

# RNN 言語モデルのためのドメイン属性情報を用いた ゼロショット学習

芦川 博人<sup>1</sup> 森岡 幹<sup>1</sup> 俵 直弘<sup>1</sup> 小川 厚徳<sup>2</sup> 岩田 具治<sup>2</sup> 小川 哲司<sup>1</sup> 小林 哲則<sup>1</sup>

**概要:** 学習に用いたことのないドメインのデータに対しても、これまでに学習に用いたドメインに共通に含まれる属性情報を頼りに単語予測を行うリカレントニューラルネットワーク言語モデルを提案する。単語分布や文脈構造といった言語的な特性はドメインごとに異なるが、それらは、スタイルやトピックといった属性の組み合わせにより特徴づけられると仮定する。本研究では、それら属性に固有な特徴と属性に依らない特徴を分けて扱うことができるようリカレントニューラルネットワーク言語モデルを拡張することで、未知のドメインであってもどんな属性をもつかという事前情報を利用して信頼性の高い単語予測を行うことを試みる。スタイルやトピックの異なる複数ドメインのデータを用いた評価実験の結果、提案モデルとドメインの属性情報を用いることの有効性が示唆された。

キーワード: 言語モデル, リカレントニューラルネットワーク, ゼロショット学習, ドメイン適応

## Zero-shot learning using domain attributes for recurrent neural network language modeling

ASHIKAWA HIROTO<sup>1</sup> MORIOKA TSUYOSHI<sup>1</sup> TAWARA NAOHIRO<sup>1</sup> OGAWA ATSUNORI<sup>2</sup>  
IWATA TOMOHARU<sup>2</sup> OGAWA TETSUJI<sup>1</sup> KOBAYASHI TETSUNORI<sup>1</sup>

### 1. はじめに

ニューラルネットワーク言語モデルは、音声認識や機械翻訳の分野において欠かせない要素技術として近年盛んに研究されている。特に、リカレントニューラルネットワーク言語モデル (RNNLM) [1], [2] は、現在のニューラルネットワーク言語モデルの主流となっている。言語データは記述スタイルやトピックといったドメインの違いにより頻出語彙や言い回しは大きく異なるため、適用対象のドメイン (目標ドメイン) ごとにモデルを構築することが望ましい。しかし、目標ドメインごとに大規模な学習データを用意することはコストが高く、非現実的である場合が多い。それに対し本研究では、ドメインはスタイルやトピックとい

た属性 (ドメイン属性と呼ぶ) の組み合わせとして表現されるとし、これを陽に扱う RNNLM を提案する。それにより、目標ドメインのデータを未学習であっても、入力データのスタイルとトピックがわかっているならば、データの言語的特性を推測し単語予測に活かすことができると考えられる。本研究では、これを言語モデルにおけるゼロショット学習と呼ぶ。ゼロショット学習とは本来学習時に見たことのない、未知のラベルに対して推定を行うものであるが、本研究ではその考え方を拡張し、未知のドメインにおける単語分布や文脈構造に対して推定を行うものとする。

ゼロショット学習は元々画像認識の分野における課題設定である [3]。一般物体認識におけるゼロショット学習では、未知のクラスラベルを推定するためにセマンティックな属性情報を補助情報として用いる属性ベースのモデル [4], [5] が提案されており、その有効性が示されている。自然言語処理の分野においてもゼロショット学習は注目されている。異なるドメインにまたがって共通な質問タイプ

<sup>1</sup> 早稲田大学  
Waseda University

<sup>2</sup> NTT コミュニケーション科学基礎研究所  
NTT Communication Science Labs. NTT Corp.

や文章のシーケンス構造を考慮した自然言語理解のゼロショット学習 [6] や、ドメイン間で再利用可能なコンセプトを利用したスロットタギングにおけるゼロショット学習 [7] などが提案されている。我々が対象とする言語モデルのゼロショット学習においても、ドメイン間で共通の特性や補助情報を用いることが有効であると考えられる。

ドメイン適応の分野において、ドメイン固有の特性とドメイン共通の特性を分けて利用する枠組みの有効性が報告されている [8], [9], [10]。本研究ではこの枠組みを応用し、ドメイン属性固有の特性とドメイン属性に依存しない特性を分けて利用することを考える。具体的には、ドメイン固有、非固有の内部表現を保持する RNNLM [9] を参考に、RNNLM のネットワークにおいて単語表現と文脈表現にあたる部分にドメイン属性固有、非固有の表現を組み込む。

本稿の以降の構成は以下の通りである。まず 2 でリカレントニューラルネットワーク言語モデルについて説明し、3 においてドメイン属性情報の定義と、ドメイン属性情報を扱う RNNLM について述べる。続いて、4 で複数ドメインデータを用いた評価実験により、提案モデルの有効性とドメイン属性情報を用いることの有効性を調査する。最後に、5 において本稿のまとめと今後の課題を述べる。

## 2. リカレントニューラルネットワーク言語モデル

RNNLM [1] は、1 時刻前の潜在層の出力を入力層にフィードバックする構造を持っており、過去の文脈情報を伝播させることが可能である。語彙の大きさが  $V$  のとき、ある時刻  $t$  の単語は 1-of- $K$  表現で  $\mathbf{w}(t) \in \mathbb{R}^V$  として表される。行列  $\mathbf{L} \in \mathbb{R}^{E \times V}$  は単語  $\mathbf{w}(t)$  を低次元の連続ベクトル表現  $\mathbf{p}(t) \in \mathbb{R}^E$  に写像する。

$$\mathbf{p}(t) = \mathbf{L}\mathbf{w}(t) \quad (1)$$

また、潜在層の出力  $\mathbf{h}(t) \in \mathbb{R}^H$  は以下のように計算される。

$$\mathbf{h}(t) = f(\mathbf{p}(t), \mathbf{h}(t-1); \mathbf{U}, \mathbf{R}) \quad (2)$$

$H$  は潜在層のユニット数を表す。行列  $\mathbf{U} \in \mathbb{R}^{H \times E}$  は入力時の重みパラメータを表し、行列  $\mathbf{R} \in \mathbb{R}^{H \times H}$  は再帰時の重みパラメータを表す。 $f(\cdot)$  は活性化関数で、本研究では双曲線正接関数を用いる。文脈表現  $\mathbf{h}(t)$  が与えられたとき、RNNLM は次の時刻に出現する単語の確率分布  $\mathbf{y}(t) \in \mathbb{R}^V$  を出力する。

$$\mathbf{y}(t) = \text{softmax}(\mathbf{V}\mathbf{h}(t)) \quad (3)$$

$\mathbf{V} \in \mathbb{R}^{V \times H}$  は回帰の重みパラメータを表す。また、RNNLM の学習には truncated back propagation through time (BPTT) を用いる。

より長い文脈を捉えるために本研究では、Long short-term memory (LSTM) に基づく言語モデル (LSTM language model; LSTMLM) [2] を用いた。LSTMLM では、

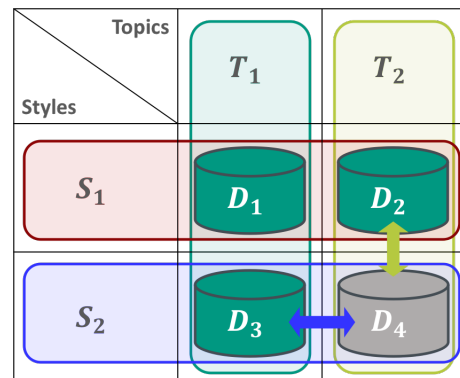


図 1 スタイルやトピックの異なる複数ドメインの例

RNNLM の潜在層の各ユニットが LSTM ユニットに置き換わる。LSTM ユニットはそれぞれ独立にメモリセル  $\mathbf{c}$  という記憶領域を保持しており、入力ゲート  $\mathbf{i}$ 、忘却ゲート  $\mathbf{f}$ 、出力ゲート  $\mathbf{o}$  という 3 つのゲートを用いることでメモリセルに保存する情報を制御する。

## 3. ドメイン属性情報の利用

### 3.1 ドメイン属性情報

テキストデータにおけるドメインの違いの例として、その文章が論文であるのか記事であるのかといった記述スタイルの違いや、何について書かれた文章であるのかといったトピックの違いが挙げられる。本研究では、ドメインごとに固有な言語的な特性は、スタイルとトピックの組み合わせによって特徴付けられると仮定する。このとき、目標ドメインのデータを学習に用いずとも、そのドメインがどんなスタイルでトピックが何であるかが既知であれば、ドメインに特徴的な単語分布や文脈構造を予測して単語の予測に活用することができると考えられる。このようなドメインの言語的特性を形作るスタイルやトピックのことを、本研究ではドメイン属性と呼ぶこととする。

図 1 に 2 種類のスタイル  $S_1, S_2$  と 2 種類のトピック  $T_1, T_2$  からなる 4 つのドメインの例を示す。ドメイン  $D_4$  を目標ドメインとしたとき、そのスタイルが  $S_2$  でトピックが  $T_2$  であるということが分かれば、 $S_2$  と  $T_2$  のそれぞれの固有の特徴をもとにドメイン  $D_4$  固有の言語的特性を予測することができる。このとき、目標ドメイン  $D_4$  のデータを直接学習に用いずとも、 $S_2$  固有の特徴はドメイン  $D_3$  のデータから、 $T_2$  固有の特徴はドメイン  $D_2$  のデータからそれぞれ学習されることが期待できる。

### 3.2 ドメイン属性情報を扱う RNNLM

ドメイン属性に特有の言語的特徴を RNNLM において陽に扱うために、入力データの属性に対応して活性化する属性固有のベクトル表現と入力の属性に依らず活性化する属性共通のベクトル表現を、内部表現として別々に持つモデルを提案する。図 2 の (a) に従来の RNNLM を、(b)

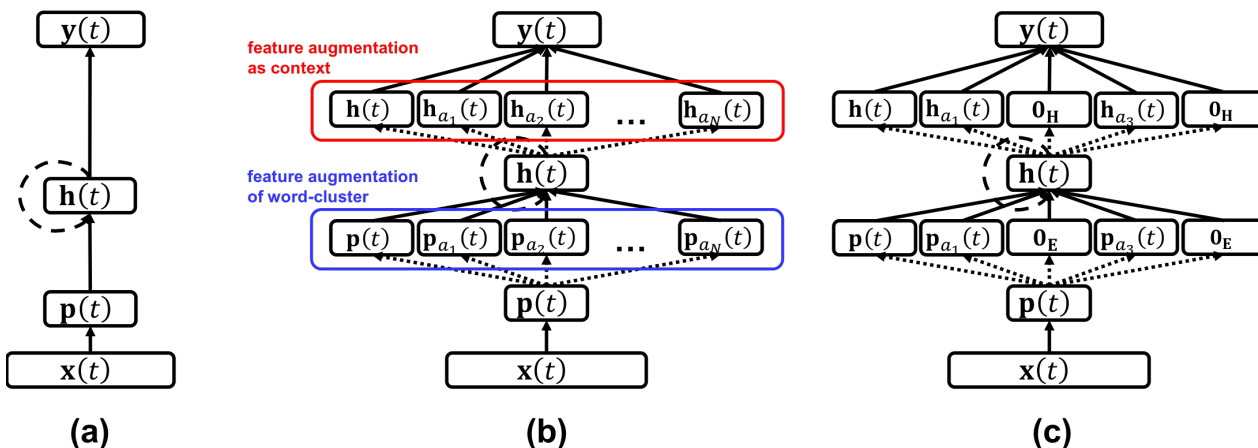


図 2 (a) 従来の RNNLM, (b) 属性共通と属性固有な内部表現を保持する RNNLM, (c)  $N = 4$  で入力データのドメイン属性が  $a_1$  と  $a_3$  のときのネットワークの内部状態

に属性共通, 固有の表現を組み込んだ RNNLM をそれぞれ示す. 破線部はフィードバック接続を表し, 点線は中間層の状態をコピーするか, もしくはゼロパディングすることを表す. 提案モデルでは, 単語のベクトル表現  $\mathbf{p}(t)$  (式 (1)) と文脈のベクトル表現  $\mathbf{h}(t)$  (式 (2)) に属性共通, 固有の表現を組み込む. つまり, 入力テキストが目標ドメインのスタイルやトピックにおいて特徴的な単語や言い回しであれば, ドメイン属性固有の内部表現が予測に寄与し, ドメイン属性に依らず頻出する単語や言い回し, もしくは未知の属性を持つ入力であれば, 属性共通の内部表現が予測に寄与する枠組みになっている.

本研究では, ドメインごとに表現を拡張するアプローチ [8] を RNNLM に適用した先行研究 [9] において提案された, ドメイン固有, 非固有の内部表現を保持する RNNLM を応用し, 複数のドメイン属性固有の情報に扱えるようにする. いま, 用いる属性の数を  $N$  とすると, 単語のベクトル表現  $\mathbf{p}(t)$  は以下のように  $N + 1$  倍の大きさのベクトル表現に拡張される.

$$\mathbf{p}'(t) = [\mathbf{p}(t)^T \mathbf{p}_{a_1}(t)^T \mathbf{p}_{a_2}(t)^T \cdots \mathbf{p}_{a_N}(t)^T]^T \quad (4)$$

上式において, 入力データのドメイン属性が  $a$  であるとき,  $\mathbf{p}_a(t)$  ( $a \in \{a_1, a_2, \dots, a_N\}$ ) は  $\mathbf{p}(t)$  と等しく, それ以外は  $\mathbf{0}_E$  となる. また, RNNLM のパラメータ  $\mathbf{U}$  は,  $\mathbf{U}' \in \mathbb{R}^{H \times (N+1)E}$  に拡張される. 同様に, 文脈のベクトル表現  $\mathbf{h}(t)$  は  $\mathbf{h}'(t)$  に拡張される.

$$\mathbf{h}'(t) = [\mathbf{h}(t)^T \mathbf{h}_{a_1}(t)^T \mathbf{h}_{a_2}(t)^T \cdots \mathbf{h}_{a_N}(t)^T]^T \quad (5)$$

このとき, 入力データのドメイン属性が  $a$  であるとき  $\mathbf{h}_a(t)$  ( $a \in \{a_1, a_2, \dots, a_N\}$ ) は  $\mathbf{h}(t)$  と等しく, それ以外は  $\mathbf{0}_H$  となる. パラメータ  $\mathbf{V}$  は,  $\mathbf{V}' \in \mathbb{R}^{V \times (N+1)E}$  に拡張される. 図 2 の (c) に,  $N = 4$  であり, 入力データのドメイン属性が  $a_1$  と  $a_3$  であるときのネットワークの状態を示す.

表 1 ICE の各ドメインの単語数と語彙数

Style	Topic	# of words	Vocab size
Academic Writing	Humanities	42107	8012
	Technology	41975	5562
Popular Writing	Humanities	42107	8109
	Technology	42107	7198

## 4. 評価実験

提案手法の有効性を調査するため, スタイルとトピックが異なる複数のドメインデータを用いてパープレキシティにより評価を行った. 比較するモデルは以下の 4 種類である.

- *LSTMLMs*: 目標ドメイン以外のデータを用いて学習した LSTMLM (従来モデル). 属性情報は用いない.
- *LSTMLMs w/ shared and specific repr. (in)*: LSTMLM の単語表現にドメイン属性共通, 固有の表現を組み込んだモデル
- *LSTMLMs w/ shared and specific repr. (out)*: LSTMLM の文脈表現にドメイン属性共通, 固有の表現を組み込んだモデル
- *LSTMLMs w/ shared and specific repr. (in, out)*: LSTMLM の単語表現と文脈表現の両方に, ドメイン属性共通, 固有の表現を組み込んだモデル

### 4.1 実験条件

実験データとして, International Corpus of English (ICE) [11] を用いた. ICE は世界各国の英語資源を収集したコーパスで, 話し言葉コーパスと書き言葉コーパスの両方が存在する. 本研究では, ICE に含まれるアメリカ (ICE-USA) とカナダ (ICE-CAN) の 2 カ国分のサブセットにおける書き言葉コーパスのうち, スタイルとトピックの組が, (Academic Writing, Humanities), (Aca-

表 2 穴抜きデータセットの各ドメインの単語数と語彙数

Style	Topic	# of words (training)	# of words (dev.)	# of words (test)	Vocab size
Academic	Humanities	112928	12395	42107	4922
Writing	Technology	113152	12303	41975	5125
Popular	Humanities	114685	12277	40468	4925
Writing	Technology	112361	12189	42880	4913

表 3 使用した LSTMMLM のパラメータ

# of projection layer units $E$	100, 200, 300
# of hidden layer units $H$	50, 100, 150
BPTT	4, 8, 16
initial learning rate	0.1
dropout rate	0.5
$l^2$ -regularization parameter	$1.0 \times 10^{-6}$

Academic Writing, Technology), (Popular Writing, Humanities), (Popular Writing, Technology) である 4 つのドメインを用いた。

各ドメインの単語数と語彙数を表 1 に示す。このデータをもとに、評価対象のドメインを学習に用いず、他のドメインのみで学習したモデルで評価を行うためのデータセット (穴抜きデータセットと呼ぶ) を作成した。学習データは評価対象ドメイン以外の 3 つのドメインのデータをあわせたものとし、そのうち 90% を訓練データ、10% を検証データとした。ドメインごとの訓練データ、検証データ、および評価データそれぞれの単語数と語彙数を表 2 に示す。また、各データセットで用いる語彙は、訓練データ中に現れた単語のうち、3 回以上出現した単語のみを用いた。

実験で使用した LSTMMLM のパラメータを表 3 に示す。学習率は初期値を 0.1 とし、検証データに対する対数尤度比が 1.003 を下回った場合に半減させた。また、正則化手法として潜在層への入力と潜在層からの出力のネットワークに対してドロップアウトを適用した。

## 4.2 実験結果

まず 4.2.1 において、属性情報を用いない従来のモデルとドメイン属性情報を扱う提案モデルを比較することで、属性情報の有効性を検証する。次に 4.2.2 において、提案モデルに対して与えるドメイン属性の組み合わせを変化させて評価を行うことで、提案モデルがドメイン属性固有の情報を扱っているのかどうかを検証する。最後に、4.2.3 では、実験に用いるドメイン属性をスタイルのみ、もしくはトピックのみに制限して提案モデルの学習、評価を追加で行い、ドメイン固有の言語的特性がドメイン属性の組み合わせにより表現されるという仮説について検証する。

### 4.2.1 実験 1: ドメイン属性情報を用いた RNNLM の有効性

ドメイン属性情報を扱う提案モデルの有効性を検証した。ドメイン属性情報を扱うモデルの訓練、検証、テストでは、

スタイルとして Academic Writing と Popular Writing, トピックとして Humanities と Technology の合計 4 種のドメイン属性を用いた。各モデルのテストセット・パープレキシティを表 4 に示す。この結果から、学習に用いることのできないデータに対しては 1 つのドメインを除き、属性情報を用いずただデータ全体で学習する従来の LSTMMLM (*LSTMMLMs*) よりも、単語表現にドメイン属性共通、固有の表現を組み込んだモデル (*LSTMMLMs w/ shared and specific repr. (in)*) が有効であることがわかる。しかし、文脈表現にドメイン属性共通、固有の表現を組み込むモデル (*LSTMMLMs w/ shared and specific repr. (out)*, *LSTMMLMs w/ shared and specific repr. (in, out)*) は、パープレキシティが悪化した。

Academic Writing の Humanities に関しては属性情報を与えても改善が見られなかった。この原因の一つとして、データ規模の小ささが挙げられる。特に語彙のカバレッジが不十分であったため、属性固有の性質を捉えることができなかった可能性がある。

### 4.2.2 実験 2: 与えるドメイン属性情報の組み合わせを変化させる

提案モデルが属性固有の情報を扱っているかを検証するために、様々なスタイルとトピックの組み合わせを属性情報として学習済みのモデルに与えて評価実験を行なった。4.2.1 において最も有効であった、単語表現にドメイン属性共通、固有の表現を組み込んだモデル (*LSTMMLMs w/ shared and specific repr. (in)*) に対して、与える属性の組み合わせを変化させたときのテストセットパープレキシティを表 5 に示す。

表 5 において、横軸は評価対象の目標ドメインを表し、縦軸は与える属性情報の組み合わせを表す。したがって、表の対角成分が評価対象のドメインに対応する正しい属性の組み合わせとなる。評価結果をみると、4.2.1 において提案モデルがベースラインを上回った 3 つのドメインに関しては、対角成分、すなわち正しいドメイン属性情報の組み合わせを与えたときに最も良い性能を見せた。このことから、提案モデルは属性ごとの性質の違いを捉え、単語予測に活用できていると言える。

逆に、4.2.1 においてベースラインを上回ることはできなかった Academic Writing の Humanities のドメインに対しては、(Academic Writing, Humanities) という正しい属性情報の組み合わせよりも、(Popular Writing, Human-

表 4 従来の LSTMLM とドメイン属性情報を扱う LSTMLM のテストセットパープレキシティ

Model	Feature augmentation		Academic Writing		Popular Writing	
	word repr.	context repr.	Humanities	Technology	Humanities	Technology
LSTMLMs			<b>265.61</b>	343.25	260.81	357.37
LSTMLMs w/ shared and specific repr. (in)	✓		279.86	<b>334.98</b>	<b>255.37</b>	<b>355.91</b>
LSTMLMs w/ shared and specific repr. (out)		✓	305.75	352.55	284.76	391.34
LSTMLMs w/ shared and specific repr. (in, out)	✓	✓	309.45	347.98	281.03	386.18

表 5 与える属性情報の組み合わせを変化させたときのテストセットパープレキシティ

Given Attributes		Target Domain			
Style	Topic	Academic Writing		Popular Writing	
		Humanities	Technology	Humanities	Technology
Academic Writing	Humanities	279.86	381.50	262.91	378.52
	Technology	371.55	<b>334.98</b>	373.26	429.63
Popular Writing	Humanities	<b>258.22</b>	385.37	<b>255.37</b>	371.89
	Technology	296.62	366.06	302.66	<b>355.91</b>

表 6 与える属性情報の種類を変化させたときのテストセットパープレキシティ

Model	Given Attributes		Academic Writing		Popular Writing	
	Style	Topic	Humanities	Technology	Humanities	Technology
LSTMLMs			265.61	343.25	260.81	357.37
LSTMLMs w/ shared and specific repr. (in)	✓		368.98	371.95	297.22	364.45
LSTMLMs w/ shared and specific repr. (in)		✓	<b>259.23</b>	356.61	262.64	456.99
LSTMLMs w/ shared and specific repr. (in)	✓	✓	279.86	<b>334.98</b>	<b>255.37</b>	<b>355.91</b>

ities) という組み合わせを与えたほうがパープレキシティを改善し、ベースラインである従来の LSTMLM のスコア (265.61) を上回る結果となった。これは提案モデルが、Academic Writing の Humanities に対して有効であるはずの属性情報 Academic Writing と Humanities をうまく扱うことができなかつたことを示唆しており、これにより表 4 においてもベースラインを越すことができなかつたと考えられる。

#### 4.2.3 実験 3: 与える属性情報の種類を変化させる

ドメイン属性情報の組み合わせによりドメイン固有の言語的特性が表現されるという仮説を検証するために、用いるドメイン属性情報をスタイルのみ、もしくはトピックのみに制限して学習、評価を行った。ドメイン属性情報を扱うモデルとしては、4.2.1 において最も有効であった単語表現に属性共通、固有の表現を組み込んだモデル (LSTMLMs w/ shared and specific repr. (in)) を用いた。

表 6 に、モデルに与える属性情報の種類を制限したときと全て用いたときのテストセットパープレキシティを示す。4 つのうち 3 つのドメインにおいてスタイルのみ、もしくはトピックのみを属性情報として与えた場合に比べ、スタイルとトピックの組み合わせの情報を与えたほうが良いスコアを示していることがわかる。このことから、学習に用いたことのない目標ドメインに対しては、ドメイン属性の組み合わせの情報を与えることが有効であることがわかる。ただし、Academic Writing の Humanities のドメインに対

しては、トピックのみの属性を与えることが最も効果的であり、次にスタイルのみ、最後にスタイルとトピックの組み合わせとなっている。この原因としては、4.2.2 において述べたように、今回、Academic Writing の Humanities のドメインを穴抜きにした際に有用な属性情報をうまく扱うことができなかつたためではないかと考えられる。

## 5. まとめと今後の課題

本研究では、学習に用いたことのないドメインのデータに対しても、これまでに学習に用いたドメインと共通する属性情報を頼りに単語分布や文脈構造を推定して単語予測を行う RNNLM を提案した。文章/テキストのスタイルとトピックをドメイン属性として、ドメインの言語的特性はドメイン属性の組み合わせにより表現されると仮定し、ドメイン属性ごとの内部状態を分けて保持できるように RNNLM を拡張した。具体的には、RNNLM における単語表現と文脈表現に対して属性共通、固有の表現を組み込み、入力データのドメイン属性に応じて伝播経路が変化するようにした。評価実験の結果、1 つのドメインを除く 3 つのドメインにおいては、単語表現に対してドメイン属性共通、固有の表現を組み込んだモデルがパープレキシティを改善した。さらに、ドメイン属性情報はスタイルのみ、もしくはトピックのみではなくスタイルとトピックの組み合わせとして用いることの有用性が明らかとなった。今後は、より効率的なドメイン属性情報の扱い方やモデル構造

について検討する予定である。

#### 参考文献

- [1] Mikolov, T., Karafiát, M., Burget, L., Cernocký, J. and Khudanpur, S.: Recurrent neural network based language model, *Proc. INTERSPEECH 2010*, pp. 1045–1048 (2010).
- [2] Sundermeyer, M., Schlüter, R. and Ney, H.: LSTM Neural Networks for Language Modeling, *Proc. INTERSPEECH 2012*, pp. 194–197 (2012).
- [3] Larochelle, H., Erhan, D. and Bengio, Y.: Zero-data Learning of New Tasks, *Proc AAAI 2008*, pp. 646–651 (2008).
- [4] Lampert, C. H., Nickisch, H. and Harmeling, S.: Learning to detect unseen object classes by between-class attribute transfer, *Proc. CVPR 2009*, pp. 951–958 (2009).
- [5] Lampert, C. H., Nickisch, H. and Harmeling, S.: Attribute-Based Classification for Zero-Shot Visual Object Categorization, *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 36, No. 3, pp. 453–465 (2014).
- [6] Sadamitsu, K., Homma, Y., Higashinaka, R. and Matsuo, Y.: Zero-Shot Learning for Natural Language Understanding Using Domain-Independent Sequential Structure and Question Types, *Proc. INTERSPEECH 2017*, pp. 3306–3310 (2017).
- [7] Lee, S. and Jha, R.: Zero-Shot Adaptive Transfer for Conversational Language Understanding, *CoRR*, Vol. abs/1808.10059 (2018).
- [8] Daumé III, H.: Frustratingly Easy Domain Adaptation, *Proc. ACL 2007* (2007).
- [9] Morioka, T., Tawara, N., Ogawa, T., Ogawa, A., Iwata, T. and Kobayashi, T.: Language Model Domain Adaptation Via Recurrent Neural Networks with Domain-Shared and Domain-Specific Representations, *Proc. ICASSP 2018*, pp. 6084–6088 (2018).
- [10] Bousmalis, K., Trigeorgis, G., Silberman, N., Krishnan, D. and Erhan, D.: Domain Separation Networks, *Proc. NIPS 2016*, pp. 343–351 (2016).
- [11] Greenbaum, S. and Nelson, G.: The international corpus of English (ICE) project, *World Englishes*, Vol. 15, No. 1, pp. 3–15 (1996).