§</sup>tnagai@ee.uec.ac.jp

## 1 はじめに

本研究では、人の動作の時系列データを Kinect センサを用いて取得し、次元圧縮を通じて得られた動作パターンを識別する識別器を構築する。また、動作を説明する自然言語文からバイグラムモデルを構築し、識別された動作に対するバイグラムモデルから、動的計画法を用いて、尤度の高い学習の単語の組み合わせを獲得することによりテキスト生成を行う。本稿では先行研究 [1][2] で提案された手法の汎用性を検証することを目的として、時系列データからの言語生成およびテキスト生成のための言語資源の転移について調査を行う。

## 2 時系列データからの言語生成

今回対象としているのは、図1に示す意味的な構成からなる簡単な人の動作である。先行研究 [1] における動作の意味的な構成の「前から」「横から」を「速く」「遅く」に変更したものにす。



図 1: 動作カテゴリの構成

対象とする動作は、「右/左/両」「手/足」を「速く/遅く」「上げる/下げる」の6つの動作から構成されると定義する。動作カテゴリは、例えば「両」「手」「速く」「上げる」の組み合わせにより「両手を速く上げる」というカテゴリが構成される。それぞれの動作カ

テゴリに対して、テキストデータを収集し、それを元にバイグラムモデルを構築する。その際、文長が長い文は尤度が低くなってしまいうため、図2のように文長を揃えるために仮想の単語 null を導入し、文長に左右されないテキスト生成を実現する。この言語資源 full から構築されるバイグラムを用いて、テキスト生成を行う。「両手を速やかに上げる」という動作に対する生成結果を表1の full に示す。



図 2: 仮想の単語 null の導入

## 3 zero-shot 学習に基づく言語資源推定

特定の意味を説明するための言語資源が存在しないことの影響を評価するために、一部の動作に対する言語資源を取り除き、最小二乗推定による zero-shot 学習を行い、他の動作に対する言語資源用いて、欠損している動作に対する言語モデルを推定する。その後、推定された言語モデルを用いて説明文の生成を行い、得られた説明文の品質を評価する。

zero-shot 学習は、動作カテゴリに対して、バイグラムを表した行列  $\Psi$  を動作カテゴリに対して、それが含んでいる動作要素を表した行列  $A$  およびそれぞれの動作要素に対して使用される言語資源の確率を表す行列  $\Phi$  とに分解できるということを前提に、特定の動作カテゴリのバイグラムが欠損している場合、行列

表 1: 「両手を速やかに上げる」という動作に対する削減された言語資源の下での生成文

動作	生成文
full	<ul style="list-style-type: none"> <li>● 両手を速やかにあげる。 null6 null7 null8 null9 null10 null11 null12 null13 null14 null15</li> <li>● 両手を速やかにあげる。 null7 null8 null9 null10 null11 null12 null13 null14 null15 EOS</li> <li>● 両手を下から上にあげる。 null6 null7 null8 null9 null10 null11 null12 null13</li> </ul>
three-quarters	<ul style="list-style-type: none"> <li>● 両手を速やかにあげる。 null6 null7 null8 null9 null10 null11 null12 null13 null14 null15</li> <li>● 両手を速やかにあげる。 null5 null6 null7 null8 null9 null10 null11 null12 null13 null14</li> <li>● 両手を下から上にあげる。 null6 null7 null8 null9 null10 null11 null12 null13</li> </ul>
half	<ul style="list-style-type: none"> <li>● 両手をさげる。 null5 null6 null7 null8 null9 null10 null11 null12 null13 null14 null15 EOS</li> <li>● 両手を速やかにあげる。 null6 null7 null8 null9 null10 null11 null12 null13 null14 null15</li> <li>● 両手を速やかにあげる。 null5 null6 null7 null8 null9 null10 null11 null12 null13 null14</li> </ul>
min	<ul style="list-style-type: none"> <li>● 両手を下から上にあげる。 null6 null7 null8 null9 null10 null11 null12 null13</li> <li>● 両手を下から上にあげる。 null5 null6 null7 null8 null9 null10 null11 null12</li> <li>● 両手を下から上にあげる。 null8 null9 null10 null11 null12 null13 null14 null15</li> </ul>

表 2: BLEU スコアおよび生成文の性能における評価結果

		full	three-quarters	half	min
BLEU	データ全動作	1.0	0.7244	0.7272	0.6817
	欠損動作	(1.0)	0.7754	0.7294	0.6773
対数尤度	min, half, three-quarters 共通欠損動作	-4.0985	-5.1324	-6.4674	-5.5602
	half, three-quarters 共通欠損動作	-3.8706	-5.2271	-6.5137	—

の一般逆行列を使って欠損部分の要素を推定する。上述した内容は以下の式で表せる。ここで  $A^+$  は  $A$  の一般逆行列を示す。

$$\hat{\Phi} = \min_{\Phi} \|\Psi - A\Phi\|^2 = A^+\Psi \quad (1)$$

## 4 実験

### 4.1 実験設定

先行研究 [2] と同様に、zero-shot 学習により、データ欠損動作の言語モデルをどの程度正確に推定可能であるかを検証するために、動作カテゴリの構成において出現していない構成要素が存在しないようにバランスを考慮しつつ、その一部を取り除いた言語資源のデータを用意した。full は全データ、three-quarters、half は全データから 1/4 および 1/2 のデータをそれぞれ取り除き、min は全構成要素が現れる最小限の言語資源とする。生成された文の定量的な評価手法として、BLEU スコアによる評価と生成文の対数尤度評価を用いた。

### 4.2 実験結果および評価結果

言語モデル、full、three-quarters、half、min に対して、それら全てに共通して推定された言語モデルで

ある動作「両手を速やかに上げる」に関するテキストの生成結果を表 1 に、評価結果を表 2 に示す。先行研究 [2] と同様、言語資源が減少するにつれて、評価が下がり、文章として説明ができていないものが増えることが確認された。

## 5 おわりに

人の動作を表す時系列データから動作速度を考慮したパターン抽出を行い、そのパターンに基づいて速さを含んだ動作に対する説明文生成を行った。また、構築した言語資源を動作の意味的なカテゴリの下で zero-shot 学習を行うことより、言語資源の量に対するテキスト生成の性能を評価した。

## 参考文献

- [1] 小林瑞季, 麻生英樹, 小林一郎, 人の動作を対象にした確率的言語生成への取り組み, 言語処理学会第 20 回年次大会, pp.920-923, 北海道大学, 2014.
- [2] 樺山絵里, 麻生英樹, アッタミムハンマド, 小林一郎, 持橋大地, 中村友昭, 長井隆行, 現象の意味的構造に基づく言語モデルの zero-shot 学習の試み, 第 29 回人工知能学会全国大会, 2F3-2, はこだて未来大学, 2015.