

Spelling Error Correction with Soft-Masked BERT

- 従来手法：BERT
 - BERTを使用して誤文から訂正すべき場所を選択する
 - Masked Language Modelingによって事前学習されているため、精度が低い
 - これは、マスク言語モデリングによるBERTの事前学習方法が、テキスト中の文字の約15%しかマスクされていないため、マスクされたトークンの分布しか学習せず、修正を行わないことを選択する傾向があるためではないかと仮説を立てた。
- 提案手法：Soft-Masked BERT
 - detection network
 - Bi-GRU
 - 単語が誤字かどうか予測する（確率を求める）。
 - correction network
 - BERT
- Soft-Masked BERT
 - モデル構造

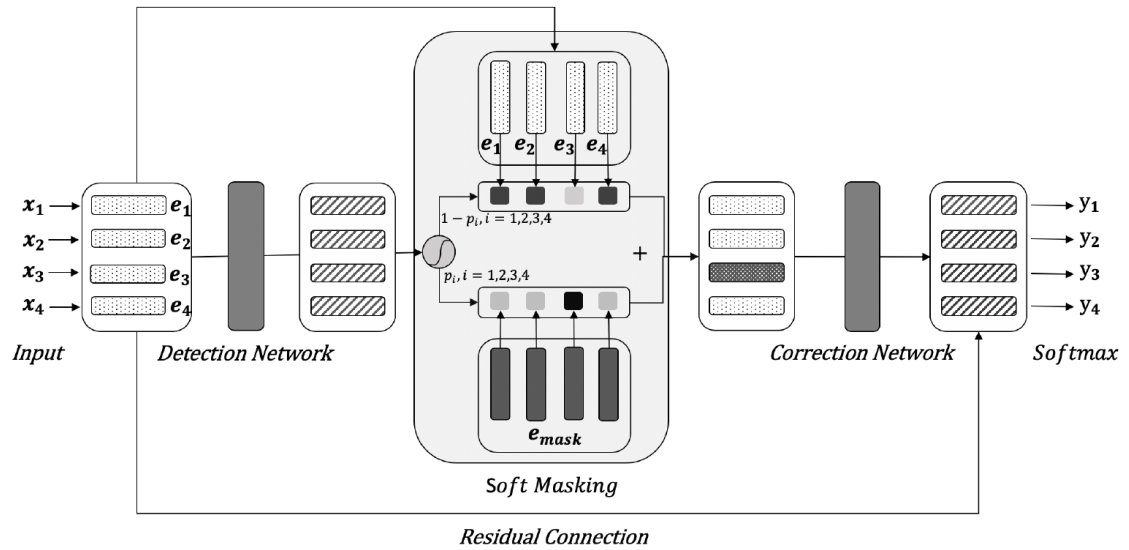


Figure 1: Architecture of Soft-Masked BERT

。 モデル詳細

- まず、入力文の各文字の埋め込みを作成する。次に、一連の埋め込みを入力とし、detection networkを用いて、一連の文字(埋め込み)の誤り確率を出力する。その後、入力埋め込み量と[MASK]埋め込み量を誤り確率で重み付けした和を計算する。計算された埋め込みから、ソフトな方法でシーケンス中の誤り部分をマスクする。そして、ソフトマスクされた埋め込みシーケンスを入力とし、最終層が全ての文字に対してのソフトマックス関数で構成されるBERTモデルであるcorrection networkを用いて、誤り訂正の確率を出力する。また、入力埋め込みと最終層の埋め込みとの間には残差結合が存在する。
- ソフトマスキングは、エラー確率を重みとする入力埋め込みとマスク埋め込みの重み付き和に相当する。