多言語対話における応答先と応答内容の選択

佐藤 元紀 [1]　大内 啓樹 [1,2]　坪井祐太 [3*]
[1] 奈良先端科学技術大学院大学 情報科学研究科
[2] 理化学研究所　革新知能統合研究センター AIP
[3] 株式会社 Preferred Networks

{sato.motoki.sa7, ouchi.hiroki.nt6}@is.naist.jp
tsuboi@preferred.jp

1 はじめに

対話生成アプローチの 1 つとして，検索型アプローチ (Retrieval-based Approach) が人気を博している [11, 8, 4, 12]．このアプローチは，相手の発話 $\boldsymbol{x}$ を入力として受け取り，リポジトリから複数の応答候補 $\mathcal{R} = \{\boldsymbol{r}_i\}_1^n$ を取り出し，ランキングモデル f に基づいて最もスコアの高い応答候補 $\hat{\boldsymbol{r}} = \underset{\boldsymbol{r}_i \in \mathcal{R}}{\operatorname{argmax}} f(\boldsymbol{x}, \boldsymbol{r}_i)$ を返す．

適切な応答を返すためにはランキングモデルが重要な役割を担うため，その部分のみに焦点を当てた応答選択タスクが注目されている [7]．このタスクは発話者 2 人の対話を想定していたが，後に複数人対話に拡張され，適切な応答だけでなく応答を返す相手 (Addressee) も予測する [10]. これらのタスクにおける高性能なモデルはニューラルネットワークに基づいており，パラメータの学習のために多くの学習データを必要とする．しかし，十分な量の対話データを入手することは困難であり，既存研究も英語や中国語のような資源が豊かな言語のみを対象としている．[1]

本研究では，**多言語応答選択**に取り組む．具体的には，以下の 3 点を行う．

- 多言語応答選択タスクの定式化
- 多言語に対応可能なモデルの提案
- 多言語対話データセットの作成・公開 [2]

まず，新たなタスクとして，多言語複数人対話における応答先 (Addressee) および応答内容 (Response) 予測タスクの定式化を行う．このタスクでは，高資源言語と低資源言語が複数与えられ，応答先・応答内容の予測を単一のモデルで行う．複数言語の応答選択を高精度で行うには，低資源言語データを補うための高資源言語データの利用法が鍵となる．そこで我々は，高資源言語の知識を転移し，低資源言語の応答も高精度で予測可能な多言語対話モデルの学習法を提案する．また，学習したモデルの性能評価のため，5 言語 (英語，イタリア語，クロアチア語，ポルトガル語，ドイツ語) からなる多言語対話データセットを作成し，公開する．

2 単言語対話における応答選択

先行研究 [10] で提案された単言語における応答先・応答内容選択 (Adressee and Response Selection; ARS) について説明する．入力 $\boldsymbol{x}$ は，応答するユーザー a_{res}，発話履歴 $\mathcal{C}$，応答候補集合 $\mathcal{R}$ とする．出力として適切な応答先 a と応答内容 $\boldsymbol{r}$ を選択する．

$$\text{入力} : \boldsymbol{x} = (a_{\text{res}},\, \mathcal{C},\, \mathcal{R})$$
$$\text{出力} : \boldsymbol{y} = (a, \boldsymbol{r})$$

応答先 a を予測するために，発話履歴に含まれるユーザー集合 $\mathcal{A}(\mathcal{C})$ から適切なユーザーを選択する．応答内容 $\boldsymbol{r}$ を予測するために，応答内容候補集合 $\mathcal{R} = \{\boldsymbol{r}_1, \cdots, \boldsymbol{r}_{|\mathcal{R}|}\}$ から選択する．

評価指標として，3 つの正解率 (ADR-RES, ADR, RES) を用いる：(1) 応答先と応答内容の両方 (ADR-RES)，(2) 応答先のみ (ADR)，(3) 応答内容のみ (RES) の正解率で評価する．

3 多言語対話における応答選択

本研究では，多言語の発話に対して，適切な応答選択を行うマルチリンガルモデル (Multilingual Model) を構築することを目的とする．

---

*この研究の一部は日本 IBM 株式会社在籍時に実施した．

[1]本稿では，資源が豊かな言語 (High-resource) を「高資源言語」，資源が乏しい言語 (low-resource) を「低資源言語」と呼ぶ．

[2]データセットのダウンロード: http://sato-motoki.com/projects/

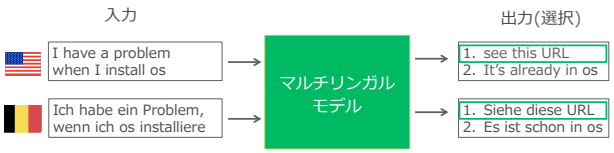

図 1: マルチリンガルによる多言語対話の例.

図 1 に多言語対話における応答選択の例を示す．この例で，マルチリンガルモデルは，英語とドイツ語どちらの言語からも入力を受け付け，それぞれの言語に適切な応答を選択している．ここで注目すべき特徴は，各言語に別個のモデルを用意するのではなく，単一のモデルで複数の言語に対応する点である．このモデルの構築のためには，高資源言語 (英語) の学習データを利用して低資源言語 (ドイツ語) を補う必要がある．

以下に多言語応答選択のタスク定義を示す．

学習

対象とする言語集合 $\mathcal{K}$ の各言語 k に対して学習データが与えられているとする:

$$\mathcal{D}^{(k)}_{\text{train}} = \{\, \boldsymbol{x}^{(k)}_i, \boldsymbol{y}^{(k)}_i \,\}^{N^{(k)}}_1, \quad k \in \mathcal{K}$$
$$\mathcal{D}_{\text{train}} = \bigcup_k \mathcal{D}^{(k)}_{\text{train}}$$

これらを用いて，モデル $\mathcal{F} : \mathcal{X} \to \mathcal{Y}$ を訓練する．

評価

対象とする全言語におけるマクロ平均を用いる:

$$\text{ADR-RES} = \frac{\sum_k \text{ADR-RES}^{(k)}}{|\mathcal{K}|}$$

ADR と RES も同様に計算し，評価する．

4 提案手法

モデル $\mathcal{F}$ は特徴抽出関数 f^E と 2 つの関数 f^A と f^R から構成される．これらの 2 つの関数は応答先，応答内容に対してスコアを返す:

$$f^A(\boldsymbol{x}, a_i) = \sigma([\mathbf{a}_{\text{res}}, \mathbf{h}_c]^{\mathrm{T}}\, \mathbf{W}_a\, \mathbf{a}_i) \quad (1)$$
$$f^R(\boldsymbol{x}, \boldsymbol{r}_j) = \sigma([\mathbf{a}_{\text{res}}, \mathbf{h}_c]^{\mathrm{T}}\, \mathbf{W}_r\, \mathbf{r}_j) \quad (2)$$

$\mathbf{a}_{\text{res}}, \mathbf{h}_c, \mathbf{a}_i, \mathbf{r}_j$ は特徴抽出関数 f^E によってエンコードされたベクトル[3]である．モデル $\mathcal{F}$ のパラメータは $\theta = \{\theta_E \cup \{\mathbf{W}_a, \mathbf{W}_r\}\}$ であり，f^E のパラメータは θ_E と表記している．

[3]本研究では先行研究 [10] で提案された Dynamic Model を用いる．モデルの詳細な説明は論文 [10] を参照.

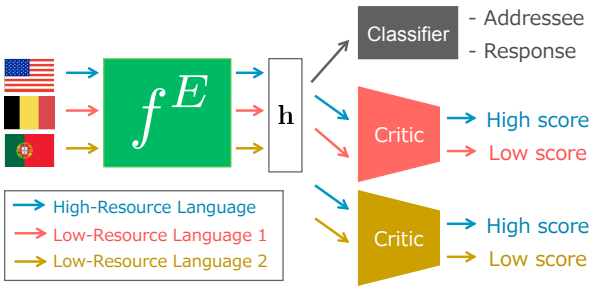

図 2: 多言語の分布を近づけるよう拡張した W-GAN.

モデルの学習のために 4 つの手法を提案する．これらの手法は，高資源言語 $\mathcal{S} \subseteq \mathcal{K}$ と低資源言語 $\mathcal{T} = \bar{\mathcal{S}}$ からなる言語集合 $\mathcal{K}$ の訓練データを用いる．

(a) 多言語単語ベクトル

モデルを高資源言語 $\mathcal{D}^{(s)}_{\text{train}}$ $(s \in \mathcal{S})$ で訓練し，低資源言語 $\bar{\mathcal{S}}$ での予測にも利用することを考える．入力言語が異なるため，多言語でベクトル空間が共通化して埋め込まれている多言語単語ベクトルを用いる．本研究では，MultiCCA[1][4]を用いた．訓練時は，高資源言語データを用いて，単語ベクトルは更新せずにモデルパラメータのみを学習する．評価時は，訓練時に用いた単語ベクトルを対象となる低資源言語の単語ベクトルに置き換えて，モデルを利用する．

(b) ファインチューニング

まず，モデルパラメータ θ を高資源言語で訓練しておく (Pre-training)．その後，訓練済みのパラメータ θ を初期値として低資源言語で再訓練 (Fine-tuning) する．

(c) 同時学習

ファインチューニングを行うと，以前に学習したタスクの知識を忘却してしまう破滅的忘却 [5] と呼ばれる現象が起きることが知られている．これを避けるため，対象となる全言語のデータを同時的に用いてパラメータの再訓練を行う．具体的には，対象言語集合 $\mathcal{K}$ が与えられ，以下の誤差関数を最適化する:

$$\mathcal{J}_{\text{joint}}(\theta) = \sum_k \mathcal{J}(\mathcal{D}^{(k)}, \theta) \quad (3)$$

誤差関数 $\mathcal{J}$ として交差エントロピーを用いた．

(d) 多言語敵対学習

言語に依存しない特徴量を得るには，複数言語間の特徴量の分布を近づける必要がある．そのため，

[4]訓練済みベクトル: http://128.2.220.95/multilingual/data/

Wasserstein-GAN (W-GAN) [2] を用いて，言語間の特徴量分布を適合させる．

図 2 に例を示す．英語を高資源言語 $s \in \mathcal{S}$，ドイツ語・ポルトガル語を低資源言語 $t \in \mathcal{T}$ とする．各言語に対して，特徴抽出関数 f^E は $\boldsymbol{x}$ を入力とし，特徴量ベクトル $\mathbf{h} = f^E(\boldsymbol{x})$ を計算する[5]．f^E を用いることで，高資源言語のベクトル $\mathbf{h}^{(s)}$ と低資源言語のベクトル $\mathbf{h}^{(t)}$ を得る．2 つのベクトル $\mathbf{h}^{(s)}$，$\mathbf{h}^{(t)}$ の分布 p が近くなるようにクリティック g_π によって Wasserstein 距離を最小化する．

$$\mathcal{W}(\, p(\mathbf{h}^{(s)}),\, p(\mathbf{h}^{(t)})\,) = \max_\pi\ \mathbb{E}_{\mathbf{h}^{(s)}\sim p(\mathbf{h}^{(s)})}[g_\pi(\mathbf{h}^{(s)})] - \mathbb{E}_{\mathbf{h}^{(t)}\sim p(\mathbf{h}^{(t)})}[g_\pi(\mathbf{h}^{(t)})] \quad (4)$$

g_π には多層パーセプトロンを用いた (π は g のパラメータ)．

式 4 は 2 つの分布のみを扱っているが，本研究では複数の分布を扱うため，以下のように W-GAN を一般化した:

$$\mathcal{J}_{\text{wgan}}(\theta) = \sum_{s\in\mathcal{S}}\sum_{t\in\mathcal{T}} \mathcal{W}(\, p(\mathbf{h}^{(k)}),\, p(\mathbf{h}^{(\ell)})\,)$$

$\mathcal{J}_{\text{wgan}}$ は同時学習の誤差関数 (式 3) に足し合わされる:

$$\mathcal{J}_{\text{adv}}(\theta) = \mathcal{J}_{\text{joint}}(\theta) + \lambda\, \mathcal{J}_{\text{wgan}}(\theta) \quad (5)$$

上式 5 を最適化することにより，言語非依存のパラメータを学習する．[6]

5 実験

5.1 データセット

多言語対話のデータセットを構築した．Ubuntu IRC Logs[7]から対話ログを収集し，前処理 (単語分割・データクリーニングなど) を行った．言語判別には言語検出器 [9] を用いた．データセットはランダムに学習 (90%)・開発 (5%)・評価 (5%) データに分割した．表 2 にデータセットの統計を示す．

5.2 実験設定

英語 (En)，イタリア語 (It)，クロアチア語 (Hr)，ポルトガル語 (Pt)，ドイツ語 (De) を用いた．英語を高資源言語，それ以外の言語を低資源言語として扱った．評価指標として，3 節の 3 つの正解率 (ADR-RES・ARD・RES) を用いた．応答内容の候補数 ($|\mathcal{R}|$) は 2 と 10 を用いた．

比較手法として以下の手法を試した:

- ENONLY: 英語のみで訓練したモデル．ただし，低資源言語に適用する際は多言語単語ベクトルを使用 (a).
- FINETUNE: 英語で Pre-training をしたモデルを低資源言語でファインチューニング (b).
- JOINT: 全言語で同時学習 (c).
- WGAN: W-GAN を用いて同時学習 (d).

	Dataset		
Language	Train	Dev	Test
English (en)	678.7 k	38.6 k	44.7 k
Italian (it)	38,511	2,561	3,873
Croatian (hr)	11,387	512	1,145
German (de)	5,500	354	569
Portuguese (pt)	5,951	285	975

表 1: 本研究で構築した多言語対話データセット．

5.3 実験結果

提案手法の効果

表 2 に実験結果[8]を示す．ほとんどの評価指標で，敵対学習 W-GAN を用いた手法が最も高い正解率を得ている．これは，W-GAN によって言語非依存の表現が獲得され，複数言語で高いパフォーマンスを発揮できたからだと解釈できる．対照的に，ファインチューニングを用いると，Pre-training で学習した高資源言語 (英語) のパフォーマンスを下げる傾向があり，破滅的忘却に陥る結果となった．5 言語と対象とした設定でも同様に，W-GAN が高いパフォーマンスを記録している．したがって，W-GAN を用いることにより，多言語対話に対応可能なマルチリンガルモデルを学習可能であることがわかった．

NMT を用いた Data Augumentation の効果

高資源言語を低資源言語に翻訳して擬似的な訓練データとする方法も考えられる．そこでニューラル機械翻訳 (NMT) による Data Augumentation の効果を調査する (表 3)．実験に使用した NMT モデルは OpenNMT[6] の学習済みのモデルを用いた．英語 → ドイツ語の翻

[5]特徴量ベクトル $\mathbf{h}$ は $\mathbf{a}_{\text{res}}$(式 1) と $\mathbf{h}_c$(式 2) を結合したものである: $\mathbf{h} = [\mathbf{a}_{\text{res}}, \mathbf{h}_c]$.

[6]ハイパーパラメータ $\lambda = 0.5$ を用いた．

[7]http://irclogs.ubuntu.com/

[8]ADR-RES の列の丸括弧内の数字は，評価に用いた各言語の ADR-RES を表す．

評価データ	手法	$\|\mathcal{R}\| = 2$			$\|\mathcal{R}\| = 10$		
		ADR-RES	ADR	RES	ADR-RES	ADR	RES
En, It	ENONLY	47.77 (50.43, 45.11)	69.66	68.12	17.93 (24.86, 11.00)	69.62	24.85
	FINETUNE	56.27 (46.91, 65.63)	73.44	75.02	25.21 (16.54, 33.88)	71.58	32.80
	JOINT	57.77 (50.13, 65.40)	73.73	76.94	29.70 (25.39, 34.00)	74.06	37.57
	WGAN	**58.60** (**50.95**, **66.25**)	**74.01**	**77.61**	**30.06** (**25.90**, **34.21**)	**74.19**	37.89
En, Hr	ENONLY	41.53 (50.43, 32.62)	66.06	62.56	16.10 (24.86, 7.34)	63.95	22.45
	FINETUNE	44.11 (46.03, 42.18)	64.27	67.65	17.73 (20.87, 14.59)	64.27	26.17
	JOINT	46.82 (**50.85**, 42.79)	65.17	70.55	21.15 (**25.27**, 17.03)	**65.87**	29.93
	WGAN	**47.42** (50.04, **44.80**)	**65.30**	**70.75**	**21.24** (25.10, **17.38**)	65.13	**30.26**
En, De	ENONLY	44.46 (**50.43**, 38.49)	66.33	66.49	16.39 (**24.86**, 7.91)	66.60	23.45
	FINETUNE	49.24 (47.15, 51.32)	67.30	71.66	22.94 (22.67, 23.20)	68.18	30.85
	JOINT	51.35 (50.33, 52.37)	68.71	71.62	23.66 (24.65, 22.67)	68.62	31.44
	WGAN	**52.59** (50.00, **55.18**)	**69.47**	**73.26**	**24.33** (24.41, **24.25**)	**69.38**	**32.21**
En, Pt	ENONLY	43.73 (**50.43**, 37.03)	**68.05**	63.75	16.64 (24.86, 8.41)	**68.31**	22.83
	FINETUNE	47.54 (47.99, 47.08)	67.81	69.18	18.35 (20.90, 15.79)	68.02	26.02
	JOINT	47.31 (49.07, 45.54)	66.73	71.04	21.06 (**25.39**, 16.72)	66.83	30.08
	WGAN	**48.94** (50.18, **47.69**)	67.96	**71.44**	**21.86** (24.13, **19.59**)	67.91	**30.40**
平均	ENONLY	44.37 (**50.43**, 38.31)	67.53	65.23	16.77 (24.86, 8.67)	67.12	23.40
	FINETUNE	49.29 (47.02, 51.55)	68.21	70.88	21.06 (20.25, 21.87)	68.01	28.96
	JOINT	50.81 (50.10, 51.53)	68.59	72.54	23.89 (**25.18**, 22.61)	68.85	32.26
	WGAN	**51.89** (50.29, **53.48**)	**69.19**	**73.27**	**24.37** (24.89, **23.86**)	**69.15**	**32.69**
En, It, Hr, De, Pt	ENONLY	40.74 (**50.43**, 45.11, 32.62, 38.49, 37.03)	67.67	60.11	11.90 (**24.86**, 11.00, 7.34, 7.91, 8.41)	67.02	16.99
	JOINT	49.81 (49.36, 61.79, 42.53, 49.21, 46.15)	68.98	70.99	22.34 (24.48, **30.21**, **16.42**, 23.37, 17.23)	69.53	30.21
	WGAN	**50.90** (49.15, **62.33**, **43.67**, **50.62**, **48.72**)	**69.67**	**71.21**	**22.55** (23.77, 29.67, 15.55, **24.08**, **19.69**)	**69.96**	**30.52**

表 2: 多言語応答選択の実験結果.

評価データ	手法	学習データ	$\|\mathcal{R}\| = 2$			$\|\mathcal{R}\| = 10$		
			ADR-RES	ADR	RES	ADR-RES	ADR	RES
En, De	JOINT	En + De	51.35 (50.33, 52.37)	68.71	71.62	23.66 (24.65, 22.67)	68.62	31.44
		En + De + De′	51.11 (50.74, 51.49)	68.79	71.14	24.11 (25.02, 23.20)	69.13	31.96
	WGAN	En + De	**52.59** (50.00, 55.18)	**69.47**	**73.26**	**24.33** (24.41, 24.25)	**69.38**	**32.21**
		En + De + De′	51.21 (50.76, 51.67)	68.98	71.13	24.19 (25.01, 23.37)	69.03	32.01

表 3: ニューラル機械翻訳による Data Augumentation の効果.

訳モデルを用いて，英語の学習データをドイツ語に翻訳した．翻訳したドイツ語データを De′ と示す．結果として，Data Augumentation は JOINT 及び WGAN で効果はなかった．これは，先行研究 [3] における「Data Augumentation と敵対学習を組み合わせても性能が向上しなかった」との報告と一致する．

6 おわりに

本研究では，多言語応答選択タスクを定式化し，多言語に対応可能なマルチリンガルモデルの学習法を提案した．評価実験の結果，特に W-GAN を拡張した提案手法が多言語に有効に対応可能であることがわかった．我々の手法とデータセットは応答生成にも適用可能であるため，多言語応答生成タスクに取り組むことが今後の課題である．

参考文献

[1] Waleed Ammar, George Mulcaire, Yulia Tsvetkov, Guillaume Lample, Chris Dyer, and Noah A. Smith. Massively multilingual word embeddings. *CoRR*, 2016.

[2] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein gan. *arXiv preprint arXiv:1701.07875*, 2017.

[3] Konstantinos Bousmalis, Alex Irpan, Paul Wohlhart, Yunfei Bai, Matthew Kelcey, Mrinal Kalakrishnan, Laura Downs, Julian Ibarz, Peter Pastor, Kurt Konolige, et al. Using simulation and domain adaptation to improve efficiency of deep robotic grasping. *arXiv preprint arXiv:1709.07857*, 2017.

[4] Zongcheng Ji, Zhengdong Lu, and Hang Li. An information retrieval approach to short text conversation. *arXiv preprint arXiv: 1408.6988*, 2014.

[5] James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, et al. Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences of the United States of America*, pages 3521–3526, 2017.

[6] G. Klein, Y. Kim, Y. Deng, J. Senellart, and A. M. Rush. OpenNMT: Open-Source Toolkit for Neural Machine Translation. *ArXiv e-prints*.

[7] Ryan Lowe, Nissan Pow, Iulian V. Serban, and Joelle Pineau. The ubuntu dialogue corpus: A large dataset for research in unstructured multi-turn dialogue systems. In *Proceedings of SIGDIAL*, pages 285–294, 2015.

[8] Zhengdong Lu and Hang Li. A deep architecture for matching short texts. In *Proceedings of NIPS*, pages 1367–1375, 2013.

[9] Shuyo Nakatani. Language detection library for java, 2010.

[10] Hiroki Ouchi and Yuta Tsuboi. Addressee and response selection for multi-party conversation. In *Proceedings of EMNLP*, pages 2133–2143, 2016.

[11] Hao Wang, Zhengdong Lu, Hang Li, and Enhong Chen. A dataset for research on short-text conversations. In *Proceedings of EMNLP*, pages 935–945, 2013.

[12] Mingxuan Wang, Zhengdong Lu, Hang Li, and Qun Liu. Syntax-based deep matching of short texts. In *Proceedings of IJCAI*, pages 1354–1361, 2015.