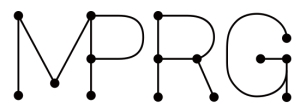


第2回ミーティング

Transformer, BERT の追加調査, モデル比較

EP20050 小池正基

担当先輩：史，張，井上



MACHINE PERCEPTION AND ROBOTICS GROUP

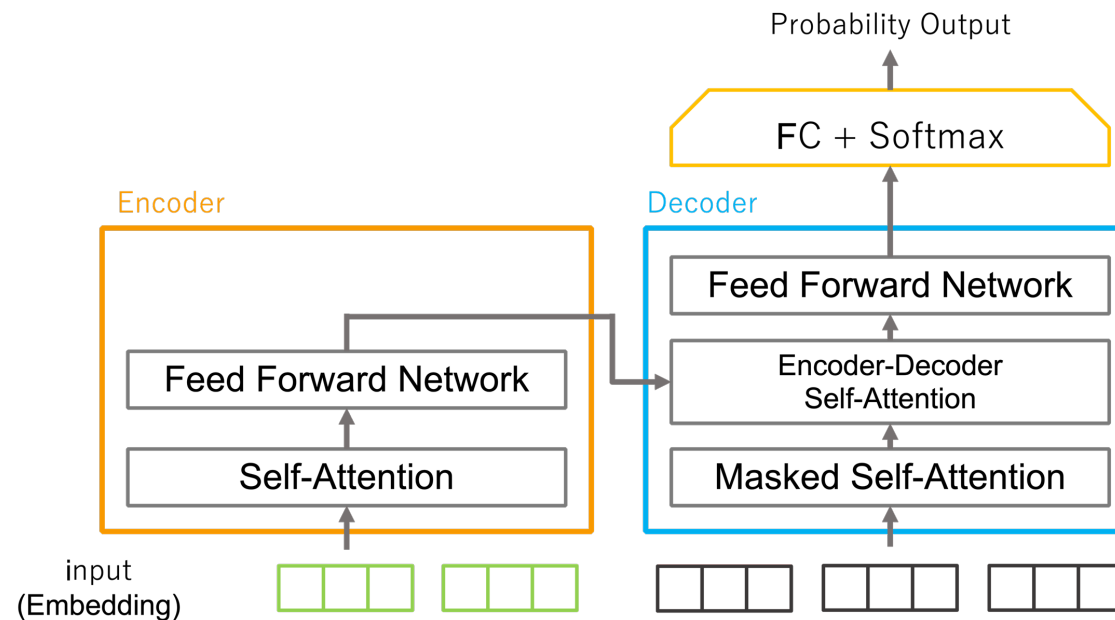
- 研究テーマ
- 進捗
- Transformer
 - Self-Attention
 - Multi-Head Attention
 - Feed Forward Network
- Bidirectional Encoder Representations from Transformers (BERT)
 - 事前学習
 - Fine-tuning
- 日本語モデルの比較

- 自由記述文からの学習行動分析
 - 文書から記述者の学習行動や学習課題を把握する研究
- 必要な技術や知識
 - 自然言語処理 (NLP)
 - テキストマイニング

- 環境構築
 - サーバー割当まで完了
- 研究
 - Transformer 論文読了
 - BERT 論文読了
- 今後の予定
 - NLPモデル制作, 比較

Transformer [A.Vaswani+,NIPS'17]

- Attentionのみを用いたEncoder-Decoderモデル
 - 広範囲な依存関係を把握
 - 計算量が減少



Self-Attention

- 文章内の単語の内積から注目度を算出

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

Q : Query

K : Key

V : Value

d_k : Q, K が持つ次元数

$$Q = XW^Q$$

$$K = XW^K$$

$$V = XW^V$$

X : 入力単語

W : 重み

Multi-Head Attention

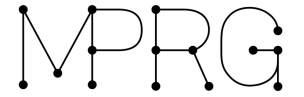
- Q, K, V をheadの数用意（論文内では8）
 - それぞれのheadでAttentionの計算し，統合

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W^O$$

where $\text{head}_i = \text{Attention}(XW_i^Q, XW_i^K, XW_i^V)$

W^O : 重み
 h : head数（論文内は8）

Feed Forward Network

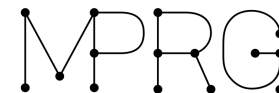


- 単語ごとに独立した2層のニューラルネットワーク

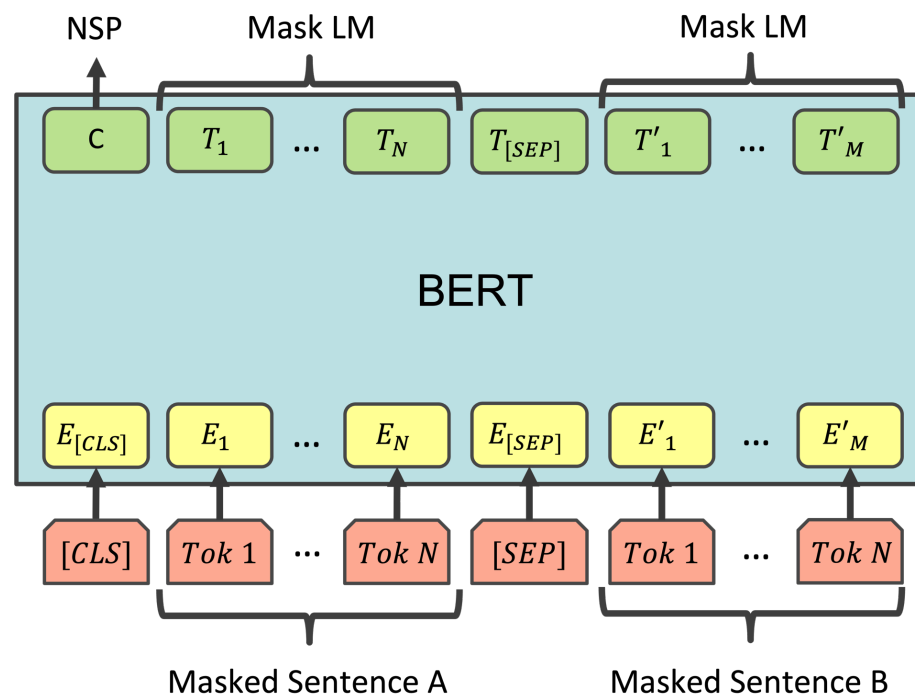
$$FFN(x) = \max(0, xW_1 + b_1)W_2 + b_2$$

W : 重み
 b : バイアス

Bidirectional Encoder Representations from Transformers (BERT) [J.Devlin+,NAACL'19]



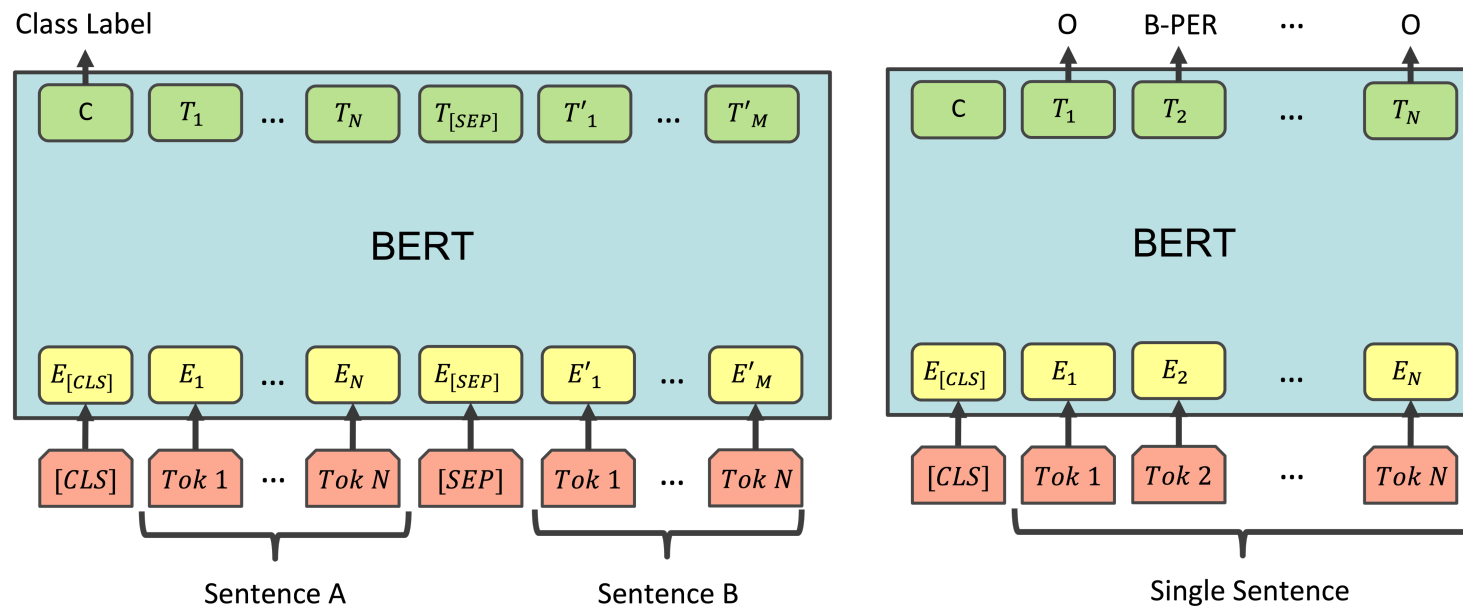
- TransformerのEncoderのみを利用したモデル
- あらゆるNLPタスクにfine-tuning可能
- 事前学習に注力
 - 両方向からの事前学習を行わせるため、2種の事前タスクを実行



- 入力文の先頭に[CLS]トークン，文の間に[SEP]トークンを追加
- MLM
 - Input の15%を[Mask]トークンで置換し，元のトークンを当てるタスク
 - Fine-tuning との差異をなくすため，一定確率でランダムなトークンに置換
- NSP
 - 2文の入力からそれが連続したタスクか推察
 - NSP専用の出力 C を用いて予測

Fine-tuning

- タスクに合わせてデータセットを入力
 - C は識別タスク
 - 感情分析, ジャンル分類
 - T_i はトークンレベルのタスク
 - Q&A, 品詞のタグ付け



日本語モデルの比較 [1/2]

- 評価指標
 - 最終更新日
 - 個人利用の可否
 - 性能評価
 - 学習元
- 候補
 - 早大RoBERTa 最終更新：2021/09/20
 - XLM RoBERTa 最終更新：2023/04/08
 - 日本語LUKE 最終更新：2022/11/09
 - 日本語DeBERTa 最終更新：2023/03/18

日本語モデルの比較 [2/2]

- JGLUE日本語理解ベンチマークを用いて評価
 - MARC -ja : 文章分類タスク
 - JSTS : 意味的類似度計算
 - JNLI : 自然言語推論タスク
 - JCommonsenseQA : 5択のQA問題

モデル名	MARC-ja	JSTS	JNLI	JCommonsenseQA
早大RoBERTa	0.969	0.890	0.928	0.900
XLM RoBERTa	0.964	0.918/0.884	0.919	0.840
日本語LUKE	0.965	0.932/0.902	0.927	0.893
日本語DeBERTa V2	0.968	0.892	0.919	0.890

-
- 研究テーマ
 - 進捗
 - Transformer
 - BERT
 - 日本語モデルの比較
-
- 今後の予定
 - NLPに関する調査
 - モデル作成, 検証