音声対話における実例に基づく未知語属性推定

高橋 康博　堂坂 浩二　相川 清明
NTTコミュニケーション科学基礎研究所

1　はじめに

音声対話システムとは，人とコンピュータが音声対話を通して，何らかの仕事(以降，タスクと呼ぶ)を行うシステムのことである[1]．音声対話システムが保有する知識に制限があり，ユーザがシステムの知識の限界を知っているとは限らないことから，ユーザ発話中にシステムにとって未知の語(以降，未知語と呼ぶ)が含まれることがある．未知語の処理を行わないシステムでは，システムが持つ知識だけを用いて未知語を強引に認識するので，ユーザに対して誤った対応をしてしまう．この場合，ユーザは，システムの誤った対応が音声認識誤りによるものなのか未知語を含むことによるものなのかを区別できないので，タスクを完了できないだけでなく，無駄な対話を続けることになる．そこで，自然な対話によりタスクを遂行する音声対話システムの実現には，未知語知識獲得が重要となる．未知語知識獲得とは，ユーザとの対話を通して，未知語を検出し，システムの知識構造と関連づけ，システムの持つ語彙として利用できるようにすることである．

本稿では，未知語知識獲得の第一歩として，ユーザ発話理解結果に現れた未知語の意味的な属性推定に注目する．未知語の属性推定を行うためには，ユーザ発話中の未知語を検出し，文字列として切り出す未知語切り出し機能が必要となるが，本稿では，未知語切り出し機能を持つと仮定した音声対話システムを考える．未知語の属性推定は，未知語が切り出された後の対話戦略の決定に必要である．例えば，人名を獲得できるビデオ録画予約システムとユーザとの対話の中で未知語が切り出された場合，未知語の属性が人名かどうかによって，その後の対話戦略は変わる．しかし，未知語が切り出されるたびに人名かどうかをユーザに質問したのでは無駄な対話が多くなる．そこで，ユーザへの質問を極力少なくして，未知語の属性推定を行うことが必要になる．

本稿では，実例に基づいて未知語属性推定を行う方法について述べる [1]．この方法は，システム発話と直後のユーザ発話からなる実例をシステムが蓄積し，ユーザ発話に未知語が現れた場合は，そのユーザ発話と直前のシステム発話からなる組に最も類似する実例を選び，その実例に基づいて未知語の属性推定を行うことを基本とする．[1] における類似度は，未知語を含むユーザ発話が行われた直前のシステム発話と未知語の前後の句に基づいている．しかし，本稿における類似度は，[1] で扱われている情報だけでなく，システム発話が行われたときのスロット状態や音声認識誤りも考慮したものである．ただし，[1] で扱われている未知語の前後の句パターンについては考慮していない．実例を利用するという方法としては，[3, 4] において，実例に基づく機械翻訳の方法が提案されている．

[1]本稿で対象とする音声対話システムは，ユーザ発話の理解状態をスロットと値の組のリストで表現するものとする．

2　実例に基づく未知語属性推定

本稿における実例について説明する．ユーザとシステムの対話において，システムは，現在のスロット状態に応じて発話し，その発話に対するユーザ発話によりスロット状態を変化させ，そのスロット状態に応じてまた発話をする，ということを繰り返す．このときの

スロット状態 → システム発話 → スロット変化

というシステムの内部状態の変化を 1 組のデータ (以降，局所的対話遷移と呼ぶ) とみなし，このデータにスロット変化を引き起こしたユーザ発話の書き起こしを加えたデータを実例とする．

ビデオ録画予約システムを例にとり，提案方法を適用してみる．現在の局所的対話遷移が次のようになっているとする (例 1)．

スロット状態　：　(曜日：火)
システム発話　：　「番組名は何ですか」
スロット変化　：　未知語

すなわち，曜日スロットの値が「火」であるとき，システムが「番組名は何ですか」と発話し，それに対するユーザ発話を，「ユーザは未知語 1 語のみ発話した」と

システムが理解したとする．このときシステムは，現在の局所的対話遷移と最も高い類似度を持つ実例を蓄積された実例の中から探す．例 1 では，次の条件を同時に満たすデータを探す．

- 曜日スロットが値を持っている
- システムは番組名についての質問をしている
- ユーザ発話によりスロットが 1 つ変化している
- スロット変化が音声認識誤りによるものではない

スロット変化が音声認識誤りによるものではないことは，実例中のスロット変化と書き起こしとの類似度をみる．上の条件の下で，次のような実例が選ばれたとする．

スロット状態　：　(曜日：水)
システム発話　：　「番組名は何ですか」
スロット変化　：　(番組名：街角クイズ)
書き起こし　：　街角クイズです

このとき，現在の局所的対話遷移の中にある未知語は，実例中の「街角クイズ」に対応するので，未知語は街角クイズと同じ「番組名」という属性を持つと推定する．

3　関連研究

音声対話における未知語の属性推定を行っている研究として [2] がある．[2] では，未知語を含むユーザ発話の解析結果だけから未知語の属性を推定しているが，本稿で提案する方法は，ユーザとシステムの対話を利用したものである．未知語知識獲得の研究としては [5] がある．[5] では，離散発声された単語列を入力として行動を出力し，その行動に対するユーザの反応をもとに，単語を学習する方法を提案している．しかし，我々の目指している未知語知識獲得は，自然な対話によるものであり，ユーザに離散発声を強いるようなものではない．

4　類似度の定義と属性推定手続き

提案方法を実現するためには，次の 2 つの類似度の定義が必要である．

- 局所的対話遷移間の類似度 sim_A
- スロット変化と書き起こしとの類似度 sim_B

この節では，これらの定義の詳細について述べる．

4.1　局所的対話遷移間の類似度

局所的対話遷移は，スロット状態，システム発話，スロット変化からなるデータなので，これら 3 つの要素を順に並べたリストで局所的対話遷移を表すことにする．2 つの局所的対話遷移 $A1$, $A2$ に対し，それらの類似度を各要素間の類似度の和とする．すなわち，各要素間の類似度を計算する関数をそれぞれ sim_1, sim_2, sim_3 とし，リスト L の第 i 成分を L_i と表すとき，局所的対話遷移間の類似度 sim_A を次のように定義する．

$$sim\_A(A1, A2) = \sum_{i=1}^{3} sim_i(A1_i, A2_i).$$

以下で，sim_1, sim_2, sim_3 の定義を述べる．

4.1.1　スロット状態間の類似度

スロット状態として扱う情報は，どのスロットが値を持っているかということのみとする．そこで，値を持っているスロットの集合でスロット状態を表すことにする．2 つのスロット状態 $SL1$, $SL2$ に対し，同じスロットが値を持っている個数をそれらの類似度とする．すなわち，sim_1 を次のように定義する．

$$sim_1(SL1, SL2) = |SL1 \cap SL2|.$$

4.1.2　システム発話間の類似度

システム発話として扱う情報には，まず，質問か確認かの区別がある．この区別が一致しないシステム発話間の類似度は 0 とする．システム発話が質問の場合は，質問されるスロット (以降，焦点スロットと呼ぶ) とそれらが発話される順番，システム発話に含まれる語が属するスロット (以降，非焦点スロットと呼ぶ) とそれらが発話される順番という情報を使う．そこで，システム発話が質問の場合は，次のように表すことにする．

(焦点スロットのリスト,非焦点スロットのリスト)

システム発話が確認の場合は，確認されるスロット (以降，焦点スロットと呼ぶ) とそれらが発話される順番という情報を使い，質問の場合と同様に表すことにする[2]．

システム発話間の類似度を定義するために 2 つの補助関数 $aux1$, $aux2$ を準備する．$aux1$ はスロットの 2 つのリストに対し，先頭同士，2 番目同士と順に比べて等しい個数を数える関数である．ただし，途中で等しくない組合せが現れるまでとする．$aux2$ はスロットの 2 つのリストに対し，それらが互いに切片または拡張に

[2]非焦点スロットのリストは空リストと考える．

なっていれば, 2 つのリストの長さの差の絶対値を返す関数である.

システム発話間の類似度は, 焦点スロットの類似度と非焦点スロットの類似度を $aux1$ により計算した値を基にし, 一方のシステム発話内容が他方のシステム発話内容の拡張になっている場合は, $aux2$ により計算される拡張の度合だけ類似度を低くする. すなわち, 2 つのシステム発話で, 共に質問または共に確認である $SY1$, $SY2$ に対し, 類似度 sim_2 を次のように定義する.

$$\begin{aligned} sim_2(SY1, SY2) &= \sum_{i=1}^{2} s_i \times aux1(SY1_i, SY2_i) \\ &- s_3 \times \sum_{j=1}^{2} aux2(SY1_j, SY2_j). \end{aligned}$$

ただし, s_1, s_2, s_3 は定数で, それぞれ, 焦点スロット, 非焦点スロット, 拡張の度合を類似度に強調させる程度を表す. また, sim_2 の計算結果が負のときは, sim_2 の値を 0 とする.

4.1.3 スロット変化間の類似度

スロット変化として扱う情報は, ユーザ発話をシステムが理解したことにより変化したスロットと変化した順番である. そこで, 変化したスロットを順に並べたリストでスロット変化を表すことにする. また, ユーザ発話中の未知語は, スロット変化に未知語スロットを入れることにより表すことにする. 未知語スロットは, $aux1$ や $aux2$ により他のスロットと比較された場合は, 全てのスロットと等しいと判定されることにする.

スロット変化間の類似度は, 変化したスロットと変化した順番について $aux1$ により比べた値を基にし, 一方のスロット変化が他方のスロット変化の拡張になっている場合は, $aux2$ により計算される拡張の度合だけ類似度を低くする. すなわち, 2 つのスロット変化 $CH1$, $CH2$ に対し, 類似度 sim_3 を次のように定義する.

$$\begin{aligned} sim_3(CH1, CH2) &= c_1 \times aux1(CH1, CH2) \\ &- c_2 \times aux2(CH1, CH2). \end{aligned}$$

ただし, c_1, c_2 は定数で, それぞれ, スロット変化, 拡張の度合を類似度に強調させる程度を表す. また, sim_3 の計算結果が負のときは, sim_3 の値を 0 とする.

4.2 スロット変化と書き起こしとの類似度

書き起こしとして扱う情報は, 書き起こしから判断されるユーザ意図が完全にシステムに理解されたと仮定したときに, 変化するスロットと変化する順番である. そこで, 変化するスロットを順に並べたリストで書き起こしを表すことにする.

スロット変化 CH と書き起こし TR との類似度 sim_B の計算は次のような手続きによる. まず, CH は, 現在のスロット変化の中にある未知語スロットに対応するスロット $slot_x$ を含む. そこで, $slot_x$ が TR に含まれるかを判定する. もし含まれなければ類似度 0 とする. もし含まれれば, CH と TR の中にある $slot_x$ の直前のスロット同士, 直後のスロット同士を比べる. 両方とも等しくない場合は類似度 1, 片方だけ等しい場合は類似度 2, 両方とも等しい場合は類似度 3 とする.

4.3 属性推定手続き

実例は, 局所的対話遷移と書き起こしからなるデータなので, これら 2 つの要素を順に並べたリストで実例を表すことにする. スロット変化に未知語スロットを含む局所的対話遷移 A と実例 B に対し, それらの類似度 SIM を次のように定義する.

$$SIM(A, B) = \frac{sim\_A(A, B_1) + sim\_B((B_1)_3, B_2)}{sim\_A(A, A) + sim\_B(A_3, A_3)}.$$

このとき, 本稿で提案する未知語属性推定の手続きは次のように述べられる.

1. 現在のスロット変化に未知語スロットが現れる.
2. 現在の局所的対話遷移と各実例の類似度を SIM により計算し, 最も値の大きい実例を選ぶ.
3. 2 で選ばれた実例が 1 個であれば, その実例により属性を推定する. 複数個ある場合は, 4 へ.
4. 選ばれた各実例により未知語の属性を推定し, 多数決で決める. それでも決まらない場合は, 5 へ.
5. 現在のスロット変化に既に現れている属性は選ばないことにする. それでも決まらない場合は, 候補となっている属性からランダムに選ぶ.

5 実験

我々のグループで作られた天気情報案内システム飛遊夢 (ヒューム) を使い, 提案方法による属性推定実験をした. 飛遊夢の理解できる語の属性は, 「時間」, 「場所」, 「気象」, 「意図」の 4 つである. 実例の集合を作るために, グループ員 10 人がシステムと対話し, 14 対話, 402 発話収集した. この対話によるシステムログと実際のユーザ発話内容を記録した音声ファイルを基

表 1: 現在のスロット変化に未知語を 1 個含む場合の属性推定結果 (正解数/問題数 (割合))

未知語位置	1 位に正解	1 位に正解 & 類似度 0.8 以上	1 位に正解 & 類似度 0.9 以上
第 1 番目	66/96 (0.688)	59/82 (0.720)	54/70 (0.771)
第 2 番目	18/38 (0.474)	10/14 (0.714)	3/6 (0.500)
第 3 番目	4/7 (0.571)	データ無し	データ無し
未知語位置	2 位までに正解	2 位までに正解 & 2 位類似度 0.8 以上	2 位までに正解 & 2 位類似度 0.9 以上
第 1 番目	78/96 (0.813)	61/72 (0.847)	27/32 (0.844)
第 2 番目	25/38 (0.658)	2/4 (0.500)	データ無し
第 3 番目	4/7 (0.571)	データ無し	データ無し

に, 96 個の実例を作った. 未知語を含む対話としては, 実例中のユーザ発話理解結果の一部の語を未知語と仮定することにより作った. 未知語と仮定された語の属性が推定されたとき, 属性推定正解とする. 4 節の類似度の定義の中のパラメータは, 予備実験の結果, システム発話間の類似度において, 焦点スロットを強調し, 拡張の度合が非焦点スロットの類似度の 1/2 と同等にするために s_1 を 2, s_2, c_1 を 1, s_3, c_2 を 1/2 とした.

実例 96 個のうち, システム発話が質問であるものは 52 個, 確認であるものは 44 個であった. システムが質問した属性は, 「時間」(25 回), 「場所」(24 回), 「気象」(3 回) の 3 種類であり, 1 回のシステム発話で 1 種類の質問をした. また, システムが確認した属性は, 「時間」, 「場所」, 「気象」, 「意図」の 4 種類であり, 1 回のシステム発話で, 「時間, 場所, 意図」(33 回), 「場所, 意図」(6 回), 「気象, 意図」(2 回), 「時間, 意図」(1 回), 「意図」(2 回) の確認をした. ユーザ発話により変化した属性は, 1 回のシステム発話に対し最大 3 個であり, 各属性がスロット変化の第 1 番目に現れた回数は, 時間, 場所, 気象, 意図の順に, 22 回, 66 回, 0 回, 8 回, 第 2 番目には, 1 回, 27 回, 1 回, 9 回, 第 3 番目には, 0 回, 4 回, 0 回, 3 回であった.

実験は次のようにして行った. まず, 96 個の実例を, 実例 10 個からなる 9 つのグループと実例 6 個からなる 1 つのグループにランダムに分けた. 次に, 1 つのグループを選び, その中の実例から未知語を含む対話を作った. そして, 未知語を含む対話を作ったグループ以外の全グループの実例を使って, 未知語属性推定をした. 未知語を含む対話を作るグループを換えて実験をくり返すことにより, 表 1 のような結果が得られた[3].

比較として, 現在のスロット変化の第 1 番目に未知語が現れた場合について, 簡単なヒューリスティックスによる未知語の属性推定を行った. このヒューリスティックスは, システム発話が質問である場合, 現在のスロット変化の第 1 番目にある未知語の属性は, システムが質問している属性であると推定するものである. また, システム発話が確認の場合は, 第 1 番目に確認された属性を未知語の属性として選ぶ. ただし, 選ばれた属性が既に現在のスロット変化に現れている場合は, 選ばれた属性以外の属性をランダムに選ぶ. このヒューリスティックスによる属性推定正解率は, 45.8% であった.

[3]表 1 中の未知語位置とは, 現在のスロット変化の何番目に未知語スロットが現れたかを表している.

6 おわりに

本稿では, 音声対話における未知語知識獲得の第一歩として, 未知語の意味的な属性推定に注目した. そして, 実例に基づく未知語属性推定法を提案し, この方法が簡単なヒューリスティックスよりも高い属性推定正解率を得ることを確認した. 今後は, 大量の対話データにより, 今回と同様の実験を行うとともに, より大規模な音声対話システムへの応用を考えていきたい.

謝辞 日頃より御指導頂くメディア情報研究部 萩田紀博部長, 有益な示唆を頂くマルチモーダル対話研究グループの諸氏に感謝致します.

参考文献

[1] 荒木雅弘, 堂下修司. 対話事例ベースによる発話内容の推定および未知語の解析. 情報処理学会第 49 回全国大会発表論文集 Vol. 3, pp. 155–156, 1994.

[2] 伊藤克亘, 速水悟, 田中穂積. 音声対話システムにおける未知語の扱い. 人工知能学会研究会資料 SIG-SLUD-9201-1 (4/15), pp. 1–9, 1992.

[3] 佐藤理史. MBT1 : 実例に基づく訳語選択. 人工知能学会誌 Vol. 6 No. 4, pp. 129–136, 1991.

[4] 佐藤理史. MBT2 : 実例に基づく翻訳における複数翻訳例の組合せ利用. 人工知能学会誌 Vol. 6 No. 6, pp. 75–85, 1991.

[5] A. L. Gorin, S. E. Levinson, and A. N. Gertner. Adaptive acquisition of spoken language. ICASSP-91, pp. 805–808, 1991.