1E5-4

深層学習を用いたTwitterにおける会話の応答文生成の試み

Generating a Reply on Twitter using a Neural Conversational Model

那須野薫 *1　　松尾豊 *2
Kaoru Nasuno　　Yutaka Matsuo

*1*2東京大学
The University of Tokyo

In this paper, we extend the recently proposed neural conversational model to generate a reply sentence on Twitter in Japanese especially depending on latent features. Experimental results indicate that our model can capture the latent features of inputs and replies and can generate simple replies depending on the features.

1. はじめに

会話の応答文生成は重要なタスクである．会話は重要なコミュニケーション方法の一つであり，会話の適切な応答は，そのコミュニケーションを円滑に行う上でも非常に重要である．特に，Twitter をはじめとするマイクロブログでは，ユーザにより日々多くの会話がなされており，Twitter における会話の応答文生成に関する研究はコミュニケーションが苦手な個人や日常的な会話を自動化したい法人等が抱える具体的な課題に対しても役に立つ可能性がある．

Twitter における会話の応答文生成は難しい．Twitter 上の会話はくだけたものが多く，単語の表記が一定ではなかったり，顔文字が利用されたりする．また，適切な応答文は応答対象の文だけでなく，相手との関係，時間帯，場所，話題など他の文脈にも依存している．従来は人がルールを作り，それに基づいて，応答文が生成されていたが，表記ゆれした語や顔文字等，ルールの適用が難しいものを多く含み，また，会話の文脈も多様な Twitter の会話の応答文生成は難しい．

近年，Neural Conversational Model（NCM）[Vinyals 15] という深層学習を用いて会話の応答文を生成するモデルが提案された．このモデルは，従来手法で人が作成していたルールを用いずに，単語の系列をそのまま入力して，その単語の系列からなる文への応答文をそのまま単語の系列として得るというものである．実験は英語の会話データを対象に行われ，比較的簡単な会話の応答文であれば生成できることが報告された．

本研究では，NCM を日本語の Twitter の会話に適用できることを確認し，また，応答文の生成を制御する入力文以外の潜在特徴量を分析し，応答文生成に関する知見を得ることを目的とする．そこで，潜在特徴量を分析できるように NCM を拡張し Twitter の実データを対象に分析を行う．

本研究の貢献は下記の通りである．

- 26 万以上の Twitter における会話の実データを対象に実験を行った．
- 日本語の会話の応答文生成，特に，表記ゆれや顔文字の利用が多い会話の応答文生成おいて，モデルの入出力の次元を文字ごとにすることで，十分可読な日本語が生成されることを示した．
- 本研究で拡張したモデルの潜在特徴空間の分析より，潜在特徴量の値により生成される応答文を制御できること，潜在特徴空間において同じ応答文が連続する空間を占めていること，投稿する状況が類似する 2 つの入力文の潜在特徴空間が類似すること示した．

連絡先: 那須野薫，東京大学，nasuno@weblab.t.u-tokyo.ac.jp

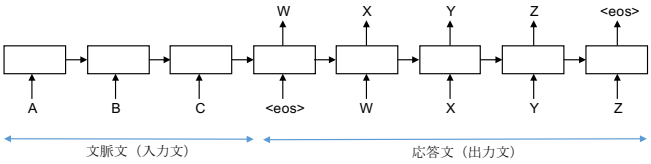

図 1: A Neural Conversational Model

本研究の構成は下記の通りである．2 章では関連研究について述べる．NCM と本研究でも用いる Recurrent Neural Networks である Gated Recurrent Neural Networks について述べる．3 章では，分析手法について説明し，4 章では実験で用いる Twitter の会話ログのデータセットについて述べる．5 章では，実験結果について述べるが，特に，潜在特徴空間と対応して生成される応答文を多く紹介する．6 章で考察し，7 章でまとめる．

2. 関連研究

ここでは，本研究で拡張する Neural Conversational Model（NCM）[Vinyals 15] と本研究で用いる Recurrent Neural Networks である Gated Recurrent Neural Networks（GRNN）について述べる．

2.1 A Neural Conversational Model

NCM は，seq2seq フレームワーク [Sutskever 14] を用いて，会話の応答文を生成するモデルである．ネットワーク構造を図 1 に示す．入力文を 1 語ずつ入力して，得られた RNN の最後の隠れ層を出力用の RNN の $t = 0$ の隠れ層に利用する．

その後，文末符号（End of Sequence）を入力し，学習時には，出力文を 1 語ずつ入力していき，同時に次の語を予測する．生成時には，前の時刻で最も生成確率が高かった語を入力していき，同時に次の語を予測する．

2.2 Gated Recurrent Neural Networks

GRNN は Gated Recurrent Unit [Cho 14] というゲート付き活性化関数を用いる RNN のことで，GRU は LSTM [Hochreiter 97] のように，長期的な表現と短期的な表現を捉えるために提案された活性化関数である．GRNN は下記の式により定

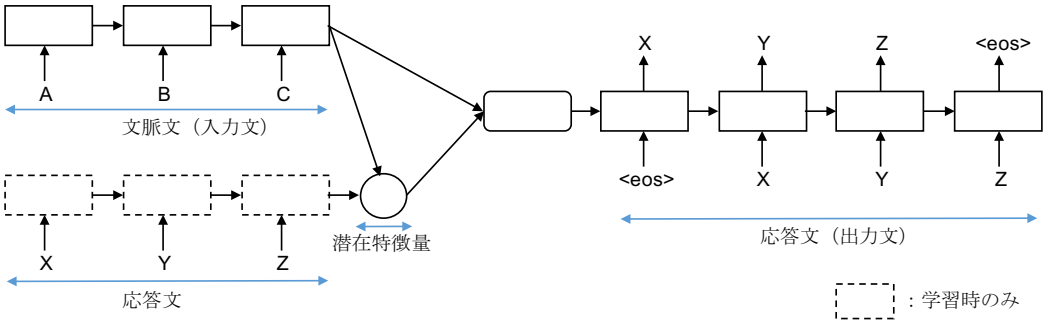

図 2: 潜在特徴量により応答文の生成を制御できるように拡張した Neural Conversational Model のネットワーク構造

義される．

$$\mathbf{r}_t = \sigma(\mathbf{W}_{xr}\mathbf{x}_t + \mathbf{W}_{hr}\mathbf{h}_{t-1} + \mathbf{b}_r) \quad (1)$$

$$\mathbf{z}_t = \sigma(\mathbf{W}_{xz}\mathbf{x}_t + \mathbf{W}_{hz}\mathbf{h}_{t-1} + \mathbf{b}_z) \quad (2)$$

$$\tilde{\mathbf{h}}_t = \tanh(\mathbf{W}_{xh}\mathbf{x}_t + \mathbf{W}_{hh}(\mathbf{r}_t \odot \mathbf{h}_{t-1}) + \mathbf{b}_h) \quad (3)$$

$$\mathbf{h}_t = \mathbf{z}_t \odot \mathbf{h}_{t-1} + (1 - \mathbf{z}_t) \odot \tilde{\mathbf{h}}_t \quad (4)$$

$$\mathbf{y}_t = \sigma(\mathbf{W}_{hy}\mathbf{h}_t + \mathbf{b}_y) \quad (5)$$

ここでは，$\mathbf{W}_{xr}, \mathbf{W}_{hr}, \mathbf{W}_{xz}, \mathbf{W}_{hz}, \mathbf{W}_{xh}, \mathbf{W}_{hh}$ は重み行列で，$\mathbf{b}_r, \mathbf{b}_z, \mathbf{b}_h$ はバイアス項である．

3. 分析手法

分析手法について述べる．分析では潜在特徴量を抽出できるように NCM を拡張し，抽出した潜在特徴量に基づいて応答文を生成し，潜在特徴空間を可視化・分析するというものである．拡張したネットワーク構造を図 2 に示す．ネットワーク構造はモデルの学習時と生成時で異なる．

学習時は入力文に加えて，応答文も入力する．T を入力文の最後の時刻，$\mathbf{h}_{inT}$ を入力文の時刻 T の隠れ層，$\mathbf{h}_{repT}$ を応答文の時刻 T の隠れ層，とすれば，潜在特徴量 $\mathbf{l}$ は $\mathbf{l} = \tanh(\mathbf{W}_1\mathbf{h}_{inT} + \mathbf{W}_2\mathbf{h}_{repT} + \mathbf{b}_1)$，出力層の時刻 $t = 0$ の隠れ層（図 2 の角が丸い四角）は $\mathbf{h}_{out0} = \tanh(\mathbf{W}_3\mathbf{h}_{inT} + \mathbf{W}_4\mathbf{l} + \mathbf{b}_2)$ と表現される．潜在特徴量 $\mathbf{l}$ の次元 D が十分小さい時，潜在特徴量のみから応答文を生成することはできず，したがって，入力文も生成に寄与するはずである．また，逆に入力文が生成に寄与する条件下で抽出した潜在特徴量も有効であれば，その特徴量は，入力文だけからでは予測が困難であり，かつ，会話文生成に影響を与える重要な会話の要素を表現している可能性がある．生成時は図 2 の点線部分は用いずに直接，潜在特徴量を入力する．本研究では，潜在特徴量の次元数 $D = 2$ とし，生成時に入力文と $-1 < l_i < 1$ を満たす l_1, l_2 を入力して，応答文を得る．

本研究は，NCM で Twitter の会話の応答文を生成できることを実験的に示し，また，こうして定義された潜在特徴量を分析することで，入力文では表現されない応答文生成を支配する潜在的な文脈に関する知見を得ようというものである．Twitter の文章の表記ゆれや顔文字が含まれるという特性を考慮して，入力次元には [Vinyals 15] とは異なり単語ではなく文字を利用する．

潜在特徴空間の分析には，投稿する状況が類似する複数の入力文に対して，異なる潜在特徴量の値を入力した場合どのような応答文が生成されるかを定性的に評価する．可視化は，入力文に対して $-1 < x < 1$ においてランダムに生成された 200 の潜在特徴量の値を用いて生成された応答文 2 次元空間上にプロットする．加えて，生成された文字列の分布領域をより明確化するために，ランダムに生成された潜在特徴空間上の座標から，入力文と生成された応答文よりモデル用いて算出した潜在特徴空間上の座標にエッジを引く．

4. データセット

データセットは Twitter における会話ログを用いる．ここでいう会話とは，あるユーザのツイートとそれに対する他のユーザのリプライによるツイートのペアである．データセット作成の流れは下記の通りである．

1. ツイートデータの取得
   2015 年 4 月から 12 月までの約 9 ヶ月間に生じた日本語のツイートを Streaming API を用いて取得する（すべてのツイートを取得するということではない）．結果，2 億弱のツイートを得た．

2. 会話抽出
   取得したツイートから in_reply_to_status_id の項目を用いて会話データを抽出する．結果，40 万弱の会話データを得た．

3. クレンジング
   一見して計算機によりツイートを自動生成させていると考えられるユーザ（ボット）の会話や URL を含む会話等は対象から除去する．結果，33 万弱の会話データを得た．

4. 入力次元の決定
   得られた会話データに含まれる文字のうち，出現頻度が高い K 個の文字をモデルの次元に用いる．モデルの次元に含まれない文字は全て $< unk >$ とする．また，応答文に $< unk >$ を含む会話は対象から除去する．K の決定は，モデルの表現力とモデルの学習に必要な計算時間やデータ数とのトレードオフになっていることを考慮する．ここでは，$K = 1500$ とし，26 万強の会話データを得た．これは，クレンジング後の会話データ全体の 80%以上の会話を網羅している．

5. 実験

実験について述べる．実験では，潜在特徴空間を可視化し，日本語の応答文を生成できること，潜在特徴量空間を変えることで生成される応答文が変わることを確認する．なお，生成された応答文のうち，特に，Twitter の最大文字列長である 140 文字以上同じ文字が連続するものについては，表記の都合，3 文字目以降の表記を省略した．

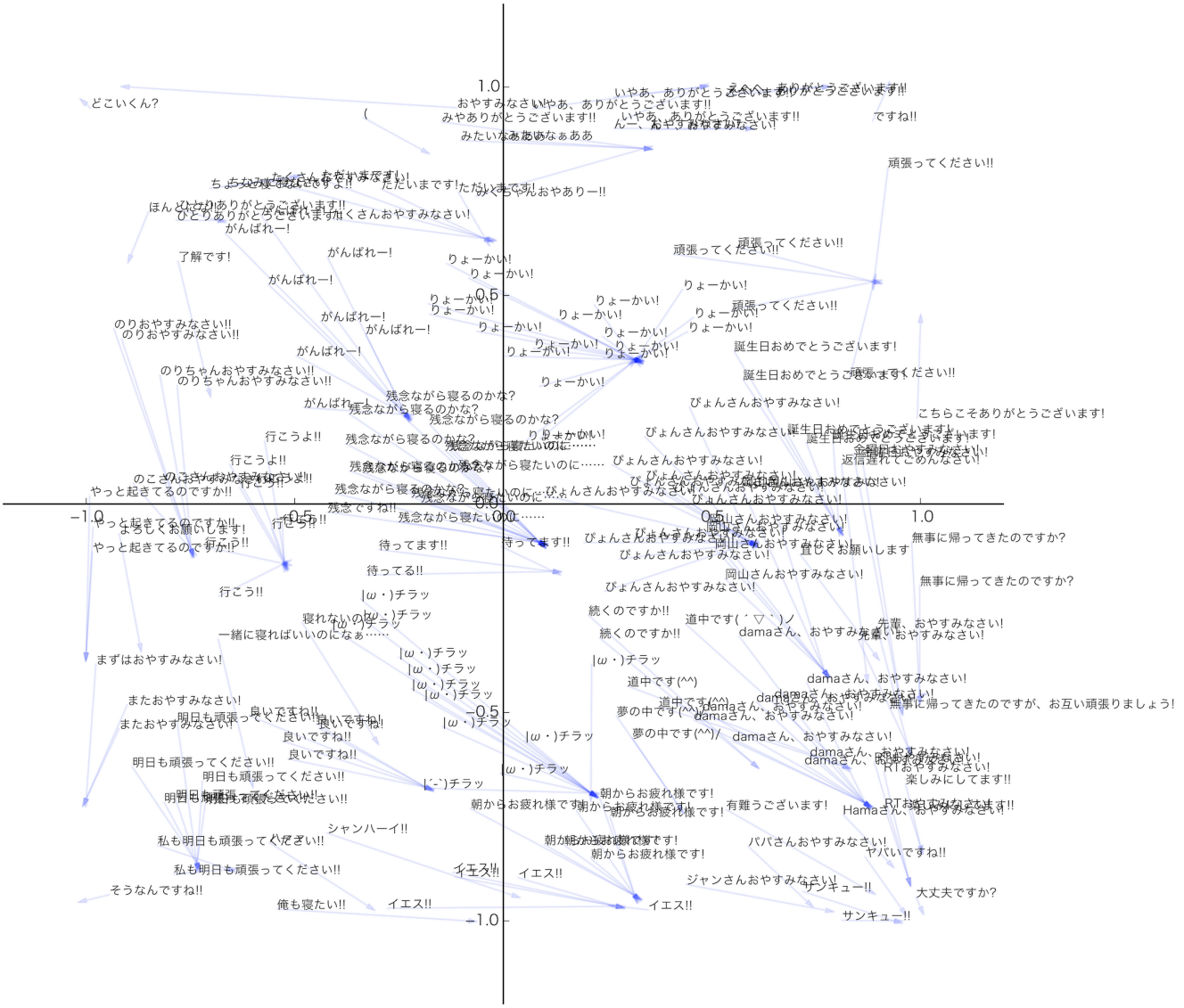

図 3: 入力文「おやすみー」に対する潜在特徴量空間の値と生成された応答文

潜在的な特徴量が抽出できているかを確認するため，複数の同じ入力文に対して，異なる潜在特徴量の値を入力した場合どのような応答文が生成されるかを定性的に評価する．ここでは，一日の活動の終わりに投稿する可能性のある 2 つのツイートである「おやすみー」と「今日はもう疲れた，寝る (>_<)」を入力文として用い，$-1 < x < 1$ においてランダムに生成された 200 の潜在特徴量の値を用いて生成された応答文を 2 次元空間上にプロットした．

生成されたそれぞれの応答文をそれぞれ図 3 と図 4 に示す．まず，生成された応答文について，概ね日本語として読める文が生成されていることがわかる．次に，潜在特徴量の値と生成された応答文について，潜在特徴量の値に応じて生成された応答文が異なり，潜在特徴量の値によって応答文が制御されていることが分かる．さらに，異なる 2 つの入力文ともに $(-0.2, -0.2)$ の付近では顔文字を含む文字列「|ω・) チラッ」等が生成されており，$(0.5, 0)$ の付近では「<相手の名前>おやすみなさい」かあるいはそれに類する文字列が生成されていること等を踏まえると，類似した状況下で投稿される異なる 2 つの入力文は類似した潜在特徴量空間を持つことが推察される．また，このことは，モデルの入出力の次元を文字ごとにすることは，Twitter のテキスト解析に際して問題となる表記ゆれや顔文字の問題に対してある程度有効であること示している．最後に，潜在特徴空間に記載した単語とそのエッジの分布は，生成された応答文は潜在特徴空間を連続して分布している可能性が高いことを示していると考えられる．

6. 考察

本研究では，日本語においても NCM を用いることで会話の応答文をある程度生成できること，を確認した．一方で，紙面の都合で掲載できなかったが，人間と会話させ違和感の程度を定性的に評価する実験では，[Vinyals 15] でも指摘されている通り，1 回限りの短文の応答文生成と異なり，人間との会話には違和感を抱くところが多かった．モデルの入出力を文字ごとにする NCM は，応答文の生成において，表記ゆれや顔文字の利用など言語処理の問題点に対してある程度有効であることが確認されたが，依然として，違和感のない会話の生成という点では問題が多く，モデルに違和感を減らす機構を導入する必

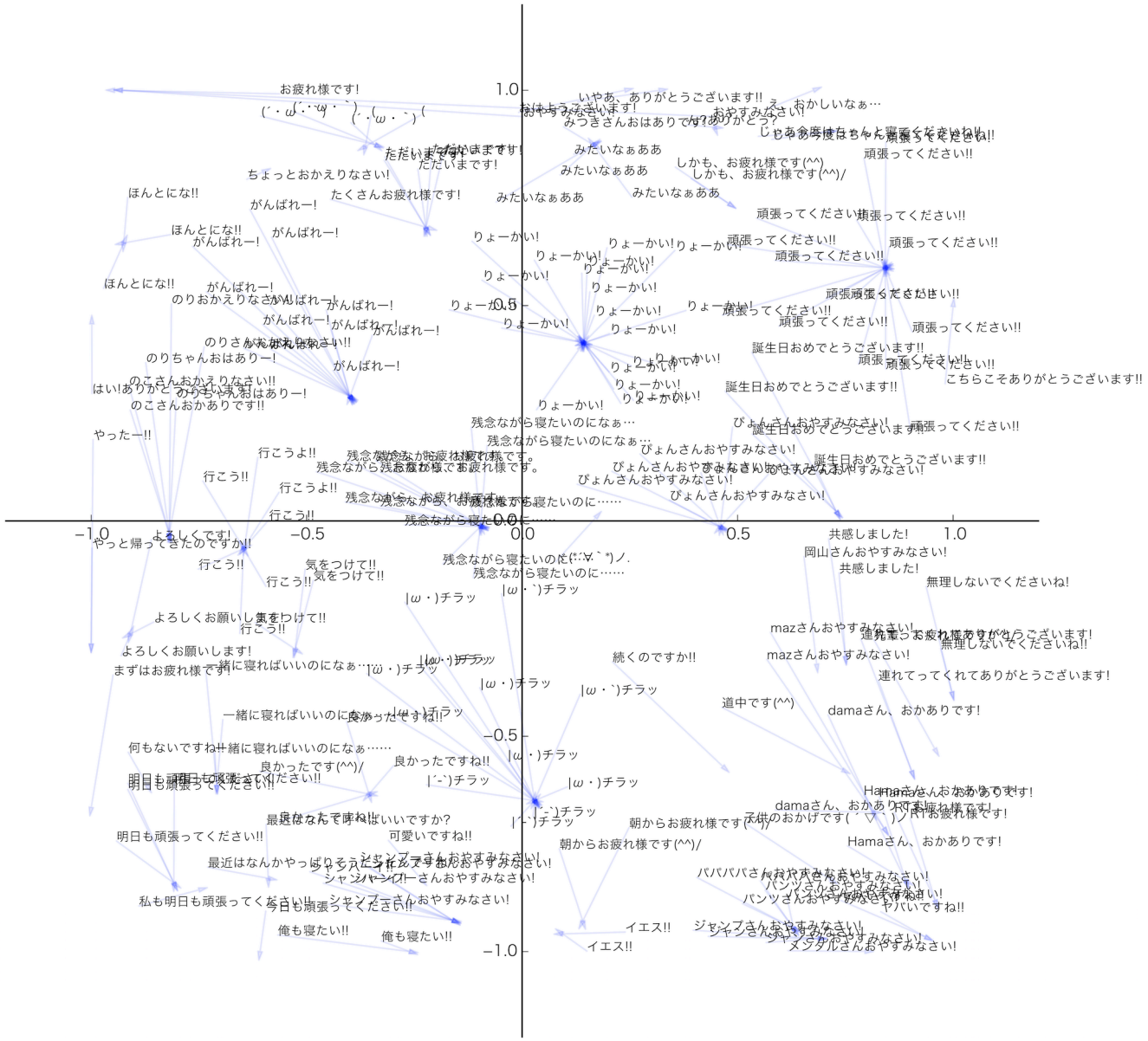

図 4: 入力文「今日はもう疲れた，寝る (>_<)」に対する潜在特徴量空間の値と生成された応答文

要があると考える．違和感を減らす機構は潜在特徴量空間の座標のずれの大きさにも関係がある可能性があるかもしれない．

7. まとめ

本研究では，Neural Conversational Model を潜在特徴量により応答文を制御できるように拡張した．26 万以上の Twitter における会話の実データを対象に行った実験より，表記ゆれや顔文字の利用が多い会話の応答文生成おいて，モデルの入出力の次元を文字ごとにすることで，十分可読な日本語が生成されること，潜在特徴空間の分析より，潜在特徴空間において同じ応答文が連続して分布していること，投稿する状況が類似する 2 つの入力文の潜在特徴空間が類似すること示した．

参考文献

[Cho 14] Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y.: Learning phrase representations using rnn encoder-decoder for statistical machine translation, *arXiv preprint arXiv:1406.1078* (2014)

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

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

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