

# 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

松本友里<sup>†</sup> 中村真吾<sup>‡</sup> 橋本周司<sup>‡</sup>

早稲田大学 先進理工学研究所<sup>†</sup> 早稲田大学 理工学術院<sup>‡</sup>

## 1.はじめに

近年、音声により操作可能な情報家電に関する研究開発が盛んに行われている[1]。しかしながら、命令音声として使用できる言葉は予め決められているため、ユーザは決められた単語を使用して家電操作をしなければならない。また、ユーザは「暖房をつけて」のように明確に命令を指示しなければならない。しかし、「寒いな」のように暗に状況改善の要求を意味した曖昧な発話で制御を行うことはできない。そこで本研究では、システムがこれまでの状況から発話意図やユーザモデル等を経験的に獲得し、ユーザがより自由な音声入力によって意図する家電操作を行えるシステムの構築を目指している。その第一段階として、因果関係を確率により記述できる Bayesian Network を用い、行動履歴や家電の状態などから家電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に対応可能なシステムを提案した[2]。このシステムは状態やユーザに合わせて学習し様々な行動の出力が可能であった。しかし、音声入力において入力できる言葉を限定し、認識率を 100%に近い状況で実験を行っていた。これは、音声の入力を自由になると、誤認識して正しい結果が出ないことがあり、さらに音声ノードの状態数を可変にして誤認識語を随時登録しても学習に時間がかかってしまうことが予想されたからである。しかし実際は実環境において音声の誤認識は頻出するものであり、提案したシステムは実用性に欠けていた。

本稿では複数の音声認識候補を考慮することで、不安定な認識においても効率よく学習と推論を行い、より自由な音声入力を可能とするシステムを提案する。

## 2.システム概要

発話された言葉は音声認識部で N-best 認識候補の上位 5 位までがテキストに変換され、各家電の状態とともにシステムに入力される。入力

An appliance control system applicable to ambiguous utterance input using Bayesian Network

<sup>†</sup> Yuri Matsumoto, Graduate School of Advanced Science and Engineering, Waseda University

<sup>‡</sup> Shingo Nakamura, Shuji Hashimoto, Faculty of Science and Engineering, Waseda University

された情報により 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行動を出力する(図 1)。

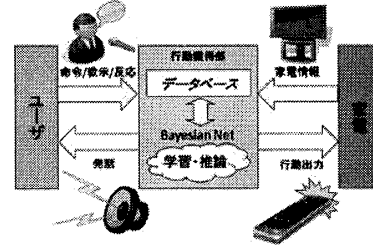


図 1 システム概略図

### 2.1 学習

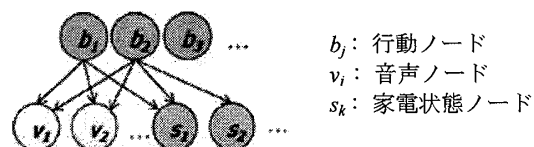
因果関係のある事象同士をノードとし図 2 のような有向グラフで表す。本手法では行動ノード  $B$  は  $B_1$ (暖房 ON であるか)、 $B_2$ (暖房 OFF であるか)など 25 種類、家電状態ノード  $S$  は  $S_1$ (暖房の ON/OFF 状態)、 $S_2$ (暖房温度の高/低状態)のように観測状態ごとに 13 種類である。音声ノード  $V$  は 1 つであり、「暑い」や「寒い」などの状態を 20 種類持ち、未知語の入力によりその個数は随時可変することとする。

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(式(1))。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 N-best 認識候補の上位 3 つ全ての単語において条件付き確率を更新する。

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

$N_j$ :  $b_j$  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,  $\alpha$  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,  $BEL(B) = \{0.0, 0.2, 0.7, \dots, 0.0\}$  のように各行動において算出される. また, 音声においては認識候補が 5 つあるため, 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する.

## 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」等と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. ただし以下の場合, システムはユーザに教示を求める.

- (i) 最高確信度が 0 である場合
- (ii) 出力がユーザに否定された場合
- (iii) 最高確信度が 2 つ以上算出された場合

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, または肯定された場合は, その時の行動結果から 2.1 節のように音声, 家電状態との関係を随時更新していく.

## 3. 実験

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[3]を用いた. 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, 家電状態を見てから発話を開始する.

提案システムの他に, 認識候補を考慮せずに入力された言葉を学習するシステムを用意し, 学習速度と精度の二点について比較した. 表 1 は認識率が平均 58% である「寒い」という発話を 60 回学習した時の実験結果である. 出力開始回数とはシステムがユーザに対して初めて何らかの行動を提示できた時の回数であり, 全体出力率は全試行の中でユーザの望む出力 (暖房: オン, 高温, 冷房: オフ, 低温, 扇風機: オフ, 弱風) がされた回数の割合である. この結果か

ら, 誤認識しやすい言葉において, 提案システムを使用した方が比較システムより高い精度で学習できていることがわかる.

表 1 「寒い」を学習した時

	出力開始	全体出力率
比較システム	16回目	45.0%
提案システム	14回目	60.0%

## 4. 考察

音声認識候補 5 位までの確信度を算出して推論することで, 1 位に正しい認識語がなくても正しい行動を出力できるようになったため精度が向上したと考えられる. また認識候補上位 3 つを全て学習に加えることで, 発音の類似している言葉を同時に学習できるので, 不安定な認識に対しても効率良い学習が実現できた. 認識候補の何位までを学習や推論に用いるかという点については精度への影響を考慮し, 今後適度な値を決定していく必要がある.

## 5. まとめと今後の展望

曖昧な発話の意図を推定可能な家電システムにおいて, 複数の音声認識候補を考慮した学習と推論を行うことで, 誤認識に対応でき, 自由な音声入力を実現することができた. 今後は, ユーザの違いに対応できる等, より実環境下で使用できるシステムを目指していく.

## 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] 内田, 常盤, 西, 高木, 麻生, 橋本, 森, 中島, 伊東, 小林, 八名, 「情報家電の操作のための対話インタフェースの開発」, 情報処理学会研究報告, 2004
- [2] 松本, 中村, 橋本, 「ベイジアンネットワークを用いた曖昧な発話の意図を推定可能な家電制御システム」, 情報処理学会研究報告, 2009
- [3] 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/>