4P-5

曖昧な発話に対応可能な音声制御システムの開発

松本友里† 中村真吾‡ 橋本周司‡

早稲田大学 先進理工学研究科† 早稲田大学 理工学術院‡

1. はじめに

近年，様々な音声操作システムに関する研究開発が行われている．しかしながら，命令音声として使用できる言葉の語彙が予め制限されているため，ユーザ独自の言葉や，要求意図が明確ではない曖昧な発話で制御を行うことはできない．筆者らは，未知語や曖昧語を含むユーザの自由な発話から適切に機器操作を行えるシステムの構築を試みている．まず，入力発話を随時学習するシステムを構築した．次に，Bayesian Network を用いて，ユーザの操作履歴や機器の状態などから機器操作の入力語と出力動作を結びつけることで曖昧な発話の意図を推定できる手法を提案し，家電機器の操作に適用してその有効性を検証した[1]．その後、さらにシステムが獲得した新しい単語について，単語間の意味的なつながりによって分類して推論に用いるアルゴリズムを実装したその結果，未知語の学習においてより少ない学習で適切な行動出力を行うことができたので、その概要を報告する．

2. システム概要

発話された言葉は音声認識部でテキストに変換され，各機器の状態とともにシステムに入力される．ここで，重要なことは，誤認識されても認識部の出力テキストがそのまま学習に利用されることである．入力された情報を Bayesian Network によって学習・推論を行い，推論結果から発話とともにシステムの動作を行動として出力する(図 1)．

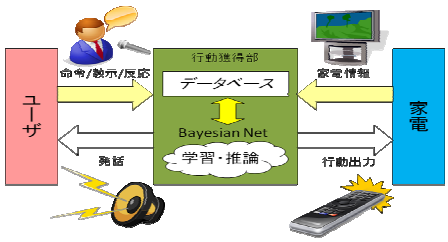


図 1 システム概略図

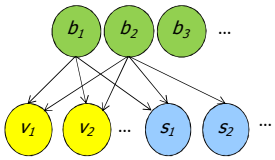
Voice control system applicable to ambiguous utterance input
† Yuri Matsumoto, Graduate School of Advanced Science and Engineering, Waseda University
‡ Shingo Nakamura, Shuji Hashimoto, Faculty of Science and Engineering, Waseda University

2.1 学習

家電状態，発話単語，行動などの事象をノードとして、それらの間の因果関係を図 2 のような有向グラフで表す．本手法では行動ノード B は B_1(暖房 ON にする)，B_2(暖房 OFF にする)など 25 種類，家電状態ノード S は S_1(暖房の ON/OFF 状態)，S_2(暖房温度の高/低状態)のように観測可能な 13 種類を用いた．音声ノード V は，「暑い」や「寒い」などの発話語の状態を持ち，未知語の入力によりその個数は随時可変することとする．行動 $b_j \in B$ が観測されたときに $s_i \in S$ や $v_k \in V$ である確率を $P(s_i|b_j)$，$P(v_k|b_j)$と表記し，全てのノード間の関係を条件付き確率で表す．条件付き確率の算出には累積頻度度数を用いる．$P(V|B)$についても同様に計算する．これらの条件付き確率はリアルタイムで更新される．また，音声については 1 回の入力に対し音声認識候補上位 3 つの単語を取得し，それぞれにおいて条件付き確率を更新する．

$$P(S|B)=\begin{pmatrix} P(s_1|b_1) & P(s_2|b_1) & \cdots & P(s_n|b_1) \\ P(s_1|b_2) & P(s_2|b_2) & \cdots & P(s_n|b_2) \\ \vdots & \vdots & \ddots & \vdots \\ P(s_1|b_n) & P(s_2|b_n) & \cdots & P(s_n|b_n) \end{pmatrix} \qquad P(s_i|b_j)=\frac{n_{ij}}{N_j} \quad (1)$$

N_j : b_jが観測された回数
n_{ij} : b_jが観測されたうち s_iが観測された回数



b_j： 行動ノード
v_i： 音声ノード
s_k： 家電状態ノード

図 2 Bayesian Network

2.2 推論

発話時の S と V の観測情報と学習時に得られた条件付き確率から行動 B の確信度 $BEL(B)$を以下の式より算出する．

$$BEL(B)=\alpha P(B)\prod_{i}^{n}P(s_i|B)\prod_{j}^{m}P(v_j|B) \quad (2)$$

$P(B)$は B の事前確率，α は正規化定数，n は音声ノード数，m は家電状態ノード数を表す．確信度の最高値は 1 であり，各行動において算出される．また，音声においては音声認識候補上位 5 つを利用し，それぞれが観測された場合に

ついて，5 通りの確信度を求め，その中で確信度が最も高い行動を出力とした．

2.3 教示と応答

本システムでは，システムとユーザの全てのやりとりは音声によって行われる．ユーザの望む出力であった場合は，「はい」等の肯定の発話を行うことでシステムは正しい動作であると認識する．ただし，最高確信度が 0 である場合や，出力がユーザに否定された場合は，システムはユーザに音声で教示を求める．ここで，教示の際，ユーザはシステムが予め意味を知っている単語のみを用いる．教示や出力の肯定がされた場合は，その時の行動結果から 2.1 節のようにノード間の確率を随時更新していく．

2.4 獲得した単語の自動分類

システムに入力された全ての単語において，同義語であると考えられるものは同じグループに分類され，グループごとに行動履歴を統合して推論に用いる．ある単語 v_i において，各行動が起きたときにその単語が観測された累積回数のベクトルを $\boldsymbol{X_i}$ とすると，$\boldsymbol{X_i}$ と $\boldsymbol{X_j}$（v_j は v_i 以外の全ての単語）のコサイン距離を計算したものを単語間類似度とし，閾値以上の単語を同グループとする．単語間類似度は，毎試行算出される．

3. 実験と結果

本システムを用いて，家電操作シミュレーションを行った．音声認識エンジンには Julius for SAPI[2]を用いた．認識用の辞書には「あ」から「ん」までの日本語かな文字を基準に登録した．また，教示用にシステムが予め意味を知っている言葉を 27 語設定した．操作対象とした家電は暖房，冷房，扇風機，テレビ，コンポ，照明の 6 種類であり，制御動作はスイッチや温度，音量，風量，明暗，チャンネルである．一回の試行につき家電状態をランダムに設定し，家電状態を観測してから発話を開始する．ユーザの望む結果が出力されるか，またはユーザが教示を行うか，のどちらかが達成された時を一試行の終了とする．

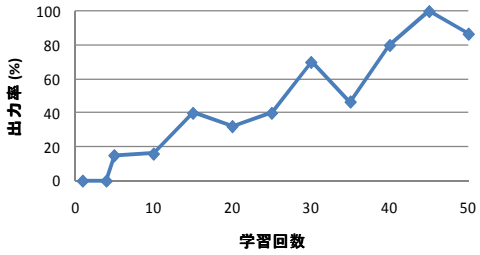


図 3　「さむい」と発話した時の適切な行動の出力率

3.1 未知語の学習

未知語である「さむい」という発話の学習を 50 回試行し，試行ごとの出力率を算出した結果を図 3 に示した．出力率とは，「暖房つける」等の，状態と発話に見合った適切な応答がシステムから出力された確率である．この結果から，未知語が正しく学習されていることがわかる．

3.2 単語分類の有効性

提案システムの他に，単語分類を行わない比較システムを用意し，未知語の学習について比較した．「さむい」という発話を 50 回学習させた後，同義語である「冷えるなぁ」という発話を新たに学習させたところ，出力率の様子は図 4 のようになった．提案システムの方が少ない学習回数で高い確率で適切な応答ができることがわかる．これは，未知語の「冷えるなぁ」が既知語の「さむい」とグループ化されたためだと考えられ，このように単語分類をすることで，未知語が入力されても学習初期から対応できる．

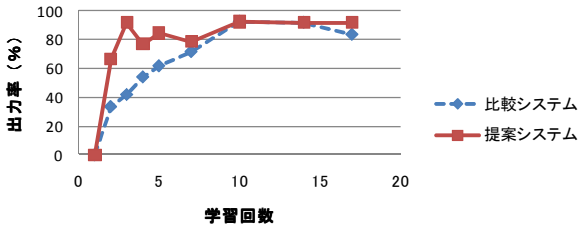


図 4　「冷えるな」と発話した時の適切な行動の出力率

4.まとめ

Bayesian Network による発話意図推定と獲得した単語の意味分類により，未知語の入力を可能にし，より少ない学習で正しい行動出力を行うことができた．

謝辞

本研究の一部は，早稲田大学ヒューマノイド研究所，独立行政法人科学技術振興機構戦略的創造研究推進事業 CREST「人を引き込む身体的メディア場の生成・制御技術」，日本学術振興会グローバル COE プログラム「グローバルロボットアカデミア」の支援を受けて行われた．

参考文献

[1]松本, 中村, 橋本, ”ベイジアンネットワークによる曖昧な発話入力を考慮した家電制御システム”, 情報処理学会全国大会講演論文集，2010

[2]名古屋工業大学 Julius 開発チーム, ”連続音声認識ソフトウェア Julius”
http://julius.sourceforge.jp/