

文脈を考慮した対話システムにおける Reminder の改良

佐藤航 長名優子

東京工科大学 コンピュータサイエンス学部

1 はじめに

対話システムやチャットボットが注目されており、多くの企業が自然言語処理や Deep learning を使用して、チャットボットで自然な対話を実現させることにチャレンジしている。そのような手法のひとつとして、記憶を保持する Transformer[1] によるチャットボット [2] が提案されている。この手法では、7 回以上会話が続いたツイートを学習データとして用いて学習を行い、会話を実現している。BERT (Bidirectional Encoder Representations from Transformers)[3] を用いた Encoder へ文章が入力されると、Transformer を用いた Reminder と Decoder を通して文脈を考慮した返答文が出力されるように学習が行われる。しかしながら、この手法では話がかみ合っていない問題がある。それに対し、この手法より前に提案された Reminder を用いない手法 [4] で文脈の考慮は行っていないが、一問一答の会話は行えている。そのため、返答が不自然になっている要因は文脈を考慮するために追加した Reminder に問題があるのではないかと考えられる。

そこで、本研究では文脈を考慮した対話システムにおける Reminder の改良を提案する。従来手法 [2] の話がかみ合っていない問題に対しては、Encoder の出力の位置と記憶する会話を取得する位置を変更することで解決を目指す。

2 Transformer

Transformer[1] は、自然言語処理だけでなく他の分野でも使われている高性能な Deep learning のモデルであり、図 1 に示すように Encoder と Decoder から構成されている。再帰も畳み込みも使わず、Attention のみを使用したモデルとなっている。文章が Encoder に入力され、埋め込み表現に変換後、Attention 層と Position-wise Feed Forward Network 層を通して照応関係情報を獲得する。Decoder では、訓練時は全単語

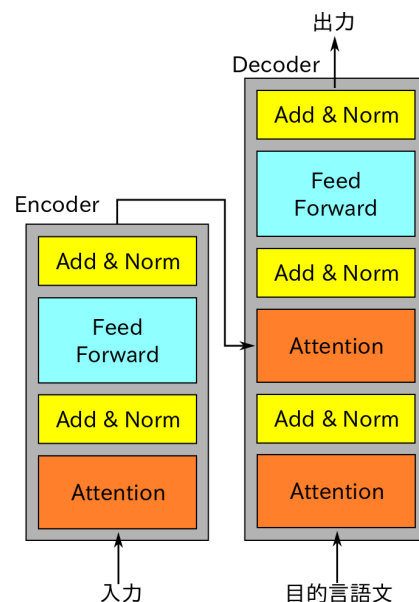


図 1: Transformer の構造

を目的言語文に同時に入力して同時に予測するが、評価・推論時は逐次的に単語列を生成する。

Attention とは、ある単語の意味を理解するとき文中のどの単語に注目すればよいかを表すスコアのことである。Transformer では Self-Attention が用いられており、1 文に含まれる単語群を使って計算する。Position-wise Feed Forward Network 層は各ブロックの Attention 層のあとに配置される。単語ごとに独立しており、他単語と干渉しない 2 層の全結合ニューラルネットワークとなっている。ただし、重みは共有され、全時刻で同じ重みを用いて変換が行われる。このようにすることで、並列処理が可能となっている。

3 BERT

BERT[3] は、Transformer の Encoder を用いているモデルであり、ラベルの付いていない文章から事前学習をすることができる。出力層を付け加えてファインチューニングを行うことで、様々なタスクに対応することができる。図 2 に示すような構造をしており、入力された文章を埋め込み表現に変換後、Transformer を通して分散表現を獲得する。

BERT は MLM (Masked Language Model) と NSP

Reminder Improvements in Dialogue System considering Context
Wataru Sato and Yuko Osana (Tokyo University of Technology, osana@stf.teu.ac.jp)

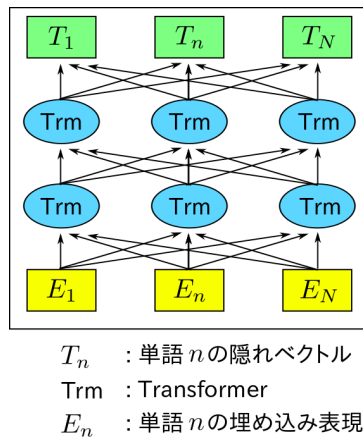


図 2: BERT の構造

(Next Sentence Prediction) で事前学習を行う、双方向型の教師なし言語表現モデルとなっている。

MLM は、入力 の 15% を [MASK] に置き換えて元の単語を予測するように学習を行う事前学習である。予測する単語を [MASK] に置き換えることで予測する単語の後ろにある文章も使用することができ、双方向からの学習を成立させている。このようにすることで、文脈に依存した分散表現を獲得することができる。NSP は、2 つの文が入力されたときにそれらが連続した文かどうかを判定する事前学習である。50% の確率で文 A と文 B が連続する文章となっており、文どうしの関係を学習させる。

4 文脈を考慮した対話システムにおける Reminder の改良

図 3 に示すように、提案システムは Encoder、Reminder、Decoder から構成されている。Encoder には BERT[3]、Reminder と Decoder にはそれぞれ Transformer を使用している。Reminder は前の会話の記憶を入力し、次の会話のために出力する層となっている。通常の入出力以外に、記憶を入出力する層を追加することで、記憶の引継ぎを可能としている。

従来手法では、Reminder で Positional Encoding を使用していなかったが、単語の位置関係をより考慮できるようにするために追加している。また、従来手法では、Reminder の入力 は Reminder の出力から取得していたが、Decoder の出力から取得するように変更している。

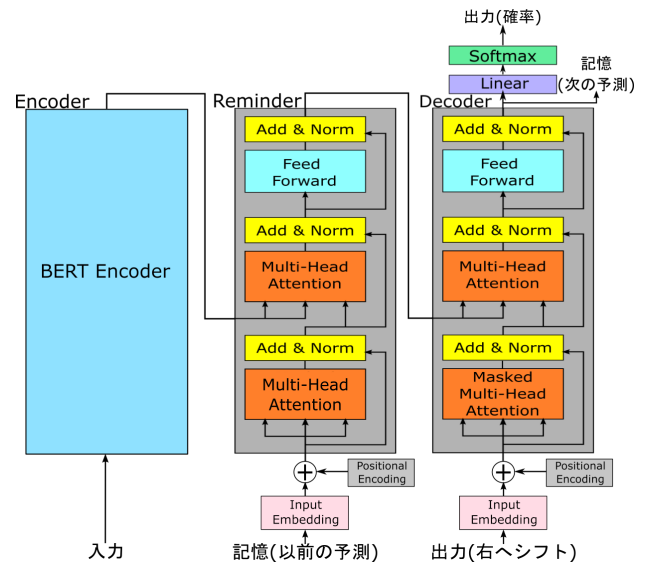


図 3: 提案手法で用いる Reminder を改良したモデル

5 計算機実験

計算機実験を行い、直前のユーザの発話を考慮した応答が行えることを確認した。

参考文献

- [1] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser and Illia Polosukhin : “Attention is all you need,” <https://arxiv.org/abs/1706.03762>, (2022 年 1 月 5 日参照).
- [2] 記憶を保持する Transformer によるチャットボット, <https://qiita.com/kasagon/items/f2b47e9b9b4d3494eeca>, (2022 年 1 月 5 日参照).
- [3] Jacob Devlin, Ming-Wei Chang, Kenton Lee and Kristina Toutanova : “BERT: Pre-training of deep bidirectional transformers for language understanding,” <https://arxiv.org/abs/1810.04805>, (2022 年 1 月 5 日参照).
- [4] BERT を Encoder とする Chatbot の作成, <https://qiita.com/reppy4620/items/e4305f22cd8f6962e00a>, (2022 年 1 月 5 日参照).