

テキスト生成における言語資源への転移学習適用

樺山 絵里 (指導教員: 小林 一郎)

1 はじめに

近年, 非言語情報からテキスト生成する研究が盛んになってきている. また, 学習用データが全く存在しない際に用いる転移学習の手法として, zero-shot 学習 [1] が注目されている. この両方を組み込んだ取り組みとして, Asoh ら [2] が行った, 観測した視覚情報である人の動作を過去に説明したことが無い場合に, これまでに見た動作を説明するのに用いた言語資源を zero-shot 学習により, その動作を説明する試みがある. 本研究では, Asoh ら [2] によって提案された言語資源の転移手法をより汎用的にするため, 言語資源の構成及び対象動作の種類をデータセットごとに変更し, テキスト生成のための言語資源の転移を行った.

2 言語モデルを用いたテキスト生成

本研究では, 収集したテキストから構築したバイグラムモデルを用いて, 尤度が高くなるような単語の組み合わせを見つけることにより文の生成を行うとする. 文長に左右されないテキスト生成を行うために, 小林ら [3] が用いた, 疑似単語 (番号付き null) をバイグラムモデルに導入することにより文長に関わらず尤度の次数を同じにするテキスト生成手法を適用する.



図 1: 仮定の単語 null の導入

3 Zero-shot 学習に基づく言語資源推定

3.1 動作の意味的な構成

本研究で用いたデータセットは人の動作を対象としており, 概要を図 2 に示す. 対象とする動作は「上げる/下げる」「速く/普通/遅く」「高く/低く」「前から/横から」「手/足」「左/右/両」の組み合わせの 120 の動作から構成されると定義する. 1 動作あたり 4 文および 6 記述要素が含まれる. 例えば「左手を前から高く速く上げる」の場合「左」「手」「前から」「高く」「速く上げる」という要素から成ると考えることができる.

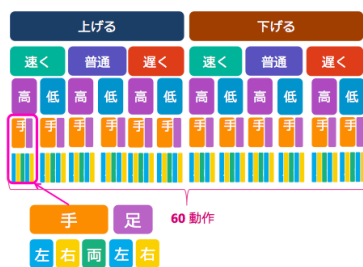


図 2: 動作カテゴリーの構成

3.2 Zero-shot 学習の方法

図 2 を例にして, Zero-shot 学習の方法を説明する. 図 2 に示す k 番目の動作に l 番目の要素が含まれていることを $a_{kl} = 1$ で表し, それを成分とする行列を A とする.

各動作に対する言語モデルとして, 2 単語のバイグラムの出現確率 $p(w_i|w_j)$ 動作 k に対する説明文集合から計算される $p(w_i|w_j)$ の値を並べたベクトルを ψ_k とし, それを各行とする行列を Ψ とする. ここで, 行列 Ψ が, $\Psi = A\Phi + \epsilon$ のように近似的に分解できることを仮定する. ここで, Φ は動作の構成要素に対する言語モデルを行とする行列である. すなわち, 各動作に対する言語モデルが, 動作の構成要素に対する言語モデルの線形の重みつき和で近似できると仮定していることになる. この仮定に基づき, 以下の手続きに示される Zero-shot 学習の方法を提案した. 以下では, 学習用データ (説明文) が存在しない動作を「データ欠損動作」と呼ぶ.

Zero-shot 学習の流れは, 以下のようになっている.

step1. Ψ の内の, 学習用データが存在する動作に対応する行だけ から成る行列を Ψ' とする. また, A の内の, 同じようにデータが存在する 行動に対応する行から成る行列を A' とする.

step2. Ψ' と A' から, 動作の構成要素に対する言語モデル $\hat{\Phi}$ を最小二乗推定する (式 1).

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

step3. 推定された $\hat{\Phi}$ を用いて $\hat{\Psi} = A\hat{\Phi}$ のように, Ψ の削除した行を復元することで, データ欠損動作に対する言語モデルを推定する.

4 実験

図 2 に示すものを実験データの例として説明する. 学習用データが存在しないことの影響を評価するために, データセットの一部の動作に対するデータを取り除き, 最小二乗推定による Zero-shot 学習を行うことにより, 他の動作に対するデータを用いて, データ欠損動作に対する言語モデルの推定を行う. その後に, 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した.

4.1 実験設定

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, 動作の意味的な構成において均等に取り除くようにして以下の 4 つの場合について検討した.

1. full (言語資源を全て使用)
2. three-quarters (4 分の 3 を使用)
3. half (半分を使用)
4. min (文生成が可能な最低限の数を使用)

生成された文の定量的な評価手法として, 以下の 2 つを考える.

表 1: 「左手を速く低く上げる」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

言語資源	生成文
full	<ul style="list-style-type: none"> • 左手を素早く上げる。 null6 null7 null8 null9 null10 null11 EOS • 左手を速く低く上げる。 null6 null7 null8 null9 null10 null11 • 左手を速く低く上げる。 null5 null6 null7 null8 null9 null10
three-quarters	<ul style="list-style-type: none"> • 左手を素早く上げる。 null6 null7 null8 null9 null10 null11 EOS • 左手を速く低く上げる。 null6 null7 null8 null9 null10 null11 • 左手を速く低く上げる。 null5 null6 null7 null8 null9 null10
half	<ul style="list-style-type: none"> • 左手を素早く上げる。 null6 null7 null8 null9 null10 null11 EOS • 左手を速く低く上げる。 null6 null7 null8 null9 null10 null11 • 左手を速く低く上げる。 null5 null6 null7 null8 null9 null10
min	<ul style="list-style-type: none"> • 左足を素早く上に挙げる。 null7 null8 null9 null10 null11 • 人が左足を素早く上に挙げる。 null7 null8 null9 • 左足を素早く上に挙げる。 null6 null7 null8 null9 null10

表 2: BLEU スコアおよび生成文の対数尤度に基づく評価結果

	推定された言語資源	full	three-quarters	half	min
BLEU	(データ全動作)	1.0	0.7949	0.6566	0.3528
	欠損動作	(1.0)	0.3940	0.3763	0.2739
対数尤度	min, half, three-quarters 共通欠損動作	-1.9464	-2.0115	-4.3354	0.2053
	half, three-quarters 共通欠損動作	-2.2846	-2.4807	-4.8463	—

• BLEU スコアによる評価

full のデータから学習した言語モデルによって生成されたテキストと Zero-shot 学習によって推定された言語モデルによって作成されたテキストとの BLEU スコアにより評価する。

• 生成文の尤度評価

Zero-shot 学習によって推定された言語モデルから生成された尤度が上位 3 件の説明文の対数尤度を full のデータから学習した言語モデルを用いて算出した際の平均値によって評価する。このとき、full の言語モデルの中に推定された言語モデルから生成された文に現れる単語ペアがない場合には、適切なスムージングを行って補う。

4.2 実験結果

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「左手を速く低く上げる」に関するテキスト生成結果を表 1 に示す。

4.3 評価結果

4.3.1 BLEU スコアによる評価

Zero-shot 学習により推定された言語モデルを用いて生成された文を、full の言語モデルにより生成した文を正解文とした場合の BLEU スコアを用いて評価した結果について述べる。表 1 に、Zero-shot 学習によって推定された言語モデルおよび取り除く対象にならなかった言語モデルの両方を用いて、全動作に対するテキスト生成を行った結果を示す。また、three-quarters, half, min それぞれのデータ欠損動作に対して Zero-shot 学習によって推定された言語モデルから生成された文と full のデータから推定された言語モデルから生成された文との一致を評価した結果を示す。表 1 より、どちらに関しても、取り除かれた言語モデルの推定に多くの学習データを使っているものほど、精度の高い文が

生成されていることがわかる。

4.3.2 生成文の性能評価

three-quarters, half, min について共通するデータ欠損動作に対して、Zero-shot 学習で推定された言語モデルから生成された文の対数尤度を、full の言語モデルで計算した。

full, three-quarters, half の 3 つのケースをより詳しく比較するため、three-quarters と half に共通するデータ欠損動作についての評価も実施した。表 2 にその結果を示す。

より多くのデータを用いて生成したものほど生成文の精度が高くなっている。全体的に、BLEU スコアに比べて生成文の尤度による評価のほうが、性能の落ち方が顕著に現れるのを観察できることがわかる。

5 おわりに

本研究では、動作の意味的な構成を利用する Zero-shot 学習によって言語モデルを生成し、そのモデルから生成された文を評価し考察した。

参考文献

[1] Larochelle, H., Erhan, D., & Bengio, Y. Zero- data learning of new tasks. AAAI Conference on Artificial Intelligence, 2008.

[2] Hideki Asoh and Ichiro Kobayashi, Zero-Shot Learning of Language Models for Describing Human Actions Based on Semantic Compositionality of Actions, The 28th Pacific Asia Conference on Language, Information and Computing, Dec. 12-14,; Phuket, Thailand, 2014.

[3] 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.