

感情極性値：

高村ら [2] によるスピンモデルによる単語の感情極性抽出の研究に基づいて作成された「単語感情極性対応表」の日本語版を使用して計算した値で、-1 から +1 の実数値で -1 に近いほどネガティブ、+1 に近いほどポジティブを表している。

本実験で用いる感情極性値の計算には、まず MeCab を用いて形態素解析を行い、自然文である 1 発話のテキストを単語に分割する。その後、感情極性対応表内の単語と一致した単語の感情極性値を式 (3) により、ネガティブとポジティブの度合いが大きい範囲のみを足し合わせて最後に割った値を用いる。1 発話内容の中で、感情極性対応表内の単語と一致する単語を w_i とし、その単語の感情極性値を $p(w_i)$ とし、 $p(w_i) \geq |0.50|$ に一致した語数を N とすると、

$$F = \begin{cases} \frac{1}{N} \sum_{i=1, N} p(w_i) & (N > 0, p(w_i) \geq |0.50|) \\ 0 & (N = 0) \end{cases} \quad (3)$$

ここでは、例えば「外 (そと)」と「外 (がい)」など同じ単語で読み方が異なる単語に関しては、形態素解析の誤りが発生しやすいため除外する。

4. モデルの評価

本章では、対立度合いの評価値を推定する提案手法の有効性を確認する為に、構築したモデルの評価実験を行う。

対立場面の認識を、対立的な発話か否かの 2 値分類タスクとして扱い、機械学習手法にはソフトマージン SVM を用いた。5 つのセッションの交渉対話がコーパスに含まれており、評価実験では 4 つのセッションから取得されたデータを訓練データとし、残り 1 セッションから取得されたデータをテストデータとして、5 分割交差検証を行った。

表 3 に対立的な発話の推測を行った機械学習の結果を示す。音声情報のみからの推定結果は、全セッションの Accuracy の平均値が 60.7% であった。また、対立的な発話と非対立的な発話の双方の推定結果を平均した結果では、Recall が 59.8%、Precision が 59.9% で、F 値が 59.8% であった。

言語情報を用いた場合、全セッションの Accuracy の平均値は 56.7% であった。また、対立的な発話と非対立的な発話の双方の推定精度を平均した結果、Recall は 60.6%、Precision は 62.6% で、F 値は 58.9%

表 3 機械学習結果

		Audio Only	Language Only	Latest Fusion
Accuracy		0.607	0.567	0.628
Conflict	Recall	0.575	0.807	0.725
	Precision	0.598	0.571	0.616
	F-measure	0.586	0.668	0.666
Non-conflict	Recall	0.622	0.405	0.557
	Precision	0.599	0.682	0.674
	F-measure	0.610	0.509	0.610
Conflict & Non-conflict	Recall	0.598	0.606	0.641
	Precision	0.599	0.626	0.645
	F-measure	0.598	0.589	0.638
Average				

となった。Accuracy は音声情報を用いた推定精度より下がったものの、いずれも音声情報と非常に近い精度であった。

音声情報と言語情報の両方を用いて、それぞれの機械学習の出力値を足し合わせる Latest fusion の手法を適用した結果（ここでは音声情報 0.8、言語情報 0.2 の割合）、全セッションの Accuracy の平均値は 62.8% となり、それぞれ単体の結果よりも上回った。また、対立的な発話と非対立的な発話の双方の結果を平均した推定結果に関して、Recall は 64.1%、Precision は 64.5% で、F 値は 63.8% となり、いずれの値もそれぞれ単体の結果よりも上回る結果であった。

5. 考察

本実験で得られた機械学習結果の 5 つのセッションそれぞれに着目したところ、セッションによって推定精度のばらつきが大きかった。音声情報の推定結果に着目するとセッション 5 の精度が特に低くなっており (Accuracy : 45.6%)、言語情報の推定結果に着目するとセッション 3 の精度が特に低くなっている (Accuracy : 35.8%)。

これらの理由として考えられることの一つに、まず音声情報の推定結果は、セッション 5 の対話内容に弱気で小さな声の非協力的な音声が多く含まれていたことが挙げられる。

次に言語情報の推定結果は、セッション 3 の対話内容に「火事にならないようにお手伝いいたしましょうか」などの「火事」というネガティブな単語を後から打ち消す場合が多かったことが挙げられる。

本提案手法の音声情報のみ・言語情報のみを用いた場合では著しく精度が下がっていたセッション (3, 5) に Latest fusion を適用した推定結果、精度が向上した。(セッション 3 Accuracy : 52.6%, セッション 5 Accuracy : 63.9%)

音声情報と言語情報を組み合わせることで 1 つずつの手法では顕著に低い精度を示していたデータの

推定精度が向上した。

また、本実験で用いた感情極性値が非協力的な態度の推定に有効な特徴量であったかを調査する為に t 検定を行った。その結果、p 値が 3.58^{-09} の有意な特徴量であることが分かった。そのため、本実験で用いた感情極性値は、非協力的な態度の推定に有効な特徴量だと考えられる。

6. 結論

本研究では、対面二者対話の交渉をマルチモーダル特徴量から推定するモデルを構築し、その評価を行った。マルチモーダル情報には、交渉対話を韻律情報に加えて、発話時間の特徴量を含む音声情報と、感情極性値を用いた言語情報を用いてモデルを構築した。ソフトマージン SVM を用いて、対立的な発話と非対立的な発話の双方の結果を平均した結果では、62.8%の精度で推定出来ることを示した (Recall 64.1%, Precision 64.5%, F 値 63.8%)。

今後の課題として、まず今回の実験では使用したデータサンプルの総数が少なかったため、機械学習の精度を上げることが難しかった。この原因はデータ収集の段階で人手による作業量が多く、十分なデータ量を集めるのに人的なコストがかかるという問題があった。そのため、この工程では既存の交渉対話ログを探してきて用いるなど、さらに改善して人手による作業量を削減する必要がある。

また、言語情報に採用出来た特徴量の種類が少なかった為、他にもモデル生成に有効な特徴量を調査して追加し、言語情報のみの推定精度を向上させる必要がある。さらに、音声情報と言語情報を組み合わせたよりよい特徴を調査する必要がある。

参考文献

- [1] M. Pantic, A. Pentland, A. Nijholt, and T. S. Huang: Human Computing and Machine Understanding of Human Behavior: A Survey, Artificial Intelligence for Human Computing, Vol. 4451, pp. 47-71, (2007)
- [2] Hiroya Takamura, Takashi Inui, and Manabu Okumura: Extracting Semantic Orientations of Words using Spin Model, Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL2005), pp. 133-140, (2005)
- [3] Sunghyun Park, Stefan Scherer, Jonathan Gratch, Peter Carnevale, and Louis-Philippe Morency: Mutual Behaviors during Dyadic Negotiation: Automatic Prediction of Respondent Reactions, Humaine Association Conference on Affective Computing and

Intelligent Interaction (ACII), (2013)

- [4] 岡田 将吾, 松儀 良広, 中野 有紀子, 林 佑樹, 黄 宏軒, 高瀬 裕, 新田 克己: マルチモーダル情報に基づくグループ会話におけるコミュニケーション能力の推定, 人工知能学会論文誌, Vol. 31, No. 6, pp. 1-12, (2016)
- [5] 野本 済央, 政瀧 浩和, 吉岡 理, 高橋 敏: 韻律的特徴と対話的特徴を用いた怒り通話音声の抽出, 電子情報通信学会技術研究報告, Vol.110, No. 143, pp. 7-12, (2010)
- [6] Kushal Dave, Steve Lawrence, and David M. Pennock, Mining the peanut gallery: opinion extraction and semantic classification of product reviews, Proceedings of the 12th international conference on World Wide Web, pp. 519-528, (2003)
- [7] R. Fernandez, R. W. Picard: Classical and Novel Discriminant Features for Affect Recognition from Speech, Interspeech, pp. 473-476, (2005)
- [8] E. Zwicker, H. Fastl: Psychoacoustics, Facts and Models, (1990)