

先端科学技術研究科 修士論文要旨

所属研究室 (主指導教員)	知能コミュニケーション (中村 哲 (教授))		
学籍番号	2011189	提出日	令和 4年 1月 19日
学生氏名	土肥 康輔		
論文題目	話し言葉文法誤り訂正に向けたMasked Language Modelに基づくアプローチ		
要旨			
<p>文法誤り検出はテキスト中の誤りを自動的に検出するタスクであるのに対して、文法誤り訂正はそれらの誤りの訂正も行うタスクである。近年では、機械翻訳や系列ラベリングに基づく手法で取り組むことが主流となっているが、文法誤り検出・訂正で利用可能な誤り情報付きデータが不足していることが課題となっている。本研究では、このデータ不足という問題に着目し、疑似誤り生成によるデータ拡張と、言語モデルによるアプローチを用いて対処する手法について述べる。</p> <p>1つ目の提案手法では、特定の誤りカテゴリについて、学習者の誤り傾向を反映した疑似誤りを生成し、そのデータでモデルをファインチューニングすることで、その誤りカテゴリの訂正性能を向上させる。書き言葉データでの実験の結果、生成する疑似誤りの割合を適切に設定すれば、モデルの訂正性能が向上することが明らかになった。</p> <p>2つ目の提案手法では、BERTにより算出される系列の確率を利用することで、誤り情報付きデータを少量しか用いずに文法誤り検出を行う。既存の言語モデルに基づく文法誤り訂正モデルは、訂正できる誤りの種類に制限があったが、提案モデルでは、テキスト中のすべての誤り種類を検出できることが確認された。F_{0.5}スコアは既存の文法誤り検出モデルに及ばなかったものの、RecallがPrecisionよりも高くなり、今後適切な制約を加えることで性能を向上させられる可能性が示された。</p> <p>また、文を書くときに比べて、話すときのほうが文法等の誤りが多くなると言われているが、話し言葉が持つ性質のため、それらの誤りを検出・訂正することは難しいタスクである。そこで、話し言葉での文法誤り検出・訂正の現状と課題を分析するための評価実験を行う。実験の結果、話し言葉に存在する非流暢性に対しては、文法誤り検出・訂正モデルとは別のモデルを用いるほうがよいことが明らかになった。加えて、上記のBERTに基づく文法誤り検出モデルの訓練に話し言葉を用いると、話し言葉の特徴をとらえた誤り検出・訂正ができるようになる可能性が示された。</p>			