マルチモーダル情報と対話履歴を用いた対話継続意欲推定

情報科学科　熊谷　直子　　指導教員：入部　百合絵

1　はじめに

非タスク指向型対話システムは，人とシステムが円滑に対話するために話題を調整することが重要である．話題の調整には，話題の切り替えタイミングを検討する必要があり，そのタイミングを測るための一つの情報としてユーザの対話意欲が挙げられる．

対話継続意欲を推定する研究は存在する[1]．しかし，多くの研究では対話履歴を考慮した時系列推定を行っていない．対話継続意欲は一過性のものではなく，連続した対話の流れから推定すべきである．また，対話意欲を適切に推定するためには，言語情報だけではなく，ユーザの非言語チャンネルも含んだマルチモーダル情報を用いることが重要である．本研究では，対話中のユーザから得たマルチモーダル情報の時系列データを用いた対話継続意欲推定を目的とする．

2　対象とする対話コーパス

本研究は，対話システムと人との対話の様子を収めたマルチモーダルコーパス Hazumi [2]を使用する．Hazumi コーパスには，対話中の被験者の映像，音声，Kinect データ，生体信号データが含まれる．本研究では被験者 26 人の計 2429 発話を用いた．Hazumi はシステムとユーザの発話対毎に，システムが現在の話題を継続すべきか否かのラベル（1 から 7 の範囲）を 5 名のアノテータが付与している．本研究では対話継続意欲の高低を識別するために，4 以下を低群，4 を超える場合を高群として，対話継続意欲を 2 クラスに判別する．

3　対話継続意欲推定に用いる特徴量

対話継続意欲を推定するため，本研究では以下の特徴量を抽出した．

3.1　音響特徴量の抽出

対話継続意欲の変化はユーザの声音変化に表出されると考えられる．そこで，音響解析ツール OpenSMILE を用いて，声音変化による感情抽出研究に用いられる， F0(基本周波数)，MFCC(Mel-Frequency Cepstral Coefficient)など計 384 次元の音響特徴量を対話音声から抽出した．

3.2　視覚特徴量の抽出

ユーザの表情や姿勢は対話継続意欲に応じて変化すると考えられる．そこで，顔分析ツール OpenFace を用いて，顔の特徴点の 2 次元座標からフレーム間速度の絶対値の最大値などを特徴量として抽出した．更に，Kinect データから得られる上半身の関節部の 3 次元座標からも同様にして特徴量を抽出し，計 87 次元の視覚特徴量を得た．

3.3　言語特徴量の抽出

ユーザ発話に含まれる単語や内容は対話継続意欲に応じて変化すると考えられる．そこで，Google Speech API よりユーザ発話の書き起こしデータを取得した．次に日本語形態素解析器 MeCab より，発話毎の名詞，形容詞などの形態素数を抽出した．次に，それらの情報をもとに発話毎に Bag-of-Words(BoW)によるベクトル表現を得た．また，対話意欲はユーザの対話内容だけではなく，その前後のシステム発話との関連も考慮すべきである．そのため，システム発話の発話意図を表す 9 種類の対話行為をもとに得た 9 次元の one-hot ベクトルなどをシステム発話から抽出した．ユーザ発話とシステム発話の言語特徴量を合わせた計 2042 次元を言語特徴量として用いる．

3.4　生体特徴量の抽出

生体信号データは外部に表出されないユーザの内部状態を取得するのに有効であると考えられる．そこで，皮膚抵抗データと心拍データから生体特徴量を抽出した．皮膚抵抗，心拍に共通して，最大値，最小値，平均値などの統計量を特徴量とした．また，皮膚抵抗では GSR(Galvanic Skin Response)として皮膚抵抗値のピーク数を特徴量に加えた．心拍では，心拍データから求めた心拍間隔より抽出した特徴量を加え，計 30 次元の生体特徴量を取得した．

4　評価実験

3 章で求めた特徴量を用いて対話継続意欲の識別を行う．過去の対話から対話継続意欲の変化を捉えた識別を行うため，時系列予測に適した RNN(Recurrent Neural Network)を用いた．また，ハイパーパラメータを Optuna により調整した．先行研究と比較するため，DNN(Deep Neural Network)を用いた識別も行った．評価方法は Day-forward-Chaining と呼ばれる交差検証であり，分割数は 5 とした．比較結果を表 1 に示す．

表 1　識別モデル毎の正解率

モデル	RNN	DNN
正解率[%]	76.35	74.49

表 1 から，DNN よりも RNN の正解率が高く，76.35%の精度を得た．RNN では前後の状態を考慮した状態遷移が行われるため，DNN と比較して前後の対話の繋がりを考慮することが，対話継続意欲の推定に有効であることが示唆された．紙幅の関係で図表は掲載できないが，RNN を用いて音響，視覚，言語，生体のモダリティ毎に識別を行ったところ，全てのモダリティを使用した場合に最も高い正解率を得た．このことから，本研究で用いた特徴量を全て組み合わせることが有効であると分かった．

5　おわりに

本研究では，対話中のマルチモーダル情報と対話履歴を用いた対話継続意欲の推定を行った．評価実験より，対話継続意欲を 76.35%で推定することができ，時系列データを考慮した識別手法が有効であることが分かった．

今後の課題は，複数モダリティの扱いに適した RNN のネットワーク構造について検討することである．

参考文献

[1] 別所他：雑談対話における話題継続願望判定の検討，人工知能学会研究会資料 SIG-SLUD-B5-1-01， pp.1-6(2015)

[2] 駒谷他：マルチモーダル対話コーパス Hazumi 公開と生体信号を含む新規データ収集，人工知能学会研究会資料 SIG-SLUD-C002-35，pp.170-177(2020)．