

# Zero-shot 学習した言語モデルによるテキスト生成結果の評価

樺山 絵里 (指導教員: 小林 一郎)

## 1 はじめに

近年, 直接観測される学習用データが全く存在しない状況における機械学習手法として, Zero-shot 学習が盛んに取り上げられている. Zero-shot 学習は, マルチタスク学習の一種であり, 対象とする学習課題に関する学習用データ無しで学習を行う手法である [1]. カテゴリ間の意味的な関係などを利用することで, 一部のカテゴリに関する学習用データが無い状態でも, 他のカテゴリに関する学習用データの情報を使った学習が可能になることが示されている. ある複数の対象を言葉で説明するために, 各対象を記述するために必要となる言語モデルによりテキストを生成することを考える. その際, 一部の対象の言語モデルが存在しない場合に, 他の対象の言語モデルを Zero-shot 学習によって推定可能であることが示されている [2]. このことを踏まえ, 本研究では, Asoh らによって提案された言語モデルの Zero-shot 学習手法において, 与えられる学習用データの量を変化させた時に生成される文に対する評価を行った結果について考察する.

## 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.

## 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

手足の上げ下げに関する人の 20 種類の動作 (「データ 1」とする. 図 1 にその概要を示す.) および, 人, 物, 場所, 動きの 4 つのカテゴリから人の動作を記述した 131 種類の動作 (「データ 2」とする. 図 2 にその一部を示す.)



図 1: 動作の意味的な構成 (データ 1)



図 2: 動作の意味的な構成 (データ 2)

### 3.2 Zero-shot 学習の方法

データ 1 を例にして, Zero-shot 学習の方法を説明する. 図 1 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 単語ペアの出現確率  $p(w_i, w_j)$  を求めることを考える. 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  $\psi_k$  とし, それを各行とする行列を  $\Psi$  とする. ここで, 行列  $\Psi$  が,  $\Psi = A\Phi + \varepsilon$  のように近似的に分解できることを仮定する. ここで,  $\Phi$  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な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.  $\Psi$  のうちの, 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  $\Psi'$  とする. また,  $A$  のうちの, 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.  $\Psi'$  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,  $\Psi$  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

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 1 およびデータ 2 の一部の動作に対する学習用データを取り除き, 最小二乗推定による Zero-shot 学習を行うことにより, 他の動作に対する学習用データを用いて, データ欠損行動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, データ欠損行動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: 「右手を上にあげる」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>● 右手, を, あげる, 。, null4, null5, null6, null7, null8, null9, null10, null11, null12, null13, null14</li> <li>● 右手, を, あげる, 。, null5, null6, null7, null8, null9, null10, null11, null12, null13, null14, EOS</li> <li>● 右手, を, 上, に, あげる, 。, null4, null5, null6, null7, null8, null9, null10, null11, null12</li> </ul>
three-quarters	<ul style="list-style-type: none"> <li>● 右手, を, あげる, 。, null4, null5, null6, null7, null8, null9, null10, null11, null12, null13, null14</li> <li>● 右手, を, あげる, 。, null5, null6, null7, null8, null9, null10, null11, null12, null13, null14, EOS</li> <li>● 右手, を, 前, に, あげる, 。, null4, null5, null6, null7, null8, null9, null10, null11, null12</li> </ul>
half	<ul style="list-style-type: none"> <li>● 右手, を, 下げる, 。, null4, null5, null6, null7, null8, null9, null10, null11, null12, null13, null14</li> <li>● 右手, を, 下げる, 。, null5, null6, null7, null8, null9, null10, null11, null12, null13, null14, EOS</li> <li>● 右手, を, 上, に, あげる, 。, null4, null5, null6, null7, null8, null9, null10, null11, null12</li> </ul>
min	<ul style="list-style-type: none"> <li>● 左手, を, あげる, 。, null4, null5, null6, null7, null8, null9, null10, null11, null12, null13, null14</li> <li>● 左手, を, あげる, 。, null5, null6, null7, null8, null9, null10, null11, null12, null13, null14, EOS</li> <li>● 左手, を, 上, に, 上げる, 。, null4, null5, null6, null7, null8, null9, null10, null11, null12</li> </ul>

表 2: 評価結果

		データ 1				データ 2			
		full	three-quarters	half	min	full	three-quarters	half	min
BLEU	データ全動作	1.0	0.9235	0.8509	0.7036	1.0	0.8959	0.7489	0.6152
	欠損動作	1.0	0.960	0.7018	0.6513	1.0	0.9061	0.7649	0.6136
対数尤度	min, half, three-quarters 共通欠損動作	-49.98	-50.16	-52.24	-56.91	-59.82	-94.11	-96.50	-98.36
	half, three-quarters 共通欠損動作	-50.25	-50.67	-52.75	—	-59.74	-94.25	-96.60	—

● 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「右手を上にあげる」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 2 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 2 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の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 学習によって推定された言語モデルから生成された文の評価を行った結果について述べた。

今後の課題としては、より多様なデータ削減方法についての評価を行い、できるだけ性能を劣化させないような言語資源の転移学習手法を考案したいと考えている。

参考文献

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

[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.