

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)	自然言語処理学 (渡辺 太郎 (教授))		
学籍番号	2311192	提出日	令和 7年 1月 21日
学生氏名	中畔 彪雅		
論文題目	パラ言語情報を用いるSpeech-to-Text対話モデル		
要旨			
<p>人の発話には言語的情報以外に韻律, 発話速度, 声量などのパラ言語情報が含まれ, 発話文が同じでも付与されているパラ言語情報が異なれば発話意図が変化する場合がある. 既存のカスケード型音声対話モデルでは, 音声をテキストに変換する過程でこれらパラ言語情報が欠落するため, パラ言語情報を考慮しない単調な応答を生成しやすい. そこで, 音声を離散整数列や埋め込み表現などの中間表現で処理することで, 感情などの情報を抽出し応答の精度を向上させる研究がされている. ただし, これらの枠組みで実際にどのようなパラ言語情報に含まれる意図がモデルの応答に反映されたのかは明確ではない. こうしたモデルのパラ言語情報に対する処理能力を計測するためには, 言語的情報を固定した上でパラ言語情報を変化させた音声データを使用することが望ましい. そこで, 同じ発話内容を異なる感情で発話してもらった音声データと, その発話の感情まで考慮した応答テキストを収集した評価データセット「paraling-data」を構築した. 入力音声に対する適切な応答を生成するにはパラ言語情報の適切な処理が必要となり, モデルが持つパラ言語情報に対する処理能力が評価可能となる. また, paraling-dataの評価性能を調査するため, 中間表現を「テキスト」・「離散整数列」・「埋め込み表現」のSpeech-to-Text対話モデルを構築し, 各モデルが生成したparaling-dataに対する応答を比較した. 結果, テキストモデルと比較して, 離散整数列モデルは人手評価において最高のスコアを獲得し, 埋め込み表現モデルは多様な応答生成が可能であることが示された. パラ言語情報の変化による各モデルの応答の差を確認できたが, 感情などの特定の情報が応答に反映された結果は確認できなかった. よりパラ言語情報に着目した評価を行うには, 感情予測を補助タスクに組み込むなど, パラ言語情報の考慮を前提とした学習設定が必要だと考えられる.</p>			