

マルチモーダル対話におけるユーザごとの心象推定のための学習データの割当て

久保 裕之輔[†]羅 兆傑[†]武田 龍[†]駒谷 和範[†][†] 大阪大学 産業科学研究所

1. はじめに

対話システムにおいて、ユーザの心象を推定できれば、それに適応した応答が可能となる。機械学習を用いて人間の発話から心象や感情を推定する研究は数多く存在する。Han ら [1] による音声から感情認識を行う研究や、Bertero ら [2] による対話システムのユーザの感情や心象を認識する研究が挙げられる。機械学習による心象推定は一般に学習用のデータをすべて用いたモデルで行うが、ユーザは多様である。十分に膨大なデータで学習を行えば問題はないが、そうでない場合は心象推定の性能が落ちてしまう。そのため、評価用ユーザと同様の分布を持つユーザのデータを使って学習することで、よりユーザに適した学習が行える。

学習データの選択を行う研究として、秋山ら [3] によるものが挙げられる。これは、音声を用いた感情認識において学習話者の選択をし、精度向上を図った研究である。この研究ではユーザ間の類似度に基づいて学習話者を選択している。類似度は二通りの手法で測られている。その一つは、各話者の特徴量の平均値をとったベクトル間の距離で類似度を比較するという手法である。

本稿では、マルチモーダル対話において、ユーザ単位の適切な学習データ集合を設定することで、心象推定の性能向上を図る。マルチモーダル情報からユーザ間の類似度を計算し、学習データを選択する。また、秋山ら [3] の研究では評価するユーザごとにモデル学習を行う必要がある。本研究ではクラスタリングを利用することで事前に心象推定モデルを複数学習しておき、評価するユーザごとにモデル選択を行う方法をとる。これによって、心象推定するユーザを変えるたびに学習する必要がなく、推定の高速化につながる。

2. ユーザのクラスタリングによる学習データ選択

想定する処理の全体像を図 1 に示す。まず学習時に、ユーザをクラスタリングし、クラスごとに、そこに属するユーザのデータを用いて心象推定モデルを学習する。評価時には、学習で得られたモデルから対象ユーザに最も適したものを選択して用いることで、心象推定の精度向上を図る。

クラスタリングの手法には k-means 法を用いる。距離尺度として秋山ら [3] による研究にならい、特徴量の平均ベクトルを用いる。ユーザ単位で各特徴量の平均をとったベクトルを用いてクラスタリングする。

実行時の学習データの選択は、各クラスターの重心と、テストデータのユーザとの距離に基づいて行う。この距離が最も短い重心を持つクラスターを特定し、それに属するユーザのデータによる機械学習モデルを用いる。

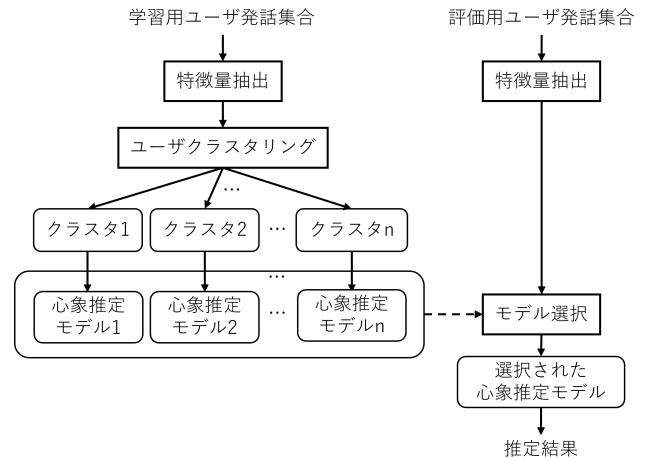


図 1: ユーザのクラスタリングを利用して学習したモデルの選択による心象推定

3. 評価実験

3.1 使用データ

実験には Hazumi1911 [4] を使用した。これはシステムと人との一対一の雑談を収録したマルチモーダル対話コーパスである。20 代から 70 代の男女 26 人分のデータが含まれる。本実験に用いるデータは公開されているダンプファイルを使用した。

特徴量には、韻律特徴量、顔表情特徴量、動作特徴量、言語特徴量、生体特徴量を用いた。それぞれ、384 次元、66 次元、20 次元、2601 次元、4 次元である。

また、ユーザの心象を表すラベルとして、第三者心象を用いた。これはアノテータ 5 名が、第三者視点からユーザの心象を推測して付与したものであり、低群、中群、高群の三段階でラベル化されている。ラベルの数は低群が 178、中群が 881、高群が 1380 である。

3.2 実験設定

本実験ではサポートベクターマシン (SVM) を用いた。カーネルは RBF カーネルであり、ハイパーパラメータは $C=1$, $\gamma=\frac{1}{n_feature \times var}$ とした。 $n_feature$ は入力の特徴量の次元数、 var は入力の分散である。また、学習データを用いて各特徴量を平均が 0、分散が 1 になるように標準化を行った。また、評価指標は発話単位での心象ラベルの正解率をユーザごとに算出し、そのマイクロ平均とする。

3.3 学習に用いるユーザ集合の事前調査

クラスタリングに基づく手法の評価の前に、まず学習に用いるユーザを適切に選ぶことで心象推定の性能が向上することを確認する。つまり、テストデータとする各ユーザごとに、心象推定の性能が上限となるユーザ単位の学習データ集合を、事後的に求める。

本調査では、このユーザ単位の学習データ集合を後ろ

train test	F2001	F2002	F3001	F3002	F3003	F4001	F4002	F4003	F5001	F5002	F6001	F6002	F6003	F7002	M2001	M2002	M2003	M4001	M4002	M5001	M5002	M6001	M6002	M6003	M7001	M7002
F2001	-	○			○	○		○		○				○		○	○				○			○	○	
F2002		-	○				○	○		○				○		○	○								○	
F3001			-			○	○		○	○				○		○	○								○	
F3002				-		○		○		○				○		○	○	○							○	
F3003					-	○	○	○						○						○	○			○		
F4001	○				○	○	-		○					○		○	○			○				○		
F4002		○	○				-		○	○				○		○	○								○	
F4003					○		○	-	○							○	○								○	
F5001		○	○		○				-	○				○				○							○	
F5002					○	○	○	○		-						○									○	
F6001	○	○							○	-						○			○							○
F6002						○	○			○		-		○			○					○		○	○	
F6003	○						○			○	○		-			○	○								○	
F7002			○	○		○	○	○						-		○	○								○	
M2001		○	○	○		○	○		○	○				○	-	○	○					○		○	○	
M2002		○	○	○	○	○	○		○					○		-									○	
M2003				○		○			○	○				○		○	-					○			○	○
M4001														○				-	○						○	
M4002									○	○				○	○	○	○		-		○	○	○	○	○	○
M5001									○					○	○		○			-	○	○	○	○	○	○
M5002		○	○						○					○							-			○	○	○
M6001						○	○					○			○	○	○		○	○		-		○	○	○
M6002												○	○	○	○	○		○	○	○	○		-	○	○	○
M6003														○		○		○	○	○	○			-		
M7001		○					○		○					○				○	○				○		-	
M7002									○		○		○					○	○	○		○		○	○	-

図 2: 正解率が最大となったときに学習データとしたユーザの一覧

表 1: 学習データの選択の有無ごとの正解率

-	なし	あり
正解率	0.780	0.810

表 2: クラスタ数ごとの正解率

クラスタ数	1	2	3	4	5
正解率	0.780	0.768	0.764	0.767	0.769

向き探索で求める。実験は以下の手順で行い、正解率が最も高くなる学習データ集合を特定する。

1. テストデータに用いるユーザ以外の全ユーザのデータで学習を行う。
2. 学習データから除いた場合に、テストデータの心象推定の結果が最も良いユーザを一人特定する。
3. そのユーザのデータを学習データから除く。
4. 2, 3 を学習データとするユーザが一人になるまで繰り返す。

結果を表 1 に示す。学習データの選択なしとは、テストデータに用いたユーザ以外の全ユーザのデータを学習に使った場合の心象推定の正解率を表す。学習データの選択ありは、学習に使うユーザ数を減らしていく中で、最大であった正解率を表す。学習データの選択なしは 0.780, ありは 0.810 であった。この結果から、適切なユーザのデータを学習に使うことで、全ユーザのデータを学習に使う場合よりも心象推定の性能が向上するということがいえる。

図 2 は、テストデータにしたユーザごとに、正解率が最大となったときに学習データに用いたユーザを示す。各行がテストデータにしたユーザ、各列が学習データにしたユーザを示しており、各行のユーザに対する心象推定モデルの学習に使ったユーザの位置に○を記載している。各ユーザは ID 5 文字で表している。この結果から、学習データに含めるべきユーザがテストデータとするユーザごとに異なることがわかった。また、学習データとするユーザ数もテストデータのユーザごとに異なり、最大は 13 人、最小は 4 人、平均は 7.6 人であった。

3.4 実験結果

クラスタリング結果を用いて自動でモデルを選択した場合の実験結果を表 2 に示す。評価方法には一人抜き

交差検証を用いた。テストデータとしたユーザ以外の全ユーザを学習データとした場合をクラスタ数 1 としている。クラスタ数は、k-means 法で設定したクラスタ数で、そのクラスタごとに学習したモデルでの心象推定の正解率を表している。これに示したとおり、平均ベクトルによるクラスタリングによって心象推定の性能は上がらなかった。クラスタリングの距離尺度や、心象推定モデルの選択方法を改善する必要がある。

4. おわりに

本稿では、事前調査として、テストユーザの心象推定の性能が上限となるユーザ単位の学習データ集合を実験によって得た。ユーザごとに算出した特徴量の平均ベクトルによってクラスタリングを行い、学習に用いるデータを選択する手法を提案した。

事前調査で心象推定の性能が上がる学習データ集合が得られたため、今後は、これに近い結果が得られるクラスタリングの距離尺度を検討する。

参考文献

[1] Kun Han, et al. Speech emotion recognition using deep neural network and extreme learning machine. Proc. Interspeech, pp.223-227, 2014.

[2] Dario Bertero, et al. Real-time speech emotion and sentiment recognition for interactive dialogue systems. Proc. of EMNLP. pp.1042-1047, 2016.

[3] 秋山大知ほか. 音声を用いた感情認識のための学習話者の選択. 日本音響学会誌, Vol.76, No.10, pp.554-561, 2020.

[4] 駒谷和範ほか. 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.