

マルチモーダル対話における興味の有無の推定と追加コーパスを用いた性能評価

Estimating Interest Level in Multimodal Dialogues and Evaluating the Performance with Newly Collected Data

西本 遥人^{1*} 武田 龍¹ 駒谷 和範¹
Haruto Nishimoto¹ Ryu Takeda¹ Kazunori Komatani¹

¹ 大阪大学産業科学研究所

¹ The Institute of Scientific and Industrial Research (ISIR), Osaka University

Abstract: We have developed a method to estimate users' interest level from multimodal information to achieve adaptive system responses to the users' interest. The multimodal information consists of dialogue-level, prosodic, utterance contents, and face image information. These four types of information are extracted as features, with which we trained a classifier for the estimation. The estimation performance was evaluated with multimodal corpus with 39 participants in total including new data with 29 participants. We also compared the performance by considering disagreement among multiple annotators.

1 はじめに

対話における人の発話には、その言語内容だけでなく、様々な有用な情報が含まれる。実際、対話分析においては、相手の声の韻律情報、顔の表情等の非言語情報も重要である [1]。それらの情報により、言語内容だけでは表現できない人の心的状態を記述できる可能性がある。

本研究では、人（ユーザ）とシステムの対話において、ユーザの心的状態を推定し、それをを用いて応答を生成する音声対話システムの構築を目指す。ユーザの心的状態の推定の一例として、ユーザの対話内容への興味の有無の推定を行う。対話コーパスから抽出したマルチモーダル情報とその興味の有無のラベルをもとに、機械学習を用いた推定を行う。

本稿では、推定の性能向上のための取り組みと、推定結果の分析について述べる。まず推定性能の向上のために、新たに利用可能となった対話コーパスを利用した場合の結果や、素性の追加改良について報告する。続いて、ラベルの一致度と推定性能の関係に着目した分析について述べる。興味の有無を第三者が判定する場合、その意見は必ずしも一致しない。そのため、興味の有無の推定が容易な発話対と困難な発話対が存在すると考えられる。そこで、アノテータが付与したラ

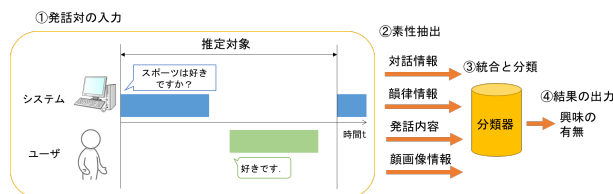


図 1: 興味の有無の推定手法の全体像

ベルの一致度を手がかりとして、推定性能との関係を示す。

2 データの追加と素性の改良

SLUD 研究会内のワーキンググループにより、昨年報告されたコーパス [2] と今年新たに報告されるコーパス [3] を使用する。対話内容はシステムが決められた話題に関する質問や情報提供を行うものである。[2] のコーパスには 6 名のアノテータ、[3] のコーパスには 6 名または 3 名のアノテータによって発話対ごとにユーザの興味の有無を示すラベルが付けられている。今回はアノテータの人数にかかわらず、過半数の意見の一致で興味あり、なしの二値の正解ラベルを作成した。過半数の意見の一致がない発話対はデータとして使用しない。対象としたデータには合計 39 名（うち 10 名は [2]、29 名は [3] のコーパス）のユーザ、2578 の発話対

*連絡先: 〒 567-0047 大阪府茨木市美穂ヶ丘 8-1
大阪大学産業科学研究所 駒谷研究室
E-mail: nishimoto@ei.sanken.osaka-u.ac.jp

が含まれる。そのうち、興味ありが1162件、興味なしが1416件存在し、全て「興味なし」と判定した場合の正解率は0.549である。

各発話対から抽出できる素性と正解ラベルをもとに機械学習で分類器を構築し、興味の有無の推定を行う。図1に機械学習を用いた興味の有無の推定手法の全体像を示す。入力として、後述のマルチモーダル情報の素性を用いる。出力となる推定結果は興味の有無の二値である。

以前の研究[4]で設計したマルチモーダル素性（対話情報、韻律情報、発話内容、顔画像情報）を改良したものをを用いた。改良点は以下の通りである。

【対話情報】システム発話の対話行為を追加。

【韻律情報】openSMILE[5]で抽出される素性から、特徴選択により素性数を10に削減¹。

【発話内容】ユーザ発話中の単語からbag-of-wordsとして作成、主成分分析で次元数を5に圧縮。

【顔画像情報】人の顔の動きを記述するAction Unit(AU)[6]に変更。代表的な17のAUの出現の有無、18のAUの強度に関する値を使用。

3 統合結果と分析

2章で述べたマルチモーダル素性を統合して推定と評価を行った。統合は次の手順で行った。まず、4つのマルチモーダル情報をそれぞれ単独で用いてサポートベクトル回帰を行った。次にそれらの回帰の結果得られる値を説明変数として、再度サポートベクトル回帰を行った。推定性能の評価は、未知のデータに対する性能を測るために、ユーザごとに分割した交差検証により行った。

素性の改良により、推定の性能が向上した。各モーダル素性のみを単独で用いた単一推定と全て用いた統合推定を行った。表1は素性の改良前後での単一推定、統合推定の結果を示している。統合の推定での性能が改良前（正解率:0.697, Macro F1:0.694）に比べ、改良後（正解率:0.711, Macro F1:0.707）の方が上回っていることが分かる。

次に、アノテータが付与したラベルの一致度と推定性能の関係を調査した。今回正解とした興味の有無は複数のアノテータによる多数意見により決定した。したがって、すべてのアノテータの意見が一致した発話対とアノテータの意見が分かれた発話対が同じラベルとして扱われる。前者を人が興味を推定することが容易なもの、後者を困難なものとして、それらに対する推

表 1: 素性改良前後での推定の結果

素性	改良前		改良後	
	正解率	Macro F1	正解率	Macro F1
対話	0.611	0.609	0.668	0.649
韻律	0.578	0.598	0.638	0.638
発話	0.694	0.693	0.701	0.699
顔画像	0.566	0.543	0.602	0.568
統合	0.697	0.694	0.711	0.707

表 2: 人の意見の一致度による推定性能の違い

発話対	データ数	正解率	Macro F1
easy	1723	0.750	0.748
hard	855	0.634	0.626
all	2578	0.711	0.707

定性能を比較する。推定に使用した発話対のうち、アノテータの意見が全員一致（アノテータが6名の場合、5名以上の一致）しているものはeasy、それ以外はhardと定めた。

全データをeasy, hardに分類したときの推定結果を表2に示す。allは全データを用いたときの結果である。easyの性能（正解率, Macro F1）はhardの結果を上回っていることが確認できる。すなわち、アノテータのラベルの一致度が、構築した分類器の推定性能に関与していることが分かる。

今後は、興味の有無が推定できる対話システムの構築に取り組み、そのシステムを用いた対話におけるユーザの印象評定実験を行う。

参考文献

- [1] Mark L. Knapp, *et al.*: Nonverbal Communication in Human Interaction, *Cengage Learning*, 2013
- [2] 荒木 雅弘 他: マルチモーダル対話データの収集と興味判定アノテーションの分析, 第8回対話システムシンポジウム, 2017
- [3] 駒谷 和範 他: 配布可能なマルチモーダル対話データの収集とアノテーション不一致傾向の分析, 第9回対話システムシンポジウム, 2018
- [4] 西本 遥人 他: 対話におけるマルチモーダル情報を用いたユーザの興味の有無の推定, 第32回人工知能学会全国大会, 2018
- [5] Björn Schuller, *et al.*: The interspeech 2009 emotion challenge, *Tenth Annual Conference of the International Speech Communication Association*, 2009
- [6] Paul Ekman, *et al.*: Facial Action Coding System: A Technique for the Measurement of Facial Movement., *Consulting Psychologists Press*, 1978
- [7] Brandon Amos, *et al.*: OpenFace: A general-purpose face recognition library with mobile applications, Technical report, CMU-CS-16-118, CMU School of Computer Science, 2016

¹scikit-learn の selectKbest を利用。