

ベイジアンネットワークを使った曖昧な 発話の意図を推定可能な家電制御システム

松本友里[†] 中村真吾[‡] 橋本周司[‡]

概要：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。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おり、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、「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ず、「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、ベイジアンネットワークを用い、ユーザモデルや環境を学習し、獲得した行動履歴からユーザの発話意図を推定することで曖昧な発話に対応可能な家電制御システムを提案する。

An Appliance Control System Dealing Ambiguous Utterance Information by Bayesian Network

Yuri Matsumoto[†], Shingo Nakamura[‡] and Shuji Hashimoto[‡]

Abstract : Most of researches about voice control appliances use only words that programmed in advance. Therefore, it can deal only obvious utterance input, but cannot do ambiguous one. In this research, we applied Bayesian Network to the appliance control system, and it makes possible to learn user's preference and deal ambiguous utterance through past user's behavior and voice information.

1. はじめに

近年の家電の多様化に伴い、ホームネットワークによって様々な家電やセンサをネットワークで接続し、家庭内の全ての家電を一つのシステムとして制御することが容

易となってきた。それに伴い、情報家電を用いた制御システムに関する研究が盛んに行われている。特に音声を入力とした家電制御手法は、人間との自然なコミュニケーションを通じて制御できるため様々な検討が行われている。対話を通して段階的に機器名や操作名を発話することで、完全な音声コマンドを覚えていなくとも操作できるインタフェースを提案しているものや[1]、入力文の構造に着目し、多様な文体を許容するための手法を提案した研究などが挙げられる[2]。しかし、これらの研究はシステムが単語の意味を予め知っているという前提があり、全くの未知語の入力には対応できないという問題がある。また、決められた単語のみによって指示することは、“自然なコミュニケーション”とかけ離れているばかりでなく、ユーザの負担ともなる。

筆者らは、家電制御システムにおいてユーザとより自然なコミュニケーションを実現するためには、音声によるコミュニケーションだけでなく、さらにいくつかの要素が必要であると考えた。まず、システムは常に知識や経験に基づきユーザや環境に適応しなければならない。特に家庭では開発者が想定できない未知の環境や状況が存在する場合が多く考えられるため、後天的に学習できる必要がある。次に、表面的な発話情報だけでなく、ユーザの曖昧な発話に隠れた意図を汲み取ることが重要である。例えば「暑いね」と発話すればそれは暗に冷房をつけて欲しいという要望だと察知し、「冷房をつける」対応ができるような「気の利く」対応のできる機能を持つ必要がある。

この様に、人間との自然なコミュニケーションを図るためには、音声による入力、状況の学習、そして曖昧な発話の中から意図を推定できる能力が、システムに要求される。そこで本研究では、音声による家電制御システムにベイジアンネットワークを組み込み、ユーザの嗜好を学習することで曖昧な表現の意図を確率的に推定することを試みる。また、実験では、仮想的な家電システムを用い、音声入力によって実際にシステムを動作させ、評価を行った。

2. 研究目的

個々の家電状態やユーザ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開発を目的とする。これにより、以下の項目の実現が見込まれる。

- 家電状態に依存した行動出力（個別家電の ON/ OFF など）が可能である

[†] 早稲田大学 先進理工学研究科

[‡] 早稲田大学 理工学術院

- ユーザの嗜好が考慮された行動出力が可能である

そこで、発話情報と家電状態から意図する動作を獲得するベイジアンネットワークをシステムに組み込むこととした。ベイジアンネットワークは複数の事象の因果関係を確率により表現する推論モデルであり、人間がある行動に至った理由を環境や状況から推論しつつ行動を獲得するのに適していると思われる。

3. 提案手法

3.1 ベイジアンネットワーク

ベイジアンネットワークは不確実性を含む事象の予測や合理的な意思決定、観測結果から原因を探る障害診断などに利用することができる。本研究ではベイジアンネットワークを家電制御システムに組み込むことにより、家電操作における行動履歴から、ユーザが意図する行動を推論するシステムを実現する。

ベイジアンネットワークにおいて、状態などを表す変数をノードとし、変数間の依存関係は向きを持つ有向グラフで表現される。原因となるノードを親ノード、矢印の先の結果となるノードを子ノードと呼ぶ。このグラフ構造をもとにして、観測された子ノードの変数の値から推定した親ノードの確率（確信度）を求めることがベイジアンネットワークにおける確率推論である。この確信度の値が最も大きい状態をその変数の予測結果として意思決定の指針に用いる。本研究では親ノードである行動ノード $B = \{b_1, b_2, \dots, b_l\}$ 、子ノードである音声ノード $V = \{v_1, v_2, \dots, v_m\}$ と家電状態ノード $S = \{s_1, s_2, \dots, s_n\}$ を設定し図 1 のようなネットワークを構築する。1つの親ノードはすべての子ノードと連結している。

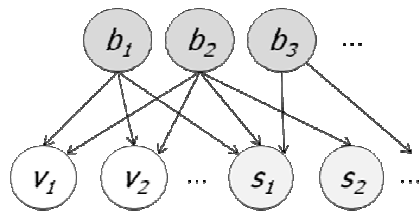


図 1 ベイジアンネットワークモデル

3.2 学習

行動 $b_k \in B$ が観測されたときに家電状態 $s_i \in S$ である条件付き確率を $P(s_i | b_k)$ と表記し、これを全ての事象間において表す(式(1))。この条件付き確率表を CPT(Conditional Probabilities Table)と呼び、次のように表す。

$$CPT_{S|B} = [P(s_i | b_k)] = \begin{pmatrix} P(s_1 | b_1) & P(s_2 | b_1) & \cdots & P(s_n | b_1) \\ P(s_1 | b_2) & P(s_2 | b_2) & \cdots & P(s_n | b_2) \\ \vdots & \vdots & \ddots & \vdots \\ P(s_1 | b_n) & P(s_2 | b_n) & \cdots & P(s_n | b_n) \end{pmatrix} \quad (1)$$

本システムではこの条件付き確率 $P(s_i | b_k)$ を行動履歴から累積頻度度数によって以下のように求める。

$$P(s_i | b_k) = \frac{n_{ik}}{N_k} \quad (2)$$

ここで N_k は b_k が観測された回数であり、 n_{ik} は b_k が観測されたうちに s_i が観測された回数を表す。行動 b_k が観測されたとき音声 v_j が観測された条件付き確率 $P(v_j | b_k)$ も同様に算出する。

$$P(v_j | b_k) = \frac{m_{jk}}{N_k} \quad (3)$$

m_{jk} は b_k が観測されたうちに v_j が観測された回数を表す。新しい情報が入力され行動履歴が更新されるたびに条件付き確率も更新されていく。この CPT を更新することで、システムは状況に応じたユーザの嗜好を学習していく。

3.3 推論

次に確率推論の方法について述べる。まず、ある 2 つのノード X, Y が X の時 Y であるという関係を持っているとする。子ノード Y について観測値（証拠）がある場合、親ノード X の確信度 $BEL(X)$ の求め方はベイズの定理を用いて以下ようになる。

$$\begin{aligned} BEL(X) &= P(X | Y) = \frac{P(Y, X)}{P(Y)} \\ &= \frac{P(Y | X) \cdot P(X)}{P(Y)} = \alpha P(Y | X) \cdot P(X) \end{aligned} \quad (4)$$

ここで $P(X)$ は X の事前確率であり、また、確信度の総和は 1 であることから、 α は正規化定数となる。先ほどの式(1)の条件付き確率を用いると、

$$BEL(X) = \alpha CPT_{Y|X} \cdot P(X) \quad (5)$$

と表せる。

これを本システムに適用し、ある状態 S や音声 V が観測された時の行動 B の確信度を $BEL(B)$ とすると式(5)より以下のように計算できる。

$$BEL(B) = \alpha P(B) \prod_i^n \lambda_i(B) \prod_j^m \mu_j(B) \quad (6)$$

n は家電状態ノード総数, m は音声ノード総数を指す. ここで λ_i は s_i の状態のうち観測された結果から求められるベクトルで, 観測された方を 1, それ以外を 0 としたベクトル s_i を用いると,

$$\lambda_i(B) = P(s_i|B) s_i \quad (7)$$

と表される. 例えば $s_f = \{\text{暖房 ON}, \text{暖房 OFF}\}$ のうち暖房 ON という状態が観測されたとすると $s_f = \{1, 0\}$ となる. 同様に,

$$\mu_j(B) = P(v_j|B) v_j \quad (8)$$

と表せ, 式(6)より

$$BEL(B) = \alpha P(B) \prod_i^n P(s_i|B) s_i \prod_j^m P(v_j|B) v_j \quad (9)$$

となる. $BEL(B)$ の総和は 1 なので, 確信度の最高値は 1 である. $BEL(B) = \{0.0, 0.2, 0.7, \dots, 0.0\}$ のように算出され, 最も確信度の高い行動を出力とする. しかし, 最高確信度があまりにも小さいものは信頼度が低いので, 閾値以下の場合はユーザーに教示を求めるものとする.

3.4 システム概要

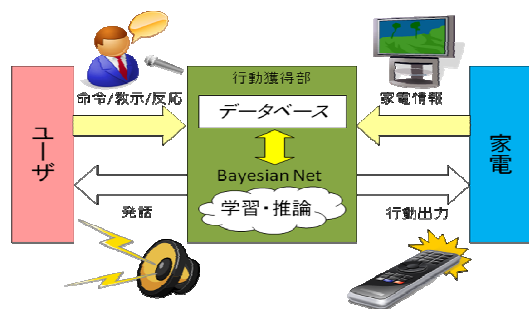


図 2 システム概要図

提案する家電制御システムのシステム概略図を図 2 に示す. 発話された言葉は音声認識部でテキストに変換されシステムに入力される. また各家電のスイッチや温度, 音量などの情報もシステムに入力される. 入力された情報により, ベイジアンネットワークの手法によって行動履歴から確率推論を行い, 行動を決定する. 決定した行動は発話によってユーザーに応答し, 同時に家電に行動を出力する. 確信度が低く行動を決定できない場合はユーザーに教示を求める. 間違った行動を出力してしまった場合も, ユーザーが教示し直す. これらのやりとりはすべて音声によって行われる. 正しい出力行動はその時の発話と家電状態とともにデータベースに保存され, 学習に使用される.

3.5 教示

ユーザーがシステムに発話の意味を教示する際の手順について述べる. 教示をする場合は以下の場合がある.

- (1) システムの行動決定に関する確信度が閾値より低く, 教示を求めてきた場合
- (2) システムの行動出力が間違っていた場合

これらの場合, システムは「どうしたら良いですか?」という応答発話によってユーザーに教示を求める. それに対してユーザーは決められた語を使って応答するとシステムは正しい行動を学習することができる. 決められた語というのは, システムが事前に意味を知っている語であり, すべて「A を B して」(A:家電, B:動作)という発話形式で記述されている. 決められた語しか使えないという制約はあるが, ユーザーが実際に動作を見せて教示をする必要はない.

教示された音声を認識できたことを確認するため, システムは「A を B します」と応答発話する. このとき, システムが間違った行動を出力した場合はユーザーが「違う」や「いや」という否定の発話をすることで何度でも教示し直せる. ユーザーが「うん」「はい」「OK」「ありがとう」という肯定的な発話をするシステムは正しい行動だと判断し, 元々の発話と家電状態とともに行動履歴をデータベースに記録し学習データとする.

4. 実験

4.1 実験方法

提案システムの評価を行うため, ユーザーとシステムの発話インタラクションによる家電操作の実験を行った. 家電は計算機上で仮想的に動作し, マイクを通じてユーザーの音声を入力する. これにより, ユーザーの発話意図が学習できているかどうか, 発話

とシステムの行動が一对一对応ではなく状態を考慮した結果が得られているか、またユーザの嗜好が結果に表れているか調べた。

まず、PC モニタ上で家電のオンオフ状態と制御する変数値の大小をランダムに取得する。その後、ユーザはランダムに設定された家電状態を見ながら、「暑い」等のシステムへの要求を発話する。それに対してシステムは、推論結果から決定された行動を「～します」と応答するか、または「どうしたら良いですか?」と教示を求めてくる。教示を求められた場合、ユーザは 3.5 で記述した手法に沿って「～して」と応答する。システムがユーザの望む応答をした場合はそれを肯定する言葉「ありがとう」や「はい」等と発話するとシステムは正しい動作であると認識する。逆に「違う」等と否定するとシステムは教示を求める。これをユーザの望む結果になるまで繰り返し、このやり取りを 40 回行った。

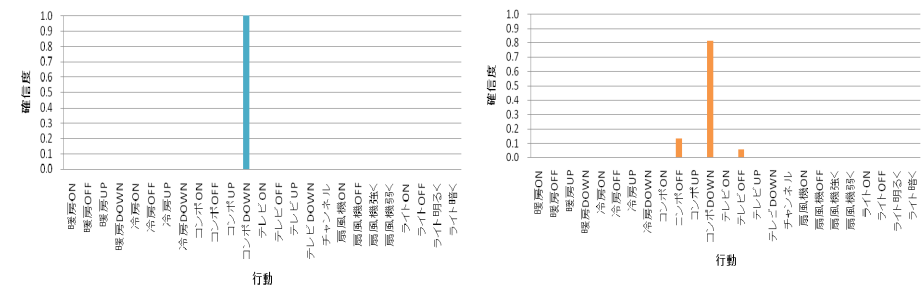
4.2 各種パラメータ

- (1) 家電と動作: 日常生活で一般的に使用される家電 6 種類において家電操作シミュレーションが行えるよう設定した。対象とした家電と動作については表 1 に示す。
- (2) 入力可能単語: システムへの要求をする際の使用可能単語を限定した。表 2 に示した 20 単語とそれらを組み合わせたものが使用できる。
- (3) 教示用単語: 教示用として、システムが予め意味を知っている言葉の設定を行う。表 3 のように、行動教示用の単語が 16 単語、また、システムの応答が正しいか正しくないかを教示するための単語が 6 単語、合計 22 単語を設定した。行動教示用の単語は、1 つの行動に対し、2 単語で構成される 1 つの文を用意した (例: “暖房をつける” という行動に対して「暖房／つけて」という教示文)。
- (4) ノードの設定: ペイジアンネットワークにおけるノードの設定数は行動ノード 25 種類、家電状態ノード 13 種類、音声ノード 20 種類とした。各ノードの具体的な事象は表 4 に示す。
- (5) 確信度の閾値: 教示を求める際の確信度の閾値は経験的に 0.6 とした。この値は、あまり高すぎると出力ができず、教示ばかりを求めるシステムになり、逆に低すぎるとユーザの望まない行動を出力する回数が増える。
- (6) 音声認識／合成: 本システムにおける音声処理は既存のプラグインソフトウェアを使用した。音声認識には Julius for SAPI[5]、音声合成には FUJITSU 音声合成 FineSpeech ケイコを使用した。

5. 実験結果

5.1 学習回数による出力の変化

学習回数によって各行動の確信度がどのように変化するか調べるために、状態と発話を同じ条件にし、学習回数だけ異なるデータを比べた。発話は「うるさい」「つまらない」「寒い」「暑い」「上げて」を使用した。図 4 は、20 回学習した場合と 40 回学習した後、家電状態が「暖房 OFF / 冷房 OFF / コンボ音量大 / テレビ音量小 / 扇風機 OFF / ライト OFF」の時の「うるさい」という発話に対する行動確信度である。図から分かる通り、学習を重ねることで出力行動のパリエーションが増加した。同様に、他の発話に対しても同じ傾向が見られた。これにより、システムはユーザの教示を次第に学習し様々な出力ができるようになったと言える。また行動によって確信度に大きな差があることから、ユーザの行動選択モデルが表れているとも言える。



(a)20 回学習時

(b)40 回学習時

図 3 【うるさい】と発話した時の行動確信度

5.2 家電状態による出力の変化

家電状態によって各行動の確信度がどのように変化するか調べるために、学習回数と発話を同じ条件にし、家電状態だけ異なるデータを比べた。図 5 は「暑い」と発話した時の行動別確信度を表したグラフである。ただし、2 つのグラフは暖房の状態が異なる。同じ「暑い」という発話であっても状態が異なるため、システムは状態に応じてユーザの意図を汲み取った対応を行えることが確認できる。他の発話の場合に関しても、家電状態が異なれば、それぞれの家電状態に沿った行動を出力していることが確認できた。

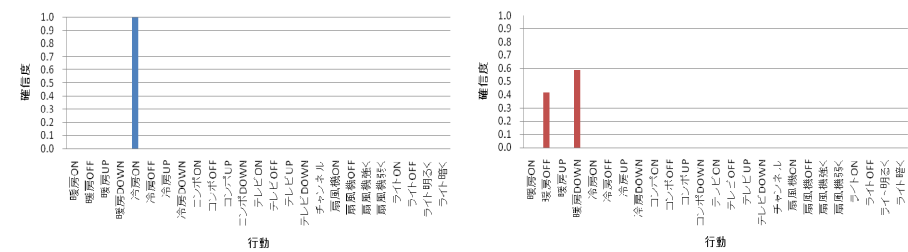


図 4 【暑い】と発話した時の行動別確信度

(左) 暖房 OFF / 冷房 OFF / テレビ OFF / コンボ ON / 扇風機 OFF / ライト ON
(右) 暖房 温度高 / 冷房 OFF / テレビ OFF / コンボ ON / 扇風機 OFF / ライト ON

6. おわりに

ベイジアンネットワークを用いることで行動履歴と現在の家電状態からユーザの発話意図を推定する家電制御システムを提案した。シミュレーション環境での実験を行い、ユーザとインタラクションをすることでユーザの嗜好を学習させた。学習回数が増すにつれて状態やユーザに依存した様々な行動の出力が可能であったことから、使い続けるほどユーザに特化した便利なシステムになると考えられる。また、曖昧な発話に対しても、システムは家電状態を観測することによってユーザの意図を推定することができた。

今後は、家電の状態だけでなく音声から感情情報などを取り込むことで、さらにユーザの意図を高い精度で推定できるシステムを目指したい。また、実用化に向け、より自由な発話に対応するために、ノード数が可変となるシステムの構築を考えている。

謝辞 本研究の一部は、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「デジタルメディア作品の製作を支援する基盤技術」、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) 榊原,中村,井垣,松本, ”ホームネットワークシステムにおける家電状態を利用した音声操作インタフェースの改善”, 電子情報通信学会 2 種研究会サイバーワールド (CW) 第 9 回研究会,2008
2) 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
3) 稲邑,稲葉,井上, ”PEXIS:統計的経験表現に基づくパーソナルロボットとの適応的インタラクションシステム”,電子情報通信学会論文誌,Vol.J84-D-I,No.6,pp.867-877.2001
4) 本村,岩崎, ”ベイジアンネットワーク技術”,東京電機大学出版局,2006
5) 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/>

付録

表 1 家電と動作

	家電	動作
1	暖房	電源オン／オフ、温度 高／低
2	冷房	電源オン／オフ、温度 高／低
3	扇風機	電源オン／オフ、風量 強／弱
4	テレビ	電源オン／オフ、音量 大／小、チャンネル
5	コンボ	電源オン／オフ、音量 大／小
6	照明	電源オン／オフ、明／暗

表 2 ユーザが使用可能な言葉

	単語		単語
1	暑い	11	弱い
2	寒い	12	つまらない
3	うるさい	13	とめて
4	大きい	14	まぶしい
5	小さい	15	見えない
6	聞こえない	16	電気
7	静かに	17	上げて
8	暗い	18	下げて
9	明るい	19	温度
10	強い	20	音量

表 3 教示用の言葉

	言葉		言葉
1	暖房	12	強く
2	冷房	13	弱く
3	扇風機	14	変えて
4	コンポ	15	明るく
5	テレビ	16	暗く
6	照明	17	うん
7	チャンネル	18	はい
8	つけて	19	いや
9	消して	20	違う
10	上げて	21	ありがとう
11	下げて	22	OK

表 4 行動と音声と家電状態の事象

	行動 B	音声 V	家電状態 S
1	暖房 ON	暑い	暖房:ON/OFF
2	暖房 OFF	寒い	暖房温度:高/低
3	暖房の温度上げる	うるさい	冷房:ON/OFF
4	暖房の温度下げる	大きい	冷房温度:高/低
5	冷房 ON	小さい	扇風機:ON/OFF
6	冷房 OFF	聞こえない	扇風機:強/弱
7	冷房の温度上げる	静かに	コンポ:ON/OFF
8	冷房の温度下げる	暗い	コンポ音量:大/小
9	扇風機 ON	明るい	テレビ:ON/OFF
10	扇風機OFF	強い	テレビ音量:大/小
11	扇風機の風量強める	弱い	テレビチャンネル:1～6
12	扇風機の風量弱める	つまらない	照明:ON/OFF
13	コンポ ON	とめて	照明:明/暗
14	コンポ OFF	まぶしい	
15	コンポの音量上げる	見えない	
16	コンポの音量下げる	電気	
17	テレビ ON	上げて	
18	テレビ OFF	下げて	
19	テレビの音量上げる	温度	
20	テレビの音量下げる	音量	
21	テレビのチャンネル変え		
22	照明 ON		
23	照明 OFF		
24	照明明るくする		
25	照明暗くする		