1B1-OS-25a-1

非明示的な発話状況を考慮したニューラル対話モデルの検討

A Neural Dialogue Model Considering Implicit Utterance Situations

佐藤 翔悦 *1 Shoetsu Sato　吉永 直樹 *2 Naoki Yoshinaga　豊田 正史 *2 Masashi Toyoda　喜連川 優 *2,*3 Masaru Kitsuregawa

*1東京大学大学院 情報理工学系研究科
Graduate School of Information Science and Technology, The University of Tokyo

*2東京大学 生産技術研究所
Institute of Industrial Science, The University of Tokyo

*3国立情報学研究所
National Institute of Informatics

We make utterances under various situations depending on time, place, topic, the relation to the listener, and the like. However, it is difficult to take into consideration the utterance situations in a dialogue system, and few data-driven dialogue systems explicitly exploit these situations. We propose a neural dialogue model that integrates a global dialogue model learned from the entire dialogue data and local dialogue models learned from dialogue data under specific utterance situations. In the experiment, we verify the effectiveness of the proposed method by response selection test using dialogue data we have collected from Twitter and investigate how the proposed utterance situations influence the response.

1. はじめに

我々が他人と会話を行う際，その応答には相手の発話だけではなく時間や場所といった様々な発話状況が大きく影響する．例えば「どこかに出かけたいね」という発言に対し，夏であれば「海で泳ごう」，冬であれば「スキーに行きたいな」といったように，会話を行う時節によってその応答は異なる．また，同じ発言に対しても応答者がどのような性格・好みを持つかによって応答は違ったものとなるだろう．このように，発言の背後にある様々な要因によって実際に行われる応答は変化する．

従来のタスク指向型対話システムにおいてはそれぞれの発話状況に対するパターンを作り込むことで状況に適した応答が可能であると考えられるが，一方で本研究で対象とする雑談対話システムでは多様な発話状況が想定される事からそうしたアプローチが取りづらい．そのため，マイクロブログをはじめとした多様な発話状況を含む対話コーパスを利用した統計的対話モデルが有望であると考えられる [Ritter 11][Vinyals 15]．

しかし我々の知る限り既存研究における教師あり学習に基づく対話応答モデルでは，応答の内容に影響すると考えられる前述の発話状況は陽に考慮されていない．その結果，ある発話に対する応答の自由度は非常に高く，正解が一意に定まらないことから適切なモデルの学習が困難であると考えられる．この問題に対し，我々は会話データに付随するタイムスタンプや発話内容のクラスタリングから獲得した季節や現在の話題といった発話状況を導入したニューラル対話モデルを提案し，応答選択テストによって提案手法の有効性を確認した [佐藤 17]．

本研究ではこれに加えて，複数種類の発話状況を同時に考慮すべくモデルの拡張を行い，応答性能の変化の検証を行った．用いた発話状況としては [佐藤 17] で用いた季節に関する情報に加えて，応答により影響を及ぼすと考えられる発話・応答者（ユーザ）の性質を発話状況として導入する．具体的には，会話データに付随するユーザのプロフィール文をクラスタリングし，獲得したユーザタイプを発話状況として用いた．

実験では，Twitter から収集した対話データを用いて提案手法の有効性を検証する．応答選択テストを用いた評価により，提案手法の有効性を確認した．

2. 関連研究

対話応答に関する研究のうち，我々の研究と大きく考え方を共通するものとしては Hasegawa らの研究が存在する [Hasegawa 13]．Hasegawa らは発話が聞き手に対し喚起させる感情の種類に着目し，人手で作成した少数の規則によって Twitter から取得した大規模な対話データを怒り，喜び，悲しみなどといった 9 つのカテゴリに分類した感情タグ付き対話コーパスを構築している．実験では，そのコーパスから統計的対話モデルを学習することで特定の感情を喚起するような応答の生成を試みている．彼らの研究では Ritter らの研究 [Ritter 11] にならって統計的機械翻訳モデルの利用によって対話モデルを構築し，また感情ごとに学習した局所的なモデルと大域的なモデルをハイパーパラメータによって重み付けを定めた上で組み合わせている．一方，我々のシステムでは Seq2Seq [Vinyals 15] を対話モデルとして採用し，局所的・大域的なモデル，さらにそれらの間の重み付けについても同時に学習を行っている．また，発話状況についてもデータから獲得可能なラベルを用いるか，もしくはクラスタリングによって教師なしで獲得している点が異なる．

また Li ら [Li 16] は，統計的対話モデルにおける応答の一貫性の低さを問題とし，ユーザの発話を外部から制御する手法を考案した．彼らは発話・応答者の性別や居住地を元に会話データをいくつかのユーザタイプへとクラスタリングし，それらの情報を Speaker-embedding と呼ばれる実数値ベクトルとしてニューラル対話モデルに与えることで，特定のユーザタイプを想定して一貫した応答を得ることに成功している．我々の研究と Li らの研究では共に会話データ中の発話・応答者の性質に着目している．しかし，Li らの提案手法は発話状況を特徴量として外部から与える手法であるのに対し，我々の手法は発話状況に応じてモデルの一部を独立に訓練する手法である点が大きく異なる．また，彼らがクラスタリングのための特徴量として性別・居住地などの情報を用いているのに対し，我々はユーザのプロフィール文を用いることで個人の性格や好みといった情報を対話モデルに取り入れる事を試みた．

また，ドメイン適応に関する Daume III らの研究

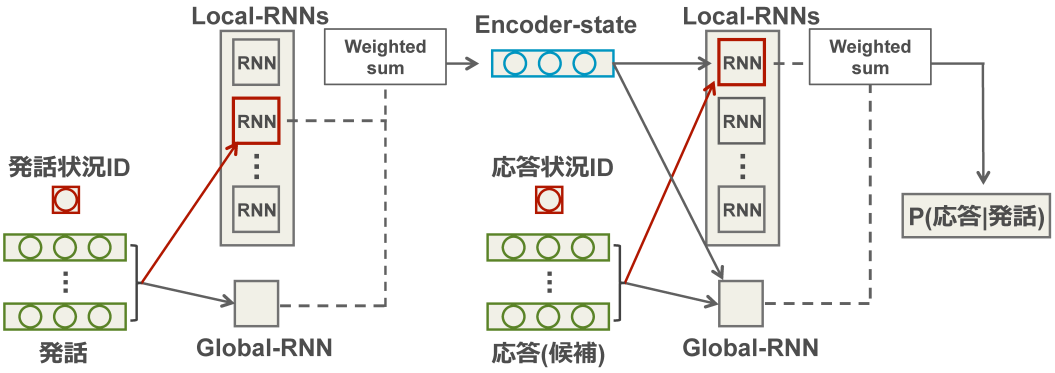

図 1: Local/Global Seq2Seq

[Daume III 07]，Kim らの研究 [Kim 16] も我々の研究と共通点を持つ．Daume III らの提案した手法は学習データが 2 種類のドメイン s，t のどちらかに属し，入力として特徴量ベクトル $\mathbf{x}$ が与えられた時に，$\mathbf{x_1} = \mathbf{x_2} = \mathbf{x}$ とした上で $\Phi^s = (\mathbf{x_1}, \mathbf{x_2}, \mathbf{0})$，$\Phi^t = (\mathbf{x_1}, \mathbf{0}, \mathbf{x_2})$ とモデルへの入力を拡張する事で，$\mathbf{x_1}$ に対応する部分についてはドメイン共通の要素を，$\mathbf{x_2}$ に対応する部分についてはドメイン固有の要素を学習することが可能になる．実験においては固有表現抽出をはじめとする様々なタスクによって，提案手法の有効性を確認している．

また Daume III らの研究に基づき，その手法をニューラルネットワークによるものに拡張したものが Kim らの研究である．彼らはドメインに対してそれぞれ学習を行った独立なネットワークと，ドメイン共通のネットワークによる出力を併用する事でニューラルネットワークにおいて Daume III らの手法を実現した．

我々の研究とこれらのドメイン適応の手法との大きな違いは，ドメインの差異を既知とするか否かにある．こうした既存のドメイン適応の多くは主にコーパスの差異をドメインの違いとして扱っているのに対し，我々の研究は文書要約タスクにおける談話構造の利用 [Louis 10] や，機械翻訳タスクにおける句構造情報の利用 [Eriguchi 16] に関する研究のように，陽に与えられてはいないが発話・応答の背後に存在すると考えられる差異についての情報を取り入れることで応答性能の向上を行うものである．

3. 提案手法

我々の提案する対話モデルは Sutskever らの Seq2Seq モデル [Sutskever 14] をベースに，発話に付与された発話状況を学習・テスト時に考慮する機構を導入したものである．以下で，Seq2Seq について簡単に導入した後，提案手法，及び考慮する発話状況について述べる．

3.1 Seq2Seq 対話モデル

Seq2Seq 対話モデルは発話を入力とし，まず Encoder と呼ばれる Recurrent Neural Network (RNN) によって発話の単語列を順次読み込み，発話内容を表現する実数値ベクトルに変換する．続いて，得られた実数値ベクトルを Decoder と呼ばれる RNN の初期状態とし，RNN の内部状態に基づいて単語を一単語ずつ入出力する．本研究では RNN として Long-Short Term Memory (LSTM) [Hochreiter 97] を採用し，応答選択テストの際は発話・応答ペアに対する Softmax 損失を擬似的なスコアとして用いて応答候補の順位付けを行う．また Softmax 関数は Jean ら [Jean 15] の Sampled Softmax を用いた．

3.2 発話状況を考慮した対話モデル

本研究における提案手法を図 1 に示す．[佐藤 17] では Encoder, Decoder 共にそれぞれ 2 種類の RNN を同時に訓練する事で発話状況を考慮するモデルを提案した．1 つは Local-RNN と呼ぶ発話状況ごとに独立な RNN であり，それぞれの発話・応答に加えて入力として与えられた発話状況 ID・応答状況 ID によってどの RNN を用いるかを決定する．それぞれの RNN はある発話・応答状況下での会話データのみから学習が行われるため，応答はより状況を考慮したものとなると考えられる．しかし，それには個々の RNN に与えられる学習データのサイズが減少してしまうというデメリットも存在する．

もう 1 つは発話状況によらず全てのデータから学習される Global-RNN と呼ばれるネットワークであり，その 2 つの出力の平均によって全体の出力を得る．このようにモデルを設計することで Local-RNN によって発話状況が考慮される一方で，Global-RNN によって先に述べたデータサイズの減少による影響が緩和されると考える．また，例えば時間と場所などといった複数種類の発話状況に注目する場合は，時間で分割した Local-RNN と場所で分割した Local-RNN のそれぞれの出力を用いる．

[佐藤 17] ではそれぞれの RNN の出力の平均を取っていたが，対象とする発話状況の種類によって Local-RNN の内部状態・出力の適切な次元数は異なると考えられる．そこで，本研究では異なる次元数を持つそれぞれの RNN の出力を併用すべく，RNN に線形写像を行う層を追加し，それぞれの RNN の出力の重み付け和を取った．それぞれの発話状況に対する適切な次元数の検証は今後の課題とする．

3.3 発話状況

season　発話・応答のタイムスタンプを元に，3-5 月, 6-8 月, 9-11 月, 12-2 月の 4 種類（春夏秋冬）へと会話データの分割を行った．本実験で用いた会話データはそれぞれの発話・応答の組は同じ日に行われたものを用いているため，発話・応答で発話状況は共通する．

user-profile　我々は発話・応答者のプロフィール文をクラスタリングすることで発話・応答がどのようなユーザによって為されたかを表す発話状況を獲得した．具体的にはまず，形態素解析を行ったプロフィール文中の動詞・名詞・形容詞のそれぞれの単語について，会話データから事前訓練した Mikolov らの word2vec [Mikolov 13] による skip-gram の平均を取る

表 1: モデルと注目した発話状況ごとの 1 in t P@k.

Model	Domain	1 in 2P@1	1 in 5P@1	1 in 5P@2	1 in 20P@3
Baseline	-	64.5%	33.9%	56.6%	28.9%
Local Seq2Seq	season	61.2%	30.3%	53.0%	25.2%
	user-profile	64.7%	34.6%	57.6%	29.5%
Local/Global Seq2Seq	season	65.2%	35.2%	58.1%	31.1%
	user-profile	66.1%	37.0%	59.3%	32.4%
	both	**66.7 %**	**37.9 %**	**60.3%**	**32.7%**

表 2: Baseline から応答が改善した（正答を選ぶようになった）例 (1 in 5P@1)

発話状況	発話	応答 (Baseline)	応答 (Local/Global Seq2Seq)
season (春)	寒いと思ったら雪降ってた	おつかれ!	まじ?この時期にやめてほしいサクラサケ
season (夏)	なにやらコミケの話で盛り上がってる。帰りのゆりかもめはトラウマ物	おつかれさまだった	今年の夏は参加しないんだっけ?
user-profile (研究)	大学の方が研究のレベル高かったな	いいなー	お前のところは科研費めっちゃあったからな
user-profile (音楽)	路上やっぱレパートリーいるなぁ	大黒摩季ですね!	箱よりもアドリブ力試されるだろうしね

ことでそれぞれのプロフィールを表すベクトルを構築する．その上で，k-means クラスタリングを適用する事で会話データを k 種類に分割している．本実験では $k = 10$ と設定し，これにデータが手に入らなかった発話・応答者についてのクラスタを加え計 11 種類へと分割した．

4. 実験

提案手法の有効性を検証するため，応答選択タスクによって手法の評価を行う．

4.1 設定

学習・テストデータ　学習・評価のためのデータセットとしては，我々の研究室において 2011 年 3 月より継続的に収集している Twitter のデータセットを元に構築した [*1]．具体的には，各投稿（ツイート）を発話とみなし，あるツイートとそれに対するリプライを 1 組の発話・応答ペアとした上で，学習データとして 2014 年から約 23,563,865 組，テストデータとして 2015 年の各月から 500 組ずつ合計 6,000 組を抽出し，それに加えてダミー応答候補として同年のツイートからランダムに 114,000 応答を抽出した．続いて MeCab[*2] (辞書には mecab-ipadic-neologd[*3] を用いた) を用いて各発話・応答を形態素に分割した．テストに用いたデータはすべて 20 単語以下のものに制限している．また発話は他のツイートに対するリプライとなっていないものに限定することで，リプライの連鎖の中に存在する文脈の不足によって回答不能となるケースを可能な限り減らしている．

評価手順　前述した学習データを用いて訓練した各モデルについて応答選択テストを行う．具体的にはあるツイートに対し，実際に行われたリプライとダミー応答候補から最大 19 応答，合計最大 20 応答候補から応答としての妥当さを順位付けする．評価尺度としては Wu ら [Wu 16] の **1 in t P@k** を用いる．これは，t 個の応答候補の中から k 応答選択し，その中に実際に行われたリプライが含まれている割合を意味する．

比較モデル　実験では以下の 3 つのモデルを比較する．

Baseline (Seq2Seq) [Sutskever 14] 第 3.1 節参照.

Local Seq2Seq (提案手法) 第 3.2 節のモデルから, Global-RNN を除いたもの．各 RNN が発話状況ごとに独立に学習され、発話状況に依存しないモデルを持たない.

Local/Global Seq2Seq (提案手法) 第 3.2 節参照.

以上のモデルを TensorFlow[*4] を用いて実装した．

RNN は 3 層の多層 LSTM を採用し，最適化は Adam [Kingma 14] によって行った．主要なハイパーパラメータは語彙: 100,000 word，dropout 率: 0.25，学習率:1e-4, バッチサイズ:800 に設定した．また，単語 embedding やそれぞれの RNN の内部状態・出力についてはすべて 100 次元に設定した．

4.2 結果

結果を表 1 に示す．まず, それぞれの発話状況について Local Seq2Seq と Local/Global Seq2Seq の結果を比較すると，Local Seq2Seq では発話状況として **season** の精度は低いものとなった．一方で **user-profile** は教師無しで獲得した発話状況であり，**season** よりも分割数が多いためモデルあたりのデータの減少が大きい．それにも関わらず Global-RNN 無しでも Baseline を僅かに上回る結果が得られたことから，発話・応答者の性質が応答に与える影響は非常に大きなものであると考えられる．

Li らの研究 [Li 16] では性別・居住地をはじめとする情報によって発話・応答者の性質を獲得していたが，そうした情報は必ずしも陽には与えられず，またそれらのラベルのみで発話・応答者の性質を十分に表現するのは困難である．そのため発話・応答者のデータセットから陽に獲得不可能な性質について，それらを低コストに推測する手法に加え，またどのような情報が獲得できれば発話・応答者の性質をより適切に表現可能であるかについての検討が今後の大きな課題であると考える．

また，Local/Global Seq2Seq ではどの発話状況を用いた場合も Baseline, Local Seq2Seq よりも優れた結果が得られた．この事から Global-RNN による補完が効果的であり，特に Seq2Seq のような学習に大規模なデータを必要とするモデルにおいては有効であると考えられる．

次に, 提案手法 (Local/Global Seq2Seq) によって応答選択結果が改善された例を表 2 に示す．左端の「発話状況」の項については，**season** を用いている場合は応答が行われた季節を

*1　2011 年 3 月に 30 名程度の著名な日本人ユーザを選択し，そのユーザのタイムラインを公式 API で継続的に収集するとともに，それらのユーザがメンション・リツイートを行ったユーザのタイムラインも収集対象として追加することで拡大していった．

*2　http://taku910.github.io/mecab/

*3　https://github.com/neologd/mecab-ipadic-neologd

*4　https://www.tensorflow.org/

表記する．一方で **user-profile** は発話・応答者のプロフィール文中の単語に基づくクラスタリングによって獲得した発話状況であることから多様な単語が混在し，あるクラスタを代表するような単語は一概には定められない．そのため表 2 の例では発話から筆者が連想した話題に相当する単語について，クラスタごとにその頻度を求め，その頻度が最も高かったクラスタと応答者が実際に所属していたクラスタが同一であった例を採用した．結果，ある話題（音楽，研究）がその話題についての会話を頻繁に行うと考えられる応答者の会話データから学習したモデルによって回答されることで，応答の改善に成功している．

5. おわりに

本研究では多様な発話状況を持つ対話データに対し，教師無しで得られる発話状況に注目し，発話状況を考慮した対話モデルを提案した．実験ではマイクロブログから抽出した対話データにおいて発話・応答が行われた季節や発話・応答者の性質を考慮した提案モデルが，ベースラインとして用いた SEQ2SEQ 対話モデルに比べて高い応答性能を発揮することを確認した．今後の課題としては，それぞれの発話状況の種類ごとに適切な local-RNN の次元数の検討や，より効果的な発話・応答者の性質の表現手法の検討が挙げられる．

謝辞

本研究の一部は JSPS 科研費 16K16109 と 16H02905 の助成を受けたものです．

参考文献

[Daume III 07] Daume III, H.: Frustratingly Easy Domain Adaptation, in *Proceedings of ACL*, pp. 256–263 (2007)

[Eriguchi 16] Eriguchi, A., Hashimoto, K., and Tsuruoka, Y.: Tree-to-Sequence Attentional Neural Machine Translation, in *Proceedings of ACL*, pp. 823–833, Berlin, Germany (2016)

[Hasegawa 13] Hasegawa, T., Kaji, N., Yoshinaga, N., and Toyoda, M.: Predicting and Eliciting Addressee's Emotion in Online Dialogue., in *Proceedings of ACL*, pp. 964–972 (2013)

[Hochreiter 97] Hochreiter, S. and Schmidhuber, J.: Long short-term memory, *Neural computation*, Vol. 9, No. 8, pp. 1735–1780 (1997)

[Jean 15] Jean, S., Cho, K., Memisevic, R., and Bengio, Y.: On Using Very Large Target Vocabulary for Neural Machine Translation, in *Proceedings of ACL-IJCNLP*, pp. 1–10 (2015)

[Kim 16] Kim, Y.-B., Stratos, K., and Sarikaya, R.: Frustratingly Easy Neural Domain Adaptation, in *Proceedings of COLING*, pp. 387–396 (2016)

[Kingma 14] Kingma, D. and Ba, J.: Adam: A method for stochastic optimization, *arXiv preprint arXiv:1412.6980* (2014)

[Li 16] Li, J., Galley, M., Brockett, C., Spithourakis, G., Gao, J., and Dolan, B.: A Persona-Based Neural Conversation Model, in *Proceedings of ACL*, pp. 994–1003 (2016)

[Louis 10] Louis, A., Joshi, A., and Nenkova, A.: Discourse indicators for content selection in summarization, in *Proceedings of SIGDIAL*, pp. 147–156, Tokyo, Japan (2010)

[Mikolov 13] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J.: Distributed representations of words and phrases and their compositionality, in *Advances in NIPS*, pp. 3111–3119 (2013)

[Ritter 11] Ritter, A., Cherry, C., and Dolan, W. B.: Data-driven response generation in social media, in *Proceedings of EMNLP*, pp. 583–593 (2011)

[Sutskever 14] Sutskever, I., Vinyals, O., and Le, Q. V.: Sequence to sequence learning with neural networks, in *Proceedings of NIPS*, pp. 3104–3112 (2014)

[Vinyals 15] Vinyals, O. and Le, Q.: A neural conversational model, *arXiv preprint arXiv:1506.05869* (2015)

[Wu 16] Wu, B., Wang, B., and Xue, H.: Ranking Responses Oriented to Conversational Relevance in Chatbots, in *Proceedings of COLING*, pp. 652–662 (2016)

[佐藤 17] 佐藤 翔悦, 吉永 直樹, 豊田 正史, 喜連川 優：暗黙の発話状況を考慮したニューラル対話モデル, 言語処理学会第 23 回年次大会 (2017)