4Y-5

感性会話ロボットのためのベイジアンネットワークを用いた対話者感情推定

趙章植　加藤 昇平　伊藤 英則

名古屋工業大学 情報工学科

1　はじめに

近年，多くのエンタテイメントロボットが開発されており，人間との円滑なコミュニケーションを目的としたロボットの研究が盛んに行なわれている [1]．ロボットと人間とのより豊かなコミュニケーションのためには，お互いの感情や情動を把握する必要がある．その実現には，人間が後天的に学習し獲得している「対話者感情を理解する知能」の推論モデルをロボットに持たせることが必要であると考える．一方，ベイジアンネットワークとよばれる確率的な知識表現と推論の枠組が人工知能の分野で研究されてきており，不確実な知識の下での知識表現能力と推論効率の高さから，近年，様々な分野へ応用されるようになってきた．そこで本研究では，感性会話ロボット Ifbot（図 1）の対話者感情推定手法として発話音声を用いたベイジアンネットモデルを提案する．

図 1: Ifbot

2　ベイジアンネットワーク

ベイジアンネットワーク（BN）は，複数の確率変数の間の定性的な依存関係を非循環有向グラフ（DAG）により表現し，個々の変数の間の定量的な関係を条件付確率で表した確率モデルである [2]．確率変数をノードとし，変数間に確率的依存関係が強いと判断される場合に対応するノード間に有向リンクを付ける．依存関係を確率的相関と同一視した場合，N 個の確率変数（$X_1, ..., X_n$）の同時確率分布 P は次式で表現される．

$$P(X_1, ..., X_n) = \prod_{i=n}^{n} P(X_i|Pa(X_i)). \qquad (1)$$

ここで，$Pa(X_i)$ は確率変数 X_i の親ノードを表す．式 (1) は，各ノード X_i が $Pa(X_i)$ にのみ依存し，X_i から辿って到達できるノードを除いた他のノードとは条件付独立となることを表している．

親ノードがある状態 $Pa(X_i) = x$（x は親ノード群の各値で構成したベクトル）の下での n 通りの離散状態（$y_i, ..., y_n$）を持つ変数 X_i の条件付確率分布は $p(X_i = y_1|x), ..., p(X_i = y_n|x)$ となる．これを各行として，親ノードがとりうる全ての状態 $Pa(X_i) = x_1, ..., x_m$ のそれぞれについて列を構成した表の各項目に確率値を定めたものが X_i についての条件付確率表（CPT）である．これにより，確率変数間の確率的な依存関係がモデル化できる．ベイジアンネットワークを用いて知識をモデル化することで，知識の記述量及び計算量が大幅に削減される．また部分的な証拠からでも確率的に推論できる長所を持つ．このため本研究では，ロボットに搭載する感情推定のための知識モデルとして効率とロバスト性を得ることが可能なベイジアンネットワークを応用する．

3　音声の韻律特徴を用いた感情推論器の学習

本研究では，対話者が発話した音声から対話者感情を推定する為に音声に対する感情推定知識をベイジアンネットワークとしてモデル化した．本節では，モデル構築の流れについて概説する．

3.1　音声資料

使用する音声資料は，感情表現がなされている必要がある．本研究では TV ドラマ，映画から俳優が感情を込めて発話したフレーズを抽出し，「怒り」「悲しみ」「嫌悪」「恐怖」「驚き」「喜び」の 6 種類（以降，6 感情）に分類した．それらの中から，聴取実験により感情が適切に表現されていると判断された音声資料をサンプルデータとする．

3.2　特徴量の抽出

音声は，3 つの要素（韻律，音質，音韻）から成り立っている．この中で，韻律的特徴が人間の感情表現に最も関連することが過去の様々な研究から明らかになっている（例えば文献 [3]）．そこで本研究では，音声資料からピッチ構造を反映する「基本周波数」（$F0$），振幅構造を反映する「短時間パワー」（PW），及び時間構造を反映する「1 モーラあたりの発話継続時間」（Tm）をそれぞれ計測する．$F0$ 及び PW に関しては，平均，最大，最小，標準偏差を抽出した．このとき，短時間分析におけるフレーム長を 23ms（250 samples），フレーム周期を 11ms とし，窓関数として Hamming 窓を使用した．以上により求めた 9 個の特徴量に加えて，発話者の性別（SE）を加えた 10 個の特徴量をベイジアンネットワークの確率変数とする．

3.3　音声特徴の量子化

3.2 節で述べた音声資料から抽出した各特徴量の統計量は連続値を取るため，離散状態を扱うベイジアンネットワークに適用するためには特徴量の量子化が必要である．量子化数はベイジアンネットワークの学習に必要なデータ量や学習データにおける各特徴量の存在分布により適切に設定する必要がある．

3.4　モデルの構造決定

本研究では，学習データに含まれる目標属性（6 感情）と属性（音声韻律特徴）との間の依存関係を表現す

*Bayesian-Based Inference of Dialogist's Emotion for Sensitivity Robots, Jangsik Cho, Shohei Kato, and Hidenori Itoh, Department of Computer Science and Engineering, Nagoya Institute of Technology, Gokiso-cho, Showa-ku, Nagoya 466-8555, Japan.

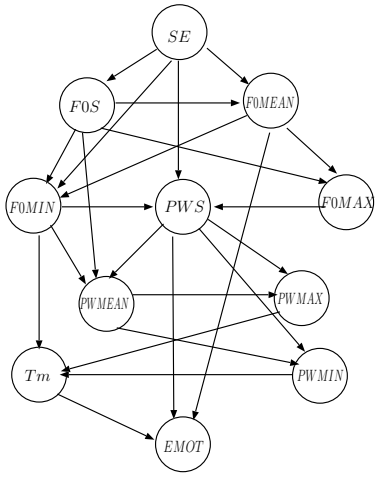

図 2: ベイジアンネットワークモデル

るために属性間の結合とその強さ（CPT）を学習することでベイジアンネットワークの構造を決定する．学習方法として，本研究では情報理論的妥当性がありデータへの過度なフィットを回避することで予測精度の高いモデルが学習可能な BIC（Bayesian Information Criterion）に基づくモデル選択手法を採用する．M をモデルとし，$\hat{\theta_M}$ を M を表すパラメータ，d をパラメータ数とすると M の評価値 $BIC(\hat{\theta_M}, d)$ は次のように定義される．

$$BIC(\hat{\theta_M}, d) = \log^N_{\hat{\theta_M}} P(D) - \frac{d \log N}{2} \qquad (2)$$

ここで D は学習データ，N は D のデータ数を表す．$\hat{\theta_M}$ は最尤法により求めた．D が部分観測の場合には EM アルゴリズムを用いて推定し CPT を補間する．本研究では，BIC が最大となるモデルを求めこれを対話者の感情推定のための知識としてロボットに与える．BIC を最大にするモデルの探索には K2 アルゴリズムを用いた．K2 アルゴリズムではノード間の親子順序を事前知識として与えることで探索空間を制限することが可能だが，ここでは音声の振幅（PW），ピッチ（$F0$），時間（Tm），及び，性別（SE）の 4 グループに分け，グループ内のノードのオーダリングを固定することにより探索空間を軽減させた．そして 4 つのグループの順列組合せについて学習を行い，準最適な構造を決定する．

4 感情推論のアルゴリズム

ベイジアンネットワークの確率推論は十分な情報がないときでも，合理的な意思決定を行うことが可能である．これにより一部の証拠のみが与えられた場合でも，確率推論ができる．本研究では，ネットワーク構造に復結合を持つ場合でも効率的に推論を行うことができる Junction Tree を感情推論のアルゴリズムに採用する．

5 感情推定実験

3 節で述べたように音声資料を 1591 事例用意し，そこから任意に 1387 の学習事例と 204 のテスト事例を作成した．学習データの各特徴量の量子化数はすべて 5 とした．図 2 は学習事例から作成されたベイジアンネットワークモデルを示す．ここで得られた変数グループの順序は $F0S$，$F0MEAN$，$F0MAX$，$F0MIN$，PWS，$PWMEAN$，$PWMAX$，$PWMIN$，Tm であった．

表 1: 感情推定実験結果

感情	正答率（%）		
	提案（BN）手法		PCA
	全証拠（10）	6 証拠	
怒り	70.0	65	20
悲しみ	64.5	38.7	9.7
嫌悪	61.3	45.2	3.2
恐怖	53.8	69.2	12.1
驚き	**51.5**	39.4	16.7
喜び	63.9	22.2	7.7

5.1 情報欠損に対する頑健性評価

評価実験はテスト事例から 10 属性すべての証拠が与えられた場合と 6 証拠（SE，$F0S$，$F0MAX$，PWS，$PWMEAN$，Tm）のみが与えられた場合の 2 つについて行った．表 1（左，中央）に感情推定の正答率を示す．全証拠を用いた実験では 6 感情すべてが認識された．特に従来の研究 [4] では「驚き」の感情が認識されなかったが，今回の実験では学習データを増やし，性別確率変数を追加することで認識された．なお，「恐怖」と「驚き」の音声データを十分収集できなかったため，これらの感情の正答率が低下した．なお，今回の実験では，様々な俳優から音声資料を抽出したため，俳優による音声特徴の格差が正答率に影響している．一方で 6 証拠のみを用いた実験では「悲しみ」「嫌悪」「驚き」「喜び」の正答率が大きく低下したものの，6 感情を無作為に回答した場合（16.7%）を下回ることはなく，6 感情すべてが認識されたといえる．このことから提案手法の頑健性が確認された．

5.2 多変量解析手法との比較

評価実験の有効性を確認するため，マハラノビス距離を用いた主成分分析（PCA）との比較実験を行った．表 1（右）に結果を示す．6 感情すべてにおいて，PCA より本手法の正答率が高いことがわかる．このことから提案手法の有効性が確認された．

6 おわりに

本研究では，音声ベイジアンネットワークを用いて感性ロボットのための対話者感情の推定法を提案した．本手法により，音声情報から対話者の感情を推定することが可能となった．今後の課題として，発話者毎の発話特徴の差異を考慮した学習により感情推定を改善したい．そして，言語や表情などを含めた総合的な感情推論器を構築し ifbot への実装を行う．

参考文献

[1] 竹内将吾，酒井あゆみ，加藤昇平，伊藤英則：対話者好感度に基づく感性会話ロボットの感情生成モデル，第 11 回ロボティクスシンポジア，pp．74-79，2006．

[2] K. B. Korb and A. E. Nicholson，Bayesian Artificial Intelligence，Chapman & Hall/CRC，2004．

[3] 重永 寛：感情の判別分析からみた感情音声の特性，電子情報通信学会論文誌，Vol．J83-A No．6，pp．726-735，2000．

[4] 杉野 良樹，加藤昇平，伊藤英則：ベイジアンネット混合モデルを用いた感性ロボットのための対話者感情の推定法，情報処理学会全国大会，Vol．4，pp．119-120，2006．