2I5-OS-08b-6

トピックの組み合わせ構造を利用した言語モデルの転移学習

Transfer Learning of Language Models based on Combinatorial Structure of Topics

麻生 英樹 *1　小林 瑞季 *2　小林 一郎 *2
Hideki ASOH　Mizuki KOBAYASHI　Ichiro KOBAYASHI

*1独立行政法人産業技術総合研究所 知能システム研究部門
Intelligent Systems Research Institute, National Institute of Advanced Industrial Science and Technology

*2お茶の水女子大学大学院 人間文化創成科学研究科 理学専攻 情報科学コース
Advanced Sciences, Graduate School of Humanities and Sciences, Ochanomizu University

In this study, we propose a framework for transfer learning of n-gram language models. We assume that a language model of sentences which describes a complex phenomenon can be linearly decomposed into language models of elementary phenomena. Based on this assumption, we propose a simple method to learn language models of unseen target phenomena from sentences describing other phenomena which are composed of common elementary phenomena with the target phenomena. Experiments using sentences which describe several kinds of human motions demonstrate that the proposed method can learn language models of unseen motions.

1. はじめに

表現としての自然言語の最大の特徴の一つは，その豊かな生成性にある．すなわち，単語の組み合わせ構造を駆使して，実質的に無限な多様性を持つ事物を意味するような表現を生成し，理解することができる．この性質によって，私たちは，今までに見たことがないような複雑な事物についても，自然言語で記述することができる．

このことは，われわれの言語獲得過程に関する大きな疑問とも関係している．すなわち，われわれは，ごく限られた経験とそれに対応する言語表現という乏しい学習データだけから，ほぼ無限に近い表現を生成し，理解する能力を獲得しているように見えるが，それはどのようにして可能なのであろうか？

これはまた，言語の獲得という学習問題が，通常のパターン識別などの学習とは異なる次元を持つことを意味している．パターン識別の問題は，言語の獲得でいえば，個々の単語の意味の獲得に対応するが，文の意味はそれを構成する単語の意味の組み合わせであり，単語の数に比してはるかに多くの，原理的にはほぼ無限のバリエーションを持つ．したがって，個々の事物ごとに対応する文章を学習するという形態の学習は不可能であり，そこには何らかの追加的な仕掛けが必要である．

この問題に対して，これまでに様々なアプローチから研究が行われてきているが，本稿では，n-gram 確率の転移学習という観点から，学習データに含まれないような現象についての自然言語文の言語モデルを獲得する方法について述べる．具体的には，現象要素を共有するような複数の現象を記述する自然言語表現に対して，現象ごとの言語モデルを学習する問題を取り上げる．

ある現象には複数の現象要素が含まれていることが普通であるため，それらの現象を記述する自然言語文には，現象要素に対応する意味要素（トピック）が含まれる．このうち一部の現象要素は，他の現象にも含まれるため，その現象を記述する自然言語文にも，対応する意味要素が含まれるであろう．

連絡先: 麻生英樹，独立行政法人産業技術総合研究所 知能システム研究部門，〒 305-8568 つくば市梅園 1-1-1 中央第 2，h.asoh@aist.go.jp

このように，複数の現象が要素的な現象を共有することから，それぞれの現象を記述する自然言語表現の言語モデルは相互に関連していると考えられる．したがって，それらの複数の言語モデルの学習を，転移学習の枠組みで捉えることで，より効率のよい学習が可能になり，その結果として，これまでに見たことがない現象に対する自然言語文の言語モデルを獲得できると期待される．

以下では，簡単な現象に対して，現象要素が既知の場合に，そうした転移学習が可能であることを示す．まず 2 節では，本稿で扱う問題をもう少し形式的に述べる．3 節では関連研究について述べる．4 節では言語モデルに対する仮定と，それに基づいた具体的な転移学習の手法について述べる．5 節では人の動作を記録した動画を記述する文章群をデータとした簡単な実験の結果について述べる．6 節はまとめと今後の課題である．

2. 問題設定

複数の現象 $x_i(i = 1, \cdots, N)$ に対して，それぞれの現象を記述する自然言語文の集合 S_i が与えられているとする．ここでは，各 S_i に対する n-gram 言語モデル M_i を構築する問題を考える．こうした現象ごとの n-gram 言語モデルは，たとえば，各現象を記述するための自然言語文の生成や，各現象を記述する音声の認識などに利用することができる．

一般に，ある現象には複数の現象要素が含まれている．たとえば，ある人が右手を体の横を通して上にあげている場合，その現象は「右」「手」「上に」「横から」「あげる」といった現象要素から構成されている．そうした現象を表現する自然言語文としては，「右手を体の横を通して上にあげる」，「右側の手を横から上の方向にあげる」 などいろいろなものが考えられるが，そうした表現には，現象要素に対応する意味要素（トピック）が含まれると考えられる．そうした現象要素の一部は，他の現象，たとえば，ある人が右足を上にあげるような現象にも含まれるため，その現象を記述する自然言語文にも，対応する意味要素が含まれるであろう．

このように，複数の現象が現象要素を共有することから，それぞれの現象を記述する自然言語文はトピックを共有することになり，その言語モデルも相互に関連していると考えられる．

したがって，複数の言語モデル M_i を学習する問題は，相互に類似した，あるいは関連性のある学習課題を同時に学習するという問題であり，転移学習の枠組みで捉えることができる．そうすることで，それぞれの S_i から M_i を個別に学習するよりも，より少ない学習データから効率のよい学習を行うことが可能になる．

3. 関連研究

この研究の直接的なモチベーションになっているのは，杉田と谷らによる研究 [Sugita 05] である．そこでは，回帰型ニューラルネットワークを組み合わせた言語獲得過程のモデルが提案され，限られた学習用データを使って学習させることで，学習データに含まれない事象についても，言語表現と理解が可能になることが示されている．

複合的な現象を記述する自然言語文の生成についての研究はそれほど多くは行われていないが，近年，複雑な画像や動画を対象として，その内容を記述する自然言語文を生成するような研究が盛んになってきている．たとえば Barbu ら [Barbu 12] は，ショートビデオの説明文を生成するシステムを構築している．また，小林ら [Kobayashi 13] や Yu ら [Yu 13] は人などの動きを含むビデオとそれを記述する文章のペアから，相互の関係を学習するという問題を扱っている．

言語モデルのトピック適応は，音声認識や音声対話などの分野で研究が行われた．そこでは，新聞記事などの一般的な大規模コーパスから学習した言語モデルを，対象トピックや利用シーンに応じた少量の文章データを用いて適応させることがよく行われている．しかし，複数の相互に関連する現象を対象とした研究は我々の知る限りでは行われていない．

相互に類似した複数の学習課題を同時に学習させることにより，個別に学習するよりも学習性能を向上させるという問題は，転移学習，マルチタスク学習などと呼ばれて研究されてきている [神嶌 10]．

4. 転移学習の方法

4.1 n-gram の線形和の仮定

ここで提案する転移学習の方法は，複合現象を記述する自然言語文の言語モデル (n-gram) が，現象要素に対応する言語モデルの線形和になっているという単純な仮定に基づく．文章中の n 個の隣接する単語を $w_{i-n+1}, \cdots, w_i$ とするとき，n-gram 言語モデル M は，通常，条件付き確率 $p(w_i \mid w_{i-1}, \cdots, w_{i-n+1})$ で表されるが，以下では，条件付き確率の代わりに n 個の隣接する単語の同時確率 $p(w_i, w_{i-1}, \cdots, w_{i-n+1})$ をすべての単語組み合わせについて並べた確率ベクトルによって言語モデルを表す．この確率ベクトルの次元（n 単語の組み合わせの数）は W であるとする．

現象要素を y_k $(k = 1, \cdots, K)$ とし，現象 x_i が現象要素 y_k を含むか否かを表すインデックスを a_{ik} とする．このとき，$a_{ik} / \sum_k a_{ik}$ を要素とする $N \times K$ 行列を A とし，N 個の現象に対応する自然言語表現から得られた N 本の確率ベクトル $\psi_1, \ldots, \psi_N$ の転置を行として並べた $N \times W$ 行列を Ψ とする．さらに，K 個の現象要素に対応する言語モデルのベクトルを $\phi_1, \cdots, \phi_K$ として，その転置を行として並べた $K \times W$ 行列を Φ とする．このとき，

$$\Psi = A\Phi + \varepsilon \quad \varepsilon \sim N(0, \sigma^2)$$

が成り立つと仮定する．

現象要素の数 K が対象とする複合現象の数 N よりも小さいと仮定すると，このことは，現象を記述した自然言語文の集合から得られる行列 Ψ が，よりランクの低い行列 A と Φ の積の形に分解できるということを意味している．この関係を使うことで，Ψ の一部を使って Φ を推定し，Ψ の残りの部分を復元することなどが可能になる．これは，対応する自然言語文データが存在しない現象に対しても，現象要素を共有する他の現象に対する自然言語文データから，言語モデルが構築できるということである．

4.2 最小二乗推定

いま，行列 A が既知である，すなわち，各現象がどのような現象要素から構成されているかが既知であるとすると，前節で述べた仮定にもとづいて，Ψ から Φ を最小二乗法によって簡単に推定することができる．

$$\hat{\Phi} = \underset{\Phi}{\operatorname{argmin}} ||\Psi - A\Phi||^2 = A^+\Psi$$

ここで，A^+ は A の一般化逆行列である．

5. 実験

5.1 実験データ

実験に用いたデータは，小林らによる動画からの文章生成の研究 [Kobayashi 13] で構築されたものを用いた．そこでは，簡単な動作をする人を撮影した動画を被験者に見せて，動作を記述する自然言語文を収集している．実験に使われた動作は 20 種類であり，それぞれの動作は 図 1 に示すような 9 種類の現象要素の組み合わせと考えられる．言語モデルとしては 2-gram（順序を考慮した隣接単語ペアの出現確率ベクトル）を用いた．

5.2 実験手順

20 種類の動作のうち 1 つの動作に対する言語データを除外し，残りの 19 種類の動作に対する言語データから言語モデルを推定し Ψ を得る．図 1 の組み合わせ構造から得られる A と Ψ から，最小二乗法によって Φ を推定する．得られた現象要素の言語モデル Φ と，学習データから除外した動作に含まれる現象要素とから，除外した動作を記述する言語表現の言語モデルを推定する．

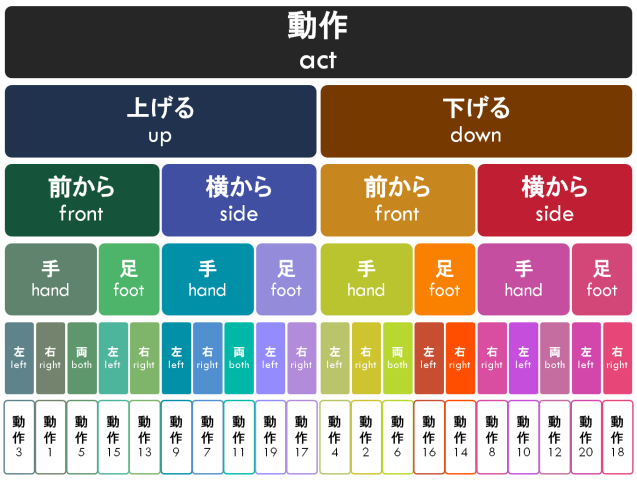

図 1: 実験に用いた動作に関する現象要素とその組み合わせ

表 1: 他の動作のデータのみを用いた場合の生成文の例

動作	生成文	尤度
1	• 右手, を, あげる, 。, null5, null6, null7, null8, null9, null10,	2.38e-31
	• 右手, を, あげる, 。, null4, null5, null6, null7, null8, null9,	1.66e-31
	• 右手, を, 上, に, あげる, 。, null5, null6, null7, null8,	1.69e-32
2	• 右手, を, 下げる, 。, null4, null5, null6, null7, null8, null9,	2.33e-31
	• 右手, を, 下げる, 。, null5, null6, null7, null8, null9, null10,	1.00e-31
	• 右手, を, 上, から, 下げる, 。, null4, null5, null6, null7,	6.85e-33
3	• 左手, を, 上, に, あげる, 。, null6, null7, null8, null9,	1.47e-31
	• 左手, を, あげる, 。, null5, null6, null7, null8, null9, null10,	9.02e-32
	• 左手, を, 上, に, あげる, 。, null5, null6, null7, null8,	1.48e-32
⋮	⋮	⋮
18	• 右足, を, 下げる, 。, null4, null5, null6, null7, null8, null9,	1.86e-32
	• 右足, を, 下げる, 。, null5, null6, null7, null8, null9, null10,	3.79e-33
	• 右足, を, 横, に, 下ろす, 。, null4, null5, null6, null7,	1.63e-33
19	• 左足, を, 横, に, あげる, 。, null6, null7, null8, null9,	5.92e-32
	• 左足, を, あげる, 。, null5, null6, null7, null8, null9, null10,	1.93e-32
	• 左足, を, 横, に, あげる, 。, null4, null5, null6, null7,	2.55e-33
20	• 左足, を, 下ろす, 。, null4, null5, null6, null7, null8, null9,	2.89e-32
	• 左足, を, 下ろす, 。, null5, null6, null7, null8, null9, null10,	6.53e-33
	• 左足, を, 横, に, 下ろす, 。, null4, null5, null6, null7,	6.43e-34

推定された n-gram 言語モデルを用いて，Viterbi アルゴリズムを用いて，尤度の高い文章を生成する（文の長さを補正するために null 項を導入している）．この手続きを, 除外する動作を変えながら，すべての動作に対して行った．さらに，学習用データから除外する動作の種類を増やしての評価も行った．

5.3 実験結果

1 つの動作に対するデータを除外して学習し，除外した動作について隣接単語ペア出現確率ベクトルの推定を行った場合，20 種類の動作に対する推定の二乗誤差の平均値は 0.0029 であった．

20 動作のうち 6 動作に対して，転移学習を用いて推定した言語モデルを用いて生成された文上位 3 件の例を表 2 に示した．これを見ると，おおむね適切な文章が生成できていることがわかる．すなわち，転移学習によって，これまで見たことのない動作に対する言語表現を学習することができている．

また，学習用データから除外する動作の種類を増やしても，推定誤差はそれほど増加しなかった．ただし，特定の現象要素が学習用データにまったく現れない場合には，推定が困難になるため，そうしたことが無いように除外する動作を選ぶ必要がある．たとえば，6 種類の動作を除外した場合でも，除外した動作に対して推定された言語モデルの二乗誤差の平均は 0.0037 程度であった．

6. まとめと今後の課題

本稿では，相互に関連する複数の現象を記述する自然言語文の言語モデル構築に，転移学習を適用する問題について述べた．非常に少数の現象要素の組み合わせから成る現象を対象とした実験では，提案手法が働くことを検証できた．しかしながら，今回用いた現象はかなり単純なものであり，前提とした仮定がより多くの複雑な現象に対して成り立つかどうかは今後よく検証してゆく必要がある．そして，線形和の仮定が成り立たない場合には，現象要素に対応する n-gram から，複合現象に対応する n-gram がどのように組み合わせ的に構成されるかについての新たな仮定を導入する必要があるだろう．

また，今回は，各対象現象について，現象要素とその組み合わせが既知の場合を扱ったが，今後の課題として，現象要素とその組み合わせもデータから推定することを検討している．具体的には，スパース行列分解や非負行列分解などの行列 Ψ を低ランク分解する手法を適用することになるが，単純に既存の分解法を適用したところではあまり良い結果は得られていない．これは，A の要素の値が 1/0 であることや，Φ の各行が確率ベクトルであることなどの制約が考慮できていないためと考えられる．そこで，そうした制約を導入した生成モデルによるベイズ的な行列分解の適用を検討しているところである．

謝辞：本研究の一部は 文部科学省科学研究費補助金 25120011 の助成を受けて行われた．

参考文献

[Barbu 12] Barbu, A. *et al.*: Video in sentences out, *Proceedings of Conference on Uncertainty in Artificial Intelligence (UAI)* (2012).

[神嶌 10] 神嶌 敏弘: 転移学習, 人工知能学会誌, vol.25, no.4, pp.572-580 (2010).

[Kobayashi 13] Kobayashi, M., Kobayash, I., Asoh, H., Guadarrama, S.: A probabilistic approach to text generation of human motions extracted from Kinect videos, *Proceedings of World Congress on Engineering and Computer Science 2013* (2013).

[Sugita 05] Sugita, Y. and Tani, J.: Learning semantic combinatoriality from the interaction between linguistic and behavioral processes, *Adaptive Behavior*, vol.13, no.1, pp.33-52 (2005) .

[Yu 13] Yu, H. and Siskind, J. M.: Grounded language learning from video described with sentences, *Proceedings of ACL 2013* (2013).